

## ライフサイエンス統合データベースセンターの活動

### 戦略立案

近年、国内においても多くのデータベースが開発・公開されていますが、それらが広く有効に利用されているかというと、そうはいえないのが現状です。その原因のひとつには、データベースの整備戦略が明確になっておらず、個々のデータベースのあいだの連携が不十分であることが考えられます。大量のデータの存在を抜きに生命科学やバイオ産業の進歩を語れない時代となった今日、そのようなデータを有効に活用できる環境をいかに作り出すかを考えることが重要となってきているのです。

そのため当センターでは、戦略立案のための参考情報としてデータベースの現状や動向の定期的な調査を行い、調査結果を公開してきました。また、さまざまな立場かつ国家的な見地からデータベース整備戦略を議論し立案するた

めの委員会を設けており、そこにはメンバーとして、PDBj, DDBJ, KEGGをはじめとする我が国を代表するデータベースの主催者が参加しています。会議の資料などは「統合ホームページ」(<http://lifesciencedb.jp/>)において公開しています。(\*1)

また、著作権や個人情報の取り扱い、情報セキュリティといった、データベースの統合化に伴うさまざまな問題についても専門家を交えた検討を進めてきました。(\*2)

プロジェクトの成果や検討内容は、シンポジウムなど(\*3)により広く公開するとともに、毎年、登録ユーザーによるサービス評価を実施するなどして、有用なサービスの提供を心掛けています。(\*4)

### 公開サイト

- \*1 「統合ホームページ」→「統合 DB 事業」カテゴリ内 『運営委員会等の資料』
- \*2 「統合ホームページ」→右下のバナーリンク 『LSDB Social Action』
- \*3 「統合ホームページ」→「統合 DB 事業」カテゴリ内 『シンポジウム・講演会』
- \*4 「統合ホームページ」→「統合 DB 事業」カテゴリ内 『外部評価』

### 統合データベースの運用・開発

生命科学系のデータベースは国内に 500 以上、全世界では 1 万以上もあるといわれており、そのなかから目的のデータベースを探し出すのは非常に難しくなっています。また、データベースはそれぞれに作り方が異なり、その使い方を習得するにも時間がかかります。

このような状況を改善するため、当センターでは、個々のデータベースや解析ツールへの窓口となるポータルサイト「統合ホームページ」を公開し、データベースカタログや、横断検索、統合 TV など、簡単にデータベースにたどりつき使いこなすためのサービスを提供しています。また、データベース同士を統合利用するためのウェブサービスやツール、辞書やオントロジーなど、基盤技術の開発にも力を入

れています。プロジェクト全体としては、医学、薬学、糖鎖科学の研究機関と連携して統合データベースの開発を進めています。こうした各分野での統合化を進め、ポータルや基盤技術を充実させることによって、データベースへのアクセスは飛躍的に容易となり、それが私たちのめざす統合データベースへの第一歩になると考えています。さらに、国内の研究機関で作成されたデータベースの寄託を受けてデータの保全と共有を図るなど、これまでになかった課題にも取り組んでいます。

個々のサービスの詳細については「統合ホームページ」をご覧ください。サービス内容については動画やリーフレット（サービス名の隣のリンク）でも紹介しています。

### 普及啓発・広報活動

質の高いデータベースの構築に不可欠な専門職として、アナテーター（データに生物学的な解釈を加える）、キュレーター（データベースの編集作業に従事する）、システムデータベース管理者を養成するため、当センターでは、プロジェクトの参画機関と協力して専門的な教育活動を行うとともに、その教材をオンライン上に公開しています。また、全国各地の大学や研究機関の協力のもと、プロジェクトの提供するサービス、生命科学系のデータベースやツールの利用法を紹介する出張講習会を行っており、講習会の資料も公開しています。(\*1)

上記の講習会の一部をはじめ、学会などでの講演やセミナーの動画なども配信しています。(\*2) 動画は USTREAM で配信することもあります。

プロジェクトのさまざまな活動は、随時「統合ホーム

ページ」の『新着情報』でもお知らせしており、twitter への配信も行っています。また、学会での発表や展示などを積極的に行うことにより、直接、ユーザーの意見や要望を聞くよう努めています。



講習会の様子とポスター



### 公開サイト

- \*1 「統合ホームページ」→「教材・人材育成」カテゴリ内 『MotDB』
- \*2 「統合ホームページ」→「教材・人材育成」カテゴリ内 『統合 TV』

## ライフサイエンス統合データベースセンター

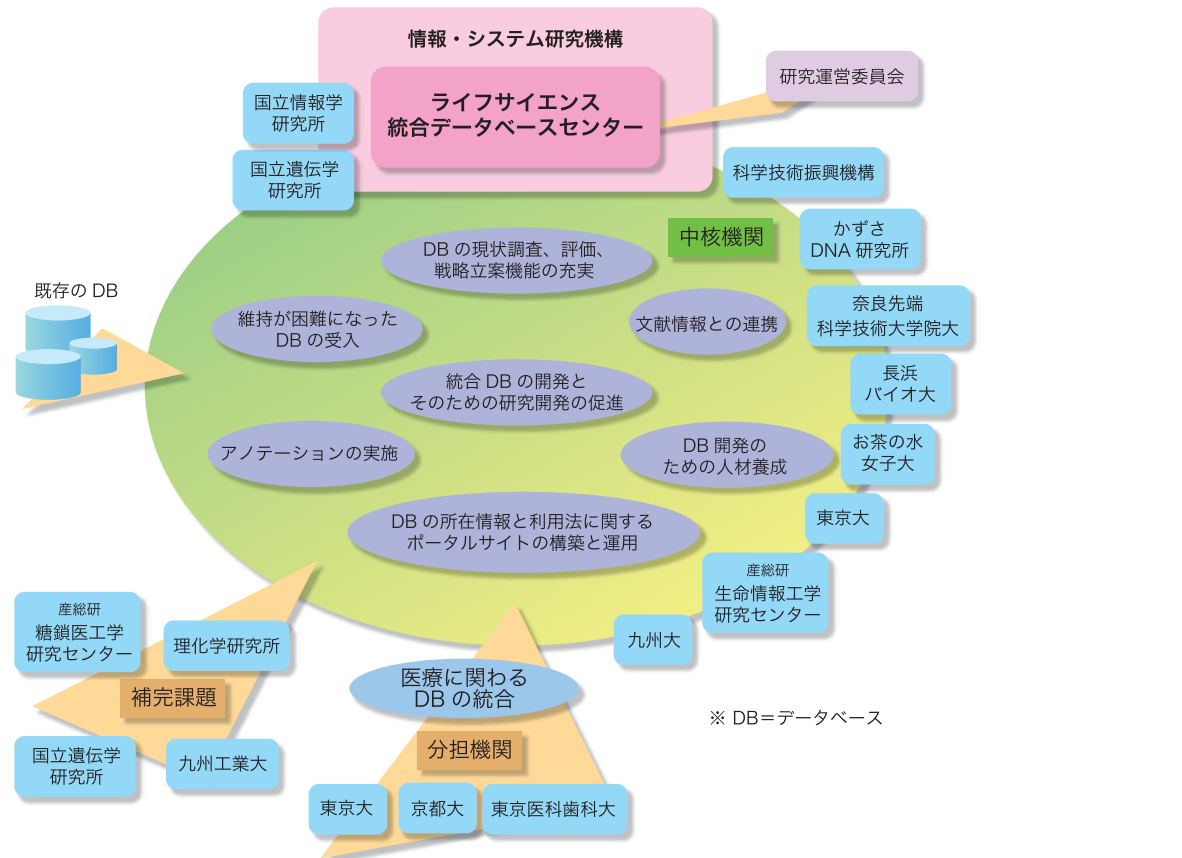
<http://dbcls.rois.ac.jp/>

ライフサイエンス統合データベースセンター (DBCLS: Database Center for Life Science) は、日本のライフサイエンス分野におけるデータベース統合化の拠点を形成することを目的として、平成 19 年 4 月、大学共同利用機関法人 情報・システム研究機構内に設立されました。

ライフサイエンス分野では、急速に進化したゲノムプロジェクトやポストゲノムプロジェクトにより、大量の情報が産み出されるようになりました。これらの情報は、多種多様なデータベースに蓄積されており、これらを「うまく使いこなすことが研究の進展にもつながります」が、利便性の点でさまざまな問題を抱えています。

当センターは、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」の中核機関として、自らをライフサイエンス分野のデータベースに関するサービスの専門機関であると位置づけ、関連機関と協力して事業を推進し、データベースにおける諸問題に取り組んでいます。

また、国内のデータベースを中心に、データベースの統合化と保全に努め、利用者の利便性を高めるための「情報技術の開発やポータルサイトの整備」を行っており、ライフサイエンス分野の研究者のみならず、データベースに関する窓口として十分に利用していただけるサービスを提供することを目指しています。



大学共同利用機関法人 情報・システム研究機構

**DBCLS** ライフサイエンス  
Database Center for Life Science 統合データベースセンター

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

TEL : 03-5841-6754  
FAX : 03-5841-8090  
URL : <http://dbcls.rois.ac.jp/>

本パンフレットの内容は改良のため予告なく変更されることがあります  
2010年5月現在

東京大学  
工学部 12 号館  
(1F に生協売店)

○ 銀座線 ○ 日比谷線  
■ JR 線 ■ 京成線 上野駅 (20分)  
↑ 不忍池

千代田線 根津駅 (5分)  
[ 浅野地区 ]  
[ 弥生地区 ]

東京大学  
[ 本郷地区 ]  
安田講堂

大江戸線 本郷三丁目駅 (20分)  
丸の内線

正門 赤門

南北線 東大前駅 (10分)

<http://lifesciencedb.jp/>




<http://lifesciencedb.mext.go.jp/>

### 生命科学関連情報のカタログサービス

生命科学に関する種々の情報をカタログ化

生命科学系のデータベース、学会や協会、主要プロジェクトがカタログ化されており検索が行えます。データベースと学協会はタイトル状に表示され、個別にコメントを記入することもできます。主要プロジェクト一覧では、プロジェクトの成果として公開されているデータベースやそのダウンロードサイトへのリンクも掲載しています。

このほか、生物種のアイコンデータ（200種以上）や、これまでに作成されたポータルサイトもここからリンクされています。登録数：データベース>800件、学協会600件、プロジェクト67件(2010年5月現在)



プロジェクトカタログ

データベースカタログ

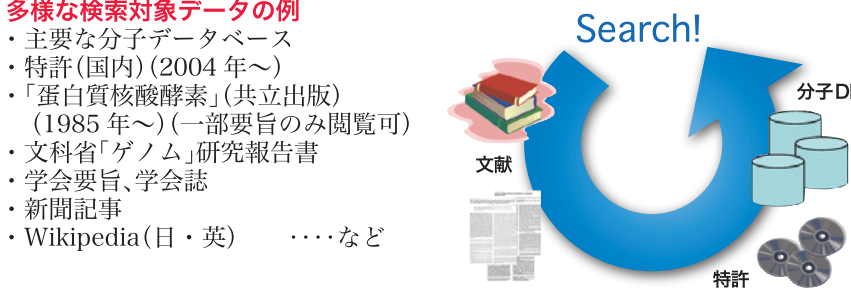
### 生命科学データベース横断検索

国内外の生命科学データベースや文献を横断的に検索

キーワードを一度入力するだけで散在する生命科学の分子データベースを一括して、特許や文献とあわせて検索できる、Google ライクなサービスです。入力には日本語でも英語でもOK、双方向に翻訳して検索します。検索対象データベース数：>250件(2010年5月現在)

**多様な検索対象データの例**

- ・主要な分子データベース
- ・特許(国内)(2004年～)
- ・「蛋白質核酸酵素」(共立出版)(1985年～)(一部要旨のみ閲覧可)
- ・文科省「ゲノム」研究報告書
- ・学会要旨、学会誌
- ・新聞記事
- ・Wikipedia(日・英) ……など



Search!

分子DB

文献

特許

### 生命科学系データベースアーカイブ

あのデータベースが丸ごとダウンロード可能に

研究者からさまざまなデータを寄託していただき、アーカイブ(ダウンロード可能な形)として提供しています。

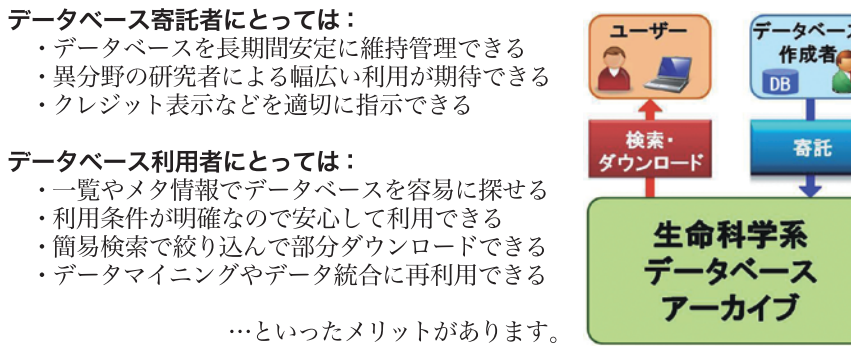
**データベース寄託者にとっては：**

- ・データベースを長期間安定に維持管理できる
- ・異分野の研究者による幅広い利用が期待できる
- ・クレジット表示などを適切に指示できる

**データベース利用者にとっては：**

- ・一覧やメタ情報でデータベースを容易に探せる
- ・利用条件が明確なので安心して利用できる
- ・簡易検索で絞り込んで部分ダウンロードできる
- ・データマイニングやデータ統合に再利用できる

…といったメリットがあります。



ユーザー

データベース作成者

検索・ダウンロード

寄託

生命科学系データベースアーカイブ

### MEDLINE 情報を利用したテキスト検索

OReFIL — 学術論文に引用されているデータベースやソフトウェアなど、オンライン資源を検索します。

Allie — 文献中に出現する略語とその正式名称の組合せ、および、その出典情報などが検索でき、検索結果を一括ダウンロードできます。

inMeXes — 文献に頻出する英語表現を文字の入力に合わせて検索します。検索結果を共有することもできます。



LSDB 生命科学系データベースプロジェクト

統合データベースプロジェクト

検索

アーカイブ

ツールと解析サービス

基礎技術開発

教材・人材育成

統合DB事業

連携

### BodyParts3D/Anatomography

解剖学用語を標準人体モデルに対応づけた解剖辞書とその利用環境

語彙は Terminologia Anatomica に準拠、用語の構造化には FMA (The Foundational Model of Anatomy Ontology) を用いています。各臓器の3次元モデルデータをそれぞれダウンロードして利用できるほか、作成した解剖図譜を画像あるいは動的 URL として使用できます。



### データベースの構築に関する基盤技術を提供

統合データベースプロジェクトでは、データベースを利用するために必要な技術だけでなく、それらを構築するためのさまざまなツールやシステムを提供しています。また、データベースに関する様式や決まりごとは国際的に標準化されることが非常に重要であるため、国際会議を主催してデータベース構築の周辺について環境を整備する努力もしています。

### 国際データバンクの新しい閲覧システム

キーワードや配列検索以外の方法でデータに到達、登録データの全貌把握が可能に

DNA データベース総覧と検索-検索でしか利用できなかった配列バンク(DDBJ/EMBL/GenBank)レコードをプロジェクトごとにまとめ、生物種と研究タイプで分類しています。絞り込み検索の結果も分類表示されます。

遺伝子発現バンク(GEO) 目次- NCBI (米国立生物工学情報センター)のGEO(Gene Expression Omnibus)データを高速に検索し、生物種や実験手法に分類して表示します。

### 参加機関による成果をすべて公開

統合データベースプロジェクトには、医療や糖鎖、化合物など、さまざまな専門性をもった機関が参加しています。それぞれが成果としてデータベースやツールなどを公開しており、すべて統合ホームページから紹介されています。

### 試験的公開サイト LSDB Lab.

本格公開前のサービスをここから公開しています。みなさんご意見・ご感想をお待ちしています。

<http://lifesciencedb.jp/>

LSDB で 検索



第3期「科学技術基本計画」(平成18年3月28日閣議決定)にもとづいて、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では、戦略重点科学技術の一つとして、「世界最高水準のライフサイエンス基盤整備」が掲げられています。また、連携施策群「生命科学の基礎基盤」データベース統合の為の調査研究(内閣府総合科学技術会議)の中間報告でも、**公共のためにデータを保存し統合利用のための最低限必要な環境を提供する**NCBI相当の機能を我が国に作る必要と指摘されました。

生命情報のデータベース統合化は、ライフサイエンス研究を支える基盤であり、その整備を進めるために必要な戦略の検討と技術開発を行うため、文部科学省では平成18年度より「統合データベースプロジェクト」を推進しています(平成22年度まで)。

本プロジェクトでは、特に我が国のライフサイエンス関係のデータベースの利便性の向上を図るため、ライフサイエンス関係データベース整備戦略の立案・評価支援、データベース基盤技術開発、ポータルサイトの整備等を行い、統合化を推進します。

本プロジェクトを通じて整備される「生命情報の統合化データベース」は、個々のライフサイエンス研究において蓄積されたデータが戦略的に統合され、付加価値の高いデータベースとして整備されるもので、ライフサイエンス分野の研究者のみならず、産業界、医療従事者などが幅広くこれを活用し、今後の我が国における**ライフサイエンス分野の科学技術の進展に大きく貢献**していくことが期待されます。

参考：「我が国におけるライフサイエンス分野のデータベース整備戦略のあり方について」ライフサイエンス委員会 データベース整備戦略作業部会 [http://www.lifescience.mext.go.jp/download/news/report\\_DB.pdf](http://www.lifescience.mext.go.jp/download/news/report_DB.pdf)

### 生命科学分野におけるデータベースの問題点と課題解決のための方策

<ul style="list-style-type: none"> <li>・所在や利用法がわからない</li> <li>・日本語化されていない</li> <li>・データベースを使いこなすための技術開発の不足</li> <li>・データベース構築、維持を担う人材の不足</li> <li>・各機関やプロジェクトで構築されたデータベースがなかなか公開されない</li> <li>・プロジェクト終了後はデータベースが維持されない</li> <li>・実験系と情報系の研究者間の相乗効果が小さい</li> <li>・実験データと文献の知識との連携が弱い</li> <li>・データベース整備の中長期的戦略を担う体制が未整備</li> </ul>	<ul style="list-style-type: none"> <li>・データベース利用窓口であるポータルサイトの構築</li> <li>・日本語による研究情報流通の促進</li> <li>・データベース利用のための基盤技術開発</li> <li>・ライフサイエンスデータベース構築を担う人材の育成</li> <li>・データベース受入/維持のための体制構築</li> <li>・アーカイブによるデータの価値向上</li> <li>・文献データの活用</li> <li>・戦略の立案および体制の構築</li> </ul>
---	--

### 統合化の年次計画

国内DB(約250)	~H19	H20	H21	H22(年度)
配列 ~100	250DB			DBポータル(所在情報)
発現 ~60		20DB	200DB	横断検索(キーワード検索)
蛋白 ~30				統合検索(高度な検索)
表現型 ~10				受入
二次DB ~50		5DB	15DB	30DB

産業・臨床・育種従事者の課題解決への展開

4省統合データベース

環境変化と知識データ増大への絶え間ない対応