

文部科学省・経済産業省関連の生命科学系データベースの相互検索が可能に ー省別にまとめられていた「統合データベースプロジェクト」の統合がより進むー

平成 22 年 5 月 24 日

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)
独立行政法人 産業技術総合研究所
社団法人 バイオ産業情報化コンソーシアム

■ ポイント ■

- ・ 省を超えて文部科学省関連・経済産業省関連どちらのサイトからも相互に利用可能
- ・ キーワードの提案機能や同義語の検索など、横断検索の新しい機能を経産省関連検索サイトから公開
- ・ 生命科学情報の一括検索が可能になることから生命科学・医学系の研究の促進を期待
- ・ 4 省統合データベースの構築を視野に、新センター設置へ布石

■ 概 要 ■

大学共同利用機関法人 情報・システム研究機構【機構長 堀田 凱樹】(以下「ROIS」という)ライフサイエンス統合データベースセンター【センター長 高木 利久】(以下「DBCLS」という)と、独立行政法人 産業技術総合研究所【理事長 野間口 有】(以下「産総研」という)バイオメディシナル情報研究センターおよび社団法人 バイオ産業情報化コンソーシアム【会長 秋草 直之】(以下「JBiC」という)は、それぞれがまとめた生命科学系データベースを相互に検索できる機能を公開した。

DBCLS は今回、平成 20 年度より公開していた「生命科学データベース横断検索」(<http://lifesciencedb.jp/dbsearch/>)において経済産業省関連の 21 件のデータベースも検索できるようにした。産総研と JBiC は、その「横断検索」に独自の機能を追加した検索サービス「MEDALS 横断検索」(<http://medals.jp/xsearch/>)を公開し、どちらのサイトからでも全データベースを検索できるようにした。これらにより、両省のプロジェクトにおいて個別にまとめていた、生命科学系データベースの情報の統合化が推進され、生命科学・医学系の研究が促進されることが期待される。

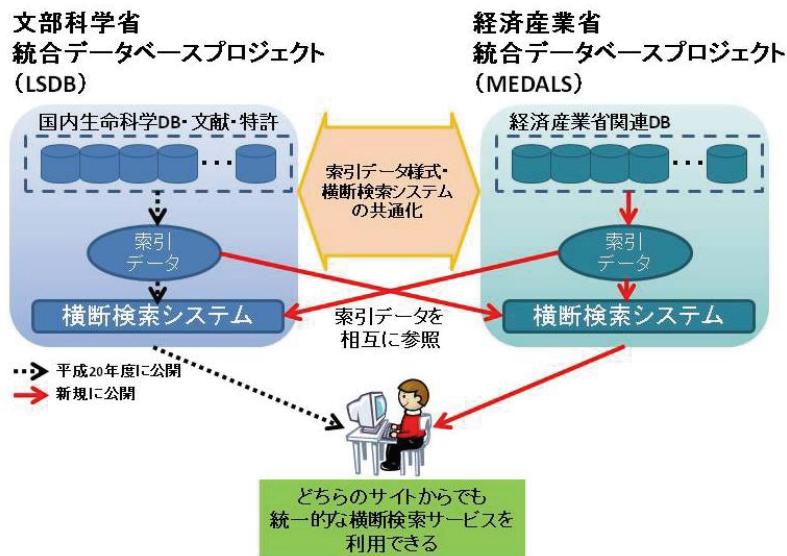


図 相互的な横断検索の仕組み
それぞれが作成した索引データを相互に利用できる

■ 開発の社会的・技術的背景 ■

わが国の生命科学系の大型研究プロジェクトは、文部科学省、厚生労働省、農林水産省、経済産業省の各省によって実施されてきた。これらの大型プロジェクトの成果物であるデータベースには、個別に作成・公開されてきたため、個々のデータベースについての所在や利用方法が利用者にはわかりにくいなどの問題点が存在した。また、より効果的な治療薬の開発研究を目指す製薬業界をはじめとする産業界からは、さまざまなプロジェクト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構築への要望が高まっていた。

そのため、わが国では第 3 期科学技術基本計画（平成 18 年 3 月 28 日閣議決定）にもとづき、総合科学技術会議が策定した、ライフサイエンス分野の推進戦略における戦略重点技術の 1 つとして「世界最高水準のライフサイエンス基盤整備」が掲げられ、また、科学技術連携施策群「生命科学の基礎・基盤」補完的課題「生命科学データベース統合に関する調査研究」（※1）では各省で実施された生命科学系の大型研究プロジェクトにおけるデータベースを個別にまとめたのち、4 省統合データベースを構築する必要があるとの報告がなされた。これを受け、文部科学省、厚生労働省、農林水産省、経済産業省では、4 省統合データベースの構築を目指して検討を進めていた。

一方、一部のデータベースはインターネット上にも公開され、Google や Yahoo のような世界中のウェブサイトを巡回する検索エンジンによって検出されることもあるが、これらの検索エンジンは、生命科学系の専門分野に特化していないため、検出の精度は高くない。また、これまでも個々のデータベース内部を検索するサービスは存在したが、研究者はそれぞれのサイトに行ってもそのつど検索を繰り返さなければならず、データやデータベースの増加に対応しきれなくなっていた。このため、複数のデータベースや文献を一度に併せて検索する仕組みの開発が求められていた。

■ 研究開発の経緯 ■

このような背景のもと、ROIS/DBCLS では、文部科学省委託研究開発事業「統合データベースプロジェクト」（※2）の一環として生命科学系の国内外の主要なデータベースを特許や日本語文献と併せて一括して検索できる「生命科学データベース横断検索」を平成 20 年度から「統合ホームページ（LSDB）」において提供してきた。これは国産の検索エンジン HyperEstraiier (<http://hyperestraier.sourceforge.net/>) を採用することで、生命科学系の専門分野に特化させた検索サービスを実現したもので、平成 21 年度にはこのサービスを用いて 232 データベースの 780 万件にのぼるデータや文書が検索可能となっていた。

一方、産総研と JBiC は、ヒトの遺伝子と転写産物を対象とした統合データベース H-InvDB（※3）をはじめとする遺伝子や分子の情報基盤の提供を行ってきた。また、平成 20 年度からは経済産業省産業技術研究開発「統合データベースプロジェクト」（※4）を、五條堀 孝 プロジェクト・リーダー、今西 規 分子システム統合チーム長のもとで実施し、生命科学分野における経済産業省関連のデータベース・解析ソフト情報を提供するポータルサイト MEDALS[メダルズ](METI database portal for life science)を構築・運用している。さらに、生命科学の専門用語が同時に出現した文献を効率的に検索できる文献データベース LEGENDA（※5）も開発している。

■ 研究開発の内容 ■

今回、ROIS/DBCLS は、離れたサーバー間で通信し索引データを共有することができる HyperEstrailer の特徴を活かし、経済産業省関連の 21 にのぼるデータベースを「生命科学データベース横断検索」においても既公開分と併せて検索できるようにした（図参照）。この結果、現在は 253 件のデータベースを一括して検索できるようになった。この仕組みを使えば利用者は省庁や研究機関の垣根を気にすることなく多数のデータベースを一度に検索することができる。

産総研と JBIC は無駄な重複開発を避け、操作性の統一感をもたせるため「生命科学データベース横断検索」で用いられている仕組みを「MEDALS 横断検索」に採用し、同じデータベースが検索できるようにした。また、「生命科学データベース横断検索」の特徴としてキーワードを自動で日英変換して検索する機能があるが、「MEDALS 横断検索」ではその機能をさらに発展させ、追加候補となるキーワードの提案機能や同義語の検索を実現した。これらキーワードの提案機能には LEGENDA（※5）の辞書を、同義語の検索にはライフサイエンス辞書（ライフサイエンス辞書プロジェクト <http://lsd.pharm.kyoto-u.ac.jp/ja/index.html>）を用いている。

■ 今後の予定 ■

ROIS/DBCLS、産総研と JBIC は、今後、扱うデータベースの数を増やし、他省庁関連のデータベースも含めて検索サービスそのものの統合運用を目指す。また、利用者の意見をさらに取り入れ、利便性の向上を図る。3 者は、検索システムの研究開発における連携に加えて人的あるいは、経済的な理由から公開・維持が困難になったデータベースの受け入れ作業においても協力しており、さらなる連携の強化を目指している。

このような連携は、データ管理、データ解析、データ公開、および論文整理や論文執筆を支援する環境など、生命科学研究を促進する知的情報環境の整備・実現への第一歩となり、これが 4 省統合データベースへと展開することによって、ひいてはわが国の生命科学研究、医薬・創薬研究の発展へとつながるものと期待される。

なお、わが国のライフサイエンス分野の統合データベース整備は、文部科学省と経済産業省の「統合データベースプロジェクト」（※2、4）終了後の平成 23 年度に、独立行政法人 科学技術振興機構に設置予定の新センターがその機能を継承するとの方針が、総合科学技術会議 ライフサイエンス PT（プロジェクトチーム）統合データベースタスクフォース（※6）において取りまとめられた。本連携成果は将来の 4 省データベース構築に向けた布石であり、ナショナルセンターを目指す新センターの事業に活かされる予定である。

■ 本件問い合わせ先 ■

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

特任准教授 川本 祥子 〒113-0032 東京都文京区弥生 2 丁目 11 番 16 号

東京大学工学部 12 号館

TEL : 03-5841-6754 FAX : 03-5841-8090

E-mail : shoko@dbcls.rois.ac.jp

独立行政法人 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 分子システム情報統合チーム

研究チーム長 今西 規 〒135-0064 東京都江東区青海 2 丁目 4 番 7 号 臨海副都心センター

TEL : 03-3599-8800 FAX : 03-3599-8801

E-mail : t. imanishi@aist.go.jp

社団法人 バイオ産業情報化コンソーシアム JBIC 研究所

担当部長 岸本 晃彦 〒135-8073 東京都江東区青海 2 丁目 4 番 32 号 TIME24 ビル 10F

TEL : 03-5531-8557 E-mail : kishimoto@jbic.or.jp

【プレス発表／取材に関する窓口】

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 箕輪 真理・高祖 歩美

〒113-0032 東京都文京区弥生 2 丁目 11 番 16 号 東京大学工学部 12 号館

TEL : 03-5841-6754 FAX : 03-5841-8090

E-mail : public_relations@dbcls.rois.ac.jp

独立行政法人 産業技術総合研究所 広報部 広報業務室 吉田 清香

〒305-8568 茨城県つくば市梅園 1 丁目 1 番 1 号 中央第 2

つくば本部・情報技術共同研究棟 8F

TEL : 029-862-6216 FAX : 029-862-6212 E-mail : presec@m.aist.go.jp

【用語の説明】

※1 科学技術連携施策群 「生命科学の基礎・基盤」補完的課題 「生命科学データベース統合に関する調査研究」

平成 17 年度～19 年度にわが国の生命科学系データベースの現状について行われた調査研究。
(研究代表者: 国立遺伝学研究所 生命情報・DDBJ 研究センター 大久保 公策)
(参照 <http://lifesciencedb.jp/sciencepolicy/report/200510102007rr.pdf>)

※2 文部科学省委託研究開発事業「統合データベースプロジェクト」

大型プロジェクトの成果として公開・作成されているものの、散在してしまっている種々の生命科学系データベースや文献などをユーザーにとってわかりやすく、使いやすくしていこうという文部科学省の事業。平成 18 年度～22 年度まで ROIS/DBCLS を中核的な機関として、国内の 15 の大学や研究機関とともに実施している。プロジェクトの一環として開発しているさまざまなサービスを「統合ホームページ(LSDB)」(<http://lifesciencedb.jp/>)より無償で提供している。

※3 H-InvDB (H-Invitational Database)

ヒトの遺伝子と転写産物を対象とした統合データベース。ヒトの転写産物を、あらゆる手法で解析することにより、ヒト遺伝子の構造、選択的スプライシング変異体、機能性 RNA、タンパク質としての機能、機能ドメイン、細胞内局在、代謝経路、立体構造、疾病との関連、遺伝子多型 (SNP、マイクロサテライト等)、遺伝子発現プロファイル、分子進化的特徴、タンパク質相互作用、遺伝子ファミリーなどの精査されたアノテーション (注釈付け) 情報を提供している。
(参照 <http://h-invitational.jp/>)

※4 経済産業省産業技術研究開発「統合データベースプロジェクト」

経済産業省の関連機関により実施された生命科学分野の研究開発プロジェクトの成果のうち、データベースに関する情報提供サイトを構築し、運用する経済産業省の事業。平成 20 年度～22 年度まで産総研と JBiC が実施している。プロジェクトの一環としてさまざまなサービスを「MEDALS (METI Database portal for Life Science)」(<http://medals.jp/>)より無償で提供している。

※5 LEGENDA (Literature-Extracted GENE-Disease Associations)

MEDLINE 文献に書かれた遺伝子、遺伝子機能、疾患、または基質を示す用語のうち、2 つが同時に出現 (共起) した文献を効率的に探すことができる産総研のデータベース。遺伝子 2 つのよ様な同じタイプでも検索できる。独自の遺伝子名辞書をもっている。
(参照 <http://h-invitational.jp/legenda/top.htm>)

**※6 総合科学技術会議 ライフサイエンス PT(プロジェクトチーム) 統合データベースタスク
フォース報告書**

総合科学技術会議 ライフサイエンス PT (座長 本庶 佑 議員)において、平成 20 年 12 月～平成 21 年 4 月にかけて統合データベースタスクフォースが設置され、ライフサイエンス分野におけるわが国の恒久的かつ一元的な統合データベースの具体的仕組みと、その整備に向けたロードマップについて取りまとめられた。

(統合データベース タスクフォース報告書について)

<http://www8.cao.go.jp/cstp/project/bunyabetu2006/life/14kai/siryo1-1.pdf>

<http://www8.cao.go.jp/cstp/project/bunyabetu2006/life/14kai/siryo1-2.pdf>

<http://www8.cao.go.jp/cstp/project/bunyabetu2006/life/14kai/siryo1-3.pdf>