

平成 23 年度産業技術研究開発

「ライフサイエンスデータベースプロジェクト」
に関する成果報告書

平成 24 年 3 月

独立行政法人 産業技術総合研究所

目次

1.	総括	1
2.	研究開発成果	2
2. 1	ポータル構築連携	2
2. 1. 1	経済産業省ライフサイエンスポータルサイトの拡張作業及び運営	2
2. 1. 2	統合データベースセンターとの連携	15
2. 2	横断検索連携	16
2. 3	アーカイブ構築連携	19
2. 4	国内外の最新動向調査及び普及に向けた取り組み	22
2. 4. 1	国内外の最新動向調査	23
2. 4. 2	広報・普及活動	24
2. 4. 3	ユーザーニーズ調査	31
2. 5	運営委員会	33
2. 6	成果発表	34
3.	謝辞	38
4.	参加者名簿	39

1. 総括

平成 23 年度経済産業省ライフサイエンスデータベースプロジェクトでは、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に関するポータル構築連携、横断検索連携、アーカイブ構築連携を行うことを目的とした。統合データベースセンターとの緊密な連携により、本事業の成果を広く産業界や社会へ提供することをめざして、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が実施した。

主な成果としては、まず、前年度までの経産省統合データベースプロジェクトによって構築されたポータルサイト MEDALS (<http://medals.jp/>) の公開を継続し、またその内容の大幅な充実を図った。特に、経産省関連の各種研究開発プロジェクトで作成されたデータベースやソフトウェアの便覧については、詳細な調査を実施した結果、合計で 138 件の成果物をリストに掲載した。また、文部科学省ライフサイエンスデータベースプロジェクトと協力して 314 種類のデータベースに対する横断検索サービスを実現した。ID を手がかりとした独自の手法により、データベースの統合化を推進した。国際的な連携については、BioDBcore の活動を通じてデータベースカタログ記載方式の国際標準化を推進した。このほかにも、データベース講習会開催などの広報活動を積極的に行ったこと、ポータルサイト MEDALS に関する論文発表や学会発表を行ったことなど、多くの成果を挙げて有用な情報を社会に発信することができた。今後も利用者となる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する。

2. 研究開発成果

● 背景

ライフサイエンス分野では、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。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。

このため、総合科学技術会議では、上記の問題を解決するため、ライフサイエンス分野の研究成果をオールジャパン体制で整備するための報告書及びロードマップを策定している。(総合科学技術会議専門調査会分野別推進戦略総合PTライフサイエンス PT 第 14 回配布資料：<http://www8.cao.go.jp/cstp/project/bunyabetu2006/life/14kai/haihu14.html>) これを受け、経済産業省では平成 20 年度から平成 22 年度にかけて、統合データベースプロジェクトを実施し、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たポータルサイトの構築等を実施してきたところである。

● 目的

本事業では、独立行政法人科学技術振興機構(JST)に設置された統合データベースセンターを中核とした政府全体の統合データベースを構築するため、本事業では関係府省の分担業務であるポータル構築連携、横断検索連携、アーカイブ構築連携、を行うことを目的とする。

● 概要

本事業では、平成 20 年度から平成 22 年度に実施した「産業技術研究開発(統合データベースプロジェクト)」の成果を活用し、統合データベースセンターと連携して政府全体の統合データベースを構築するため、「1. ポータル(サイト)構築連携」、「2. 横断検索連携」、「3. アーカイブ構築連携」、「4. 国内外の最新動向調査及び普及に向けた取り組み」、及び「5. 運営委員会」について実施した。以上を実施するにあたり、文部科学省、農林水産省等、他省庁のデータベース関連プロジェクトと連携し、さらに産業界のニーズを取り込むことに積極的に取り組んだ。

2.1 ポータル構築連携

2.1.1 経済産業省ライフサイエンスポータルサイトの拡張作業及び運営

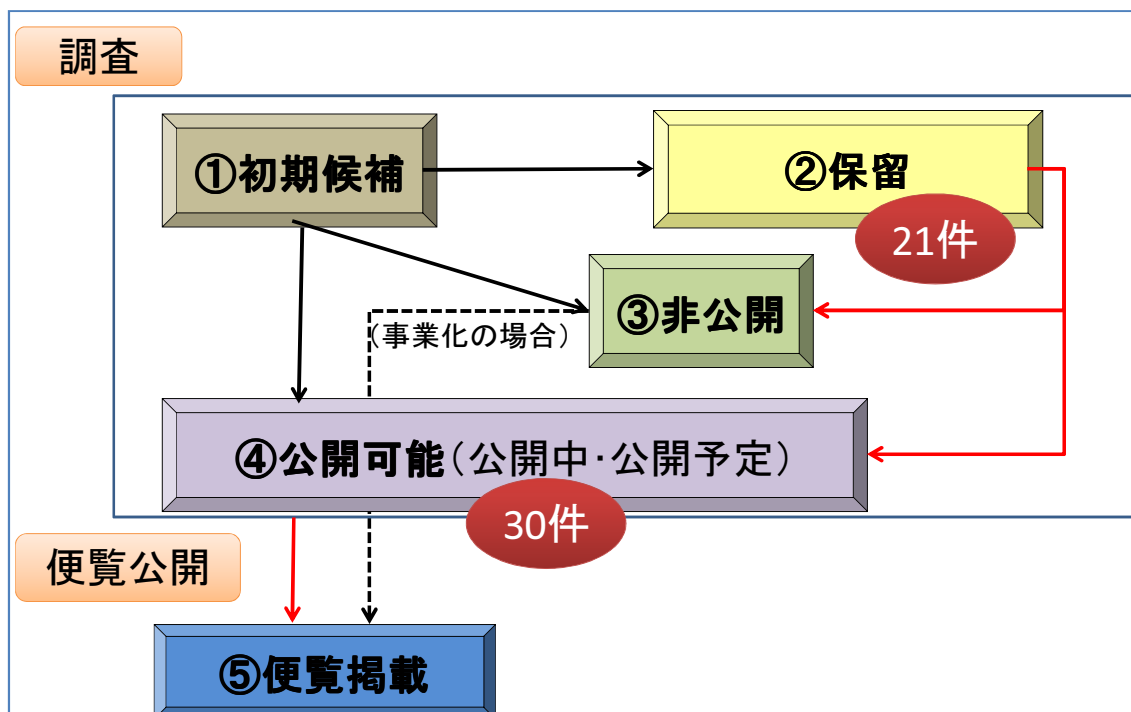
2.1.1-(1) 公開可能な成果物の調査と便覧への掲載

① 成果物の調査

平成 20 年度から平成 22 年度に実施された「産業技術研究開発(統合データベースプロジェクト)」において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によるデータベースが調査された。その結果、データベース等の成果物候補が存在するプロジェクト数は 59 件(当時進行中のプロジェクトを含む)であり、これによる成果物候補は 211 件あることが明らかになっていた。また、これらの成果物のうち、公開されたものについては、平成 23 年 3 月 31 日時点で、129 件がポータ

ル サイト MEDALS (<http://medals.jp/>) の 成 果 物 便 覧 (カ タ ロ グ) に 格 納 さ れ て い た (http://medals.jp/list/list_alldb、http://medals.jp/list/list_alltool)。残りの 82 件は便覧未掲載で、31 件が非公開データ、21 件が公開可否未確認、30 件が公開可能(公開予定を含む)という状態であった。

これらの成果物の調査からポータルサイト MEDALS の成果物便覧に載るまでの流れは図 2. 1-1 の形で行われていた。まず成果物便覧に掲載してあるもの以外で他の成果物がないか、各プロジェクトの成果報告書をもとに、それらの存在・公開状態の調査を行う。主な候補は、新エネルギー・産業技術総合開発機構(NEDO)の未格納分や製品評価技術基盤機構(NITE)によるライフサイエンス系成果物である。経済産業省の関連機関や Web 上の情報を元に、ライフサイエンスのデータベース(ツールも含む)として適切と判断したものを「初期候補」とする。候補からコンタクト先を調査し、各プロジェクトのリーダーや担当の方にコンタクトをとって、どのような具体的な成果物(データベース)があったかを問い合わせる。回答によって、「保留」、「非公開」、または「公開可能(公開している、公開予定である)」となる。「保留」の例としては、個別の担当者にコンタクトがとれないものや、担当者が不明ですぐには出てこないものなどがある。「非公開」というのは、無償公開ではないということで、製品やサービスとして商業利用している場合はこれに含まれる。また、「公開可能」には公開予定のもの、すなわち調査中の時点ではまだ公開されていないものが含まれる。



平成23年度の事業としては、「保留」21件の公開可否未確認の調査を行った。方法としては「①初期候補」に対する調査方法と同様で、これらについて再度コンタクトを取り、公開か非公開かを調査した。すぐには回答が得られないことが多く、その場合でも最低3回はメールないし電話でコンタクトをとった。また、30件の「公開可能」について進捗や詳細を調査し、公開されたものを便覧に載せ、各データベースの本家サイトへ便覧からハイパーリンクをする作業を行った。この作業が済んだ成果物は「便覧掲載」の状態になる。この次の段階としては、2.2節に記述の「横断検索」や2.3節に記述の「アーカイブ」という形での活用を検討することになる。それらについては、MEDALSのプロジェクト便覧(各成果物を出したプロジェクトについて記述している)にその情報を掲載すること、およびデータをダウンロードする形態での公開(2.3節のアーカイブ構築作業に詳述)を検討した。

「保留」21件の公開可否未確認の調査の結果をまとめたのが表2.1-1である。調査の過程で、別の報告書に書かれていた別名のものが実は同じ成果物と判明したものが4件あったため、実際には17件であった。この17件を調査した結果として、1件が公開、14件が非公開、2件が保留となった。唯一公開された1件は、既公開システムのGGDB(糖鎖遺伝子データベース)に統合されていた。非公開となった場合、その理由として多いのは、事業化済みが6件、内部システムのため公開しないものが4件であった。事業化済みのものについては、プロジェクト便覧に成果物名称と事業化の会社が存在する場合はそのコンタクト先を載せることで、アクセスしたいユーザーへの便宜を図り、活用を促進するようにした。

表 2.1-1 「保留」の公開可否未確認の調査結果

項番	名称	状態	非公開理由等
1	糖鎖合成関連遺伝子ライブラリーのデータベース	公開	GGDBに統合
2	共生微生物のメタゲノムデータ	非公開	共同研究利用中
3	XanaMetaDB(メタゲノムデータベース)	非公開	事業化済
4	XanaPGMap(SOMIによる微生物ゲノム系統分類)	非公開	事業化済
5	リアルタイム・マルチカラー・デジタル録画システム	非公開	事業化済
6	遺伝子ネットワーク解析研究支援システム(voyagene)	非公開	事業化済
7	遺伝子制御ネットワーク構築を支援するワークベンチシステム(Webgen-Net)	非公開	事業化済
8	遺伝子同定支援ワークベンチシステム(Webgen-Search)	非公開	事業化済
9	セルロース分解酵素	非公開	担当者離職
10	遺伝子機能推定システム(GenePick)	非公開	内部システム
11	糖鎖統合データベース	非公開	内部システム
12	糖転移酵素データベース(Cabos)	非公開	内部システム
13	発現頻度情報収集用ソフトウェア	非公開	内部システム
14	バーチャルスクリーニングシステム	非公開	無回答
15	糖鎖構造解析並びに機能予測システム	非公開	古い手法で価値なし
16	遺伝子発現プロファイル計測実験支援システム	保留	
17	転写因子データベース(TFDB)	保留	

「公開可能(便覧未掲載)」30件についての調査結果をまとめたものが表 2. 1-2である。調査の結果、およそ半分(16個)のデータが公開済みとなっていた(うち2件は一部が公開)。一部公開になった2件のうち、「組織別遺伝子発現データベース」については残りのデータも論文発表などで公開の可能性があるとしており、引き続き調査が必要である。その他、6件のデータは公開予定のまま進捗がなかった。また、昨年度の段階では公開予定であったものでも、結果的には非公開となったものが多いことが判明した(8件)。理由として多いのは、知的財産に関する問題(知財問題)が解決しなかった、ということである。これらのデータベースは同じプロジェクトで様々な企業が関与していた。プロジェクト開始時の計画では、企業向けに内部的公開をして、その後一般公開の予定であったが、結果的には非公開となった。この出来事は、公開への手順や期限を守ることが、ともすれば守られないか変更されるということを示しており、今後公開を予定しているプロジェクトにとって教訓になると思われる。

表 2. 1-2 「公開可能(便覧未掲載)」についての調査結果

項番	名称	状態	備考(非公開理由等)
1	FragmentAlign(GC-MS分析による検出ピーク解析ソフト)	公開済	-
2	Fuloja(ミヤコグサの完全長cDNA)	公開済	-
3	Metabolometrisc(GC-MS分析による検出ピークアラインメント)	公開済	-
4	MFSearch	公開済	-
5	PHMMTS(二次構造配列アラインメント)	公開済	-
6	PMPi-Blast(ミヤコグサやカンゾウのESTや完全長cDNAのDB)	公開済	-
7	PowerFT	公開済	-
8	PSTAG(RNAの偽結び目構造のアラインメント)	公開済	-
9	SokanProject(Pearson積率相関係数の総当たり計算ツール)	公開済	-
10	SpiceHIT(電気泳動、質量分析ピーク検出システム)	公開済	-
11	癌組織遺伝子発現データベース	公開済	-
12	細胞局在データ	公開済	HGPD(既公開)で公開
13	染色体異常データシステム	公開済	-
14	日本人ゲノム多型のデータ収集と情報データベース	公開済	-
15	組織別遺伝子発現データベース	一部公開済	-
16	ラットF344の遺伝子発現データ(発癌性化合物)	一部公開済	担当者離職
17	アルツハイマー病脳画像データ	公開予定	後継PJで公開準備
18	自己免疫疾患等SNPデータセット	公開予定	後継PJで利用中
19	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	公開予定	-
20	代謝知識情報DB	公開予定	-
21	毒性知識情報DB	公開予定	-
22	有害情報評価支援システム統合プラットフォーム	公開予定	-
23	DSstandardcompoundDB(化合物同定支援データベース)	非公開	内部用システムである
24	npdfold	非公開	改良中で公開予定なし
25	バイオインフォマティクスデータベースシステム	非公開	成果不良
26	ARTRA(シロイヌナズナのマイクロアレイ情報)	非公開	知財問題
27	カンゾウESTデータベース	非公開	知財問題
28	トチュウESTデータベース	非公開	知財問題
29	パラゴムESTデータベース	非公開	知財問題
30	ユーカリESTデータベース	非公開	知財問題

以上の2つの作業の結果、今後も継続調査が必要な成果物の数は、保留の21件については残り2件となり、公開可能30件については残り7件の合計9件となった。これら9件については、次年度以降のプロジェクトで経過を調査することが期待される。これにより「公開可能」の件数は、この残り7件と、公開済みであるが未だ便覧掲載していない5件があり、合計14件となった。図2.1-2は、これまでの調査結果で判明した各カテゴリの成果物件数を時系列で示したものである。全体的に成果物が増加し、「①候補」「②保留」であった成果物が、段階的に公開か非公開に分類されていることが分かる。また、作業をすすめていく途中で、関連研究機関からの発表で知るところとなった成果物も7件存在した。これらは、計画段階では報告者らが把握していなかったものである。これらは判明した時点で調査を行い、結果を便覧に掲載した。

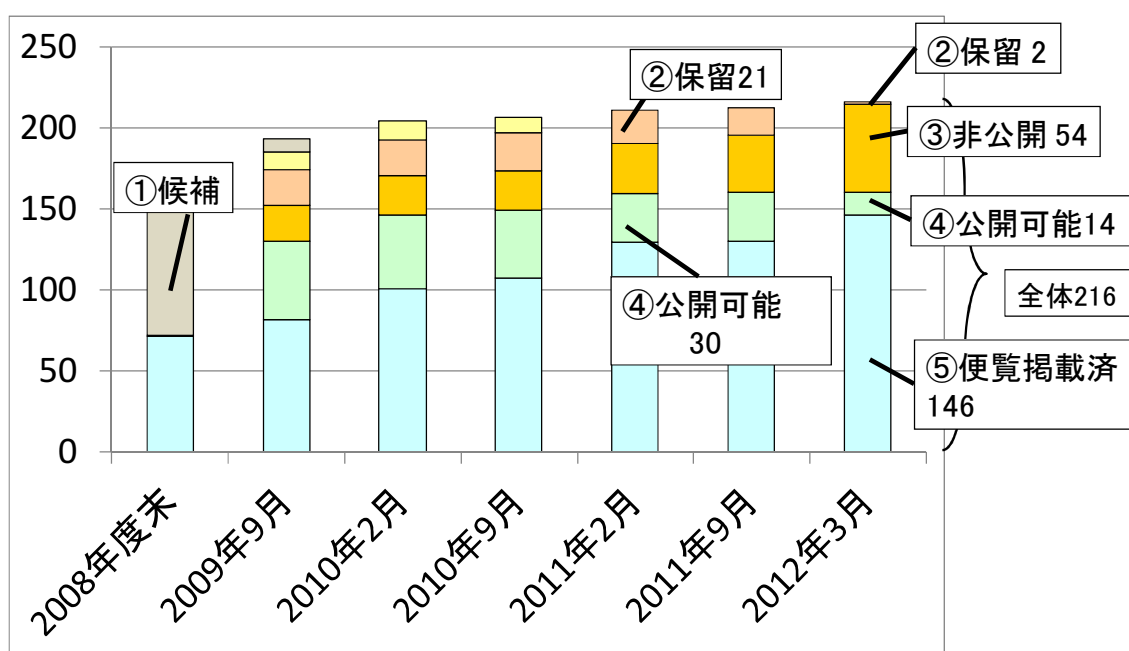


図 2.1-2 成果物調査および便覧掲載作業の進捗

縦軸は件数、横軸は年月。調査が進んだ結果、図 2.1-1 の成果物状態の件数が変化して「②保留」の件数が減り、「⑤便覧掲載」件数が増えている。これは順調な進捗を示している。

② 便覧の更新

(a) 便覧の定期更新とRSSによる更新情報の配信

既に本便覧に掲載されているデータベース詳細情報の内容の更新、改良を定期的に行った。更新にあたっては、我々が独自開発した外部のデータベースサイトに対してそのコンテンツ変更を検知する「更新検知ツール」を用いて毎日更新をチェックした。このツールをもちいると、MEDALS便覧に載っているデータベースについて、サイトのページ変更された部分が検知される。それをうけて、我々の担当者が外部サイトを目視で確認し、データ(ソフトウェア)自体が変わったか、あるいは機能が大幅に追加されたかの2つをチェックした。そのどちらかが目視で認められた場合に、

便覧を更新してニュース事項としてトップページに載せ、また RSS (RDF Site Summary) を使って広報した。RSS は検知ツールから自動で生成しており図 2. 1-3 のようなデータが生成される。これを MEDALS におき、ユーザー(RSS リーダー)の要求に応じて配信した。ユーザーのブラウザのような RSS リーダーでこの情報をみると、図 2. 1-4 のように何が更新されたのかを簡略した形式で見ることができる。

RSS の更新は、平成 23 年 6 月から平成 24 年 3 月までの 10 ヶ月にわたり、合計 60 件の更新をした。この内訳としては、MEDALS サイト(便覧等)の更新:20 件、便覧に書かれた外部データベースサイトの更新:26 件、イベント等ニュース:14 件となっている。

```
-  
<rss version="2.0">  
-  
<channel>  
<title>MEDALS</title>  
<link>http://medals.jp/</link>  
<docs>http://blogs.law.harvard.edu/tech/rss</docs>  
<language>ja</language>  
<pubDate>Feb, 06 Dec 2010 14:53:11 +0900</pubDate>  
<item>  
<title>[外部サイト更新][成果物]Evola</title>  
<link>http://medals.jp/list/detail/1</link>  
<guid isPermaLink="true">http://medals.jp/list/detail/1</guid>  
<pubDate>Feb, 03 Dec 2010 06:03:34 +0900</pubDate>  
</item>
```

図 2. 1-3 自動作成された RSS



図 2. 1-4 RSS 情報をブラウザで表示した図

(b) 生物種の表現の検討

MEDALS のデータベース便覧には、生物種という項目があり、データベースがどの生物種についてのデータを格納しているかを記載している。例えば、DOGAN(酵母、有用細菌、極限環境微生物のゲノムデータベース)では、「生物種」の項目に「微生物ー古細菌、バクテリア、コウジカビ」と記述をし、「キーワード」の項目に、個別に種の情報(この場合は学名)を記述していた(図 2. 1-5)。実はこの記載にはマイナーな問題があった。記載された種名が、生物種分類の著名なデータベースである NCBI Taxonomy データベースでどのように記述されているかを迅速に参照できないことである。このため 記載種名の NCBI Taxonomy データベースへのリンクあるいは Taxonomy ID の表示が望まれていた。また、MEDALS 便覧に格納される成果物には、産業への応用を照準としているものが少なくない。従って、学名だけではなく、用途や生息環境といった実業に近い観点に基づく生物種分類の表示が望まれていた。また、種名の各バリエーションを全て「生物種」の欄に記述すると、冗長性が生じたり、学名以外の名称がない種の場合に表記できないという不整合が生じる。そこで、検討の結果、以下の便覧記載ルールを設けることとした。

1. 便覧の「生物種」にはまず学名を入れ、さらに対応する和名・英名を入れる
2. さらに「キーワード」に、日本語・英語の「人為分類」を入れる(例えば好熱古細菌、抗生物質生産菌)

この結果、新しい便覧では、初見では図 2. 1-6 のように表示されて専門用語と和名などがわかる。ここで more ボタンをクリックすると、図 2. 1-7 が表示されて、Taxonomy ID やリンクなどの生物種情報が表示されることになった。これらは Hide ボタンで図 2. 1-6 にもどることも可能である。また、全文検索に対しても多くのキーワードにヒットすることが可能となる。この方式を DOGAN 以外のデータベースの便覧にも適応し、比較的多くの生物種を含む 2 件のデータベース(Evola、SEVENS)に対しても同様の対策を施した。

キーワード	Aspergillus oryzae Aeropyrum pernix Pyrococcus horikoshii Sulfolobus tokodaii Brevibacillus brevis Corynebacterium efficiens Desulfovibrio magneticus Gemmatimonas aurantiaca Kocuria rhizophila Rhodococcus erythropolis Rhodococcus opacus Staphylococcus aureus Staphylococcus aureus Staphylococcus haemolyticus Streptomyces avermitilis
ダウンロードデータ総量(Mbyte) データ一括取得方法	80 http://www.bionite.go.jp/dogan/DataDownload
使っている外部リソース	PubMed UniProt InterPro KEGG
主な対象データ	DNA-配列、プロテオーム
生物種	微生物-古細菌、バクテリア、コウジカビ
利用条件	なし
データ更新頻度(過去2年間)	4
最終更新日(調査日)	2010/11/26 (2010/12/06)
利用できるID	ORF_ID NF_ID
IDを使った成果物の利用方法	http://www.bionite.go.jp/dogan/ProteinInformation?ORF_ID=[ORFID]

図 2. 1-5 改良前の「生物種」(下の赤い枠)の表示画面

キーワード	Aspergillus oryzae RIB40 (黒麹菌) Saccharomyces cerevisiae Kyokai no. 7 (清酒酵母) Aeropyrum pernix K1 (好熱古細菌) ... 
ダウンロードデータ総量(Mbyte) データ一括取得方法	80 http://www.bionite.go.jp/dogan/DataDownload
使っている外部リソース	PubMed UniProt InterPro KEGG
主な対象データ	DNA-配列、プロテオーム
生物種	Aspergillus oryzae RIB40 Saccharomyces cerevisiae Kyokai no. 7 Aeropyrum pernix K1 ... 

図 2. 1-6 改良後の DOGAN データベースの便覧(初回アクセスした場合)



図 2. 1-7 改良後の DOGAN データベースの便覧(情報を展開した場合)

More ボタンを押すことにより、生物種を全件閲覧できる。図の Taxonomy ID(数字)には、NCBI Taxonomy のデータベースへのリンクが張られている。学名以外の和名や英名がある場合には同時に参照できる。

(c) ポータルサイト運用

ポータルサイトの運営については、情報セキュリティ上問題がないように運営し、アクセス情報の管理、利用状況や利用方法の分析を行った。

ポータルサイトへのアクセス数の集計については、MEDALS サーバーに Google Analytics の仕組みを設定し、集計した。アクセス(ページビュー)数は、平成 23 年 4 月から平成 24 年 2 月までの集計で 13,411 件であった。

(d) 便覧のアクセス数

表 2. 1-3 に、便覧のアクセス数トップ 20 を示す。但し、便覧以外のページ(トップページ、リストページなどのコンテンツ)は除いた統計である。期間は平成 23 年 4 月から平成 24 年 2 月である。

表 2. 1-3 便覧のアクセス数トップ 20

ランク	ページ タイトル	ページビュー数
1	Tool PubMedScan	275
2	Tool PMID-Extractor	231
3	DB 正常臓器 遺伝子発現データベース	183
4	Tool CentroidFold	108
5	DB ラットの脳断面図	107
6	Tool ID一括変換システム	99
7	DB ATTED II(植物遺伝子共発現)	78
8	Tool リンク自動管理システム	77
9	DB PPI view(蛋白質間相互作用)	76
10	DB ARCHAIC(古細菌)	64
11	Tool myPresto(分子シミュレーション)	62
12	DB CellMontage(遺伝子発現)	47
13	DB 微細藻類遺伝子工学データベース	47
14	DB MS-MS Fragment Viewer	46
15	Tool MAFFT(配列アラインメント)	42
16	DB HGPS(ヒト遺伝子蛋白)	40
17	Tool HEAT(遺伝子セット解析)	40
18	DB CELLPEDIA(細胞)	38
19	Tool PRRN(配列マルチプルアラインメント)	35
20	DB KNApSack Family	34

赤字はデータベース。黒字はツール(ソフトウェア)。便覧以外のページ(トップページ、リストページなどのコンテンツ)は除く。

(e) PubMedScan の運用

PubMedScan は PubMed 文献データベースに登録された文献データのうち、ユーザー毎の興味に応じた新規文献をアラートするシステムで、MEDALS のサブシステムである。アラートする文献対象の範囲を、キーワードで指定するのではなく、登録した文献から似たものを探すのが特徴である。平成 23 年度は、ユーザー数が 28%増加した。これは、2. 4. 2 節広報普及活動に記載して

いる講習会や様々な場面で紹介しているので、広報活動の成果の 1 つと考えられる。また時折、ユーザーからの登録データの修正の要望があったので、その都度対応した。

(f) myPresto の更新

myPresto は、医薬品開発支援のために作成された分子シミュレーション計算のプログラム集である。2008 年から MEDALS のサイトからダウンロードできる。便覧の URL は <http://medals.jp/list/detail/61> である。他にも一般社団法人バイオ産業情報化コンソーシアム(JBIC)、大阪大学からでもダウンロードできるが、それらのサイト URL はいずれも MEDALS に掲載している。今年度は 2 回のバージョンアップがあったので、MEDALS においてもデータの同期・更新を行った。

2.1.1-(2) ID を用いたデータ統合:「リンク自動管理システム」の開発

産総研・バイオメディカル情報研究センターでは、持続可能なデータベース統合の方式を模索すべく、データベース間のリンクを自動で更新するための「リンク自動管理システム」の開発を進めている(<http://biodb.jp/>)。これは、多数の公開データベースからデータ管理用 ID を集めてその対応関係を自動で解析し、データベース間のリンクを常に最新かつ網羅的に提供するシステムである。また、データ ID の一括変換サービスや、データ ID 対応表のダウンロードなどのサービスも提供している。本プロジェクトにおいては、主に MEDALS データベース便覧に掲載されている経産省関連のデータベースをリンク自動管理システムに対応させ、データ ID によるデータベースの統合的な利用ができるしくみを整備することを目的とした。

① 新規データベース追加

今年度の成果として、MEDALS データベース便覧を参考にしつつ、新たなデータベースをリンク自動管理システムに追加した。まず、MEDALS データベース便覧からヒト等の分子情報を扱うデータベースを選別し、その中から DNA 塩基配列の Accession Number や、UniProt の Accession Number、H-InvDB の HIT や HIX などのデータ ID を用いているものを選び出した。そして、対象データベースごとにデータ ID 一覧表の自動取得プログラムを作成して、リンク自動管理システムに組み込んだ。MEDALS 便覧にあるデータベースのうち追加されたデータベースは、以下の通りである。

- ・ 日本人のサンプルを用いて検出されたゲノム多型情報を集めた JSNP database (NEDO プロジェクトによる日本人 786 人の遺伝子頻度データを含む)
- ・ 生物種-代謝物データベースの KNApSack core system (薬用植物データベース、自然活性、metabolite 活性、漢方薬、インドネシアのハーブ、食用データベース、ハーブデータベースを含む、NEDO プロジェクトの成果が含まれる)

このほか、ヒト分子情報を扱う有用なデータベースと、創薬研究のための実験動物として有用なラットの分子情報に関するデータベースとして、以下を追加した。

- ・ ヒトのタンパク質に関する情報が蓄積されている Wikipedia human proteins
- ・ マウス遺伝子の公式名称である Mouse gene symbol
- ・ マウスに関し、タンパク質立体構造の PDB、突然変異を持つマウス系統の情報を集めた BRC mouse strain、MGI の遺伝子変異情報である MGI allele
- ・ ラットに関し、Rat Genome Database、NCBI の Entrez Gene、Nucleotide、GenBank、PubMed、Ensembl の Gene および Transcript、KEGG の Gene および Pathway、UniProt、Evola

以上の開発により、リンク自動管理システムが対応するデータベースの種類はヒトの分子情報は 35 種類、マウスは 16 種類、ラットは 10 種類、ホヤは 11 種類、化合物は 12 種類となった。そして、取り扱うデータ ID の種類は 69 種類、件数は約 1 億 3000 万件となり、世界最大級のデータ ID の統合データベースとなった。

② 新規追加機能:新着データのハイライトおよびメール配信

今年度は、リンク自動管理システムに二つの便利な機能を新たに開発して公開した。ひとつは、全データベース検索に追加した「新着データのハイライト機能」である(図 2. 1-8)。これは、リンク自動管理システム的全データベース検索において、前週の月曜日以降に新たにリンクがつけられたデータに“New”のマークを表示したものである。これにより、研究者が最も興味を持つ新着データを容易に発見できるようになり、非常に便利になった。本機能は 2011 年 12 月に公開した。

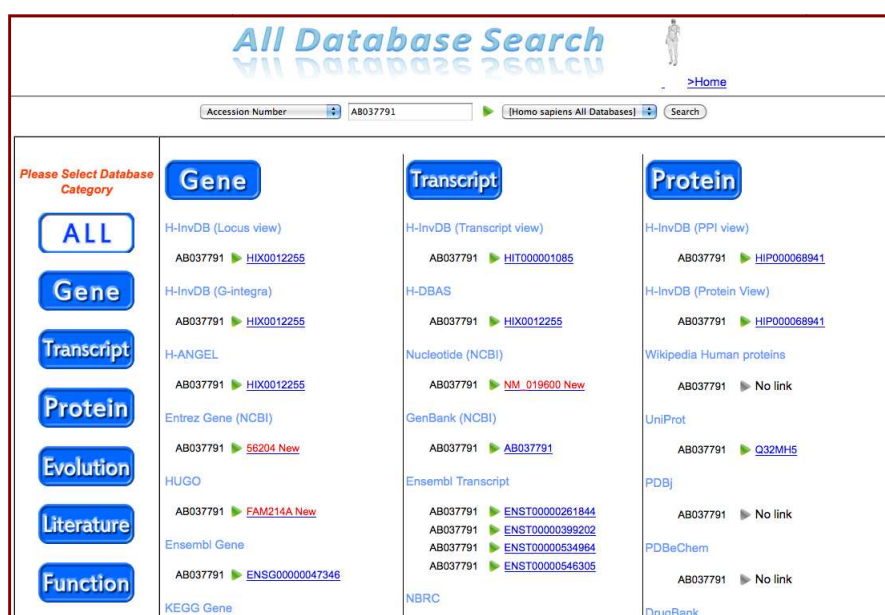


図 2. 1-8 リンク自動管理システムの新着データハイライト機能

新着データは赤字の“New”が点滅して表示される。

もう一つの新機能は、新着データお知らせツール BioDBScan である。BioDBScan は、上記の

「新着データのハイライト機能」をウェブサービスで利用し、利用者があらかじめ登録しておいたデータ ID に対して新たにリンクされたデータの情報を毎週一度電子メールで配信する画期的なサービスである。世界に分散しているさまざまなデータベースから、利用者が興味を持っている分子等に関連する新着データの情報を容易に取得できるため、非常に便利である。本機能は PubMedScan の新機能として、平成 24 年 3 月に公開した。

BioDBScan の使い方は、まず PubMedScan & BioDBScan のトップページ (<http://medals.jp/pubmedscan/>) にアクセスし、[New Registration] でアカウントを作る。ログイン後に、上部にある“move to BioDBScan”のアイコンをクリックし、BioDBScan の登録画面に移動する(図 2. 1-9)。

The screenshot shows the BioDBScan registration interface. It includes a header with the BioDBScan logo and a message '5 record(s) was deleted.' Below this is a section titled '[What is BioDBScan?]' which describes the service as a free web-based application that alerts users to new data releases. The main section is '[Register IDs of your interest]', which contains a form for selecting a taxonomy type (Homo sapiens, Mus musculus) and an ID type (HUGO gene symbol, KEGG Pathway ID, OMIM ID, Mouse Gene Symbol). There is a text box for entering IDs and an 'Add IDs' button. Below the form is a table titled '[Registered IDs]' showing the list of registered IDs. At the bottom, there are buttons for 'Delete checked IDs' and 'Send an E-mail report now', along with contact information for BioDBScan and MEDALS.

検索クエリIDの登録

検索クエリIDの一覧

検索の試行

Taxonomy type	ID type	ID	Registered date	
☐ Homo sapiens	HUGO gene symbol	PPARG	2012-02-14	Edit
☐ Homo sapiens	KEGG Pathway ID	hsa04210	2012-02-15	Edit
☐ Mus musculus	Mouse Gene Symbol	Glis1	2012-02-15	Edit

図 2. 1-9 BioDBScan の画面構成

次に、テキストボックスの中に利用者が興味を持っている任意の分子の ID を入力し、[Add IDs] のボタンを押す。例えば以下のような ID の登録ができる(表 2. 1-4)。ライフサイエンス分野の(特にウェットの)研究者には、特定の分子に関するあらゆる情報を迅速に集めたいというニーズがあり、本システムはこれに十分に答えることができる。

表 2. 1-4 ID の指定例

Taxonomy type	ID type	ID (ID の意味)
Homo sapiens	HUGO gene symbol	PPARG (ヒトの遺伝子名)
Homo sapiens	KEGG Pathway ID	hsa04210 (ヒトの apoptosis パスウェイ)
Homo sapiens	OMIM ID	180300 (OMIM の関節リウマチ)
Mus musculus	Mouse gene symbol	Glis1 (マウスの遺伝子名)

2. 1. 2 統合データベースセンターとの連携

経産省関連のデータベースについてのデータベース便覧(カタログ)を公開しているが、NBDCへ情報を提供し、合同のポータルサイトで統合されたカタログを公開した。ここから各サイトへリンクが張られている。

NBDC とはデータベースのカタログ作成活動の他、横断検索の構築運用やアーカイブ活動も連携している。これら連携活動をいっそう規模を広げて進めるため、および省庁連携の活動を広報するために、合同のポータルサイト(integbio)を立ち上げて 2011 年 12 月に公開した。図 2. 1-10 に integbio のトップページを示す。URL は <http://integbio.jp/> である。これは、科学技術振興機構(JST)、医薬基盤研究所(NIBIO)、農業生物資源研究所(NIAS)、および産業技術総合研究所(AIST)の合同の成果である(2. 4. 2 節 広報・普及活動参照)。このページでは、連携の活動や成果を説明するテキスト情報以外に、カタログ、横断検索、アーカイブの 3 つ(図 2. 1-10、右上)を中心に紹介している。MEDALS のデータベース便覧については 10 月時点の MEDALS 上の便覧データを NBDC へ情報提供した。これは、他のデータベース(ポータルサイト)のカタログとともに統合・再編集されクリエイティブコモンズのライセンスのうち、著作権がほぼフリーなもの(CC0)で公開された(図 2. 1-11)。



図 2. 1-10 省庁連携の活動を紹介する integbio のトップページ

右上の「データベースを探す(カタログ)」は、カタログ統合ページへのリンクである。左下に MEDALS ポータルの紹介がある。



図 2. 2-1 MEDALS トップページでの横断検索ワードの入力箇所

2つの赤い枠から横断検索のページに遷移または検索ができる。



図 2.2-2 横断検索のページ

検索キーワード「プロテアーゼ」を入力すると、右下のような検索結果が表示される。

横断検索の検索インデックスの作成、ノイズのクリーニング等の作業においては、まず有用性やデータベースの選定を行い、以下のデータベースを候補とした。

クローリングと ID リストの調査で検索インデックス作成が可能なもの (7 件)

1. H-Exp (ヒト遺伝子発現)
2. PPI view (H-InvDB、ヒト蛋白質間相互作用)
3. 微細藻類遺伝子工学データベース (遺伝子組み換え関連情報)
4. TraP (古細菌代謝経路)
5. ConfC (蛋白質立体構造動的情報)
6. ARCHAIC (古細菌の DNA、遺伝子)
7. Arch GeNet (古細菌の遺伝子蛋白発現)

更に個別なクリーニング処理が要るもの (5 件)

1. KNApSACk (植物 2 次代謝産物)

2. KNApSAcK Family（上記を種、漢方等に拡張）
3. H-ANGEL（H-InvDB のうち、ヒト遺伝子発現に関するデータ）
4. ATTED II（シロイヌナズナ遺伝子発現）
5. DNAProbeLocator（DNA プローブの ID と遺伝子の対応関係）

本プロジェクトでは横断検索に追加する対象データベースの件数として、5 件以上が目標であったが、上記の 12 件が実行可能であると見積もり、3 月までに横断検索対象として追加した。

運用でも NBDC と連携を行った。迅速かつ正確な連絡をとることを目的に、平成 23 年 8 月に横断検索連携専用のメーリングリストを新設した。これにより、停止予定やインデックスデータ更新等の連絡を年度内で 48 回の連絡を行った。

2.3 アーカイブ構築連携

各省統一形式でのデータダウンロードを可能とするため、NBDC で提供しているアーカイブ作成ガイドラインに沿ってデータベースのアーカイブ化を実施した（ガイドライン URL は http://dbarchive.biosciencedbc.jp/files/nbdc_dbarchive_guidelines.pdf）。全体の作業は NBDC と作業内容や進捗について協議を行い、NBDC の担当者と作業を分担しつつ進めた。

図 2.3-1 は、MEDALS のトップページにおけるダウンロードページへのリンク位置を示す。MEDALS からダウンロードできる成果物は、データ、ツール（ソフトウェア）、成果報告書となっている。H23 年度からは NBDC においてアーカイブページが出来てサービスが始まっているため、データベースについては NBDC から提供するようにし、その他のツールや報告書については MEDALS から提供するようにした。



図 2. 3-1 MEDALS トップページでのアーカイブへの入り口

ガイドラインにあるアーカイブの作業内容を簡略化して示すと図 2. 3-2 のようになる。開発者に対して、アーカイブの意義、作業手順、公開・利用条件などの説明を行い、その後にデータを統一的に扱えるように整形し(紫部分)、非専門分野の研究者でもなるべく内容がわかるような説明を検討して説明文を記述する(緑部分)。最後に公開形態やデータ自体を開発者に確認いただき、承諾書面をいただいてから公開した。

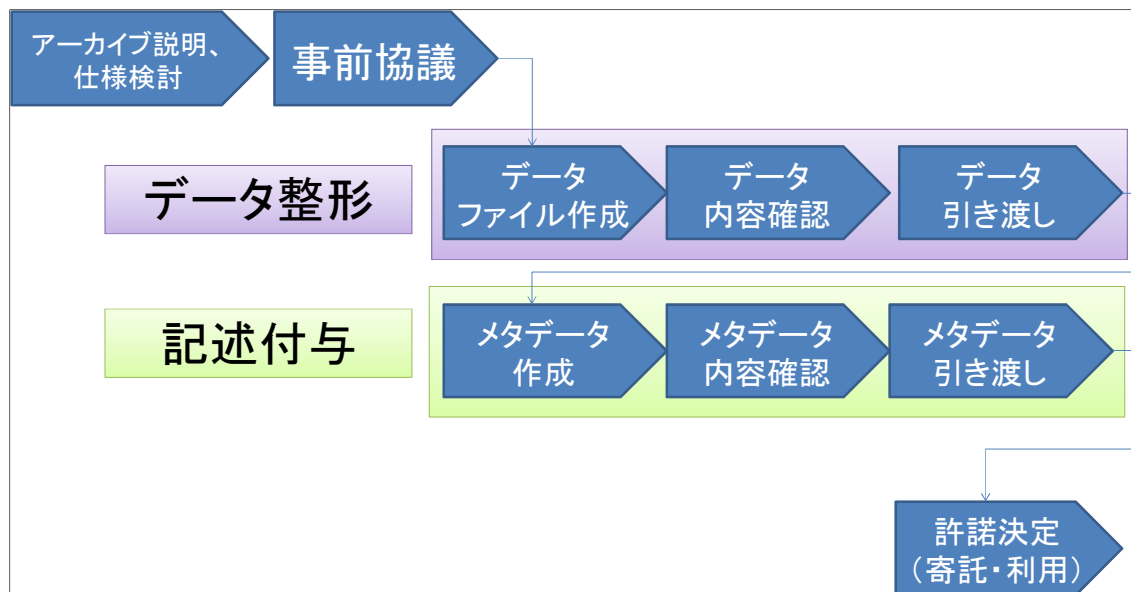


図 2. 3-2 アーカイブガイドラインによるデータアーカイブ化の流れ

アーカイブする対象のデータベースの選定については、MEDALS 便覧に記載があるものから、いくつかの選定基準をもって選定した。選定基準としては、1) 独自性が高く活用促進が望まれるもの、2) 新規なもの、3) 網羅性が高く、データをセットとして利用すると著しく効果が高いもの（例えばゲノムを対象としているもの）、4) 担当者が異動したなど、提供の安定化をいそぐもの、という4つを設けていずれかに該当するものを選定した。その結果、NITE 酵母画像（独自性が高い）、KNAPSAcK core（独自性が高い）、GDBS（ゲノムベースのデータ）、GenoBase（ゲノムベースのデータ）、および INOH（平成 22 年度に「オントロジーによるパスウェイの高度化および国際標準化」プロジェクトが終了して新規。パスウェイデータは注目度が高い）の5件のデータベースについてアーカイブ化を行うこととした（上記の括弧内は選定の主な理由）。これらについて作業が完了したデータは NBDC に納めた。3月14日時点ではこのうち一部が未公開であるが、NBDC の確認作業が完了しだい、NBDC から公開されることになっている。図 2. 3-3 は、公開されている INOH の画面である。このような画面でデータの内容説明およびデータが提供された。

図 2. 3-3 NBDC のサイトで公開したアーカイブの例 (INOH データベース)

図 2. 3-3 NBDC のサイトで公開したアーカイブの例 (INOH データベース)
左はデータ項目の説明、右は個別エントリと属性値の一覧を示す。

2. 4 国内外の最新動向調査及び普及に向けた取り組み

本年度の国内外の最新動向調査及び普及に向けた取り組みとして、国内外の最新動向調査、広報・普及活動、ユーザーニーズ調査の3項目を実施した内容を以下に報告する。

2.4.1 国内外の最新動向調査

ライフサイエンス分野のデータベース統合に関する知見を、本事業に活かすため、関連学会等に参加や研究会等に参加し、国内外の動向について調査した。また、関連する技術や研究内容の把握のための資料収集を行った。具体的には、国際学会に4人回、国内学会等に30人回参加した。表 2.4-1 および表 2.4-2 参加した国内学会等に、参加した国際会議と国内会議をそれぞれ記した。

表 2.4-1 参加した国際学会等

日付	会議名	開催場所
2011/7/17-19	ISMB-ECCB2011	ウィーン
2011/7/26-31	SMBE2011/日本進化学会	京都
2011/6/29-7/2	第10回 国際バイオ EXPO	東京
2011/8/21-26	BioHackathon2011(BH11)	京都
2011/10/5-7	BioJapan 2011	横浜
2011/11/8-10	CBI/JSBi 2011 合同大会	神戸

表 2.4-2 参加した国内学会等

日付	会議名	開催場所
2011/5/28	圧縮センシングとその周辺	東京
2011/10/5	トーゴの日シンポジウム 2011～ライフサイエンス分野のデータベース統合の”カタチ”を探る～	東京
2011/10/27	ゲノムテクノロジー委員会シンポジウム	東京
2011/11/21-25	統合データベース技術情報交換ワークショップ (BH11.11)	修善寺
2011/11/4	WebDB Forum 2011	東京
2011/12/13-16	第34回日本分子生物学会年会(MSBJ2011)	横浜
2012/1/31-2/1	平成23年度第11回 産総研・産技連 LS-BT 合同研究発表会	つくば
2011/1/26	オミックス医療研究会交流フォーラム	東京
2012/3/8	セマンティック Web コンファレンス 2012	東京

調査事例① ISMB-ECCB2011

- ・ ELIXIR というヨーロッパでの統合データベースプロジェクトがある。これは 2008 年に開始した。3 年の準備段階で 6.8 億円。集中・分散、機密保持などでクラウド化を議論している。

- ・ データベースの世界共通なカタログを作成維持することによって、情報交換や高度な解析をしやすくするために、データベースの説明データ(メタデータ)の項目や形式をコミュニティベースで決定しようとする「BioDBCore」という活動がある。連携するメンバーは、拡大している模様である。

調査事例② CBI/JSBi 2011 合同大会

- ・ データを RDF 化して公開する DB が行われている(例:PDBj)。
- ・ 理研スーパーコンピュータ「京」を用いた進展が期待されている。
- ・ 統合データベースによる疾患関連遺伝子の絞り込みが行われるようになった(例: TargetMine)。
- ・ 次世代シーケンシングのマッピングの高度なアルゴリズムを実装するのに、学生実習で 200 行ほどのプログラムで可能である。
- ・ Aberystwyth 大学(英)は、機械学習により実験すべき仮説を生成、選択して、その後はロボットを使って Wet な実験を行い、仮説を検証するシステムを開発した。

調査事例③ BioHackathon (BH11.11)

- ・ UniProtをはじめ、バイオ系のデータベースだけでも、約 20 個のデータベースやツールがデータの RDF 化に対応しており利用されている。
- ・ オントロジーの作成にあたっては、BioOntology から提供しているオントロジーが標準的であるので、これをたたき台として利用するのが適当である。

調査事例④セマンティック Web コンファレンス 2012

- ・ 省庁が公開する情報の再利用しやすい方向で工夫がされている。CSV 形式などで、再利用促進の必要性が語られた。
- ・ 電子カルテでの利用にむけ臓器部位など医学オントロジーを開発している。
- ・ MEDALS を含めたバイオデータベース統合活動について、川本 祥子 副センター長(ライフサイエンス統合データベースセンター)が招待講演し、活動を紹介した。
- ・ 複数の発表で、CSV データを入れると簡易的に RDF 化して Linked open data (LOD) 対応サーバーに載せるシステムが紹介されており、LOD への基盤技術やツールが日本でも整いつつあることがわかった。

2.4.2 広報・普及活動

本事業で構築した成果物の利用普及のため、2.4.3-(1)から 2.4.3-(3)で調査を実施した学会やイベント等において、ポスター発表、口頭発表等の機会を利用し、10 回広報・普及活動を行った。

2.4.2-(1) パンフレット作成

図 2.4.2-1 に示したパンフレットを増刷して、学会や発表会において配布し、広報に役立てた。MEDALS について、刷新した主な箇所は「MEDALS 横断検索」の追加、便覧件数の更新である。また、リンク自動管理システムを紹介するパンフレットを新規に作成した。



図 2.4-1 今年度に刷新し広報に用いたパンフレット(A4 サイズ)

2.4.2-(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 およびヒト遺伝子統合データベース H-InvDB に関連するポスター発表・口頭発表を行い、その機会を利用してデータベースの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。第 10 回 国際バイオ EXPO、BioJapan 2011、第 34 回日本分子生物学会ではブースを出展し、バイオ関連企業も対象とした広報活動を行った(BioJapan 2011 では産総研連携ブース、第 34 回日本分子生物学会では 4 省連携ブースに出展)。全体では、国際学会発表 3 件、国内発表 10 件を行った。例として、平成 23 年 12 月 13 日～16 日、横浜において開催された第 34 回日本分子生物学会(MBSJ2011)でのポスター・ワークショップ発表、ブースの様子を示す(図 2.4-2 学会発表風景)。

第 34 回日本分子生物学会の 2 日目(12 月 14 日)には、村上研究員とライフサイエンス統合データベースセンターの川本氏をオーガナイザーとするワークショップ「役に立つデータベース、役に立たないデータベース」を開催した。このワークショップでは、膨大なデータを産出できるようになり多くのデータベースが林立するようになったライフサイエンスの現状を踏まえ、データベースの利用や開発に関わる新たな課題を発掘し、次世代のデータベースのあり方を議論した。ワークショップには約 150 人が参加するとともに、ワークショップの様子を Twitter で配信するという先端的な実験要素も含まれた。

ポスター発表(12/14)



4省連携ブースでのデモ



図 2. 4-2 学会発表風景

第 34 回日本分子生物学会(MBSJ2011、平成 23 年 12 月 13 日～16 日)においてポスター・ワークショップ発表、ブースでのデモを実施した。

2.4.2-(3) 講習会開催

MEDALS および H-InvDB に関する講習会を全国各地で計 4 回開催し、内容の講義・デモンストレーションおよび実習を実施した(表 2. 4-3)。各回には、国内の様々なデータベースの開発者を講師として招聘した。各回のプログラムを以下に示す。

平成 23 年度データベース講習会「創業における統合データベースの活用」

第 1 回(平成 23 年 5 月 19 日、東京)

1. 創薬研究における統合データベース
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 今西 規
2. ヒト遺伝子統合データベース H-InvDB の活用
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 山崎 千里
3. 遺伝子発現データベース H-ANGEL、H-EXP の活用
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 野田 彰子
4. ヒト選択的スプライシングデータベース H-DBAS の活用
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
5. 統合データベースポータルサイト MEDALS の活用
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
6. ヒトプロテオーム研究における HGPD の活用
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 丸山 友希夫、五島直樹
7. ヒト細胞・細胞分化データベース CELLPEDIA/CellMontage の活用

産業技術総合研究所生命情報工学研究センター 千葉 啓和、幡野 晶子、藤
渕 航

8. 理研サイネス(SciNetS)統合データベースと PosMed の活用
理化学研究所生命情報基盤部門 蒔田 由布子

第2回(平成24年1月20日、大阪)

1. 創薬研究における統合データベースの現状とヒト遺伝子統合データベース H-InvDB
の活用
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 今西 規
2. 比較ゲノムデータベース Evola、G-compass の活用
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 原 雄一郎
3. 創薬にむけた KEGG パスウェイ解析
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 五斗 進
4. 医薬基盤研究所のデータベースと創薬研究: Open TG-GATEs と TargetMine を中心
として
医薬基盤研究所 水口 賢司
5. 統合データベースポータルサイト MEDALS の活用
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

第3回(平成24年3月9日、つくば)

1. ヒト遺伝子統合データベース H-InvDB の活用
産総研・バイオメディシナル情報研究センター 山崎千里
2. ヒト選択的スプライシングデータベース H-DBAS の活用
産総研・バイオメディシナル情報研究センター 武田淳一
3. 統合データベースポータルサイト MEDALS の活用
産総研・バイオメディシナル情報研究センター 村上勝彦
4. ヒト疾患と対応したマウス表現型リソースデータベースの活用
理化学研究所 バイオリソースセンター 榎屋啓志
5. 創薬にむけた化合物データベース myPresto の活用
HPC システムズ、産総研・バイオメディシナル情報研究センター 佐藤大介、福西
快文

第4回(平成24年3月16日、東京)

1. 創薬研究における統合データベース
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 今西規
2. ヒト遺伝子統合データベース H-InvDB の活用

- 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 山崎千里
3. 統合データベースポータルサイト MEDALS の活用
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
 4. ヒトタンパク質間相互作用の統合データベース、PPI view の活用
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 世良実穂
 5. 相互作用ネットワーク・パスウェイがさくさく書けるフリーツールの紹介
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 長井陽子
 6. ヒトタンパク質の立体構造予測データベース、SAHG の活用
産業技術総合研究所 生命情報工学研究センター 本野千恵
 7. 疾患解析における Human Variation DB の活用
日立製作所 小池麻子

外部からの講師には、MEDALS 成果物の作成者だけでなく、4 省連携に関連するデータベースの開発者も招聘し、データベースの統合システムだけではなくライフサイエンス研究者への啓蒙においても連携を試みた。平成 24 年 1 月 20 日に大阪（産総研・関西センター）で開催したデータベース講習会の開催風景を図 2. 4-3 に示す。

表 2. 4-3 今年度で開催した講習会

日付	講習会名	開催場所	参加人数
平成 23 年 5 月 19 日	第 1 回講習会	東京（産総研・臨海副都心センター）	18 名
平成 24 年 1 月 20 日	第 2 回講習会	大阪（産総研・関西センター）	18 名
平成 24 年 3 月 9 日	第 3 回講習会	つくば（産総研・つくば中央）	10 名
平成 24 年 3 月 16 日	第 4 回講習会	東京（産総研・臨海副都心センター）	16 名



図 2. 4-3 データベース講習会開催風景（平成 24 年 1 月 20 日、大阪）

2.4.2-(4) プレスリリースおよび新聞報道

今年度は、平成 23 年 12 月 12 日に文部科学省、厚生労働省、農林水産省、経済産業省が取り組む生命科学系データベースの統合化の方針や成果を紹介する合同ポータルサイト「integbio.jp」(インテグバイオ)を共同で開設した。これについて科学技術振興機構(JST)、医薬基盤研究所(NIBIO)、農業生物資源研究所(NIAS)、産業技術総合研究所(AIST)が共同で 1 件のプレスリリースを報道し、それぞれの関連 Web サイトへの掲載(9 件)およびメール配信(4 件)を行った(表 2. 4-4)。公開したプレスリリースの一部を表 2. 4-4 に示す。この結果、新聞誌面での一次報道は3件あった。更に、表 2. 4-4 に示したようにオンラインで 13 件の二次報道がみられた。また、メールマガジンでの配信や各種ニュースサイトへの引用掲載も多数行われた。

表 2. 4-4 integbio.jp の開設を報道したメディア一覧

色つき部分は、科学技術振興機構(JST)、医薬基盤研究所(NIBIO)、農業生物資源研究所(NIAS)、産業技術総合研究所(AIST)が配布したメディア。白部分は、それをうけた二次報道の掲載メディア。

No.	区分	内容
1	プレスリリース	AIST プレスリリース
2	プレスリリース	JST プレスリリース
3	プレスリリース	NIAS プレスリリース
4	サイト掲載	AIST 主な研究成果
5	サイト掲載	AIST ニュース
6	サイト掲載	JST ニュース
7	サイト掲載	NBDC ニュース
8	サイト掲載	NIBIO お知らせ
9	サイト掲載	NIAS ニュース
10	サイト掲載	BIRC 公式サイトニュース
11	サイト掲載	MEDALS ニュース
12	サイト掲載	H-InvDB 公式サイトニュース
13	メール配信	JST ニュースメールマガジン No.149
14	メール配信	農業生物資源研究所メールマガジン 53 号
15	メール配信	[産総研 Topics!] Vol.10 No.1
16	メール配信	[YAKUJINIPPO Mail News] 2011/12/15
17	新聞報道	日刊工業新聞 19 面
18	新聞報道	化学工業日報 9 面
19	新聞報道	薬事日報 2 面
20	新聞等サイト	日経プレスリリース

No.	区分	内容
21	新聞等サイト	薬事日報オンライン
22	サイト掲載	日経バイオテク online
23	サイト掲載	日経 BP 社 Pharma Business
24	サイト掲載	J-NET21:中小企業ビジネス支援サイト
25	サイト掲載	engineer 技術者向けのものづくり情報検索ポータル
26	サイト掲載	カレントアウェアネス・ポータル(国立国会図書館サイト)
27	サイト掲載	PR Times プレスリリース配信・掲載サービス
28	サイト掲載	行政の「今」がわかる更新統合サイト
29	サイト掲載	マイナビニュース
30	サイト掲載	BioMarket.jp
31	サイト掲載	ターゲットタンパク研究プログラム:タンパクニュースウォッチ
32	サイト掲載	YAHOO ニュース/テクノロジー






平成23年12月12日

科学技術振興機構 (JST)
Tel:03-5214-8404 (広報ポータル部)

医薬基盤研究所 (NIBIO)
Tel:072-641-9832 (戦略企画部)

農業生物資源研究所 (NIAS)
Tel:029-838-8469 (広報室)

産業技術総合研究所 (AIST)
Tel:029-862-6216 (広報部)

4省の生命科学系データベース合同ポータルサイトintegbio.jpを開設 ～オールジャパン体制へ～

科学技術振興機構(JST)、医薬基盤研究所(NIBIO)、農業生物資源研究所(NIAS)、産業技術総合研究所(AIST)は、内閣府総合科学技術会議の議論を受けて平成23年12月12日(月)に文部科学省、厚生労働省、農林水産省、経済産業省が取り組む生命科学系データベースの統合化の方針や成果を紹介する合同ポータルサイト「integbio.jp」(インテグバイオ)を共同で開設しました。

生命科学におけるデータベースは、例えばiPS細胞作製の成功には遺伝子データベースの情報が不可欠であったように、研究のインフラとして世界中で活発に開発、利用されています。その結果、現在では多数のデータベースが乱立しており、それを使いやすくするための統合化が重要な課題となっています。日本でも、生命科学に関係した各省で統合化に向けた取り組みが平成18年頃から始まりました。当初は各省で独自に進めていたデータベースの統合化をオールジャパンの体制で進めるために、内閣府の主導のもと各省の協力体制を構築して、平成23年度からは4つのステップ(カタログ、横断検索、アーカイブ、再構築)に沿って進めています。

図 2. 4-4 integbio.jp 開設に関するプレス発表

A4 で 2 ページの解説をプレスリリースとして、科学技術振興機構(JST)、医薬基盤研究所(NIBIO)、農業生物資源研究所(NIAS)、産業技術総合研究所(AIST)から配布した。

2.4.3 ユーザーニーズ調査

MEDALS ポータルサイト(および一般的バイオデータベース)のユーザーニーズを把握するため、Web アンケート、ヒアリング、およびモニターの3つの方式で調査をおこなった。

2.4.3-(1) アンケート調査

専用のアンケートサイトを設置し、ライフサイエンス分野データベース・MEDALS ポータルサイト・H-InvDB とその連携データベースを調査項目として Web アンケート調査を実施した。アンケート揭示期間は平成 23 年 6 月 27 日から平成 23 年 8 月 31 日までの約 2 カ月で、201 件の有効回答を収集した。

2.4.3-(2) ヒアリング調査

上記のアンケートよりも具体的かつ詳細なニーズを把握するため、4 件のユーザーグループにご協力いただいて対面でのヒアリング調査を行った。ヒアリング先は製薬企業(2 社)と研究機関(2 法人)である。

2.4.3-(3) モニター調査

MEDALS ポータルサイトのコンテンツを一定期間利用した、MEDALS の内容をよく理解したレベルのユーザーからのデータベースやポータルサイトのニーズを抽出するため、モニター調査を行った。モニターには以下の内容の作業をしていただいた。

- ・ ポータルサイト MEDALS でデータベース・ツール・プロジェクト便覧などを 1 週間利用
- ・ MEDALS 横断検索でキーワード検索を 1 週間利用
- ・ 新規文献お知らせツール PubMedScan へ登録して 1 週間利用
- ・ 10 問程度のアンケートに回答

モニター協力者は 57 名で、うち 16 名から終了時アンケートの回答を得た。

2.4.3-(4) 調査結果で判明した改善点のポータルサイトへの反映

上記 2.4.3-(1)から 2.4.3-(3)の調査で判明した問題点を検討し、そのうち(1)ポータルサイトの使い方の例を示して欲しい、(2)産業に役立つ情報を増加して欲しい、という要望について対策を講じた。

1 つ目については、“MEDALS Q and A”ページを設けて、様々な利用方法の解説を記述することとした(図 2. 4-5)。2 つ目の産業に役立つ情報掲載については、リンク集のページに国内・海外の特許サイトへのリンクを追加することにした(図 2. 4-6)。

12のよく質問される項目について、使用例なども含めた日本語のわかりやすい解説ページを公開



図 2. 4-5 新設した“MEDALS Q and A”ページ

特に要望があった
国内・海外の特許サイトへのリンクを追加

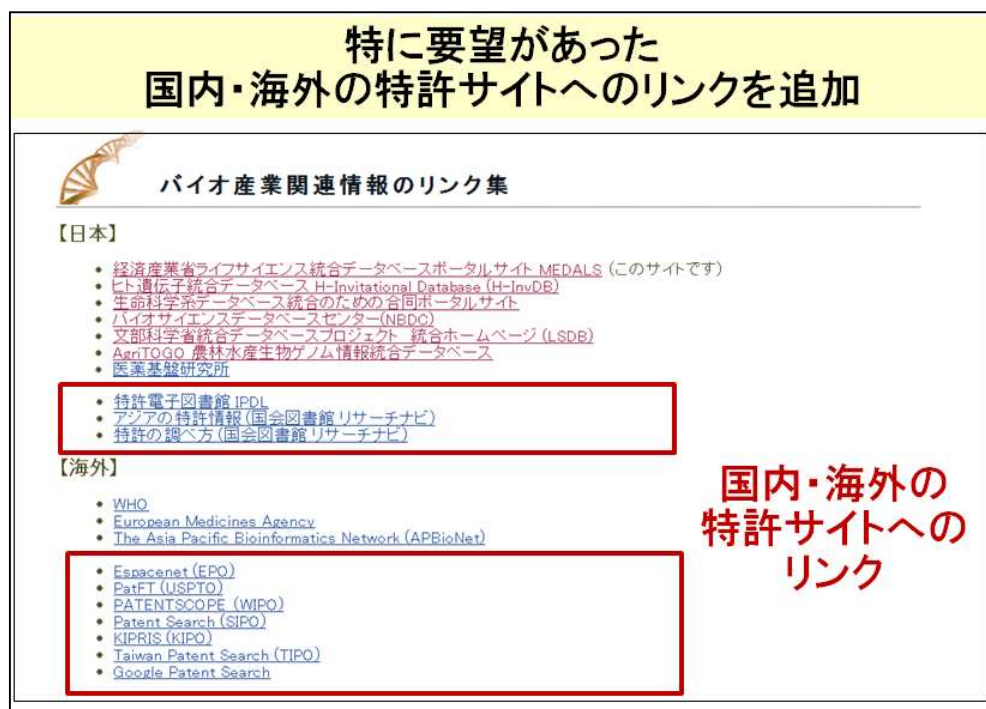


図 2. 4-6 “MEDALS リンク集”へのリンク

2.5 運営委員会

経済産業省ライフサイエンス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、平成 23 年 9 月 6 日と平成 24 年 2 月 8 日に計 2 回の運営委員会を産総研臨海副都心センターにて開催した。アドバイザー、研究協力者、ならびにバイオサイエンスデータベースセンターなどからのオブザーバーを含め、第 1 回、第 2 回運営委員会ともに 34 人が参加した。運営委員会では、プロジェクトの実施テーマごとに進捗報告とアドバイザー等からのコメントがあり、続いて総合討論を行った。ポータルサイト MEDALS の運営・開発、ならびに成果物の調査が順調に進んでいることが確認された。また、横断検索、アーカイブ作成、学会での広報といった NBDC との連携も滞りなく緊密に行われていることが確認された。リンク自動管理システムはアドバイザーから高い評価を受け、村上研究員が行う MEDALS 内、外部データベース間の生物学的なつながりを可視化する研究との統合が期待された。第 2 回運営委員会では議論の時間をこれまで以上に多く確保し、アドバイザー、オブザーバーと活発に議論した。便覧への掲載が困難だったプロジェクトの例から、成果を統合データベースに公開するための仕様設定やガイドライン、公開ポリシー、公開のための予算設定、データ提供側のインセンティブなど、経済産業省関連プロジェクトのデータ統合を滞りなく進めていくための議論が多面的に行われた。

2.6 成果発表

平成 23 年度の本プロジェクトの成果に関する発表を以下に示す。

2.6-(1) 論文

1. Toshiaki Katayama, Mark D Wilkinson, Rutger Vos, Takeshi Kawashima, Shuichi Kawashima, Mitsuteru Nakao, Yasunori Yamamoto, Hong-Woo Chun, Atsuko Yamaguchi, Shin Kawano, Chisato Yamasaki, Riu Yamashita, Noriyuki Satoh and Toshihisa Takagi (2011)「The 2nd DBCLS BioHackathon: interoperable bioinformatics Web services for integrated applications. Journal of Biomedical Semantics2:4 2011.8
2. Takayuki Taniya, Susumu Tanaka, Yumi Yamaguchi-Kabata, Hideki Hanaoka, Harutoshi Maekawa, Chisato Yamasaki C, Roberto Barrero, Boris Lenhard, Milton Datta, Mary Shimoyama, Imanishi T, and Gojobori T (2011) A prioritization analysis of disease association by data-mining of functional annotation of human genes. Genomics 99 (1):1-9. 2012.1

2.6-(2) 総説など

1. 村上勝彦、今西規(2011)「MEDALS:経済産業省ライフサイエンスデータベース・解析ツールのポータルサイト」実験医学 29(15):54-59. 2011.9
2. 山崎千里(2011)「BioDBCore における DB メタデータの RDF による共有」日本バイオインフォマティクス学会ニュースレター 23 号 4. 2011.10

2.6-(3) 発表

1. 今西 規(2011)「創薬研究における統合データベース」平成 23 年度第 1 回データベース講習会 @お台場「創薬研究における統合データベースの活用」2011.5
2. 山崎 千里(2011)「ヒト遺伝子統合データベース H-InvDB の活用」平成 23 年度第 1 回データベース講習会 @お台場「創薬研究における統合データベースの活用」2011.5
3. 小倉 彰子(2011)「遺伝子発現データベース H-ANGEL、H-EXP の活用」平成 23 年度第 1 回データベース講習会 @お台場「創薬研究における統合データベースの活用」2011.5
4. 武田 淳一(2011)「ヒト選択的スプライシングデータベース H-DBAS の活用」平成 23 年度第 1 回データベース講習会 @お台場「創薬研究における統合データベースの活用」2011.5
5. 村上 勝彦(2011)「統合データベースポータルサイト MEDALS の活用」平成 23 年度第 1 回データベース講習会 @お台場「創薬研究における統合データベースの活用」2011.5
6. 今西 規(2011)「ヒト遺伝子統合データベース H-InvDB を用いた発見的データマイニング」第 10 回国際バイオ EXPO 2011.6
7. Katsuhiko Murakami and Tadashi Imanishi(2011)「Correlation between annotation in the human gene database H-InvDB」ISMB/ECCB 2011. 2011.7
8. 村上 勝彦、松矢 明宏、間宮 健太郎、山崎 千里、原 雄一郎、今西 規(2011)「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:ポータルサイト MEDALS」トーゴの日シンポジウム 2011 2011.10
9. 今西 規(2011)「リンク自動管理システムと ID 一括変換システム」トーゴの日シンポジウム 2011 2011.10
 10. 山崎 千里、武田 淳一、羽原 拓哉、世良 実穂、小倉 彰子、村上 勝彦、今西 規、五條堀 孝(2011)「ヒト遺伝子の統合データベース H-InvDB と関連データベース・ツール群」トーゴの日シンポジウム 2011 2011.10
 11. 村上 勝彦、今西 規(2011)「Information Richness in Gene Annotation Revealed by Correlation Between Annotations in a Gene Database」CBI/JSBi2011 合同大会 2011.11
 12. 村上 勝彦、今西 規(2011)「The 7 Truths of Highly Effective Biological Databases」第 34 回日本分子生物学会年会 2011.12
 13. 村上 勝彦、山崎 千里、原 雄一郎、世良 実穂、今西 規(2011)「MEDALS : A Portal Site For METI Life Science Integrated Database Project.」第 34 回日本分子生物学会年会 2011.12
 14. 今西 規(2011)「Hyperlink Management System for Creating Maintenance-Free Hyperlinks among Major Biological Databases」第 34 回日本分子生物学会年会 2011.12
 15. 山崎 千里、武田 淳一、羽原 拓哉、世良 実穂、原 雄一郎、小倉 彰子、村上 勝彦、今西 規、五條堀 孝(2011)「ヒト遺伝子の統合データベース H-InvDB と関連データベース・ツール群」第 34 回日本分子生物学会年会 2011.12
 16. 今西 規(2012)「創薬研究における統合データベースの現状とヒト遺伝子統合データベース H-InvDB の活用」平成 23 年度第 2 回データベース講習会@大阪「創薬研究における統合データベースの活用」 2012.1
 17. 原 雄一郎(2012)「比較ゲノムデータベース Evola、G-compass の活用」平成 23 年度第 2 回データベース講習会@大阪「創薬研究における統合データベースの活用」 2012.1
 18. 村上 勝彦(2012)「統合データベースポータルサイト MEDALS の活用」平成 23 年度第 2 回データベース講習会@大阪「創薬研究における統合データベースの活用」 2012.1
 19. 今西 規(2012)「世界のバイオデータベースの統合化をめざしたリンク自動管理システム」平成 23 年度第 11 回産総研・産技連 LS-BT 合同研究発表会 2012.1
 20. 村上 勝彦、山崎 千里、原 雄一郎、松矢 明宏、間宮 健太郎、今西 規、五條堀 孝(2012)「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:ポータルサイト MEDALS」平成 23 年度第 11 回産総研・産技連 LS-BT 合同研究発表会 2012.1
 21. 山崎 千里、武田 淳一、羽原 拓哉、世良 実穂、原 雄一郎、小倉 彰子、村上 勝彦、今西 規、五條堀 孝(2012)「ヒト遺伝子統合データベース H-InvDB と関連データベース・ツール群」平成 23 年度第 11 回産総研・産技連 LS-BT 合同研究発表会 2012.1
 22. 山崎 千里(2012)「ヒト遺伝子統合データベース H-InvDB の活用」平成 23 年度第 3 回データベース講習会@つくば「創薬研究における統合データベースの活用」 2012.3
 23. 武田 淳一(2012)「ヒト選択的スプライシングデータベース H-DBAS の活用」平成 23 年度第 3

- 回データベース講習会@つくば「創薬研究における統合データベースの活用」 2012.3
24. 村上 勝彦(2012)「統合データベースポータルサイト MEDALS の活用」平成 23 年度第 3 回データベース講習会@つくば「創薬研究における統合データベースの活用」 2012.3
 25. 今西 規(2012)「創薬研究における統合データベース」平成 23 年度第 4 回データベース講習会「創薬研究における統合データベースの活用」 2012.3
 26. 山崎 千里(2012)「ヒト遺伝子統合データベース H-InvDB の活用」平成 23 年度第 4 回データベース講習会「創薬研究における統合データベースの活用」 2012.3
 27. 村上 勝彦(2012)「統合データベースポータルサイト MEDALS の活用」平成 23 年度第 4 回データベース講習会「創薬研究における統合データベースの活用」 2012.3
 28. 世良 実穂(2012)「ヒトタンパク質間相互作用の統合データベース、PPI view の活用」平成 23 年度第 4 回データベース講習会「創薬研究における統合データベースの活用」 2012.3
 29. 長井 陽子(2012)「相互作用ネットワーク・パスウェイがさくさく書けるフリーツールの紹介」平成 23 年度第 4 回データベース講習会「創薬研究における統合データベースの活用」 2012.3

2.6-(4) データベース公開・更新

平成 23 年度の本プロジェクトに関するデータベースの公開・更新は以下の通り実施された。この他に成果物便覧の更新は随時行われ、年間 34 回の更新作業によって延べ 68 の成果物が更新された。

【2011 年】

- | | |
|----------|--|
| 4 月 13 日 | リンク自動管理システム更新
“Ciona intestinalis”の正式公開を開始。対応データベースに Ghost、Aniseed を追加。 |
| 5 月 10 日 | リンク自動管理システム更新
対応 ID に Mouse Gene Symbol を追加。 |
| 6 月 26 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 更新
平成 22 年度産業技術研究開発「統合データベースプロジェクト」に関する成果報告書を統合データベースプロジェクトに関するプロジェクト便覧に追加
リンク集更新 |
| 8 月 1 日 | リンク自動管理システム更新
対応データベースに Wikipedia Human proteins を追加。 |
| 9 月 1 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 更新
リンク集更新 |
| 9 月 14 日 | リンク自動管理システム更新
対応データベースに JSNP DATABASE を追加。 |
| 10 月 4 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 更新 |

	便覧フォーマット仕様変更
10 月 26 日	リンク自動管理システム更新 対応データベースに Mouse PDB、BRC Mouse Strain、MGI Allele を追加。
12 月 8 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 更新 便覧「生物種」、「キーワード」項目の仕様変更
12 月 12 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 更新 4 省の生命科学系データベース合同ポータルサイト integbio.jp を開設
12 月 12 日	リンク自動管理システム更新 全データベース検索に「新着データのハイライト機能」を追加。

【2012 年】

1 月 19 日	4 件のデータベース (SokanProject、FuLoJa、Database of Genomic Variants、MCG CNV Database) を新規追加
1 月 26 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 更新 MEDALS Q and A 新規公開 4 件のデータベース (PSTAG、PHMMTS、FragmentAlign、SpiceHIT) および 1 件のプロジェクト (基礎研究から臨床研究への橋渡し促進技術開発 アルツハイマー病総合診断体系実用化プロジェクト) を新規追加
2 月 2 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 更新 リンク集更新
2 月 7 日	リンク自動管理システム更新 “Rattus norvegicus”の正式公開を開始。対応データベースに NCBI,Ensembl、Uniprot,RGD、Evola を追加。
3 月 9 日	リンク自動管理システム更新 対応データベースに KEGG を追加。(Rattus norvegicus)
3 月 31 日	BioDBScan 新規公開

2.6-(5) プレス発表・新聞記事

1. 「4 省の生命科学系データベースへの合同ポータルサイト integbio.jp を開設」プレス発表 2011.12.12
2. 「生命科学系 DB 幅広い領域で閲覧可能 JST など 4 研究機関ポータルサイト開設」記事掲載 化学工業日報 9 面 2011.12.13
3. 「生命科学系 DB 統合へ 合同サイト開設 JST など」記事掲載 日刊工業新聞 19 面 2011.12.13
4. 「4 省の生命科学 DB 合同ポータルサイト開設」記事掲載 薬事日報 2 面 2011.12.16

3. 謝辞

本プロジェクトの実施にあたり、多数の方のご協力、ご貢献を頂きました。ここに深く感謝申し上げます。

- 便覧・横断検索・アーカイブの作成にご協力いただいた各データベース・解析ツール開発者の方々
- アンケート、モニター調査にご協力いただいた方々
- ヒアリングにご協力いただいた企業、研究所の方々

4. 参加者名簿

4-(1) プロジェクト実施者

役職	氏名	所属機関
プロジェクトリーダー	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
プロジェクト顧問	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
研究員	村上 勝彦	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
	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
	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
	長井 陽子	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
	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
	原 雄一郎	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
	世良 実穂	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
	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
	小尾 信男	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

4-(2) 運営委員(アドバイザー)

油谷 浩幸	東京大学先端科学技術研究センター 教授
長洲 毅志	エーザイ株式会社 理事・CSO 付 担当部長
田中 博	東京医科歯科大学 難治疾患研究所 教授

4-(3) 研究協力者

高木 利久	独立行政法人 科学技術振興機構(JST) バイオサイエンスデータベースセンター(NBDC)
-------	--

畠中 秀樹	独立行政法人 科学技術振興機構(JST) バイオサイエンスデータベースセンター(NBDC)
川本 祥子	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)
浅井 潔	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(CBRC)
光山 統泰	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(CBRC)
藤田 信之	独立行政法人 製品評価技術基盤機構(NITE) バイオテクノロジー本部
森田 弘一	独立行政法人 新エネルギー・産業技術総合開発機構 (NEDO)バイオテクノロジー医療技術開発部