

平成 24 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 25 年 3 月

(委託先)独立行政法人 産業技術総合研究所

目次

1.	総括	3
2.	研究開発成果	4
2.1	ポータル構築連携	4
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および運営	4
2.1.2	統合データベースセンターとの連携	21
2.2	横断検索連携	24
2.3	アーカイブ構築連携	27
2.4	国内外の最新動向調査及び普及に向けた取り組み	30
2.4.1	国内外の最新動向調査	30
2.4.2	広報・普及活動	32
2.4.3	ユーザーニーズ調査	36
2.5	運営委員会	41
2.6	成果発表	42
3.	謝辞	46
4.	参加者名簿	47

1. 総括

平成 24 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、日本全体のライフ分野の統合データベースセンターとしての役割が期待される科学技術振興機構(JST)バイオサイエンスデータベースセンター(NBDC)との緊密な連携のもとで、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で実施された。

今年度の主な成果は、まず、前年度までの経産省ライフサイエンス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/>)を継続的に改良しつつ運営し、データベースやソフトウェア等の成果物に関する情報発信を行った。ポータルサイトの内容は随時更新し、新しいデータベース等の紹介も行った。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については、詳細な調査を実施した結果、合計 154 件の成果物を「便覧」に掲載した。また、多数のデータベースを一括検索できる「MEDALS 横断検索」については対象データベースをさらに充実させたほか、データ ID による検索との連携を実現した。データベースのアーカイブ作成も、NBDC との連携のもとで実施した。広報活動では、データベース講習会の開催やアンケートによるニーズ調査のほか、幅広い利用者への情報提供を実現するために新しい情報発信の手段として Facebook や Twitter の利用を開始した。

以上のように、本プロジェクトはライフ分野のデータベース統合化において多くの成果を挙げ、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、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らる。

平成 24 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・リーダー 今西 規

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 23 年度に、ライフサイエンス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 22 年度において実施した「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)」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および運営

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 216 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、2012年3月31日時点で、145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71件は便覧未掲載で、53件が非公開データ、3件が公開可否未確認、15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、産業技術総合研究所(AIST)、および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。

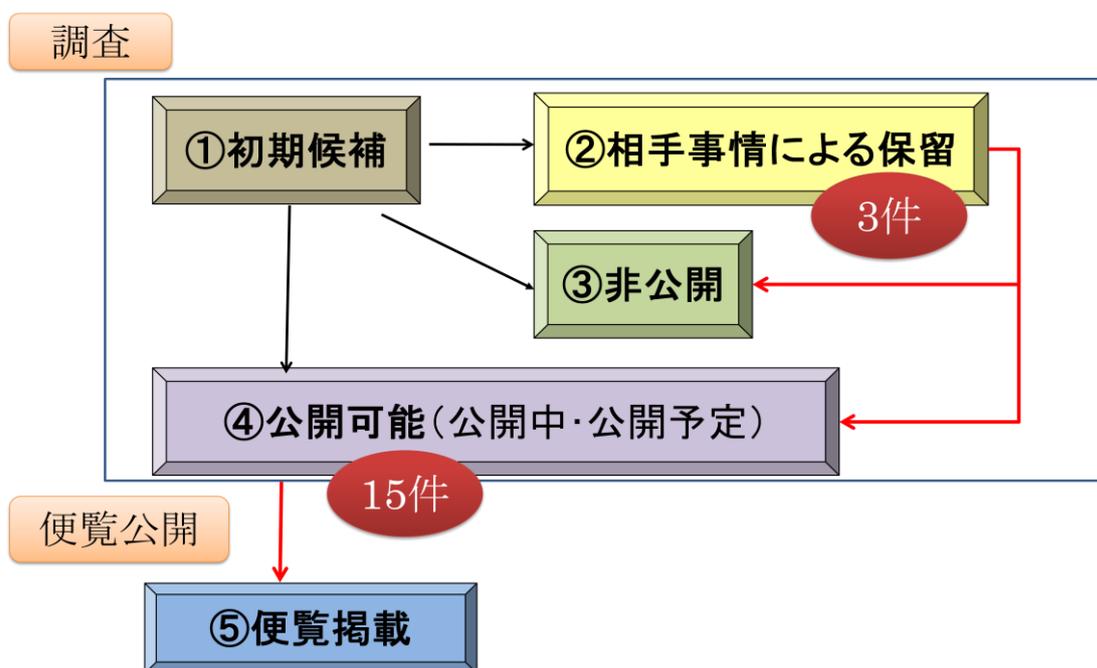


図 2.1-1 成果物候補の選定から便覧掲載までの状態遷移

平成 23 年度の事業としては、「保留」3 件の公開可否未確認の成果物および「公開可能」15 件の成果物について、再調査を行った。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、15 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連携に詳述)を検討した。

「保留」3 件の公開可否未確認の対象成果物を表 2. 1-1 に示した。この 3 件を調査した結果、3 つとも 2002 年度に同じプロジェクト終了していた。これは、2008-2009 年度にまとめられた成果物 6 件のうちの 3 つであり、うち 3 件は事業化済であったため、2011 年度の(本プロジェクトと同様の活動である)データベースプロジェクトにおいて、PJ 便覧へ掲載がすんでいた。残っている該当の 3 件について何度も電話で問合せをした。その結果として、該当組織が各担当の企業に新規担当者の確定を依頼することになった。これによって、公開への道筋が新たに出来たことになる。この 3 件については、公開や非公開などは未定であるが、今後も引き続き連絡をとっていく必要がある。

表 2. 1-1 「保留」の公開可否未確認の対象成果物(3 件)

1	転写因子データベース(TFDB)
2	遺伝子発現プロファイル計測実験支援システム
3	発現頻度情報収集用ソフトウェア

「公開可能(便覧未掲載)」15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、2件を除いて13件のデータが公開済みとなっていた。ただし、公開されたうち2件は、別の1件と統合されて公開」という形となっている(13, 14, 15番)。未公開の成果物は11, 12番の2件だけとなった。また、独自の成果物としては公開していないものが先の統合公開も含めて7件(5-9番、14番、および15番)あったため、これらは個別の便覧(カタログ)作成の対象外とした。ただし、関連するプロジェクト便覧に、どうなっているかについて載せることにし、データのアクセス先や所在がわかるようにした。

表 2. 1-2 「公開可能(便覧未掲載)」についての調査結果

番号	内容	公開状態	提供場所(サイト)*	成果物 便覧掲載	プロジェク ト便覧掲
1	Paraclu:配列データセットからクラスターを発見するツール	公開	CBRC	○	○
2	Fdur:系統樹モデルの統計量を計算するツール	公開	CBRC	○	○
3	Rentropy:RNAの二次構造のボルツマン分布からエントロピー等を計算するツール	公開	CBRC	○	○
4	Rchange:塩基置換におけるRNA二次構造エネルギー変化を計算するツール	公開	CBRC	○	○
5	ラットF344の遺伝子発現データ(発癌性化合物)	公開	発がん性予測システムおよびNEDO 報告書	対象外	○
6	予測に用いた遺伝子に関するToxArrayIII データ	公開	発がん性予測システムおよびNEDO 報告書	対象外	○
7	16S rRNAデータベース	公開	NCBI	対象外	○
8	癌組織遺伝子発現データベース	公開	NCBI	対象外	○
9	細胞局在データ	公開	HGPD	対象外	○
10	集団遺伝学解析ソフト(AStat, POPGAN, Clark, LDConf, RIGHT)	公開	H-GOLD	対象外	○
11	被験者の脳のMRI,PETの画像、心理検査、髄液・血液検査などのデータ。サーバーに蓄積しており、一般公開する予定	公開予定	未定	未	○
12	遺伝子情報ラット毒性データ	公開予定(H24年度)	未定	未	○
13	有害情報評価支援システム統合プラットフォーム(統合)	統合公開	NITE	○	○
14	代謝知識情報DB(統合)	13に統合公開		対象外	○
15	毒性知識情報(統合)	13に統合公開		対象外	○

*CBRC:生命工学情報センター、NITE:製品評価技術基盤機構、CERI:化学物質評価研究機構

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3件については担当者が明確になる予定となり、公開可能で便覧未掲載の15件については残り2件となった。これらについては、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。

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8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。うち4件は公開済のため、データベース便覧に記載した。残り4件は便覧に載せる

には至らなかった。これを追加すると「④公開可能」の件数は、表 2. 1-2 の 2 件に、未掲載 4 件が追加された 6 件となった。

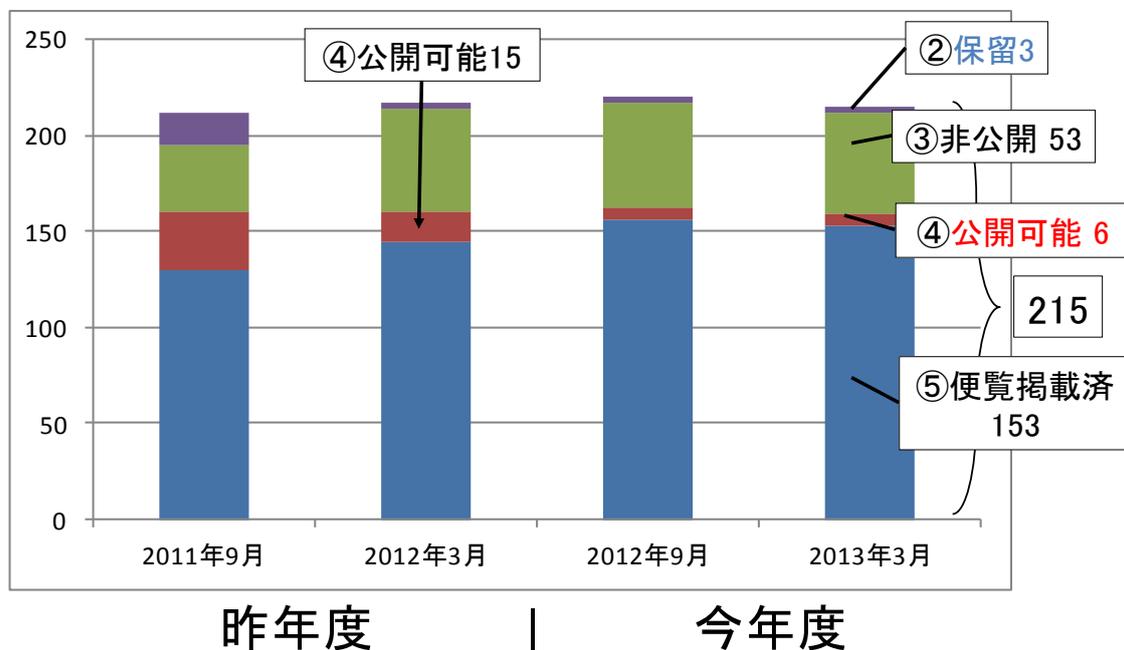


図 2. 1-2 成果物調査および便覧掲載作業の進捗

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。調査中に成果物の統廃合があったので、トータルの件数は 3 月には減った(2013 年 3 月予定)。

② 便覧の更新

(a) 新規便覧の作成と公開

新規に 9 件のデータベースおよびツール便覧を作成し、公開を行った(2013 年 3 月 12 日時点)。9 件のうち 4 件(発がん性予測システム、BioDBScan、RAvariome、PCDq)は、上述した関連研究機関からの発表で知るところとなった成果物である。また、11 件の成果物については、関連するプロジェクトの便覧に名称の記載を行った(表 2. 1-2)。

(b) 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いる。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ーのような RSS リーダーでこの情報をみると初めは図 2. 1-3 のようなサマリ画面が見られる。いずれかの行をクリックすると図 2. 1-4 のように何が更新されたのかの詳細を見ることができる。

便覧の定期更新は、2012 年 4 月から 2013 年 3 月までの 12 ヶ月にわたり、週一回を原則として実施した。また、RSS は、合計 138 件の配信をした。この内訳としては、MEDALS サイト(便覧等)の更新: 44 件、便覧に書かれた外部データベースサイトの更新: 111 件、イベント等ニュース: 16 件となっている。

MEDALS »		
☆ [外部サイト更新][成果物]FORTE - 2013 February 20 the NCBI non-redundant database: 12/02/13 (98% & filtered) 22444	2013/02/21	
☆ [外部サイト更新][成果物]SPALN - Present Version: 2.0.6d, Last update: 2013-02-18	2013/02/21	
☆ [外部サイト更新][成果物]FORTE - 2013 February 13 the NCBI non-redundant database: 12/02/13 (98% & filtered) 22431	2013/02/14	
☆ [外部サイト更新][成果物]LAST - last-274.zip 12-Feb-2013 11:54 424K	2013/02/14	
☆ [外部サイト更新][成果物]FORTE - 2013 February 06 the NCBI non-redundant database: 12/02/13 (98% & filtered) 22411	2013/02/08	
☆ [外部サイト更新][成果物]LAST - last-270.zip 04-Feb-2013 17:07 422K	2013/02/08	
☆ [外部サイト更新][成果物]MAFFT - The latest version is 7.023 (2013/Feb/03).	2013/02/08	
☆ [外部サイト更新][成果物]MassBank - 2013年02月01日 常に最新の公開データをダウンロードできるようになりました。	2013/02/08	
☆ [外部サイト更新][成果物]FORTE - 2013 January 30 the NCBI non-redundant database: 12/02/13 (98% & filtered) 22398	2013/01/31	
☆ [外部サイト更新][成果物]KOMICS - Data update (2013/01/29) Data for biological samples were added. There are data for 75	2013/01/31	
☆ [外部サイト更新][成果物]PubMedScan - PubMedScan v3.4 was released. [13.01.29] Fixed a bug the search function for	2013/01/31	
☆ [外部サイト更新][成果物]リンク自動管理システム - January 24, 2013 COXPRESdb were included.	2013/01/25	
☆ [外部サイト更新][成果物]PowerFT - PowerGet 3.3.6ベータ版を公開 LC-MSの解析ソフト、PowerGetの3.3.6ベータ版を公開しまし	2013/01/25	
☆ [外部サイト更新][成果物]MassBank - 2013年01月23日 ワシントン州立大学のデータ(2,626 スペクトル)を公開しました。	2013/01/25	
☆ [外部サイト更新][成果物]SPALN - Present Version: 2.0.6c, Last update: 2013-01-23	2013/01/25	
☆ [外部サイト更新][成果物]FORTE - 2013 January 23 the NCBI non-redundant database: 12/02/13 (98% & filtered) 22380	2013/01/25	
☆ [外部サイト更新][成果物]DoBISCUIT - 2013-01-16 Release [8] Benastatin Chartreusin Chromomycin Enterocin	2013/01/25	
☆ [外部サイト更新][成果物]Workflow - 2013/01/17 · RNA Structure Prediction Active ワークフローにRNA相互作用予測を追加しまし	2013/01/25	
☆ [外部サイト更新][成果物]FORTE - 2013 January 16 the NCBI non-redundant database: 12/02/13 (98% & filtered) 22371	2013/01/18	

図 2. 1-3 RSS Reader での RSS データの表示

☆ [外部サイト更新][成果物]Evola - [03-Dec-2010] Evola release 7.5: Orangutan orthologs have been updated, human transcripts have corresponded to H-InvDB 7.5, methods for molecular phylogeny have been improved, and dN/dS data has been available for download. Dec 6, 2010

[外部サイト更新][成果物]Evola

[03-Dec-2010]
Evola release 7.5: Orangutan orthologs have been updated, human transcripts have corresponded to H-InvDB 7.5, methods for molecular phylogeny have been improved, and dN/dS data has been available for download.

☆ Add star Like Share Share with note Email Add tags

図 2. 1-4 RSS 情報をブラウザーで表示した図

(c) 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)。この生物種の表記には一定のルールが存在していたが自由筆記も認めており、若干統一性に乏しかった。また、表記される生物種の階層(種、科、門など)も便覧ごとに異なり、生物種を問い合わせとしてデータベースを見つけることに不便があった。そこで、便覧間で生物種の表記を統一し便覧の利便性を向上させる目的で、生物種記載ルールを設け各データベースの生物種の項目を更新した。各データベースから使用生物種(株)の学名を取得し、NCBI Taxonomy ID に対応付け、NCBI Taxonomy に掲載されている common name を英名として取得した。さらに WebLSD プロジェクトの辞書から和名を取得した。また、産業的用途では、学術的な生物種名よりもより人間生活に根差す「慣用名」の閲覧や検索が便利な場合もある。たとえば、*Streptomyces antibioticus* (ストレプトマイセス・アンチバイオティカス)を、慣用名オレアンドマイシン合成菌と紐づけることで、便覧の産業的用途への利便性を向上できる。そこで、LSD (Life Science Dictionary) データおよび NITE ウェブサイトの情報から慣用名を取得した。

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。

ルール設定には前年度のパイロット的対応を反映し、NCBI Taxonomy ID の取得以降のパイプラインを半自動化した。例外処理のみマニュアルで対処した(図 2. 1-8)。この半自動化パイプラインを用い、データベース便覧 77 件のうち、53 件の生物種便覧を更新した(図 2. 1-9)。残り 24 件のうち 16 件は生物種を対象としていないデータベースであり、8 件は対象とする生物種が膨大であるため、本件では対象としなかった。対応したデータベースに含まれる生物種は 376 件で、うち 51 生物種にマニュアル対応した(図 2. 1-9)。

また、世界で最も使われている生物分類データベースである NCBI Taxonomy の ID と便覧の学名を連携させることにより、便覧から汎用性の高いバイオデータベースに簡単にアクセスできるようになった。

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」「キーワード」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(情報を展開する前)

主な対象データ	比較ゲノム
生物種	<div style="border: 1px solid #ccc; padding: 5px;"> <div style="text-align: right;">Hide</div> <ul style="list-style-type: none"> • Homo sapiens [Taxonomy_id: 9606, ヒト] • Pan troglodytes [Taxonomy_id: 9598, チンパンジー] • Pongo abelii [Taxonomy_id: 9601, スマトラオランウータン、オランウータン] • Macaca muratta [Taxonomy_id: 9544, アカゲザル] • Mus musculus [Taxonomy_id: 10090, ヲウカネズミ、マウス] • Rattus norvegicus [Taxonomy_id: 10116, ドブネズミ、ラット] • Canis familiaris [Taxonomy_id: 9615, イヌ] • Bos taurus [Taxonomy_id: 9913, ウシ] • Equus caballus [Taxonomy_id: 9796, ウマ] • Monodelphis domestica [Taxonomy_id: 13616, ハイイロシネズミオポッサム、オポッサム] • Gallus gallus [Taxonomy_id: 9031, ニワトリ] • Orzias latipes [Taxonomy_id: 8090, メダカ] </div>
キーワード	<div style="border: 1px solid #ccc; padding: 5px;"> <div style="text-align: right;">Hide</div> <ul style="list-style-type: none"> • Acetobacter pasteurianus IFO 3283-32 (酢酸菌) • Acidiphilium multivorum ATU301 (好酸性、金属耐性) • Aeropyrum pernix K1 (好熱古細菌) • Anaerolinea thermophila UNI-1 (高温排水処理リアクター) • Aspergillus oryzae RIB40 (麹菌) • Brevibacillus brevis NBRC 100599 (タンパク質生産宿主) • Caldisericum exile AZM16c01 (新門カルディセリカ門) • Corynebacterium efficiens YS-314 (アミノ酸生産) • Deferribacter desulfuricans SSM1 (深海熱水噴出孔) • Desulfovibrio magneticus RS-1 (磁性粒子合成菌) • Gemmatimonas aurantiaca T-27 (新門ジェマティモナデス門) • Kitasatospora setae KM-6054 (抗生物質生産) </div>

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・キーワードを全件閲覧できる。生物種の項目において 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おり、さらに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。キーワードの項目では、学名と慣用名を閲覧できる。

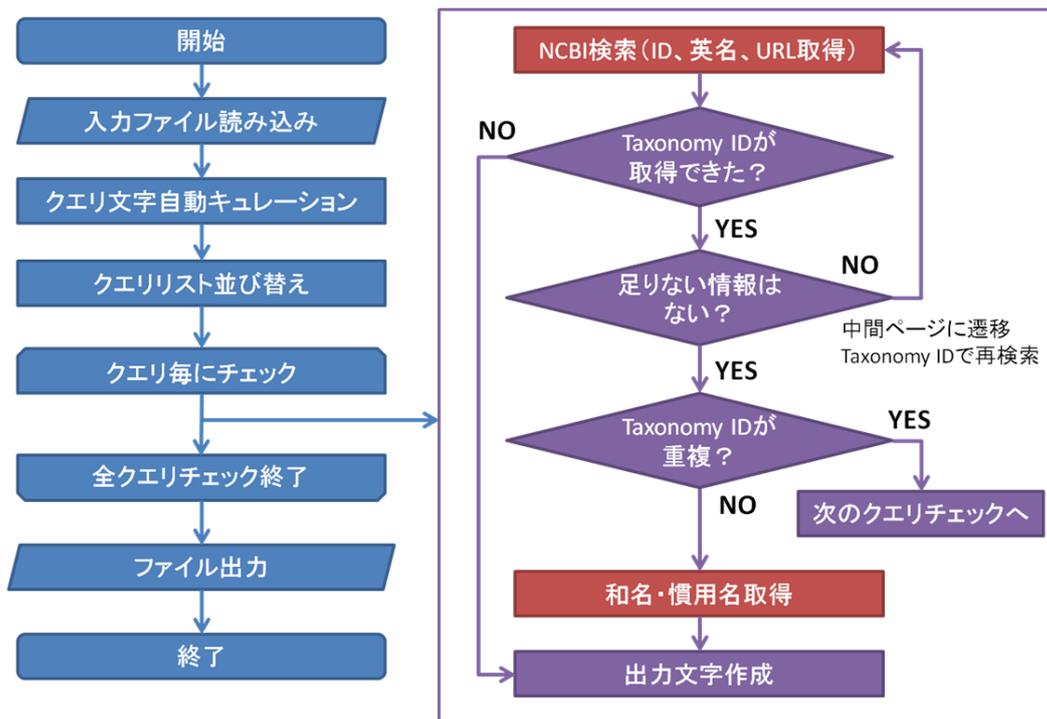
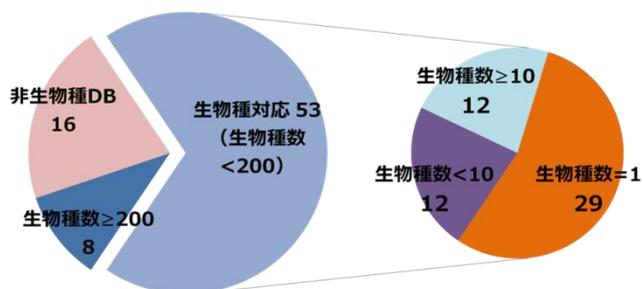


図 2. 1-8 生物種取得の半自動パイプライン

便覧数

データベース便覧数: 77



生物種数

生物種(株)数376
 ※延べ生物種(株)数520

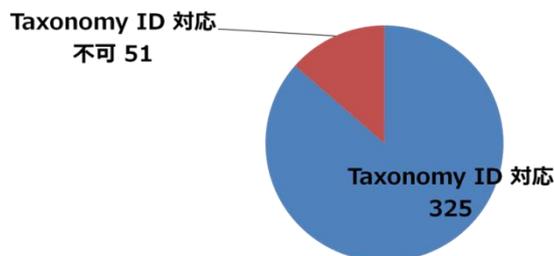


図 2. 1-9 生物種の対応を行ったデータベース便覧数と生物種数

③ ポータルの拡張

(a) 収録データベースの分類

データベースを新たな分類法を開発してユーザーに提示する。現在の MEDALS にも「DNA、RNA、タンパク質、その他」という分類が用意されている。図 2. 1-10 に 2012 年 4 月でのデータベース便覧のカテゴリを表示しているが、示されているように「その他」カテゴリに属する DB が 46 件と多数になっていて全体の把握が困難である。そこで、データの共通性・類似性に基づいてデータベース全体の地図を別途作成して、データベースの全体を俯瞰したり、データベースそれぞれの位置づけを直観的に把握しやすい形で視覚化した図を提供することとした。あらたに提供する図を図 2. 1-11 に示す。各データベースをノード、なんらかの意味で似ているデータベースに線を引き、近い位置にマップするようにした、データベース・ネットワーク図であり、これによって全体の構成を把握できる。例えば右上のデータベースは TraP, H-DBAS など選択的スプライシングデータベースであり、その下の“FLJ_Human_cDNA”や“full_length cDNA”などといった転写産物(トランスクリプトーム)のデータベースが複数あることがわかる。ユーザーが各ノードの近くをマウスでクリックすれば、各便覧にとんで詳細記述を読むことができる。



図 2. 1-10 データベース便覧のカテゴリ

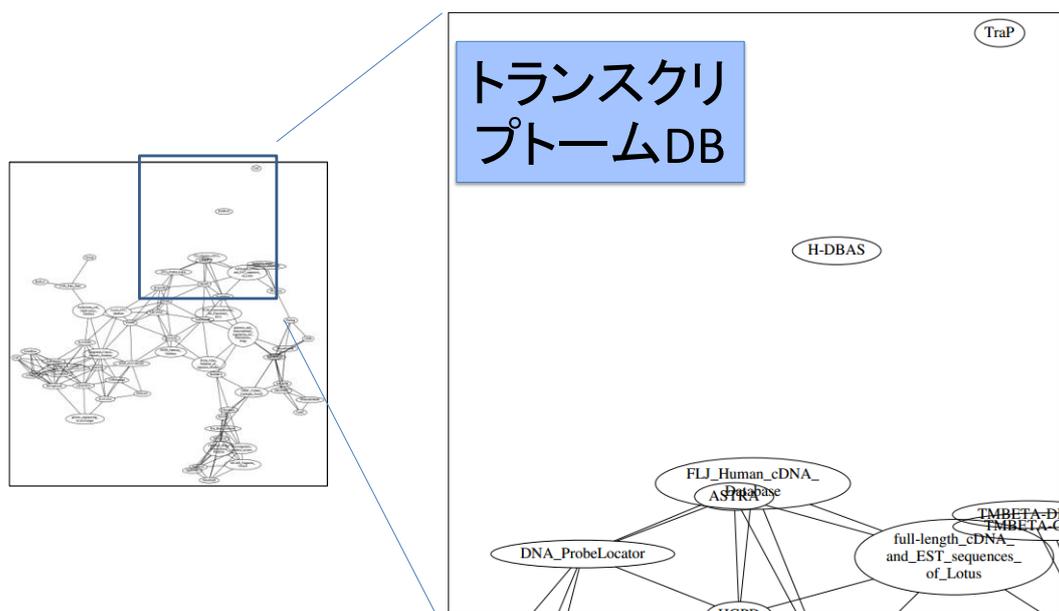


図 2. 1-11 図示化されたデータベースの位置づけや関連性

(b) 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。本年度は、MEDALS のサブシステムである文献推薦システム PubMedScan の改良を行った。具体的には、

- ユーザーのパスワード確認・再発行の機能
 - 大量の文献にも対応できる高速表示の機能追加
 - 文献でなく、データベース内の分子 ID を登録できる BioDBScan の改良を行った(図 2. 1-12)
- 。BioDBScan は、PubMedScan の拡張である。今年度は、文献を示す PubNedID を含めない選択肢を加えた。

BioDBScan

5 record(s) was deleted.

[What is BioDBScan?]

BioDBScan is a free web-based application that alerts you when any data about genes, proteins, chemicals or drugs of your interest is newly released in world-renowned databases such as NCBI, Ensembl, DrugBank and others. It delivers an E-mail message every Monday if any data is detected by the [Hyperlink Management System](#).

[Register IDs of your interest]

Select a [taxonomy] and an [ID type], and then enter IDs in the textbox below (e.g., "PPARG, TCF7L2, HNF4A, HHEX").

Taxonomy type: ID type:

IDs:

[Registered IDs]

<input type="checkbox"/>	Taxonomy type	ID type	ID	Registered date	<input type="button" value="Edit"/>
<input type="checkbox"/>	Homo sapiens	HUGO gene symbol	PPARG	2012-02-14	<input type="button" value="Edit"/>
<input type="checkbox"/>	Homo sapiens	KEGG Pathway ID	hsa04210	2012-02-15	<input type="button" value="Edit"/>
<input type="checkbox"/>	Mus musculus	Mouse Gene Symbol	Glis1	2012-02-15	<input type="button" value="Edit"/>

BioDBScan (Japanese, English) [contact: biobscan@m.aist.go.jp](mailto:contact@biobscan@m.aist.go.jp)
 MEDALS (Japanese, English)
 2008-2012 Copyright© AIST . All rights reserved.

図 2. 1-12 BioDBScan にて分子の ID を登録する画面

(c) MEDAS のメタデータの RDF 化

便覧掲載のデータベースや付与した説明(メタデータ)の活用を図ることは重要である。近年、次世代 Web 技術として「セマンティック Web」が広がっている。セマンティック Web による利点は、1) 統合や再利用が容易になる、2) データの意味を区別できるといわれており、統合活動を促進すると考えられる。そこで、本プロジェクトでは、MEDALS 便覧の記述データの RDF 化を行った。これにより研究者が MEDALS にあるメタデータを様々な形態で利用しやすくなると思われる。

図 2. 1-13 は、MEDALS で管理しているメタデータで、XML 形式で示したものである。これを XSLT によって変換をほどこし、XML/RDF の形式で作成した。結果として例えば図 2. 1-14 のようになる。RDF で提供する説明項目(要素)としては、正しく変換されているかをチェックできるような数に絞って変換を行うために全てを対象とはしなかった。ここでは、integbio(本プロジェクトにおける省庁連携統合の成果を紹介する統合サイト)の統合されたカタログにある項目のみとした。

```

<database title='Evola' update_rate='8' url='http://www.h-invitational.jp/evola/search_jp.html' id='1' mdate='2010/12/03'
alias_name='Evola: Evolutionary annotation database '>
  <specification data_type='DNA' data_volume='223.8' spec_class='DB' >
    <id_type id='HIT(transcript)'/>
    <id_type id='Accession Number'/>
    <id_type id='HUGO Gene Symbol'/>
    <related_id_type id='H-InvDB'/>
    :
    <mol_type primary='yes' type='比較ゲノム'/>
    <interface type='GUI'/>
  </specification>
  <medals id='1' site_image='http://medals.jp/images/thumbnail/Evola.ja.jpg' list_url='http://medals.jp/list/detail/1/' >
  <input_example>トップ / サーチページ 上部の "Evola Search ortholog by" で、プルダウンメニューから "Keyword
(Human)"
を選び、テキストボックスに "calmodulin" を入力して "submit" ボタンを押す。
</input_example>
  <method_download>http://www.h-invitational.jp/evola/download_jp.html </method_download>
  <resource name='H-InvDB'/>
  <resource name='DDBJ'/>
  :
  <taxonomy primary='yes' />
  <taxonomy name='Homo sapiens' />
  <taxonomy name='Pan troglodytes' />
  <taxonomy name='Pongo abelii' />
  <operation country='日本' >
  <organization name='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' />
</operation>
<legal condition='表示-継承' />
<article>17982176</article>
<keyword word='オルソログ' />
<keyword word='パラログ' />
<keyword word='比較ゲノム' />
<keyword word='脊椎動物' />
Evola (Evolutionary annotation database)はH-InvDBのサブデータベースの1つとして、ヒト遺伝子の分子進化アノテーション情報を格納しています。ヒトと14のモデル生物とのオーソログについて、アラインメントや系統樹、正負の自然選択情報を提供しています。オーソログ解析対象の生物はヒト、チンパンジー、オランウータン、マカクザル、マウス、ラット、イヌ、ウマ、ウシ、オポッサム、ニワトリ、ゼブラフィッシュ、メダカ、ミドリフグ、トラフグです。
</database>

```

図 2. 1-13 変更前の XML データ(赤字部分が変換対象)

```

<rdf:Description rdf:about="Evola">
<MEDALS:database-title>Evola</MEDALS:database-title>
<MEDALS:database-title-alias>Evola: Evolutionary annotation database</MEDALS:database-title-alias>
<MEDALS:database-url>http://www.h-invitational.jp/evola/search_jp.html</MEDALS:database-url>
<MEDALS:database-description>Evola (Evolutionary annotation database)はH-InvDBのサブデータベースの1つとして、ヒト遺伝子の分子進化アノテーション情報を格納しています。ヒトと14のモデル生物とのオーソログについて、アラインメントや系統樹、正負の自然選択情報を提供しています。オーソログ解析対象の生物はヒト、チンパンジー、オランウータン、マカクザル、マウス、ラット、イヌ、ウマ、ウシ、オポッサム、ニワトリ、ゼブラフィッシュ、メダカ、ミドリフグ、トラフグです。</MEDALS:database-description>
<MEDALS:specification-spec_class>DB</MEDALS:specification-spec_class>
<MEDALS:medals-list_url>http://medals.jp/list/detail/1/</MEDALS:medals-list_url>
<MEDALS:taxonomy>yes</MEDALS:taxonomy>
<MEDALS:taxonomy>Homo sapiens</MEDALS:taxonomy>
<MEDALS:taxonomy>Pan troglodytes</MEDALS:taxonomy>
<MEDALS:taxonomy>Pongo abelii</MEDALS:taxonomy>
<MEDALS:taxonomy>...</MEDALS:taxonomy>
<MEDALS:operation-country>日本</MEDALS:operation-country>
<MEDALS:article>
  PMID: 17982176</MEDALS:article>
</rdf:Description>

```

図 2. 1-14 RDF 化された XML データ

④ 運営

(a) 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、2012年4月から2013年1月まで(10ヶ月)の集計で月平均で1,561件(合計で15,617件)であった。

(b) 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2012年4月から2013年3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	206
2	DB 正常臓器 遺伝子発現データベース	147
3	DB 機能性RNA配列データベース	128
4	DB ATTED II	116
5	Tool BioDBScan	108
6	DB PCDq	100
7	DB ARCHAIC	95
8	DB CELLPEDIA	90
9	Tool ID一括変換システム	90
10	Tool リンク自動管理システム	89
11	DB CellMontage	81
12	Tool PMID-Extractor	80
13	DB 化学物質統合検索システム	79
14	DB PPI view	79
15	Tool myPresto	75
16	DB 微生物データベースシステム	73
17	Tool PowerFT	69
18	DB DoBISCUIT	68
19	DB GenoBase	67
20	DB KNApSAck Family	67

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。

(c) セキュリティチェック

ポータルサイトの運営においてセキュリティ上の問題が無いことを確認するため、安定した検査ソフトウェアを利用してセキュリティ上の脆弱性を点検した。広く利用されている専用ソフトウェアを

選定し、2012年2月の最新バージョンにてチェックを行い、問題がないことを確認した。

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(図2.1-15)。

今年度追加したデータベースは以下の通りである。

- ・ 2012年4月6日 対応データベースとして、KNApSAcK, JECDB, HMDBを追加した (Chemical compounds and Drugs)。
- ・ 2012年8月10日 対応データベースとして、NCBI RefSeq Proteinを追加した (Human, Mouse, Rat)。
- ・ 2012年10月3日 対応データベースとして、Database of Genomic Variants(ヒトゲノムにおける構造変化のカタログ)を追加した(Human)。
- ・ 2012年10月 (KEGGについて、APIによる全ID取得への対応)
- ・ 2013年1月24日 対応データベースとして、COXPRESdb(共発現する遺伝子とその機能)を追加した(Human, Mouse, Rat)。
- ・ 2013年3月 対応データベースに、Human Protein Atlas(ヒトのプロテオーム)等を追加予定 (Human)。



PubChem Compound ID: 6274,5280961 [Chemical compounds All Databases] Search

カテゴリDBを選択して下さい

すべて

薬

文献

化合物

タンパク質

薬

DrugBank

5280961 ▶ [DB01645](#)

6274 ▶ [DB00117](#)

Kegg Drug

5280961 ▶ No link

6274 ▶ [D00032](#)

PharmGKB

5280961 ▶ No link

6274 ▶ [PA449882](#)

文献

CiteXplore

5280961 ▶ [10741415](#)

5280961 ▶ [21848266](#)

6274 ▶ [15206581](#)

6274 ▶ [21608102](#)

6274 ▶ [21719207](#)

化合物

PubChem Compound

5280961 ▶ [5280961](#)

6274 ▶ [6274](#)

ChEBI

5280961 ▶ [28088](#)

6274 ▶ [15971](#)

Kegg Compound

5280961 ▶ [C06563](#)

6274 ▶ [C00135](#)

PDBeCem

5280961 ▶ No link

6274 ▶ [HIS](#)

KNAPSAcK

5280961 ▶ [C00002526](#)

6274 ▶ [C00001363](#)

JECDB

タンパク質

PDBJ

5280961 ▶ No link

6274 ▶ No link

UniProt

5280961 ▶ [P03372](#) Human All Databases

5280961 ▶ [P11388](#) Human All Databases

5280961 ▶ [Q14289](#) Human All Databases

5280961 ▶ [Q15596](#) Human All Databases

5280961 ▶ [Q15786](#) Human All Databases

5280961 ▶ [Q92731](#) Human All Databases

6274 ▶ [P12081](#) Human All Databases

6274 ▶ [P19113](#) Human All Databases

6274 ▶ [P42357](#) Human All Databases

6274 ▶ [Q99624](#) Human All Databases

図 2. 1-15 リンク自動管理システムに新たに追加されたデータベース

この図では、化合物版における KNAPSAcK データベースの検索結果を赤でハイライトしている。

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 38 種類、マウスは 18 種類、ラットは 12 種類、ホヤは 12 種類、化合物は 15 種類となった。そして、取り扱うデータ ID の件数は約 1 億 6000 万件、扱うリンク数は約 8 億件となり、世界最大のデータ ID の統合データベースとなった。

② MEDALS 横断検索との連携による ID 検索の実現

リンク自動管理システムにおける ID の検索は、MEDALS 横断検索におけるキーワード検索とともに、多数のデータベースに対して横断的な検索ができる便利なシステムである。前者は検索クエリとしてデータベースのデータを管理する ID を使うのに対し、後者は一般的なキーワードを用いるため、これまでは両者のサービスは別々に運用されてきた。しかし、遺伝子シンボルなどの用語は ID としてもキーワードとしても用いられるため、両者のサービスに対して同時に検索をかけることで、より効率的なデータ検索ができるようになると期待される。

そこで今年度は、リンク自動管理システムと MEDALS 横断検索を連携させ、MEDALS 横断検索で ID による検索ができるしくみを開発した。この開発は、次の 2 つの要素により実現した。

- (1) リンク自動管理システムに新たなウェブサービスを開発し、API で ID による検索リクエストを受け取り、それに対して検索結果を返すしくみを開発した。
- (2) MEDALS 横断検索の画面を改良し、ユーザーが投入したキーワードをリンク自動管理システムの API に投げ、ID 検索の結果を受け取り、画面に表示するしくみを開発した。

これらを MEDALS 横断検索に組み込み、2013 年 3 月より、MEDALS サイトで公開した(図 2. 1-16)。これにより、ID 統合の成果を MEDALS 横断検索サービスでもユーザーが意識することなく利用できるようになり、ユーザーが幅広く ID 統合の成果を活用できるようになった。



図 2. 1-16. MEDALS 横断検索と ID 検索の統合化

MEDALS 横断検索のクエリをリンク自動管理システムのウェブサービスに投げることにより、さまざまなデータ ID を検索可能にした。2013 年 3 月 18 日に公開。

2. 1. 2 統合データベースセンターとの連携

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 年に公開している。図 2. 1-17 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. 1-17、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 9 月時点の MEDALS 上の便覧データを NBDC へ情報提供した。また、統合されたカタログは昨年度にはベータ版を公開していたが、平成 24 年度は 10 月 1 日に正式版の公開となった(図 2. 1-18)。正式版では、インターフェースが改良され、「カテゴリ(対象)」及び「カテゴリ(データの種類)」が追加され、稼働状況の毎日自動更新されるようになっている。その後、MEDALS 便覧の更新情報を NBDC データベースカタログに速やかに反映する方針を検討した。それまでは、MEDALS 便覧を更新した際に、その更新情報が NBDC データベースカタログに反映するまで数ヶ月かかっていたが、連絡体制を検討した結果、2013 年 1 月以降には数日で反映できる体制を整えることができた。

図 2. 1-17 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。

-生命科学系データベースを一覧から探す-

Integbioデータベースカタログ

全条件をリセット

一覧内を検索する

一覧を絞り込む

対象を選ぶ

- ゲノム (160)
- 遺伝子 (265)
- cDNA (184)
- タグ配列 (核酸) (144)
- 多型 (94)
- その他のDNA (84)
- RNA (108)
- 蛋白質 (319)
- 酵素 (27)
- その他の生体分子 (106)
- 薬剤/化学物質 (74)
- 細胞 (58)
- 個体/種 (214)
- 健康/疾患 (177)
- その他 (68)

データの種類を選ぶ

- 配列 (464)
- 構造 (177)
- 遺伝子発現 (136)
- 相互作用/パスウェイ (92)
- 系統発生/分類 (79)
- 画像/動画 (217)
- オントロジー/用語/学名/命名法 (87)
- 論文/その他報告書/書誌 (252)
- バイオリソース (119)
- その他 (62)

稼働状況を選ぶ

- 稼働中
- 休止
- 運用終了

その他の条件を選ぶ

- LSDBアーカイブ収録のDB
- 日本国内のDB

データベースのレコード一覧

1180 件 最初へ 前へ 1 2 3 4 5 6 7 8 9 10 次へ 最後へ

追加

Mutation Database

運用機関: 独立行政法人科学技術振興機構 バイオサイエンスデータベースセンター
 生物種: *Homo sapiens*
 説明: 疾患原因遺伝子や関連遺伝子の変異に関する情報とそれに伴う疾患の臨床情報をアーカイブし共有することを目的としたレジストリデータベースです。ラボでの独自の解析結果と文脈から抽出した... [詳細へ](#)

OpenPML

運用機関: 一般社団法人バイオ産業情報化コンソーシアム
 生物種: *Homo sapiens*
 説明: ゲノムの多様性と表現型の多様性の関連を解析する研究の推進には、大規模且つ多様なデータを多くの協力者や研究者の間で共有することが不可欠です。そのためには各々の研究者がデータの互換... [詳細へ](#)

TMBETA-DISC

運用機関: 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
 生物種: *Homo sapiens*
 説明: アミノ酸配列をもとにβ-バレル型膜タンパク質を識別するためサポートベクターマシン (SVM) や統計的手法を用いた計算手法を開発しました。アミノ酸の構成、residue pair preference及びモチーフ... [詳細へ](#)

サイエンティストライブラリー特別編 ~日本の生命研究を築いた科学者~

運用機関: 株式会社生命誌研究館
 生物種: *Homo sapiens*
 説明: JT生命誌研究館が発行する季刊生命誌に掲載のサイエンティストライブラリーをウェブ用に編集した記事のアーカイブです。20世紀、独自の研究によって日本の生命研究の基礎をついた科学者が... [詳細へ](#)

生命誌ジャーナル

運用機関: 株式会社生命誌研究館
 生物種: *Homo sapiens*
 説明: 生命誌ジャーナルは、JT生命誌研究館が年に4回発行するカード季刊生命誌と連動した、Webジャーナルです。内容はトウ、リサーチ、サイエンティストライブラリーの3本立てです。巻頭トーク... [詳細へ](#)

Protein Complex Database with complex quality index

運用機関: 産業技術総合研究所 バイオメダシナル情報研究センター
 生物種: *Homo sapiens*
 説明: 予測複合体を含めたヒト蛋白質複合体、それらの相互作用、さらにヒト蛋白質相互作用情報のデータベースです。5つの異なる蛋白質相互作用 (PI) データベース (BIND, DIP, MINT, HPRD, I... [詳細へ](#)

Rheumatoid Arthritis variome

運用機関: 産業技術総合研究所 バイオメダシナル情報研究センター
 生物種: *Homo sapiens*
 説明: 自己免疫疾患、関節リウマチ (RA) に関わる遺伝子多型情報を収集し、個人の遺伝的リスク予測や分子標的創薬を支援するデータベースです。本データベースでは、GWASなどの疾患関連解析の研究... [詳細へ](#)

近藤高太郎 収集種子データベース

運用機関: 岡山大学 寛福植物科学研究所
 生物種: *Spermatophyte*
 説明: 大塚実業会農業研究所 (現・岡山大学寛福生物科学研究所) 旧所長 近藤高太郎博士により収集された種子の標本の目録です。標本名や試験管のラベルの記載をもとに作成したものです。収集日... [詳細へ](#)

メニュー

- ホーム
- 本カタログについて
- 更新履歴
- ダウンロード
- お問い合わせ
- 他カタログ系サイト

新着情報

2012/11/05: 1件のレコードを追加しました

2012/10/29: 2件のレコードを追加しました

2012/10/22: 4件のレコードを追加しました

2012/10/15: 2件のレコードを追加しました

2012/10/01: Integbioデータベースカタログを正式公開しました。

図 2. 1-18 正式公開された integbio サイトのデータベースカタログ

2.2 横断検索連携

経済産業省ライフサイエンスポータルサイトに掲載されているデータベースについて、他省のポータル間での横断的な検索を可能とするため、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を行った。なお、インデックスの作成においては、統合データベースセンター等と協議の上で統一化した。図 2. 2-1 は MEDALS の横断検索ページへのリンクを示している。これをクリックすると、図 2. 2-2 に遷移する。

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 1) 比較的新規プロジェクトの成果で注目度が高いもの、2) 独自性が高いデータで広く利用されることが望まれるもの、3) 化合物に関するもの、という 3 種類の観点から調査・検討・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(5 件)

1. KNApSAcK Family Motorcycle(植物の酵素間の反応パスウェイ)
2. KNApSAcK Family DietNavi(食材・代謝物と病気との関係性情報)
3. DoBISCUITE(細菌の二次代謝産物構成遺伝子クラスター)
4. TMBETA-GENOME(ゲノム中の β バレル型膜タンパク質データ)
5. OPEN-PML(遺伝子多様性と表現多型オブジェクトモデル)

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、3 件以上が目標であり、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、2011 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(2013 年 2 月まで)で 98 回の連絡を行った。



図 2. 2-1 MEDALS トップページと横断検索ワード入力箇所

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。

MEDALS METI database portal for life science

横断検索 キーワード [?] MEDALS 表示画面選択 MEDALS 表示件数 30 検索オプション | HELP Powered by DBCLS

データベース 横断検索 キーワード [?] プロテアーゼ

データベース (226244)

- 統合DBプロジェクト (7)
- 文献 (908)
 - 蛋白質核酸酵素 (512)
 - 文科省「ゲノム特定領域」年次報告書 (12)
 - 文科省「旧ゲノム特定領域」最終報告書 (36)
 - BIOPHYS:生物物理
 - 科学新聞記事 (3)
- 学会要旨 (1328)
- 特許関連文書 (13427)
 - 日本国特許 (X)
- 用語解説 (6687)
- ゲノム・遺伝子・RNA (64615)
- 遺伝子発現・転写制御 (51062)
- タンパク質 (29809)
- パスウェイ・相互作用・生体反応 (55512)
- 糖・脂質 (112)
- 細胞・組織 (444)
- 発生 (0)
- 医療・薬 (2420)
- 食品・栄養 (21)
- 農学・環境 (2)

データベースの詳細情報

Select a database to see more information...

検索結果 ALL

日本語/英語を内部で変換

KEGG OGENES: NC_011087_685[]
 NC_011087_685 CDS PlastidsNAME lepB, PCC_0690DEFINITION Paulinella chromatophora plastid; Peptidase ... ase S26A, signal peptidase IPOSITION 757456..758013DBLINKS RS: NC_011087 NCBI
http://www.genome.jp/dbget-bin/www_bget?ogenes:NC_011087_685

BodyMap-Xs Bombyx mori - esophagus, ... [bodymap_xs]
 Gene Name Expression Ranking in " Organism: Bombyx mori Tissue: alimentary " (#EST / #library) ... s, strongly similar to NP_001093273.1 trypsin-like protease [Bombyx mori] - search 943 972 0 0 0 0 0 943 0 0 4 ... 240 1 2 1 2 0 221 0 0 0 6 0 0 7 7 Bmo.2035 Serine protease (Sp), mRNA - search 197 226 0 0 0 0 0 197 0 0 8 0 ... , weakly similar to XP_001652121.1 serine-type endopeptidase, putative [Aedes aegypti] - search 110 111 0 0 0 0
<http://bodymap.jp/sort?organism=Bombyx%20mori&group=27%2C28%2C29%2C30&start=1&width=100&expand=&cpm=&pathology=all>

PDB-2zle: DegP12<タンパク質品質管理因子>・OMP<外膜タンパク質> - EM Navigator[]
 EM Navigator PDB-2zle DegP12<タンパク質品質管理因子>・OMP<外膜タンパク質>の低温電子顕微鏡 ... 子顕微鏡構造 Cryo-EM structure of DegP12/OMP 記述子 Protease do (E.C.3.4.21.-), Outer membrane protein C 著者 ... R. 画像 前左 キーワード HYDROLASE, DegP, HtrA, protease, chaperone, PDZ, outer membrane protein, OMP, ... roteins, DegP protease (EC 3.4.21.-), Serine Endopeptidases (EC 3.4.21.-) キーワード Bacterial Outer Membra
http://emnavi.protein.osaka-u.ac.jp/emnavi_detail.php?id=pdb-2zle&lang=ja

γ線照射処理したたんぱく食品の栄養学的健全性の評価<アイソトープ協会研究成果中間報告書>[]
 食品照射に関する文献検索 健全性 (WHOLESOMENESS):毒性・微生物学的安全性、栄養学的適格性を総合した ... 化試験には、ペプシン、トロンクレアチン、ロイシナミノペプチダーゼ・プロリダーゼ*2)を用いた。ペプシン・トロンクレアチン ... (Scheme 1)とペプシン・トロンクレアチン・ロイシナミノペプチダーゼ・プロリダーゼを用いる方法 (Scheme 2)*2)で行った。
<http://foodirra.jaea.go.jp/dbdocs/001002000010.html>

Ortholog table : Staphylococcus aureus N315 Chromosome[doqan]
 DOGAN > S. aureus N315 > Ortholog table help Ortholog table around SA1119 of Staphylococcus aure ... SA1121 hypothetical protein, similar to processing proteinase homolog Around this ORF MW1161 hypothetical protei ... n, similar to processing proteinase homolog 3 SA1122 hypothetical protein, similar to ... protein, similar to processing proteinase 57866777 peptidase, M16 family 27467874 precessino proteinase MW1162

Page 1 of 7542 検索結果226242件中 1 - 30件目を表示

MEDALSについて | お問い合わせ
 2008-2010 Copyright© JBiC and AIST.
 JBiC | AIST

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。

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。2011 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。

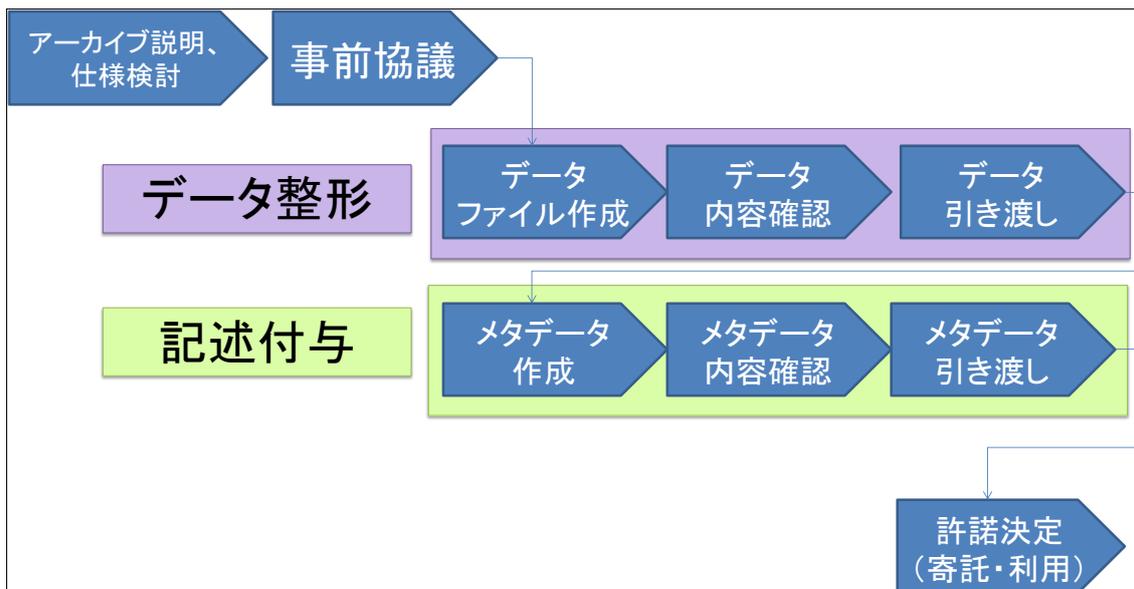


図 2. 3-2 アーカイブガイドラインによるデータアーカイブ化の流れ

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) データ公開が最近であり鮮度の高いもの 3) 担当者の異動に伴い今後の維持が困難になると予想される等で緊急性が高いもの、という 3 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、Genobase v8.0(緊急性、独自性)、LEGENDA (独自性)、CIEST(出版の緊急性、独自性)、TMBETA-GENOME (移動による緊急性)、DNAProbeLocator(維持困難の緊急性)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)。5 件の内データベースについて作業が完了したデータは NBDC に納めた。3 月 12 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 で作業が完了したデータは NBDC に納めた。3 月 12 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 CIEST 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。また、データベースではないが「更新検知ツール」、「SNP-System」の 2 件のツールについては、公開のための作業が完了したプログラムを MEDALS アーカイブから 3 月 22 日に公開予定である。



図 2. 3-3 NBDC のサイトで公開したアーカイブの例(CIEST 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6人回、国内学会等に24人回参加した。表2.4-1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 参加した国際および国内学会等

日付	国際会議名	開催場所
2012/04/02-04	The Fifth International Biocuration Conference (Biocuration2012)	米・ワシントンDC
2012/06/23-26	Society for Molecular Biology and Evolution (SMBE2012)	アイルランド・ダブリン
2012/07/15-17	20th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB2012)	米・ロングビーチ
2012/09/02-07	BioHackathon2012 (BH12)	京都
2012/12/12-14	Genome Informatics Workshop (GIW) 2012	台湾
日付	国内会議名	開催場所
2012/04/25-27	BIO tech 2012 第11回国際バイオテクノロジー展	お台場
2012/10/05	トーゴの日シンポジウム 2012～ライフサイエンスデータベース統合の医学への応用を探る～	東京
2012/10/14-17	生命医薬情報学連合大会 (2012年日本バイオインフォマティクス学会年会 情報計算法学生物学会(CBI学会)年次大会 オミックス医療研究会年会)	東京
2012/12/11-14	第35回日本分子生物学会年会 (MBSJ2012)	博多

(1) 国際会議の調査事例①: The Fifth International Biocuration Conference (Biocuration2012)

2012年4月2日～4日に米国ワシントンDCのGeorgetown University Hotel and Conference Centerで開催されたThe Fifth International Biocuration Conference (Biocuration2012)に参加しポスター発表を行った。これはUniProtの一員である米国のThe Protein Information Resource (PIR)がホストした、国際バイオキュレーション学会(the International Society for Biocuration, ISB)の第五回目の年会(第三回は東京で開催)であり、当チームの山崎がISB Executive committeeのメンバーとして参画している。17カ国から300名以上が参加し、Plenary talks 3件、Oral sessions 7件、workshops 3件が行われ、ポスター発表は約120件行われた。会議でのトピックとしては、ゲノミクス、メタゲノミクス、比較ゲノム、タンパク質構造と相互作用、機能アノテーションとパスウェイなどのアノテーション全般に関する話題をはじめ、コミュニティアノテーション、テキストマイニング手法のワークフローへの取り込み、オントロジーと標準化などの新しい取り組みが議論された。また、データベースと学術出版社との連携や、キャリアパスサーベイの結果についてのワークショップなど多岐にわたるトピックスについて発表が行われた。次回は、2013/04/07-10に英国・ケンブリッジで開催予定である。

会議後、同ワシントンDCのNational Institutes of Health (NIH)敷地内にあるThe National Center for Biotechnology Information (NCBI)を訪問し、既存NCBIデータベースのアノテーション手法の詳細や、開発中の新規データベースについて知見を得ることができた。

(2) 国際会議の調査事例②: 20th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB2012)

2012年7月19日～21日まで米国・ロングビーチで開催された20th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB2012)に参加した。ISMBは世界最大のバイオインフォマティクスに関する国際学会であり、世界のメジャーデータベースの最新情報等を収集した。例えば、2012年5月にRCSB Protein Data Bank (RCSB PDB)では立体構造ドメインに基づくアライメントデータの公開を開始した。また、データベースの新しい形態としてWiki pathway、GeneWikiなどからWiki型DBが公開されているのをはじめとして、XML, REST, HTML, RDF/SPARQL (セマンティック Web)など様々な形式での提供が進んでいた。更に、新たなデータ活用法を探索するために国際コンテスト形式で活用法を募集例も報告されていた。例えば、CAGI (Critical Assessment of Genome Interpretation; <http://genomeinterpretation.org>)からは2010年より多型から表現系を予測する表現型予測法コンテストを実施している成果が報告された。この他にも多数の国内外のデータベースについて最新情報を収集する事ができた。

(3) 国際会議の調査事例③: BioHackathon2012 (BH12)

BioHackathon (バイオハッカソン)は、2008年より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)が、2010年より独立行政法人科学技術振興機構バイオサイエンスデータベースセンター(NBDC)とDBCLSが主催してきた生命科学のデータ解析やデータベース構築に関わる実務者のための国際開発者会議である。2012年9月2日～7

日に富山で開催されたBioHackathon2012 (BH12)に参加し、最新のデータベース統合技術であるセマンティックウェブ技術、W3Cが提唱する標準規格であるResource Description Framework (RDF)、またRDFを記述するオントロジー開発について情報収集を行った。参加者は12ヶ国から合計78名、そのうち国内参加者が54名であった。今回の会議では特に、公開されている生命科学の様々なリソースの統合を実施するため、これらのリソース間での相互運用性を高めセマンティックウェブを介した仮想的な統合による各コンテンツの複合的な有効活用する手法について議論する事ができた。

(4) 国内会議の調査事例①: BIO tech 2012 第 11 回国際バイオテクノロジー展

2012年4月25日～27日に東京・お台場で開催されたBIO tech 2012 第11回国際バイオテクノロジー展に出展し、アカデミックフォーラムでの講演、ブースでのポスター発表、パンフレット3種(422部)の配布およびデモを行った。講演参加者は33名、3日間のブース来訪者数はのべ39名で、ポータルサイトやデータベース利用についてユーザーからの詳細なニーズをヒアリングする事ができた。例として、「遺伝子と特許の対応に興味がある」「疾患情報に興味がある」「質量解析に興味がある」「フリーで誰でも使えるのか?」「海外のDBでレスポンスが遅いことがある。その代わりに使ってみたい。」などの質問・コメントを頂くことができた。

(5) 国内会議の調査事例②: トーゴーの日シンポジウム 2012～ライフサイエンスデータベース統合の医学への応用を探る～

2012年10月5日に東京で開催されたトーゴーの日シンポジウム2012～ライフサイエンスデータベース統合の医学への応用を探る～に参加し、ポスター発表4件および受付にてパンフレット3種(100部)の配布を行った。本シンポジウムは、NBDCの主催で4省のプロジェクトが連携して毎年開催しており、厚生労働省プロジェクトの医薬基盤研究所、農林水産省プロジェクトの農業生物資源研究所、および経済産業省プロジェクトの産業技術総合研究所が共催で開催している。講演では、当チームの今西プロジェクトリーダーが座長を務めた。

今回のシンポジウムでは特にデータベースの医療への応用がテーマで、ヒト癌ゲノムプロジェクトでのデータベースの利用や医薬基盤研究所における安全性バイオマーカー探索のためのデータベース利用について講演が行われた。また、ヒトのゲノムデータを共有しようとする場合、その情報学的・倫理的な扱いは重要な課題になっており、文部科学省、厚生労働省、経済産業省において「ヒトゲノム・遺伝子解析研究に関する指針」の改訂が進められ議論が始まっているが、日本バイオバンクプロジェクトにおけるヒトに由来するデータ(個人情報を含む)の扱いについても講演された。また、ポスター発表ではポータルサイトやデータベース構築について情報収集と情報発信を行った。

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、2.4.1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、広報・普及活動を行った。

(1) パンフレット配布

図 2.4-1 のような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。



図 2.4-1 今年度広報に用いたパンフレット(A4)

(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。全体では、国際学会発表 4 件、国内発表件 14 件、講習会 4 回を行った。

例として、2012 年 4 月 25 日(水)～27 日(金)に東京・お台場で開催された BIO tech 2012 第 11 回国際バイオテクノロジー展のアカデミックフォーラムでの講演、ブース出展の様子を示す(図 2.4-2)。



図 2.4-2 学会発表風景(2012年4月25日(水)～27日(金)、BIO tech 2012 第11回国際バイオテクノロジー展のアカデミックフォーラムにおいて講演、ブースでのデモを実施)

(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. 4-2)。当研究チームで、創薬研究に役立つヒト遺伝子データベース H-InvDB を運用しているため、これとタイアップすることにより内容的に幅の広い魅力的な講習会プログラムを企画し、実習を含めた対面での講習会により、MEDALS ツールの操作などの応用的な MEDALS の活用法を発信した。

また、第 2 回講習会では共催した日本バイオインフォマティクス学会東北支部の協力により Ustream による中継および映像による配信を行った(<http://www.ustream.tv/channel/jsbi-tohoku2>)。更に、また第 4 回講習会でも Ustream による中継および映像による配信を行った(http://www.ustream.tv/channel/%E3%83%87%E3%83%BC%E3%82%BF%E3%83%99%E3%83%BC%E3%82%B9%E8%AC%9B%E7%BF%92%E4%BC%9A#utm_campaign=t.co&utm_source=13165776&utm_medium=social)

例として、2012 年 8 月 9 日に札幌(北海道大学)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. 4-2 今年度開催した講習会

日付	講習会名	開催場所	参加人数
2012/8/9	第 1 回講習会	札幌(北海道大学)	44 名
2012/11/3	第 2 回講習会	仙台(東北大学)	16 名
2012/12/26	第 3 回講習会	大阪(産総研・関西センター)	20 名
2013/1/24	第 4 回講習会	お台場(産総研・臨海副都心センター)	26 名

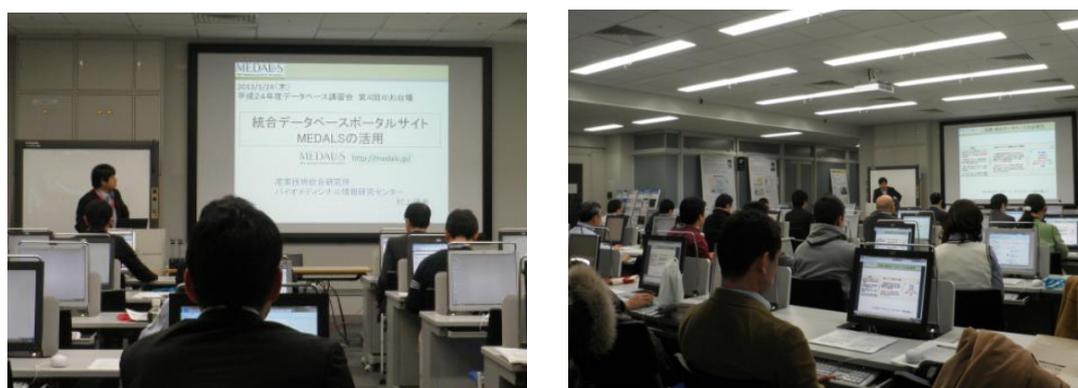


図 2. 4-3 データベース講習会開催風景(2013 年 1 月 26 日お台場(産総研))

(4) SNS(ソーシャル・ネットワーキング・サービス)を用いた情報発信

SNS(ソーシャル・ネットワーキング・サービス)はここ数年で急速に普及した新たな情報発信・情報収集の手法で、twitter や Facebook は国内外の学術出版社や主要データベース等からの情報発信に利用されるようになった。今年度は、特に新しい情報に敏感な研究者層を重点ターゲットとして MEDALS の普及を図るため SNS を用いた情報発信を新たに実施した。

具体的には、twitter および Facebook にアカウントを開設し、2012/05/08 から MEDALS のアップデートやイベントの種別ごとに区別できるハッシュタグを付けニュース配信を合計 153 件行った(表 2. 4-3)。図 2. 4-4 に実際の配信例を示す。

表 2. 4-3 今年度実施した SNS(ソーシャル・ネットワーキング・サービス)を用いた情報発信

ニュースの種別	ハッシュタグ	配信数*
成果物の更新情報	#medals_update	122
イベント情報(講習会など)	#medals_events	8
その他ニュース(停電告知など)	#medals_news	23
合計		153

*2013/03/11 時点の配信数



図 2. 4-4 SNS(ソーシャル・ネットワーキング・サービス)を用いた情報発信例

また、より多くのユーザーへ情報発信を行うため、MEDALS トップページに SNS(ソーシャル・ネットワーキング・サービス)へ誘導するためのリンクボタンの設置を行った(図 2. 4-5)。



図 2. 4-5 MEDALS トップページからの SNS(ソーシャル・ネットワーキング・サービス)への誘導

2.4.3 ユーザーニーズ調査

MEDALS の潜在ニーズを調べることを主な目的として、Web 上のアンケート調査および対面でのヒアリング調査を実施した。ニーズ調査結果で判明した改善すべき点を、2.1から2.1の作業へと反映させた。

(1) アンケート調査

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2012年の7月から8月にかけて実施し、201名の回答結果を集計した。質問項目は自由記述回答も含む全27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。実施内容の概要を表2.4-4に示す。

表 2.4-4 2012 年度アンケート調査実施概要

調査期間	2012/07/25～2012/09/07（約1か月半）
調査方法	Web アンケート(27問)
目的	・ユーザーニーズに即したより良いポータルサイトの構築 ・研究成果の連携強化 ・MEDALS の実装および経産省関連プロジェクト成果物の広報
調査項目	・MEDALS ポータルサイト ・横断検索 ・経産省関連プロジェクト新規成果物
有効回答	201名 (H-InvDB メールマガジン購読者、SNS 閲覧者、各種学会会員等)

① 調査対象者について

アンケート案内の配信先を検討した結果、5割以上の回答者の所属が公的研究機関と民間企業であった(図2.4-6)。

【アンケート案内の対象】

- 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
- バイオインダストリー協会会員
- つくばバイオビジネスネットワーク会員
- 横浜・神奈川バイオビジネスネットワーク購読者
- 近畿バイオインダストリー振興会議会員
- BIOWEB 登録者

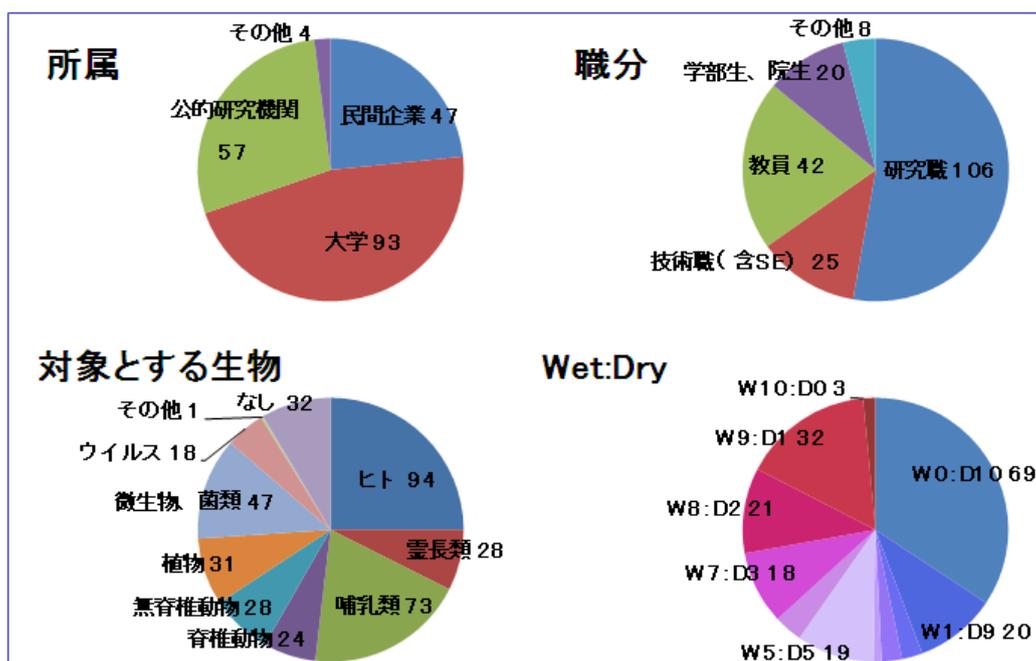


図 2.4-6 2012 年度アンケート調査 実施対象

② MEDALS ポータルサイトの利用について

MEDALS ポータルサイトの各コンテンツの利用、および国内主要バイオポータルの利用について選択式および自由記述形式で調査した。MEDALS ポータルサイトの各コンテンツについては半分以上の回答者が「使ったことがない」「初めて知った」と回答しており本アンケート調査が潜在的なユーザーに対する広報の役割も果たした事がわかった(図 2.4-7)。また、自由記述では、より見やすいインターフェース(GUI)を求める要望と、より周知を求める記述が多く見られた。

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。

- ・全体に見づらいような感じがします。慣れてないせいかもしれませんが
- ・使いやすく、汎用性が高いシステムであれば、もっと広まるのでは無いかと思う。
- ・文字が多すぎて、なかなか欲しい検索サイトにいけない。初心者にも判りやすい表示にしてほしい。
- ・学会などで使い方のセミナー(ワークショップ)等をやリ、もっと宣伝してほしい。

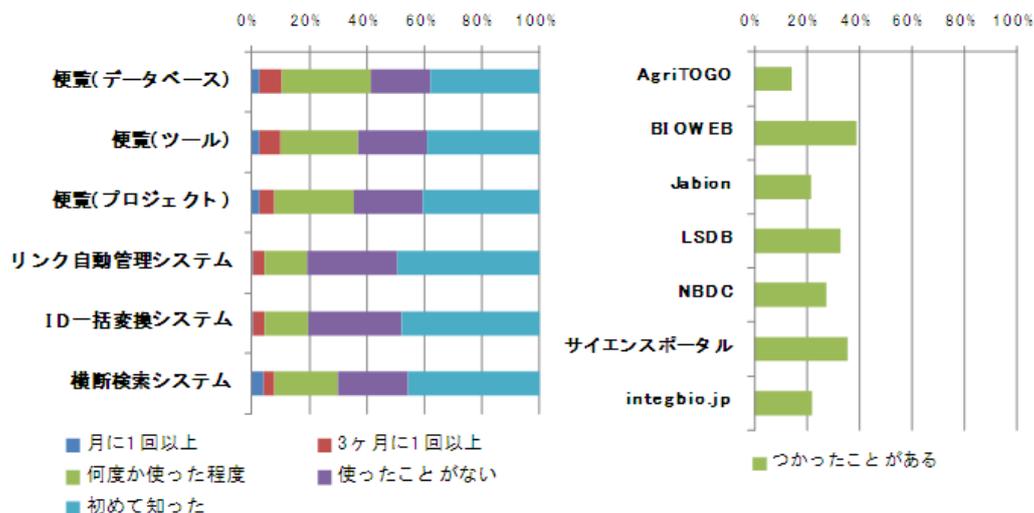


図 2.4-7 2012 年度アンケート調査 MEDALS の便覧・システム
および国内主要バイオポータルの利用度

(2) ヒアリング調査

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、データベース利用者および開発者の研究機関、大学、企業の合計 5 件実施した(表 2.4-5)。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。

表 2.4-5 ヒアリング調査

日付	ヒアリング先
2012 年 11 月 12 日(月)	バイオ IT 企業
2012 年 11 月 20 日(火)	バイオ研究機関
2012 年 12 月 18 日(火)	大学
2013 年 2 月 11 日(金)	バイオ研究機関

(3) 調査結果で判明した改善点のポータルサイトへの反映

ユーザーニーズ調査の結果の以下の意見から判明した改善点について、MEDALS ポータルサイトへの改善を行った。

まず、便覧について下記2件の意見を基に指摘されたデータベース便覧の追加を行った。

意見①: 今回産総研の方との共同研究でタイレリア原虫ゲノムDBを作製しました。日本で初の原虫全ゲノム解読になるかと思えます。ぜひ公開していただけると幸いです。

意見②: JCGGDB の糖鎖関連のデータベースは GlycoProt DB など一部しか便覧に記載されて

いませんが、他のものは収録しないのでしょうか。

次にPubMedScanについて下記意見を検討した。ユーザーが操作できるパスワード再設定機能を開発し、その他の点についても検討した。

意見③: Login ID, Password を忘れたときに自分で変更できるシステムを設置する。

意見④: アラートの間隔をユーザーが設定できるようにしてほしい。

意見⑤: 新着だけでなく、過去の情報(過去1年とか)も検索したい。

また、アンケート実施についての下記意見をもとにこれまで実施した4回(平成20年度、平成21年度、平成23年度、平成24年度)のアンケートフォームを「MEDALS Q and A」ページより公開し、アンケート中の有用な情報をユーザーに提供した(図2.4-8)。

意見⑥: このようなアンケートによって初めて知る DB が数多くあるので、今後も定期的に行って欲しい。

意見⑦: これらの LINK を一覧にしたページあるいはメールマガジンを用意してほしい

更に広報活動についての下記意見を基によりユーザーが情報を得られやすくするため、MEDALS サイト講習会サイトやメルマガ登録ページヘルプリンク設置し改善した(図2.4-9)。

意見⑨: 使い方の講習会を開催して欲しい。

意見⑩: 利用に関する説明会等を集中的に行ってほしい。

MEDALS Q and A

MEDALSについて

MEDALSとはなんですか？

MEDALSとはなんですか？

MEDALSは「METI Database portal for Life Science」の中で、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ポータルサイトです。前身プロジェクトである統合データベースプロジェクトの時代から、これまで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において、1989年以降の成果物(データベース、ツール等の情報を整理・提供しています。

MEDALSではポータルサイトとして成果物、プロジェクトの情報を整理・提供するほか、一部成果物のダウンロード提供、MEDALSツールとして新情報連携お知らせツール、PubMedScan、リンク自動管理システム、ID一括変更システム、PMD-Extractor等を提供しています。

MEDALSはどのようにして運営されていますか？

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一環として、(独)産業技術総合研究所「バイオゲノム情報研究センター」が経済産業省から委託を受けて運営しています。詳しくはMEDALSへお問い合わせください。

アンケートで紹介されていたデータベースやウェブサイトの情報が知りたいです

過去に統合データベースプロジェクト(平成20-22年度)、ライフサイエンスデータベースプロジェクト(平成23-24年度)において実施したアンケート設問のPDF版をMEDALSにて公開しています。

- 平成20年度アンケート設問
- 平成21年度アンケート設問
- 平成23年度アンケート設問
- 平成24年度アンケート設問

MEDALS Q and Aに項目を追加

図 2.4-8 アンケートフォームを「MEDALS Q and A」ページより公開

メールマガジン登録
へのリンク

講習会ページ
へのリンク

MEDALS
METI database portal for life science

ホーム データベース ツール ダウンロード

About | English | Q and A | **講習会**

RSS MEDALS on Twitter いいね! 19 送信

MEDALSや講習会に関するおしらせはH-InvDB メールマガジンでも配信しています。

お知らせ
データベース数: 66 / 解析ツール数: 64 / プロジェクト数: 52
myPrestoがV4.2081にバージョンアップしました。容量の大きなファイルがありますのでダウンロードにはご注意ください。
「創薬研究における統合データベースの活用」(第四回データベース講習会@お台場)の情報を掲載しました。[13.01.15]
ライフサイエンスデータベースプロジェクト平成23年度成果報告書を公開しました。[12.11.09]

図 2. 4-9 MEDALS サイト講習会サイトやメルマガ登録ページへリンク設置

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、2012年9月28日と2013年2月28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回運営委員会には28人が、第2回運営委員会には30人が参加した。アドバイザーならびに研究協力者は4-(2)、4-(3)参照のこと。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、ならびに本年度に新規に立ち上げた事業である生物種対応やSNSによって継続してプロジェクトを発展させて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。とくにアドバイザーからは、経産省ライフサイエンスデータベースプロジェクトを長期的視野で運営すべく、複数年の受注を可能とすべきとの意見をいただいた。

2.6 成果発表

平成 24 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. Shingo Kikugawa, Kensaku Nishikata, Katsuhiko Murakami, Yoshiharu Sato, Mami Suzuki, Md Altaf-Ul-Amin, Shigehiko Kanaya, and Tadashi Imanishi (2012) PCDq: human protein complex database with quality control index that summarizes different levels of evidences of protein complexes predicted from H-Invitational protein-protein interactions integrative dataset. BMC Systems Biology 6 (2):S7. 2012.12
2. Jun-ichi Takeda, Chisato Yamasaki, Katsuhiko Murakami, Yoko Naga1, Miho Sera, Yuichiro Hara, Nobuo Obi, Takuya Habara, Takashi Gojobori and Tadashi Imanishi (2013) H-InvDB in 2013: an omics study platform for human functional gene and transcript discovery. Nucleic Acids Research 41:D915-D919. 2012.12
3. Toshiaki Katayama, Mark D Wilkinson, Gos Micklem, Shuichi Kawashima, Atsuko Yamaguchi, Mitsuteru Nakao, Yasunori Yamamoto, Shinobu Okamoto, Kenta Oouchida, Katsuhiko Murakami, Chisato Yamasaki, et al. (2013) The 3rd DBCLS BioHackathon: improving life science data integration with semantic Web technologies. Journal of Biomedical Semantics 11;4(1):6. 2013. 2.

2.6-(2) 発表

1. 今西 規 (2012)「ヒト遺伝子統合データベース H-InvDB の新リリース 8.0」BIO tech 2012 第 11 回国際バイオテクノロジー展 2012.04
2. Katsuhiko Murakami, Tadashi Imanishi (2012)「Automatic extraction of correlation of annotation terms in databases to find similar concepts, synonyms, and multifunction」20th Annual International Conference on Intelligent Systems for Molecular Biology 2012.07
3. 村上 勝彦、今西 規 (2012)「ヒト遺伝子統合データベース H-InvDB」バイオインフォマティクス講習会 生物情報データベース入門 2012.08
4. 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」トーゴの日シンポジウム 2012 2012.10
5. 村上 勝彦、喜久川 真悟、西潟 憲策、佐藤 慶治、鈴木 満美、Altaf-Ul-Amin Md.、金谷 重彦、今西 規 (2012)「蛋白質複合体データベース: human protein complex database with quality index (PCDq)」トーゴの日シンポジウム 2012 2012.10
6. 村上 勝彦、間宮 健太郎、山崎 千里、原 雄一郎 (2012)「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・ポータルサイト MEDALS の新たな展開」トーゴの日シンポジウム 2012 2012.10

7. 長井 陽子、今西 規 (2012)「RAvariome: 関節リウマチ感受性ゲノム多型データベースと遺伝的リスク予測ツール」トーゴの日シンポジウム 2012 2012.10
8. 村上 勝彦、今西 規 (2012)「Database Maps and Meta-data Annotation for Strategic Research Planning in Data-driven and Omics Era」生命医薬情報学連合大会 2012.10
9. 山崎 千里 (2012)「Semantic web service for H-InvDB, integrated database of human genes, transcripts and proteins」生命医薬情報学連合大会 2012.10
10. 今西 規 (2012)「ヒト遺伝子統合データベース H-InvDB の活用」平成 24 年度第 2 回データベース講習会「創薬研究のためのデータベース講習会」 2012.11
11. 武田 淳一、今西 規 (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 24 年度第 2 回データベース講習会「創薬研究のためのデータベース講習会」 2012.11
12. 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 24 年度第 2 回データベース講習会「創薬研究のためのデータベース講習会」 2012.11
13. 長井 陽子 (2012)「関節リウマチに関連する遺伝子多型データベース RAvariome の活用」平成 24 年度第 2 回データベース講習会「創薬研究のためのデータベース講習会」 2012.11
14. 村上 勝彦、今西 規 (2012)「ヒト遺伝子、蛋白質における機能や立体構造をどう読むか～遺伝子データベース記述内容の潜在的な相関関係の抽出と相関を用いた GSEA 高度化へむけて～」第 35 回日本分子生物学会年会 2012.12
15. 今西 規 (2012)「ヒト遺伝子データベースの現在と未来」第 35 回日本分子生物学会年会 2012.12
16. 喜久川 真悟、西潟 憲策、村上 勝彦、佐藤 慶治、鈴木 満美、Md Altaf-Ul-Amin、金谷 重彦、今西 規 (2012)「PCDq: human protein complex database with quality index which summarizes different levels of evidences of protein complexes predicted from H-Invitational protein-protein interactions integrative dataset」Genome Informatics Workshop (GIW) 2012 2012.12
17. 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 24 年度第 3 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」 2012.12
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 24 年度第 3 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」 2012.12
19. 山崎 千里 (2012)「ヒト多型アノテーションデータベース VarySysDB と 1000 人ゲノム、ENCODE ブラウザーの活用」平成 24 年度第 3 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」 2012.12
20. 今西 規 (2013)「創薬研究における統合データベース」平成 24 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」 2013.01
21. 村上 勝彦 (2013)「統合データベースポータルサイト MEDALS の活用」平成 24 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」 2013.01

22. 村上 勝彦 (2013)「ヒトタンパク質間相互作用の統合データベース PPI view とヒトタンパク質複合体データベース PCDq の活用」平成 24 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」 2013.01
23. 山崎 千里 (2013)「ヒト遺伝子統合データベース H-InvDB の活用」平成 24 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」 2013.01
24. 原 雄一郎 (2013)「比較ゲノムデータベース Evola、G-compass の活用」平成 24 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」 2013.01

2.6-(3) データベース公開・更新

2012 年

- | | |
|----------|---|
| 4 月 6 日 | リンク自動管理システム更新
KNAPSAcK, JECDB, HMDB を対応データベースとして追加(Cheical compounds and Drugs)。 |
| 5 月 24 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
対応データベースに paraclu を追加 |
| 6 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
対応データベースに Fdur を追加 |
| 6 月 14 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
対応データベースに Rentropy を追加 |
| 6 月 22 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
対応データベースに Rcahnge を追加 |
| 6 月 28 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
対応データベースに発がん性予測システムを追加 |
| 7 月 25 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
対応データベースに BioDBScan, RAvariome, PCDq を追加 |
| 7 月 11 日 | リンク自動管理システム更新
「話題のデータ機能」を追加 |
| 8 月 10 日 | リンク自動管理システム更新
NCBI RefSeq Protein を対応データベースとして追加した |
| 10 月 2 日 | リンク自動管理システム更新
DGV を対応データベースとして追加した。 |
| 11 月 9 日 | ライフサイエンスデータベースプロジェクト平成 23 年度成果報告書を公開 |
| 2013 年 | |
| 1 月 24 日 | リンク自動管理システム更新
COXPRESdb を対応データベースとして追加した。 |

- 3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
対応データベースに TOT-DB を追加
- 3月18日 MEDALS 横断検索更新
リンク自動管理システムとの連携による ID の横断検索を実現。

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、大学の方々

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
	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
	齋藤 禎一	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
	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	
間宮 健太郎	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	

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	東京医科歯科大学大学院 疾患生命科学部 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)
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センター 上席参事官
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・医療技術部 部長