参考資料（１）

報　告　書

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

平成１８年５月１７日

科学技術・学術審議会
研究計画・評価分科会
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

目次

（注）フッターにある（ ）付き番号は、
参考資料内のページ番号です。

1. はじめに

（データベース開発の歴史）

ライフサイエンス分野のデータベースは、古くは米国国立医学図書館 NLM (National Library of Medicine) による MEDLARS (Medical Literature Analysis and Retrieval System、1964 年)にさかのぼる。これは現在ライフサイエンス分野で最もよく利用されている文献データベース MEDLINE (MEDLARS On-Line、1971 年)の前身にあたるものである。もっとも紙ベースのものまでデータベースと呼ぶことにすれば、さらに、米国 American Medical Association による CIM (Cumulated Index Medicus、1879 年)や米国 Army Medical Library による CLML (Current List of Medical Literature、1937 年)までさかのぼることができる。これらはすべて医学関連の文献集であるが、1970 年前後からそれ以外の種類のデータベース、すなわち、今日我々がデータベースという言葉から想起する種類のデータベースが次々と作られるようになる。ジョンズホプキンス大のマキューージック博士による遺伝子変異疾患データベース MIM(Mendelian Inheritance in Man、1969 年)、米国エネルギー省のブルックヘブン国立研究所によるタンパク質立体構造データベース PDB (Protein Data Bank、1971 年)などがそれである。

ライフサイエンス分野の最も基盤的なデータベースである核酸配列のデータベースに関しては 1980 年に欧州分子生物学研究所 EMBL(European Molecular Biology Laboratory)で現在の EMBL-Bank の前身が作られ、米国では 1982 年にロスアラモス国立研究所で GenBank が産声をあげた。日本では 1986 年に国立遺伝学研究所で DDBJ (DNA Data Bank of Japan)が始動し、これら三者による現在の日米欧三極体制が出来上がった。

このようにライフサイエンス分野において古くからいろいろなデータベースが作られてきたが、データベースの開発と普及の観点で最も重要な契機は何といっても 1990 年前後に世界的に開始されたヒトゲノム計画であろう。これによりライフサイエンスの大量情報化時代の幕が開いた。冒頭に述べたように、これ以前にも多くのデータベースが作られてはいたが、利用者は一部の研究者に限られていたし、その利用法も限定的なものであった。それが、ヒトをはじめとする種々の生物のゲノム計画が相次いで始まったことにより、また、インターネットの普及やデータベース技術の進歩などによりデータベースの利便性が高まり、利用が一気に拡大した。その後、配列データに加え、遺伝子の発現データ、タンパク質の立体構造や相互作用などのデータが続々とデータベース化され、さらに利用が拡大しつつあることは周知の通りである。データベースは質的にも量的にも種類的にも急成長を遂げ、ライフサイエンス研究に欠かせないものになったのである。

さて、このようにデータベースがライフサイエンスの発展の重要な鍵になるという認識は、昨日今日なされたわけではない。このような時代が到来することはゲノム計画開始当初にすでに一部の研究者には予想されていた。その当時の文献を紐解くとデータの産出よりもその管理や解析のほうが大変で重要だということがいまから 15 年以上も前から指摘されていたことが分かる。この認識は、我が国においても同様で、例えば、1991 年に東京大学医科学研究所にヒトゲノム解析センターが作られたとき、最初にゲノムデータベース分野が設置されたことからもそのことが伺える。このようなデータベースの重要性への理解を背景に、我が国においては、ゲノム関連の科学研究費補助金などにおいて 1991 年頃よりデータベース開発への積極的な取組みが行われてきた。また、これと並行して、国立遺伝学研究所生命情報研究センターや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の設置などデータベース構築を目的の一つに掲げた組織の整備も行われた。これにより
我が国においても多くのデータベースが開発され、一般の利用に供されてきた。そうした中から、関係者の献身的な努力の甲斐もあり、世界的に誇れるデータベースがいくつか生まれてきた。

しかしながら、その一方で、せっかく苦労して作られたにもかかわらず、あるいは、当初は世界でも最先端を行くものであったにもかかわらず、維持更新されずに朽ちていったデータベースも少なくない。この原因の多くは、データベースでは継続的な維持更新が最も重要にもかかわらず、多くは競争的資金で、しかも個々の研究者の個人的な努力により開発が行われていたことに求めることができる。また、データベース構築が高度に知的な活動であることに対する理解や評価の低さもその背景にあったであろう。データベース作りは個々の開発者の創意工夫が必要であるという点に着目すれば研究としてとらえることができるが、一方では組織をあげて長い時間をかけて地道に作って行くものであるという点からは事業としての性格をもつものである。データベース作りはこのような二面性をもつ特殊なものであるにもかかわらず、いわゆる研究と同じ枠組みの中で扱われ評価されてきたことが我が国においてデータベースの健全な発展を阻害してきたと言えよう。

さて、いまから 6、7 年前の 2000 年前後にデータベースの構築や維持更新に関するこのような問題が顕在化するのとほぼ時を同じくして、それを担うバイオインフォマティクス分野の人材不足も大きな社会問題として浮上してきた。データベースの問題も含め、バイオインフォマティクスのこのような状況を改善するためには、どういう取り組みをすればよいのであろうか？その当時、この件に関して、さまざまなところでさまざまな議論や取り組みが行われたが、その中でも最も具体的かつ実効性があったものの一つが、その当時の科学技術会議ライフサイエンス部会ゲノム科学委員会において行われた議論とそれに基づく提言である。その内容は 2000 年の 11 月に出された報告書「ゲノム情報科学におけるわが国の戦略について」にまとめられている。この中では、バイオインフォマティクスの人材養成、研究開発の振興、データベース整備戦略、の 3 つの課題に関して推進方策が提言されている。そして、この提言をもとに、2001 年度より科学技術振興機構(JST) にバイオインフォマティクス推進センター (BIRD)が設立され、これら３つの課題について精力的な取組みが行われてきた。

この報告書で提言された３つの課題（人材養成、研究開発振興、データベース整備戦略）の推進方策は 5 年経ったいまでも決して古くはなっていないし、BIRD での取組みも大きな成功を収めてきた。この５年間にゲノム、トランスクリプトーム、プロテオーム、メタボロームなどで大規模なデータが次々現れてくるなど、表面的には状況が大きく変化しつつある面もあるが、ここで示された提案内容の多くは基本的にいまでも十分通用するものである。しかしながら、データベース整備戦略に関して言えば、提言されたことがすべて実現されたわけではないこと（例えば、国家レベルの整備戦略を立案する機能や、それを実施する中核的な機能は実現されなかったこと、など）、複数の省庁におけるライフサイエンスに関連するデータベース開発の活動はあったものの文部科学省だけの動きに留まった面があること、BIRD はその守備範囲からして国家的な戦略を考える組織としては不十分であったこと、その後ライフサイエンス分野で続々と行われることになる大型プロジェクトのデータベースの受け皿作りが 2000 年の報告ではあまり考慮されていなかったこと、データベースが予想をはるかに超え多く作られるようになったこと（そのためのポータルサイトや統合化などの必要性が増したこと）、オントロジーや文献の知識などの重要性が

増したこと、など、更なる体制の強化を図ったり、戦略を再度見直したりすることが必要な面が出てきた。また、期せずして産業界からもデータベース整備の強化や見直しを求める声もあがってきた。

（データベース整備戦略作業部会における検討）

そこで、ライフサイエンス委員会の下に、データベース整備戦略に関する作業部会を設けて　2005年年8月12日、11月10日、2006年1月16日、2月28日、3月24日、5月11日の計6回にわたって、ライフサイエンス分野における今後のデータベース戦略をどうすべきか議論を重ねてきた。この作業部会では、ライフサイエンス分野のデータベースとしては、ゲノム、トランスクリプトーム、プロテオーム、メタボローム、などの網羅的かつ基盤的なデータ（いわゆるオームデータと呼ばれるもの）とそれを解釈するためのパウスェイやオントロジーなどの知識のデータベース化、すなわち、ライフサイエンスの基盤となるデータベースに重点をおいて議論を展開してきた。また、利用者もこれらのデータベースを使って研究開発を行う研究者、技術者をおもな対象として議論を進めてきた。当然のことながら、ライフサイエンス分野のデータベースあるいはバイオ分野のデータベースという範疇には、これら以外にもさまざまなデータベースが存在する。これには、生物資源等の研究用材料に関するもの、医療現場で用いられる臨床情報や医薬品情報、化合物の構造や毒性情報、食品の成分や安全性に関するもの、作物や家畜の育種に関するもの、産業上有用な微生物の情報、など多岐にわたるものが該当する。これらのデータの重要性は、オームデータに比べて決して劣るものではないし、本作業部会でおもな議論の対象とした基盤的データとこれらのデータとの連携・統合を図ることが重要であるが、これらの多くは、文部科学省以外の省庁で精力的に取り組みがなされており、また、現在科学技術振興調整費「科学技術連携施策群の効果的・効率的な推進」の一テーマとして調査研究が進められていることもあり、十分な配慮はしたが、検討のおもな対象とはしなかった。

このように本作業部会では、ライフサイエンス全般のデータベースを視野に入れながらも、その基盤となるようなデータベースに重点をおき、ライフサイエンス研究におけるデータベースの意義や重要性、世界的なデータベース開発の動向、我が国のデータベースの現状と問題点などについて綿密な調査を行うとともに、今後のデータベース整備戦略の基本的な考え方や推進方策あるいはその実現に向けた体制作りについて活発な議論を戦わせてきた。本報告書はそれらの結論をとりまとめたものである。ライフサイエンス研究においてデータベースの重要性が今後ますます高まることは万人の認めるところである。本報告書での提言が速やかに実施されることが、我が国におけるデータベース開発の発展、ひいては、それを基盤としたライフサイエンス研究、医療、バイオ産業の発展につながるものと確信している。なお、本報告書では、主たる検討の対象をライフサイエンスの基盤的なデータベースに絞ったが、上に述べたように、ライフサイエンス分野のデータベースは多岐にわたる。また、現在、関係各省を横断的に俯瞰したライフサイエンス分野のデータベース統合に関して、内閣府を中心に活発に議論が進められていることから、本報告書で取り上げた提言の具体的実施に際しては、これらの議論や提言と十分な整合性をもって進める必要がある。

2.ライフサイエンス研究におけるデータベースの意義および整備の必要性

1977 年のマクサムとギルバートおよびサンガーによる DNA 塩基配列決定法の開発を端緒とし、その後 1990 年頃より開始されたヒトゲノム計画の進展により、ライフサイエンスは莫大な量の DNA 塩基配列データ産生の時代を迎えた。現在はこれに加えて、いわゆるポストゲノム時代を迎えて、タンパク質の立体構造データや遺伝子の発現データも爆発的に増えている。このようなデータの洪水の中で、データベースの活用なしにライフサイエンス研究を行うことは事実上困難になってきている。また、実験データだけでなく、文献数も着実に増加しており、米国の文献データベース MEDLINE に登録されている文献の総数は 1500 万件を超え、一人の研究者が関連する分野の論文すべてに目を通すのは不可能な状況になっている。こうした情報の洪水という状況下では、一方で、データベースを活用することにより、従来不可能であったような質と量のデータを個々の研究者が利用できるようになっており、これが現在のライフサイエンス分野の研究開発効率の飛躍的向上を可能としている。今後も情報の増加が加速することが予想されるライフサイエンス研究をさらに進展させていくためには、より網羅的かつ正確で付加価値のついたデータベースを整備し、活用していくことが不可欠である。

データベースの整備は、研究開発の効率化のためばかりではない。ライフサイエンスという多岐にわたる学問体系を生命のシステムとして再統合し、俯瞰することにより、生命の理解がより深まり、ライフサイエンス研究の一層の進歩が見込まれる。データベース整備はその礎となるものであり、国内における整備が不可欠である。ゲノムネットワーク、タンパク 3000、国際Ｈａｐ Ｍａｐプロジェクトなどの近年のライフサイエンスの大規模プロジェクトに期待される成果の一つはデータベース整備にあるといってもあながち間違いではない。すなわち、データベースはライフサイエンス研究に不可欠の基盤であると同時に、次世代の研究への糸口を提供する。あわせて、医学の発展を通しての国民の健康への貢献、食糧問題、環境問題、資源問題への貢献、さらには産業利用など、ライフサイエンス研究の成果の適切な活用という観点からも、データベースの公開・整備は重要である。

一方で、データベースの構築、提供は、単なる既存の情報の提供サービスに止まるものではなく、その知識の蓄積が新たな研究分野を作っていくものであり、その構築時には予想もしていなかった研究の展開が期待できる。例えば、ヒトやマウスの完全長 cDNA 配列データ（巻末の用語解説参照）がデータベースに多数蓄積された結果、タンパク質に翻訳されないが機能をもつ RNA 分子が多数存在すること、多くの遺伝子が読み方を変えることにより複数種類のタンパク質を作りうること、などの発見により多様なＲＮＡに着目した新たな研究分野が拓けたことなどがそれである。また、累積されたデータを整理・統合するということ、すなわち、データベース開発を推進することは、とりもなおさず、生命を担う構造の空間軸、時間軸を俯瞰する能力のある人材を育成することにもつながる。これにより将来新たな視点でのライフサイエンスを拓くことが可能になる。また、データベースは、多額の税金を使って行われるライフサイエンス研究の成果を医療、民間企業、さらには一般社会に還元するための手段としても重要である。データベース化により、その成果が誰の目にも明らかにできるからである。

後述するように、このような意義や重要性をもつデータベースへの理解や整備が我が国では遅れており、学界のみならず産業界からもデータベース整備に対する強い要望が寄せられている。例えば、ゲノム分野の国家プロジェクト等の成果を広い分野で迅速に実用化研究に活用できるよ

うに、一元的に集約・統合され様々な角度からデータを参照できる、無償公開を原則としたデータベースを国が積極的に整備してほしいなどの要望がある。

以上、ライフサイエンス研究の観点からも、産業の観点からも、更なるデータベース整備の強化充実がいままさに求められている。そのためには、今後、公開性・透明性・客観性・科学性を担保しつつ国内でデータベースを戦略的に整備していくことが重要である。その際、データベースを構築する側の立場に立った整備ではなく、それを利用する側の立場に立った整備に努めることが必要である。このためには、情報系、実験系との緊密な共同作業によるデータベース整備がなされなければならない。また、データベースの構築と普及には長い年月を要するため、また、いわゆる研究とは異なる側面をもつため、上記の整備は定常的な経費をもって永続的に続ける必要があることは言うまでもない。

3. 国内外のデータベース開発の現状と動向

3-1 データベース開発の世界的動向

DNA 塩基配列の登録データ量が指数関数的に増加してことはよく言われていることであるが、英国の科学雑誌である NAR (Nucleic Acids Research)が毎年 1 月に発行しているデータベース特集号に登録されているデータベース数からみると、ライフサイエンス分野のデータベースそのものの数も指数関数的に増大していることが伺える。また、タイトルにデータベースという記載のあるライフサイエンス分野の論文数から判断すると、データベースの累計は一万にも及ぶと推定される。また、その種類も多岐にわたってきており、文献から抽出したデータやオントロジー（巻末の用語解説参照）といった知識に関するデータも急激に増加してきている。また、DNA やタンパク質の配列や立体構造といった生体関連物質の構造に関わるデータばかりでなく、遺伝子やタンパク質の発現や相互作用といった生体関連物質間の関係に関わるデータ、およびパスウェイ（用語解説参照）、疾患、表現型という機能に関わるデータも数多くデータベース化されるようになってきた。すなわち、生体関連物質の個々の部品のデータから、それらが構成するシステム全体に関わる情報のデータベース化へと、開発の重点が移りつつある。また、DNA 配列データバンクのようなデータ生産者からの一次データが登録されるデータ登録型のデータベースだけでなく、すでに登録された様々な分野のデータを加工した、あるいは、文献から抜き出したデータを収録した、知識集約型の二次的なデータベースも増加してきている。

以上紹介したように、ライフサイエンス分野のデータベースは量的にも、質的にも増加、拡大してきており、研究面でも、産業応用面でもその重要性はますます高まってきている。すなわち、データベースはライフサイエンス研究や医療、バイオ産業の国家戦略を考える上で欠かせない大きな柱の一つであるとの認識が世界的に広がっている。

3-2 欧米におけるデータベース整備の現状

本報告書の冒頭「1. はじめに」に記載したとおり、米国では、1964 年には、国立衛生研究所 NIH(National Institutes of Health)の下部組織である NLM で現在の文献データベースサービス PubMed につながる活動が始まっている。その後、1969 年に遺伝子変異疾患データベース MIM が、1971 年にタンパク質立体構造データベース PDB が、1982 年に核酸配列データベース GenBank が、というように、現在のライフサイエンス研究になくてはならないデータベースが続々と誕生した。欧州では、1980 年に欧州分子生物学研究所 EMBL において核酸配列データベース EMBL-Bank の前身が形作られた。その後、米国では 1988 年に NLM の配下に国立バイオテクノロジー情報センター NCBI(National Center for Biotechnology Information)が設立され、欧州でも、1992 年に EMBL の下部組織として欧州バイオインフォマティクス研究所 EBI(European Bioinformatics Institute)が設立された。これらの組織は、それぞれ GenBank、EMBL-Bank をはじめとするライフサイエンス分野のデータベースの開発および維持更新を専門に担うとともに、バイオインフォマティクスの研究とデータベースや情報解析のサービスの中核拠点として機能するように設けられたものである。

表 1 に NCBI と EBI の概要を示した。NCBI は、NLM の一部門として設立されたものであり、予算規模は 85 億円、人員規模は約 400 名である。サービスは GenBank を中心とする核酸配列データを中核に、標準配列である RefSeq データベースの提供、Entrez システムによる統合データベース環境の構築、および世界標準的な相同（ホモロジー）解析ソフトウェアである BLAST を中心とする各種解析ソフトウェアにその特徴がある。さらに、国立医学図書館 NLM の下部機関としての特徴を活かした文献データベースサービス(PubMed)の提供は、他にはない大きな特徴である。一方、EMBL の一部門である EBI の予算規模は 32 億円、人員規模は 300 名弱である。サービスの特徴には、UniProt や InterPro といったデータベースに代表されるタンパク質配列を対象とした機能情報の提供と、Ensembl と呼ばれるデータベースにおける真核生物のゲノムを対象とした詳細なアノテーション（データに生物学的医学的な解釈を加えること）情報の提供などがある。資金的には、英国のウエルカム財団や米国 NIH のからの資金も得て活動している。そのため、他機関との共同開発も多い。NCBI、EBI とも人員の二割から三割程度の研究部隊を抱えており、単にデータベースの整備に限定された組織ではなく、データベースの整備やサービスの提供とそれに関連する研究開発とが対になった組織が形成されている。また、これらの組織ではデータベースの開発が明確な目標をもったプロジェクト制で実施されている。

3-3 我が国におけるデータベース整備の現状

日本においても欧米同様にライフサイエンス分野のデータベースは、医学関連の文献情報にその起源を求めることができる。1903 年の医学中央雑誌がそれである。その後だいぶ時は下るが、1958 年に JST において科学技術文献速報が発行され、1976 年にはオンラインデータベースサービスが開始されている。ただし、これらはライフサイエンスだけでなく科学技術分野全般を対象としたものである。

核酸配列データに関しては、前述したように、1986 年に DDBJ が国立遺伝学研究所で産声をあげ、米国の GenBank、欧州の EMBL-Bank と合わせた日米欧の三極体制がこのときに形作られた。その後、NCBI、EBI の設立と同じような時期に、ゲノムネットと呼ぶ国際的なバイオ情報サービスが京都大学化学研究所と東京大学医科学研究との連携にもとに立ち上がった。また、東京大学医科学研究所にヒトゲノム解析センターが、国立遺伝学研究所に生命情報研究センターが、京都大学化学研究所にバイオインフォマティクスセンターが次々と設置され、データベース構築やバイオインフォマティクス研究の下地が整えられた。しかしながら、欧米の NCBI や EBI に匹敵するような中核機関の設置までには至らなかった。ちなみに、我が国におけるセンターとしては最大規模を誇り、また、日米欧との三極体制を担う DDBJ は年間予算 12 億円で、人員は事務員も含めて約 60 名で活動している。実態的には、国立遺伝学研究所の生命情報・DDBJ 研究センターの 5 つの研究室が基盤となっており、タンパク質の構造、機能に関するデータベースである GTOP や遺伝子発現に関するデータベース CIBEX など、独自のデータベースの開発も行っている。

上述のようなデータベース構築やバイオインフォマティクス研究の振興を目的としたセンターの設置の動きに加え、前述の「ゲノム情報科学におけるわが国の戦略について（平成 12 年 11 月科学技術会議ライフサイエンス部会ゲノム科学委員会）」を受けて科学技術振興機構 JST が 2001 年にバイオインフォマティクス推進センターBIRD を設立し、データベース整備やバイオインフォマティクス人材養成に関する競争的資金を拡充した。この枠組みによって、京都大学の KEGG(Kyoto

Encyclopedia of Genes and Genomes)や大阪大学のPDBj(Protein Data Bank Japan)といった世界的に定評のあるデータベース構築や国際協調によるデータベース整備への支援などが進められてきた。KEGGは、細胞レベルでの生命システムの機能に関する知識を分子間相互作用ネットワークの情報としてデータベース化したパスウェイデータベースを中心に、遺伝子カタログ情報(GENES)、生体関連化学物質情報(LIGAND)、機能情報(BRITE)などから構成される一種の統合データベースであり、論文からの引用が多いことでも知られている。また、PDBjは、米国のRCSB(The Research Collaboratory for Structural Bioinformatics)およびEBIのMSD(Molecular Structure Database)との連携のもとに運営されているタンパク質の立体構造データベースである。PDB自体は、前述したようにもともとは米国のブルックヘブン国立研究所により運営、維持されていたものであるが、現在は上記の体制で、国際的な連携のもとに運営されている。PDBjでは、XML(Extensible Markup Language)などの最新情報技術を利用した新しいデータ記述と解析ソフトの開発、およびタンパク質表面形状と物性に関するデータベースeF-siteなどの二次データベースの開発を行っている。

また、文部科学省科学研究費補助金の特定領域研究の中でも、研究成果の一環としてデータベースの構築が行われている。「ゲノム4領域」では、研究成果としてのデータベースリストを公開しており、現在68件のデータベースが登録されている。また、「発生システム」でも、ユウレイボヤなど6種類の個別生物に関するデータベースが開発され、公開されている。その他、一般の科学研究費補助金などでも多くのデータベースが開発されているものの、全体にどの程度のデータベースが構築されているかの把握はなかなか難しい。英国の科学雑誌であるNAR の2006年1月のデータベース特集号、JSTバイオインフォマティクス推進センターBIRDのデータベースディレクトリサイトWING、上記の文部科学省科研費特定領域のホームページ、知的基盤整備委員会による平成16年11月のデータベース見直しリスト、および平成16年度学術情報データベース実態調査などから総合的に判断すると、公開予定のものも含めて文部科学省関連では190件程度のデータベースがあるものと推定される。ただし、これらのうちの約三分の二は一種類の情報にしか登録されていないため、さらに多くのデータベースが水面下にある可能性は否定できない。

さらに、文部科学省以外の他省に目を転じてみると、まず、経済産業省関係では1998年以来進められてきたヒト完全長cDNAの構造解析プロジェクトで得られた成果を基盤に、タンパク質機能解析・活用プロジェクトが実施され、遺伝子の発現頻度解析により得られたデータについてデータベース化され公開されている。また、ヒト完全長cDNA構造解析プロジェクトで得られた完全長cDNAクローンの配列情報は、DDBJに登録されると同時に、世界各機関の全長cDNA配列情報とともにアノテーションがつけられ、H-Invitationalデータベースとして一般に公開されている。その他、産業技術総合研究所あるいは製品評価技術基盤機構からも20件程度のデータベースが公開されている。また、厚生労働省関係では、ミレニアムプロジェクトの一環として進められてきた「疾患データベース」プロジェクトの成果が、GeMDBJ データベースという形で公開されている。解析データは、アルツハイマー病、がん、糖尿病、高血圧、喘息からなる5疾患に関わるゲノムワイドなSNP解析（巻末の用語解説参照のこと）と、ファーマコジェネティックス（用語解説参照)に関するSNP解析、また一部の疾患等に関するチップによる遺伝子発現データを含んでいる。農林水産省関係では、1991年に開始されたイネゲノム解析研究の成果として、イネゲノム配列情報を中心に、遺伝地図情報、物理地図情報、EST 情報も含め幅広い、質の高い情報が蓄積され、それらはINEをはじめとするデータベースで公開されている。また、ブタのcDNA情報や蚕のゲノ

ム情報も解析され、それぞれ公開されている。データベースの件数としては、主要なもので 20 件程度になる。

3-4 日米欧の競争力比較

欧米では 3-2 に記載の通り、それぞれ国家戦略に基づき、データベースに関わる中核機関を設置し、この機関を中心にデータベースの開発、維持を一元管理している。さらに、NCBI、EBI とも独自の研究部隊をもっており、単にデータベースの整備だけでなく、将来のデータベースのサービスを見据えた研究開発も同時に行える体制になっている。また、ゲノム解析の時代からデータ産出（実験系）と強い連携をもってデータベース開発を進めてきた。一方、日本は表 1 に示すように、JST バイオインフォマティクス推進センターBIRD の人員は 61 名（JST 雇用者）、年間予算は 19 億円(平成 17 年度)であり、欧米の中核機関と比較して優位な状況にはない。予算の仕組みや事業内容が異なるため、厳密な比較は困難であるが、予算的には、BIRD に大学共同利用機関（例えば DDBJ）の予算を加えてはじめて欧州と対等の存在になるが、人員的には、その他の中核機関（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や東京大学医科学研究所ヒトゲノム解析センター）を含めても欧州と対等の存在になるかならないかである。米国には、予算的にも、人員的にも水をあけられたままである。ただし、上にも書いたように、何をもってデータベースに関する予算や人員であると定義するかはいろいろ難しい面があり、また、上記の見積もりには、我が国におけるデータベース構築・維持活動が網羅されているわけでもないため（欧米においてもしかり）、予算でも人員でも大幅に足りないと安易に結論付けるわけには行かない。これに関しては、現在内閣府を中心に進められている調査の結果を待つ必要があろう。ここでは、欧米と比較して予算面人員面で決して優位な状況にはないこと、むしろ遅れをとっている可能性が高いこと、米国との比較に関してはそれがより顕著であることを述べるにとどめておく。

さて、欧米との違いは予算面人員面だけではない。データベースの構築支援のあり方、戦略の立て方、統合化のレベル、データ産出側との連携、などにも大きな違いを見出すことができる。我が国では欧米と異なりプロジェクト制ではなく、研究者の創意に基づくデータベースを支援しているケースが多い。また、繰り返し述べているように日本にはデータベース整備に関わる中核機関がないなど、これまで国家戦略がなかったため、JST が資金提供してきた一部のデータベースを除いて、統合化、標準化が遅れている。また、データ産出側との連携の弱さやデータベース利用者のニーズの把握不足も大きな問題であろう。

以上、まとめると、予算面人員面の不足、国家戦略の欠如、その推進を担う体制や省庁を超えた連携の不備により、欧米と比較して、データベースの整備やその標準化・統合化が総じて遅れていると言えよう。そしてこのことが、バイオインフォマティクスの人材不足とあわせて、我が国におけるライフサイエンス研究やバイオ産業の競争力の低下の一因になっていると言えよう。

表１　日米欧の主要中核機関の概要

	日本 （中核的な機能を果たしている機関の例）		米国	欧州
	科学技術振興機構 バイオインフォマティクス推進センター	国立遺伝学研究所 生命情報・ＤＤＢＪ研究センター	国立バイオテクノロジー情報センター （NCBI）	欧州バイオインフォマティクス研究所 （EBI）
組織形態	独立行政法人科学技術振興機構（JST）の組織新しい生物情報の研究開発によるデータベースの整備等の推進と普及のための拠点　統括、副統括、事務局で構成	大学共同利用機関国立遺伝学研究所の付属施設。「生命情報学」の我が国における研究拠点。我が国を代表するDNAデータベースのDDBJを運営	米国ＮＩＨ傘下のＮＬＭの付属機関 分子生物学分野を支援する公共データや解析ソフトの提供と計算機を利用した基礎研究機関	EMBL の傘下の非営利学術機関　バイオインフォマティクスの研究とサービスの中心機関
組織の永続性	JSTの運営費交付金（バイオインフォマティクス推進事業）により運営	国立遺伝学研究所の運営費交付金により運営	根拠法：Public Law 100-607	上部機関EMBLは18カ国からの公的研究資金で運営されている。EBIの資金の半分を負担。やや不安定
予算	19億円	12億円	85億円	32億円
人員	61名（JST雇用者）	62名(事務員含)	約400名	283名
サービスの概要	生物情報データベースの高度化・標準化、バイオインフォマティクスの創造的研究開発、新しい情報生物学の創造のための起業支援センター	国際塩基配列データベースの共同構築と運営	配列情報データの標準配列(RefSeq)の提供やEntrezによる統合データベース、各種解析ソフト提供の世界的な中心 アクセス数：4000万/日	タンパク質配列を基礎とした機能情報(UniProtやInterPro)や真核生物のゲノム情報の統合サービス(Ensembl)
特色	KEGG、PDBjなど国際的データベースの開発支援	タンパク構造(GTOP)や遺伝子発現(CIBEX)など独自データベース開発	文献情報（PubMed）と配列情報、各種解析ソフトの充実、データベース間の連携	タンパク情報の充実
国内プロジェクトとの連携	国内の代表的データベースの構築･高度化を支援	ゲノムネットワークなど他プロジェクトに参画	スタッフは、塩基配列解析、遺伝子同定、遺伝子発現に関する実験的解析において、他のNIHの機関と協力	EU、NIHなどの研究資金を得てプロジェクトに参加している。英国のウエルカム財団、サンガー研究所と連携
自前の研究機能	なし	生命情報･DDBJ研究センターの5研究室で実施	基礎研究グループは、Computational Biology Branchの中にあり、70名のsenior scientist、staff scientist、research fellow、postdoctoral fellowからなる	バイオインフォマティクスの研究グループ１７（新規３）からなる。EBI独自のデータベースや機能サービスを担うグループを含む
教育機関人材養成との連携	バイオインフォマティクス分野の明日を担う人材の育成を目指した講座の開催	総合研究大学院大学傘下の研究機関として博士課程教育を実施。実習付の講習会の実施	"A Field Guide and NCBI Resources"がアメリカ全土で開催。コースは3時間の講義と2時間の実習	学位（PhD）取得を目指す学生から独立した研究者に対するコースを提供
その他特筆すべき事項			NLMには外部への研究資金配布機能がある。NIHの研究資金により実施された研究に由来する論文やデータについて、受け皿をNCBIやNLMが用意する	企業へ最先端の技術を普及することや企業からの寄付を得るための仕組みを整備

4. 我が国におけるデータベースの問題点と今後取り組むべき課題

4-1 データベースの問題点

3-3 で記載したように、我が国のライフサイエンス分野においても、これまで非常に数多くのデータベースが開発されてきている。その中には、国立遺伝学研究所の核酸配列データベース DDBJ、京都大学の統合パスウェイデータベース KEGG、大阪大学のタンパク質立体構造データベース PDBj、東京大学/科学技術振興機構の日本人一塩基多型データベース JSNP など世界に誇れるデータベースも一部にはあるが、多くのデータベースに関しては、各機関や各プロジェクトでバラバラにデータベースが作られ、所在情報が誰にでも分かるようにはなっていない、似たようなものがいくつもありどれを使ってよいか分からないなど、それらを関連付けて使おうとしたときに大変使い勝手が悪いという状況である。また、多くは十分な解析や解釈がなされず生データをただ格納したものになっており、また、臨床情報などの表現型データとの統合が十分でなく、医療や創薬その他の産業への応用が困難になっている。多くのデータベースが日本語化されていないことも広く応用展開していくことを困難にしている要因であろう。さらには、それらを使いこなして有用な生物学的あるいは医学的な知識を発見するための利用技術の開発も十分ではない。また、それ以前に、各機関や各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、仮に公開されても永続的に維持更新されないという問題がある。プロジェクト期間が終了すればそのまま放置される場合が多く、せっかくの成果が広く活かされないうちに消えてしまう。また、別の問題として、各機関や各プロジェクトで産出された実験データと文献に書かれた知識とを対応づけることが近年重要になってきているが、それらの連携が我が国では弱いという問題もある。

このような状況の背景には、制度上・予算上の問題、データベース構築への理解不足やそれを担う人材不足などがある。

まず、データベース構築、維持に関わる制度上の問題点としては以下の点が指摘できる。第一に、データベース整備は国家をあげて取り組むべき重要な課題であり、我が国のライフサイエンスの未来を決するものであるにもかかわらず、ライフサイエンス全般を見渡して、我が国のデータベース整備はどうあるべきか、目的や対象をどう設定すべきかを常に考え、戦略を立案する体制が十分には整っていないことである。JST バイオインフォマティクス推進センターBIRD や国立遺伝学研究所 DDBJ の委員会など一部にはそのような機能を期待されるものもなくはないが、それらの守備範囲は限定的であり、また、他に本業を抱える非常勤の委員からなる委員会がその役割を担っており、それらに我が国全般の戦略立案を期待することはできない。データベースの整備戦略の立案は片手間では行えない、また、高度な専門性を有する仕事である。

第二に、仮にデータベースの整備戦略がうまく立案できたとしても、それを速やかに効率よく実行に移す体制が整っていないことである。これに関しても前述の JST BIRD や国立遺伝学研究所 DDBJ その他の組織はあるが、予算の制約や制度上の問題から、多くは期待できない。これは、複数の制度、機関、研究分野を有機的に組み合わせる実施体制が整っていない、すなわち制度的、分野的縦割りを打破できる体制が整っていないことによるものである。

第三に、第二の問題とも関連するが、実験系研究者と情報系研究者間の共同研究を実施する土壌や体制が整っておらず、両分野の研究者による相乗効果が十分発揮できていないという問題である。そのため、データの解析や解釈が十分に施されず、いろいろなデータがバラバラなまま、データベース化されてしまうことにつながっている。

第四に、これが最も大きな問題かもしれないが、予算上の問題である。データベースを継続的に維持していくための予算的枠組みが十分でないこと、および期限付きの予算で作成されたデータベースの予算終了後の受け皿、すなわち、そのための予算確保の制度（例えば、プロジェクト経費の一部を必ずデータベース整備と維持管理に当てるなど）が整っていないことである。データベースは一般に長い間丹念に維持更新してはじめて大きな価値を生むという性格を有している。我が国のデータベースのほとんどが、多くの人が使う非常に基盤的なものでも、個々の研究者が行う、いわゆる研究と同じ扱いをされ、競争的資金によって維持管理されている。これが大きな問題を生み出している。

さて、このような制度上、予算上の問題以外にも、我が国において、データベースを開発することの意義、重要性が少なくとも研究の観点では必ずしも重要視されていないことが指摘されよう。新しいデータベースの開発については論文を出せるようにはなったが、データベースの維持・管理については論文化が難しく、論文出版が研究者に対する主な評価対象となっている我が国では、データベースの開発・維持に対する意欲刺激が高くならない。そのため、これらの開発や作業に従事する研究者およびデータに生物学的医学的な解釈を加える専門職員（アノテータ）やデータベースの編集作業に従事する専門職員（キュレータ）の人材不足を招いている。また、ライフサイエンス分野のことを十分に理解して、データベースシステムの開発や運用にあたるシステムエンジニアやオペレータなども不足している。その結果として、データベース自体、研究の片手間に作ればよいという風潮を招いている。これらのことが我が国で世界的に競争力のある、また、使い勝手のよいデータベースの開発を遅らせている要因になっており、また、構築されたデータベースの価値を正しく評価する目利きや仕組みが我が国に育っていないことにもつながっている。そのため、存在する個々のデータベースの重要度を評価し、支援すべきデータベースを選別することが難しくなっており、整備戦略の立案を困難にしている遠因になっている。

4-2 取り組むべき課題

4-1 で述べた問題を解決し、ライフサイエンス研究や産業応用に十分に役立ち、また、世界的な競争力を備えたデータベースを整備するには、これまでの整備方針や既存の組織やプロジェクトの見直しを図るとともに、それを踏まえて、上で指摘した問題点を解消するための制度作り、予算措置、体制作り、および、人材養成に取り組む必要がある。具体的には、以下の取組みが必要であると考える。

(1)ライフサイエンス、バイオ産業全般を見渡して、また、国家的視点に立って、データベースの整備戦略を立てること。また、そのための体制を整備すること。ライフサイエンス分野は急激に進展しているため、また、データベースは欧米との競争・協調の中で整備を進める必要があるため、戦略そのものよりも、常に我が国のデータベースの現状を質的・量的の両面から調査・評価し、それに基づいて戦略の見直しが柔軟に図れる体制作りが重要である。
(2)データベース整備は研究とは異なる事業的な側面をもつことを十分に認識し、DNA 配列やタンパク質立体構造のような、ライフサイエンスの基盤として不可欠なデータベース、あるいは、有

用性が広く認識され、かつ、世界的に競争力のあるデータベースを安定的に支援するような体制を整備すること。
(3)我が国で開発された種々のデータベースの所在情報や利用法などを漏れなく掲載したポータルサイトを構築し運用すること。
(4)高度に統合化されたデータベースを開発し、大学・研究機関等に加えて産業界における利用者の利便性を一層向上させること。また、そのための技術を研究開発すること。また、これらの推進に必要な体制を整備すること。その際、データベースを構築する側の立場だけでなく、利用する側の立場に十分に配慮すること。また、そのために、実験系の研究者や技術者がデータベース作りに参画する仕組みを確立すること。
(5)プロジェクト終了後にそこで開発されたデータベースを受け入れ、維持管理するための体制を整備すること。そのために必要となる予算をプロジェクト設置と連動させて、また、ライフサイエンスの進展状況に応じて機動的に確保する仕組みを制度化すること。
(6)種々の実験データと文献に書かれた知識とを対応させる仕組みを強化すること。
(7)未解釈、未解析のままのデータをアノテーションして付加価値を高めること。これを継続的に行う仕組みを確立すること。また、そのための、実験系、情報系の連携を図る仕組みを設けること。
(8)ライフサイエンスの展開に対応して、新しい種類の、あるいは、新しい発想に基づくデータベースの開発を支援する体制を整備すること。
(9)データベースの利用技術（バイオインフォマティクス）の研究開発を促進する方策を講じること。
(10)データベースを構築したり、そこから有用な知識を引き出したりできる人材（研究者、キュレータ、アノテータ、システムエンジニア、オペレータ、など）を養成すること。

これらの取組みに際しては、一部繰り返しになるが、データベースに対するニーズを十分に把握すること、すなわち、データベースを利用する側の意見を十分に取り入れる必要がある。そのために、データベースのおもな利用者である実験系の研究者や技術者がデータベース構築に深く関与する仕組みを作ること、また、そのために情報系と実験系が共同でデータベースを構築したり、そのための共同研究を行ったりすることを積極的に支援するための環境整備にも十分な配慮が必要である。また、新たな制度や体制の構築に際しては、既存の体制の限界を認識するとともに、ライフサイエンスの進展に応じて出てくる新しい要望に対して国をあげて柔軟に対応できる体制や制度を構築することが必要である。さらには、データベースはライフサイエンス分野の知的基盤であり、安定的に活用するという観点からの事業の継続性が最も重要である。

5. データベース整備戦略の基本的考え方

前節の後半 4-2 で述べた「取り組むべき課題」は、どれもこれも重要なものであり、一つとしてゆるがせにすることは許されない。また、我が国で作られている数多くのデータベース（3-3 参照のこと）にはそれぞれに存在意義があり、十分な評価を行った上でできるだけ支援することが望ましい。しかしながら、予算の制約もあり、また、データベースの問題は 2 節でも述べたように、我が国のライフサイエンスのあり方にも大きく関わる面があり、具体的な方策を講じるには十分な検討を加える必要がある。また、一方で、日本語化の問題を除けば、データベース開発は欧米との競争にさらされざるを得ず、そのような視点からも整備戦略を練る必要がある。3-4 で述べたように、一部のデータベースを除いて我が国のデータベースの現状は決して優位にあるとは言い難い。

このような状況を踏まえ、今後、日本がとるべきデータベース整備戦略の策定にあたっては、以下の方針で臨むことが大切であると考える。

第一に、データベースはこれまでのライフサイエンス研究の叡智をまとめた宝であると同時に、ライフサイエンスやバイオ産業にとって不可欠の基盤であるとの認識に立ち、国が構築を支援したデータベースは原則的に公開すべきである。

第二に、我が国におけるデータベース整備の見直しにあたっては、日本としての特徴（強み）を出せるように進めることが必要不可欠で、そのための新たな機能や目的の付加が必要である。単に重要なデータベースだから、あるいは、欧米でも作られているから、というだけでは支援するわけには行かない。

第三に、過去の研究資金の投資状況、国内の研究者の裾野の広がり、当該研究分野の国際的な位置づけ、民間資金における代替可能性などを総合的に分析し、データベースとして備えるべき機能や取り込むべき研究分野についてメリハリを付けること、優先順位付けが重要である。そのためには、整備戦略立案は質的・量的両面にわたる調査・評価のしっかりとした裏づけに基づくものでなければならない。

第四に、安全保障、波及効果等の競争以外の観点にも配慮が必要である。上に書いたことと一見矛盾するようだが、仮に欧米ですでに作られていても我が国独自に構築すべきと判断すれば重複や後追いを恐れず構築すべきである。

第五に、知的基盤として長期的な取組みが可能となるような枠組みの構築が必要ということである。これには人材養成、責任体制の明確化も含む。データベース構築およびそのための体制整備は一朝一夕にはできない。5 年後、10 年後を見据えた年次計画を立案すべきである。

第六に、人材養成においては人材の備えるべき能力に対する社会の要望のみならず、養成された人材の将来の処遇（キャリアパス）についても十分に配慮する必要がある。

第七に、国際的データベースを分担開発するような場合でも、単に分担するだけにとどまらず、国内のデータベースとの連携や統合に十分な配慮を払うべきである。

最後に、生物資源等の研究用材料（バイオリソース）や最先端・高性能汎用スーパーコンピュータ等、データベース以外の研究基盤の整備計画と十分な連携をとり、互いにその効果を高め合うような配慮が必要である。

なお、上記の日本としての特徴を出せる例としては、cDNA、イネゲノム、微生物ゲノムなど強みのある分野の機能を付加したデータベース、SNP や生物多様性情報等の地域性のあるデータベ

ース、パスウェイ等国際的に優位にあるデータベース、日本で開発・収集されたバイオリソースとの連携に関わるデータベース、最先端・高性能汎用スーパーコンピュータ等の高速計算機と情報解析技術の活用が期待できるデータベース、我が国の国家プロジェクトの成果を戦略的に活用する方向性のあるデータベースなどが挙げられる。

なお、狭義のデータベースではないが、我が国では文献からの知識抽出技術の開発や遺伝子辞書の構築に関して強みがある。これらを活かした整備戦略も検討に値する。もちろん、日本語での利用環境の提供は日本独自のものであり、この点も十分検討すべき項目である。

最後に、4-2 で述べたことの一部繰り返しになるが、ライフサイエンス研究の進展、欧米のデータベース整備状況の変化等により、日本としての特徴も時々刻々と変わりうるため、状況の変化に国をあげて柔軟に対応できる、すなわち時代に応じて進化可能な体制の整備やそのための制度設計が最も重要な課題の一つである。

6. 推進方策とそれを実現するための体制

6-1 推進方策

前節「5. データベース整備戦略の基本的考え方」を踏まえて、「4-2 取り組むべき課題」を具体的に実現するためには以下の推進方策を遂行する必要がある。推進すべき方策とその留意点は下記の通りである。

(1)データベースの現状調査、評価、戦略立案機能の充実

現在、データベース整備の戦略立案機能はJSTバイオインフォマティクス推進センターBIRDや国立遺伝学研究所DDBJに設けられている委員会あるいは文部科学省のライフサイエンス委員会などによって一部担われているが、それらは非常勤の委員からなる委員会活動であり十分ではない。また各機関の委員会では、その守備範囲もその組織の活動に関するものに限られ、限定的なものとなっている。そこで、専門家による日常的活動（研究者の常勤）を基盤とし、データベースの現状や動向の定常的な調査および既存の戦略や活動の弛まぬ評価に立脚して、省庁の枠を超えて国家的視野に立って、ライフサイエンス研究全般やバイオ産業全般を見渡した戦略立案する機能が是非とも必要である。

なお、これらの調査、評価、立案に際しては、以下の点に十分な配慮・検討が必要である。

- データベースだけの問題と捉えるのではなく、ライフサイエンス研究の方向性も十分に踏まえた戦略を立案すること。
- データベース構築は、個々の研究者の創意工夫による研究とは異なる事業的な側面をもつことを十分に認識し、その推進および体制の整備に努めること。
- データベースは、ライフサイエンス研究全般、医療、バイオ産業全般の知的基盤、後方支援との明確な位置づけを行い、ニーズを的確かつ継続的に把握すること。
- データベースを構築する側の立場だけでなく、利用する側（例えば、医療や産業界）の意見が十分に取り入れられるように配慮すること。また、そのための仕組みを確立すること。
- 現在、ともすれば別々の戦略をもって収集・管理が行われている医学情報や薬学情報との連携にも十分配慮すること。
- データベース間の連携強化のためのデータベースの形式や構造の標準化や知識の体系化に向けた用語の統一化（辞書作成・標準化）のための戦略もあわせて立案すること。
- また、用語の統一化やデータの記述形式の標準化などをデータベース構築の際に義務づけるための制度設計もあわせて行うこと。
- データベースの開発とそのための技術開発(研究）とを緊密に連携させる仕組みを考案すること。
- 国として支援するデータベースや国として構築するポータルサイトの厳格な評価を行うための仕組みを検討すること。具体的にはモニター制度、利用者評価等を取り入れることを検討すること。
- 文献データベースとの連携のための仕組みを検討すること。
- 既存のデータベースだけでなく、ライフサイエンスの進展に対応した、新しい種類のデータベースあるいは従来にない発想に基づくデータベースの開発の振興にも十分配慮すること。

・データベース構築だけでなく、それを利用する技術開発の促進策も検討すること。
・長期的視点に立って、人材養成の促進を図る教育体制を構築すること。
・国家プロジェクトの成果活用の方向性を検討し、効果的な情報提供に向けた連携のための施策を考案すること。
・海外との連携をさらに進める方策を立案すること。特にアジア諸国のデータ生産者、バイオインフォマティクス研究者およびデータベース運営機関との連携について留意し、積極的な交流を図ること。

(2)基盤データベースの安定的な支援

我が国のライフサイエンス研究の基盤として欠かせないデータベース、世界的競争力の確保に向けて戦略的に重要なデータベースなどについては、安定的、永続的に支援することが必要である。現在この機能の一部は国立遺伝学研究所DDBJで実施されており、その他にもJSTバイオインフォマティクス推進センターによる支援が行われているが、データベースの数も限られており、また現在支援を受けているものについても、予算や期間の制約があり十分とはいえない。今後の更なる拡充が望まれる。なお、基盤データベースの安定的な支援に際しては、以下の点に十分な配慮・検討が必要である。

・我が国が独自に保有することが不可欠のものや世界的に存在が認められる知識基盤に限定して支援すること。その際、存在意義が認められる期間、安定的に維持するための必要額を十分に精査し支援すること。そのための評価基準として、論文への引用件数、アクセス数、一次データ量などによる定量的評価、外部有識者や利用者による定性的評価、およびサービス体制の充実度等を用いること。
・データベースの存在価値を維持するためのデータの収集・精査、サービス向上に直接関連する研究開発に限定して支援すること。新たな研究開発要素などは別予算（別途審査）（下記の(8)や(9)を参照のこと）で対応すること。
・ここで支援するデータベースについては、用語の統一化、データベースの記述形式や構造の標準化などの制約を課して、我が国のデータベースの統合化に寄与することを義務づけること。
・価値の高いデータベース、世界的に競争力のあるデータベースでありつづけるためには、それに関係した研究グループと密接な関係を常に維持していなければならない。そのための配慮を十分に行うこと。

(3)データベースの所在情報と利用法に関するポータルサイトの構築と運営

ライフサイエンス関係のデータベースに関する所在情報や利用法に関するポータルサイトを構築し運営することが必要である。これに関しても、いくつかの機関（JSTバイオインフォマティクス推進センター、国立情報学研究所など）でその試みはあるが、十分とは言えない。その理由は、常勤の専門家による運営が必ずしもなされていないこと、利用者からのフィードバックを常に活かしてサイトを最新のものに更新する仕組みが整っていないことによる。その背後には、このような仕事への評価の低さと予算面の手当てのなさの問題がある（国立情報学研究の活動は予算的な裏づけがあったが、平成17年度末で終了）。3-3節で紹介したように、我が国では数多くのデータベースが日々作られている。これらを十分に活用するためには、常に最新の情報を保持したポ

ータルサイトが不可欠である。このサイトの構築・運用に際しては、以下の点に留意すべきである。

- 何といってもポータルサイトにとって重要なことは、その網羅性である。日々、新しいデータベースが作られているような今日の状況では、個々の利用者が関連するデータベースすべての所在情報や利用法を把握するのは事実上不可能である。ポータルサイトにはデータベース作成者の意向も踏まえた上で、我が国のデータベースを漏れなく収載することが欠かせない。
- 一方、ポータルサイトに掲載されるデータベースが玉石混淆ではかえって混乱を招く。これを避けるため、引用数、アクセス数、データ量等を調査し、利用者側から見て分かりやすいよう、掲載するデータベースの分類をすること。
- 使いやすさによるデータベースの評価や利用法からみた分類などによるガイダンス機能の導入など、利用者の視点に立ったポータルサイトの運用に努めること。
- ポータルサイトの自動構築や評価のための技術開発もあわせて行うこと。
- ライフサイエンス分野の研究者、技術者を主たる対象とするが、一般の医療関係者あるいは育種家といった利用者も想定し、日本語での情報提供にも十分配慮すること。

(4)統合データベースの開発とそのための研究開発の促進

データベースの統合化に関しては、我が国においてもいくつかの機関でそれぞれの取組みが行われている。それらには一長一短あり一概には評価することはできないが、いずれも我が国のデータベース全般を統合化するという視点は弱い。その理由は、そもそもそのような使命を負わされてわけでもないし、権限があるわけでもなく、そのための予算の裏づけがあるわけでもないからである。JSTバイオインフォマティクス推進センターにおいても、データベースの高度化・標準化が謳われているが、統合化は必ずしも視野には入っていない。しかしながら、上述のポータルサイトの構築・運営だけでは、我が国の様々なデータベースの価値を十分に引き出すことはできず、ライフサイエンス研究のみならず産業界からの要請にも応えることはできない。多種多様なデータが生物的医学的に整理された形で統合されなければ、膨大なデータの洪水に流されてしまうだけになってしまい、ライフサイエンスの発展が止まってしまう。逆に、バラバラだったデータベースを統合化することができれば、これまで別々のデータベースに収められていたデータ間の潜在的な関係（例えば、遺伝子と疾患と薬剤との間の新たな関係やゲノムの進化と表現型の進化の間の対応関係）を見出すことが可能になる。ポータルサイトだけでは、このような新たな知識の発見を直接的に支援することはできない。データベース構築の大きな目標の一つはそこから新たな発見をすることにあり、統合化はまさにそのためのものである。一朝一夕には無理でも、我が国のデータベースの統合化に向けた研究開発を強力に、かつ、地道に推し進める必要がある。

ただし、統合化と言っても生命階層のどのレベルの、どのような知識を発見したいのか、どのようなことに統合データベースを使いたいのかによって、その目指すところ、意味するところは異なってくる。仮に目指すところが同じでもいろいろなアプローチがありうる。そのため、我が国としてどのようなアプローチでどのような統合化を目指すべきかに関しては、将来のライフサイエンスの動向や産業界からのニーズも十分踏まえた検討を行い、その議論に基づいて推進を図るべきである。幸い、現在、科学技術振興調費「科学技術連携施策群の効果的・効率的な推進」の一テーマとして調査研究が進められているところでもあり、その結果も踏まえて、前記(1)の戦

略立案機能の中で推進策を練ることが望ましい。

ところで、どのような統合化を目指すにせよ、統合化にあたっては、そのための用語や概念の統一化、データベースの記述形式や構造の標準化が前提となる。これらの中はすでに欧米で開発が進んでいるものもあり、それらを採用することも考えられるが、5節の「データベース整備戦略の基本的な考え方」に述べたように、我が国の特徴や強みが十分に発揮できるように十分な配慮・検討が必要である。

この他にも、データベースの統合化とそのための技術開発に向けては、以下の点に十分な配慮・検討が必要である。

- 国が支援するデータベースの構築者に対し、情報提供や技術指導を行うなど十分な連携をとり、用語の統一や記述形式の標準化を図ること。
- データベースの専門家（特にバイオインフォマティクス研究者）だけでなく、実験研究者や医療やバイオ産業に従事する人でも簡単に使えるような検索ソフトの開発や日本語環境の整備にも努めること。
- 欧米の後追いにならず、次世代の統合化を先取りするためにも、最先端の情報処理技術の活用や開発を行うこと。例えば、画像情報や新しい計測機器の出力結果等、新しい形式のデータに対応した情報処理技術や、新たな情報共有の枠組みのための情報処理技術を開発すること。
- 概念や用語の統一が統合化の鍵を握ることから、また我が国独自の特徴を出す意味からも、分野毎に、実験系の研究者と情報系の研究者の双方からなる専門家集団を形成し、それらの専門家集団の知識の融合に基づく統合データベースを目指すこと。
- 上のことと関連するが、データベース構築には実験研究者も深く関与できるような体制作りが必要である。

(5)維持が困難になったデータベースの受入れ

4-1節「データベースの問題点」に述べたように、各機関や各プロジェクトで開発されたせっかくのデータベースが、予算が切れると維持更新されなくなってしまうという問題がある。これに関しては、現在は研究者あるいは研究室の自発的な努力に頼るしかない状況であり、我が国のライフサイエンスにとって由々しき問題である。当然のことながら、すべてのデータベースを管理し続けるのは意味もないし不可能であるが、存続することが重要と判断されたものに関しては十分な支援が必要である。すなわちプロジェクトや科研費などの研究費が終了するなどして維持が困難になったデータベースの受け皿を、国として用意する必要がある。もちろん、闇雲に受け入れる必要はなく、存続価値を十分に厳正に評価して受入れや支援を判断すべきである。その際、以下の点に留意すべきである。

- ライフサイエンスの進展とともに、支援しなくてもよくなるデータベースも出てくるが、その一方で新たに支援すべきデータベースも出てくる。このような変化に柔軟に対応できるような制度（例えば、データ産出プロジェクトの設置に際しては、そのプロジェクト経費の一部をプロジェクト終了後も一定期間データベースの維持更新が可能なように積んでおくことを義務付けるなど）を導入すべきである。
- 文科省以外の省庁が整備したデータベースについても受け入れを検討すること。その際、内閣府の委員会、調査なども踏まえて検討すること。

・データベースの受け皿機関への移管に関しては、権利関係、事務手続きなどに配慮すること。
・ここで支援するデータベースについても、移管する際に、可能な限り、用語の統一化、データベースの記述形式や構造の標準化などの制約を課して、我が国のデータベースの統合化に寄与することを義務づけること。

(6)文献情報との連携

3-1節「データベース開発の世界的な動向」に述べたように、機能情報のデータベース化が重要な課題になりつつある。機能情報の多くは論文の中にテキストとして記述されていることから、文献中に記述されたデータや知識と、配列や立体構造などの実験データとの連携と統合に今後取り組まなければならない。米国NCBIでは、同じ組織で実験データも文献データも管理されていることから連携は比較的スムーズであるが、我が国ではこれまで別々に扱われてきたことから、今後連携を図っていく方策を講ずる必要がある。具体的には、遺伝子名や塩基配列のアクセッション（用語解説参照）などによる共通識別キーでの統合的検索を可能とするほか、ライフサイエンス分野の知識を計算可能な形へ変換し、概念対概念の関係を自動生成することにより、増大する論文データに対応できる知識の体系化を実現する必要がある。また、これにより提示される知識体系を活用しつつ、各データベースで利用されている各種用語の標準化を図る必要がある。また、このようなことを可能とする技術開発（概念・知識の収集の自動化やデータベースからの知識発見など）を並行して進める必要がある。なお、これらは上記(4)の「統合データベースの開発とそのための研究開発の促進」と重なる部分がいくつかあり、十分な連携のもとに進める必要がある。

(7)アノテーション（情報解読による実験データの注釈付け）の実施

基盤データベースの支援や維持が困難なデータベースの受入れ、さらにデータベース統合化および文献情報との連携といった活動と連動して、我が国で産出されたデータにもかかわらず未解析、未解釈のまま放置されている種々の実験データの意味付け（生物学的、医学的な解釈）を強力に推進すべきである。また、すでにアノテーションされているものでも正確さを欠くものもあり、それらについても再度アノテーションを実施すべきである。これについては、現時点で未解析・未解釈・不正確なデータのアノテーションを実行するだけでは不十分である。今後出てくるデータに対しても常に最新の技術、知識をもった専門家によるアノテーションを施す体制の確立が望まれる。アノテーションされていないデータは統合化しても意味がないし、ポータルサイトで所在が明らかになっても利用価値は低い。アノテーションを実施する際に留意すべき点は以下の通りである。

・アノテーションは独自の基準でバラバラに行うのではなく、上記(4)「統合データベースの開発とそのための研究開発の促進」や(6)「文献情報との連携」で開発された用語やガイドラインに基づいた注釈を行うこと。これによりデータベースの統一化が可能となる。
・実験系と情報系の研究者が協力できる体制を構築して、より正確で意味のある情報解読・注釈付けを実施すること。
・cDNA、イネゲノム、微生物ゲノムなど日本の強みを発揮できるデータについては、統一基準でより信頼性の高い形での再アノテーションを実施し、それを公開データベースに反映することを検討すること。

(8)新たなデータベース構築への投資

上述の基盤的データベースや評価の確立したデータベースの安定的な支援のほかに、ライフサイエンス研究の進展に対応した新たなデータベース、新たな発想に基づくデータベースの構築にも投資すべきである。これは、現在一部科研費特定領域研究「ゲノム４領域」やJSTバイオインフォマティクス推進センターで実施されているが予算や期限が限られており十分ではない。データベースには長期的視点が必要である。今後このような観点にたった支援制度を是非とも設けるべきである。長期的視野に立つといっても、新しく作られるデータベースは玉石混淆である。5年程度の時限を設けた競争的研究資金により実施することが望ましい。そこで評価が確立したものついては、例えば、上記の(2)「基盤データベースの安定的な支援」により支援することが考えられよう。なお、新たなデータベース構築への投資を行う際には、以下の点に十分な配慮・検討が必要である。

- ・ここで支援するデータベースについては、最初から用語の統一化、データベースの記述形式や構造の標準化などの制約を課して、我が国のデータベースの統合化に寄与することを義務づけること。
- ・新たなデータベース構築は機関で行ってもよいし、数人から個人の研究者レベルで行ってもよい。個人で行う場合の支援策に関しては、特に若手研究者が行う場合には、データベース開発に対する研究者の理解が不足している現状を考慮して、任期付きあるいは終身雇用の職をどこかに用意するなどの点に配慮すること。
- ・データベースの構築そのものでなくても、それの基盤となる、分散処理、高速通信、データベースマネジメントシステム等の基盤的技術開発を支援することも必要。

(9)データベースを活用した研究（バイオインフォマティクス）の促進

当然のことながら、データベースはそこから有用な知識を発見してこそ意味がある。逆に言えば、そのことを見越してデータベース開発を進める必要がある。そこで、データベース構築への支援と並行して、それを活用する技術の研究開発、いわゆるバイオインフォマティクスの促進も図る必要がある。バイオインフォマティクスそのものは、科研費特定領域研究「ゲノム４領域」やJSTバイオインフォマティクス推進センターなどで振興が図られているが、データベース構築と一体となった研究開発は必ずしも活発には行われていない。今後この面での支援策を講ずる必要がある。また、従来の施策の更なる拡充を図る必要がある。なお、これに関しては、若手研究者の育成、そのための任期付きあるいは終身雇用の職の確保、競争的な資金制度におけるバイオインフォマティクス分野の研究と連携、などに十分配慮して遂行すべきである。

(10)データベース開発のための人材養成

いくつかの大学において、21世紀COEプログラムや科学技術振興調整費人材養成などの支援を受けながら、バイオインフォマティクス分野の研究者や技術者の養成が行われている。しかしながら、質の高いデータベース構築を行う上で不可欠の人材である、アノテータ（データに生物学的医学的な解釈を加える専門職員）やキュレータ（データベースの編集作業に従事する専門職員）を目的としたものはほとんどない。経済産業省の産業技術総合研究所生命情報科学研究センターや国立遺伝学研究所で一部実施されてはいるものの十分ではない。問題は、教育する側にあるのではなく、受け手の少なさにある。それはアノテータやキュレータの技術を身につけても我が国

にはその職がないからである。また、そのような仕事の重要さへの理解が不足しているからである。我が国で世界的に競争力のある、また、意味付けがきちんとされた、有用なデータベースを開発するには、まずはアノテータやキュレータの安定的な職を数多く確保するとともに、それに相応しい人を養成することが不可欠であり、そのための体制を早急に確立する必要がある。また、そのためにその後の将来の処遇（キャリアパス）につながるような学会の認定資格などの方策も検討する必要がある。高度に専門的な知識や技術をもったアノテータやキュレータを養成するには、振興調整費人材養成プログラムあるいは大学の専門教育との連携がなくてはならない。このことを十分に考慮した人材養成の仕組みを構築する必要がある。上記の観点は、データベースのシステム開発や運用を専門的に担ういわゆるシステムエンジニアやオペレータの育成に関しても言えることである。

6-2 推進体制

6-1 で示した推進方策(1)から(10)を具体的に、かつ、効率よく遂行するために、以下に述べるような体制の整備を提案する。ただし、この提案は、委員会として中長期的な視点から望ましいと考える姿を示したものである。

- ・関係省庁間の連携のための戦略委員会の設置（内閣府総合科学技術会議の議論を踏まえて決定）
- ・関係省庁のデータベース関係機関による連携、調整のための枠組み
- ・関係省庁のデータベース関係機関による連携、調整のための枠組みの中核的機能を担う体制の整備（文部科学省が整備）

それぞれについて、以下にその役割や機能を述べる。

(A) 連携のための戦略委員会

データベースに関する関係省庁間の連携のため、前記 6-1 節「推進方策」の(1)「データベースの現状調査、評価、戦略立案機能の充実」に述べた役割、すなわち司令塔的な役割を担う。具体的にはデータベースの現状や動向、ニーズを定常的に調査し、それに基づき、我が国のデータベース戦略や構築活動を評価し、ライフサイエンスに関わるデータベースの整備戦略を練る。さらに、後述の関係機関による連携、調整のための枠組みの種々の活動を監督・指導する。上記枠組みに属する各機関が有機的に連携しているか、効率的に予算が使われているか、真に役立つデータベースを構築しているか、などを常に監視し、必要に応じてそれらに指導を行うものとする。

なお、統合化データベースを含むライフサイエンス基盤整備は、第３期科学技術基本計画に基づくライフサイエンス分野推進戦略の戦略重点科学技術に位置づけられている。そこで、連携のための戦略委員会のあり方については、総合科学技術会議において、今後詳細を検討することが適当である。総合科学技術会議では、統合化データベースを含むライフサイエンス基盤整備を、第３期科学技術基本計画ライフサインエス分野別推進戦略の戦略重点科学技術に掲げ、今後の推進方針の検討中である。それを受けて文部科学省としては今後の方向性を踏まえて、詳細を検討することとする。

(B) 関係機関による連携、調整のための枠組み

上記の戦略委員会が立案した計画を具体的に実行に移す組織として、各省庁のデータベース関係機関の連携、調整を行う枠組みを設ける。後述するように、統合データベース構築や関係機関による連携、調整及び戦略委員会の計画を実際に遂行するためには、中核的機能を担う体制の整備が不可欠であるが、しかしながらこれだけでは十分ではない。6-1 の(2)の留意点で述べたように、データベースはそれを構築したり利用したりする研究者グループと密接な関係を常に維持することが必要である。そのため、ある種のデータベースは研究者グループのいる機関にそれぞれ分散して保有し、それを連合体として有機的に連携させることが望ましい。関係機関による連携、調整のための枠組みはまさにそのためのものである。

(C)中核的機能を担う体制

上述したように、データベースは研究者グループのいる機関にそれぞれ分散して保有し、それを連合体として有機的に連携させることが望ましい。しかしながら、我が国で作られている多くのデータベースが共通の情報処理を行っていること、また、共通に使うデータを重複して持っていることから、データベースは集中的に構築維持管理したほうが効率がよい面が多々ある。また、標準化や統合化を図るには、連合体では不十分でありそれらに関して強力な指導力を発揮できる能力をもった専門家集団が必要である。すなわち、我が国におけるデータベース構築の効率化、標準化、統合化のためには中核的機能を担う体制の整備が欠かせない。連携のための戦略委員会で立案された計画を漏れなく速やかに遂行するためにも、また、人材養成や国際対応の観点からもこのような中核となるものが不可欠である。

そこで、上記(B)の関係機関による連携、調整のための枠組みの中核的機能を担う体制を置き、6-1 節の「推進方策」の(2)から(10)すべてを担当させることとする。また、当該体制では日常的にデータベース開発の世界的動向や我が国のデータベース構築活動、および様々な分野の利用者のデータベースに対するニーズを調査・評価し、それに基づきデータベース戦略を提案するなどして、上記のデータベース連携のための戦略委員会を補佐する、すなわち、推進方策の(1)にも貢献する。なお、「推進方策」の中の(2)基盤データベースの安定的支援、(7)アノテーション、(10)人材養成、については、必要に応じて外部へその役割を委託する。また、若手研究者の受け皿となる職（任期付きあるいは終身雇用）を用意して、新たなデータベース構築(8)や利用技術(9)の振興に努める。

この中央に設置された体制には、自前でのデータベース開発や外部のデータベース構築を支援するために、一定規模の計算機資源（中央データベース・サーバ）を保有させるものとする。この中央データベース・サーバの利用に際しては、最先端・高性能汎用スーパーコンピュータ（平成 18 年度より整備）との連携も検討する。この他に、データベースの広報、普及啓発活動のために、シンポジウムや講習会・トレーニングコースの開催やホームページの整備を行わせ、あわせて国際的な連携や産業界との連携のための活動も実施させる。なお、中核的な機能を担う体制は、優秀な人員の確保や養成の面からも、データベース整備が時間のかかる作業であるという面からも、5 年から 10 年にわたる年次計画を立てそれに基づいて段階的に整備を行ってゆく必要がある。

なお、上記の(A)(B)(C)という３つの機能・制度の設置、運用に際しては、その役割分担を明確にし、透明性・公平性・客観性を十分に担保した形で進めなければならない

6-3 中核的機能を担うための体制案について

現在、想定している中核的機能を担うための体制案を図 1 に示す。これはあくまでも例示に過ぎないが、統括のもと以下の 5 つのチーム構成とする。

・ポータルサイト構築運用チーム：(3)「データベースの所在情報と利用法に関するポータルサイトの構築と運営」を主に担当。また、日常的にデータベース開発の世界的動向や我が国のデータベース構築活動を調査・評価し、それに基づきデータベース戦略を提案するなどして、前述の連携のための戦略委員会を補佐する。

・データベース運用チーム：(2)「基盤データベースの安定的な支援」、(5)「維持が困難になったデータベースの受入れ」、(7)「アノテーション（情報解読による実験データの注釈付け）の実施」および中央データベース・サーバの運用に関わる業務を主に担当。必要に応じて、外部機関にこれらの業務を委託する。

・統合データベース開発チーム：(4)「統合データベースの開発とそのための研究開発の促進」、(6)「文献情報との連携」に関わる業務を主に担当。

・技術開発チーム：データベースの運用、統合化、その他必要となる様々な技術開発を行う。(8)「新たなデータベース構築」、(9)「データベースを活用した研究（バイオインフォマティクス）」に関する業務も一部担う。

・国際対応、産学連携チーム：国際対応や産学連携の業務を担う。それに加え、(10)「データベース開発のための人材養成」のための各種教育、トレーニング、シンポジウム、ホームページ作成、大学連携、などに携わる。

なお、統括のもとに運営委員会を設けて、適宜運営に関して助言を求めることとする。

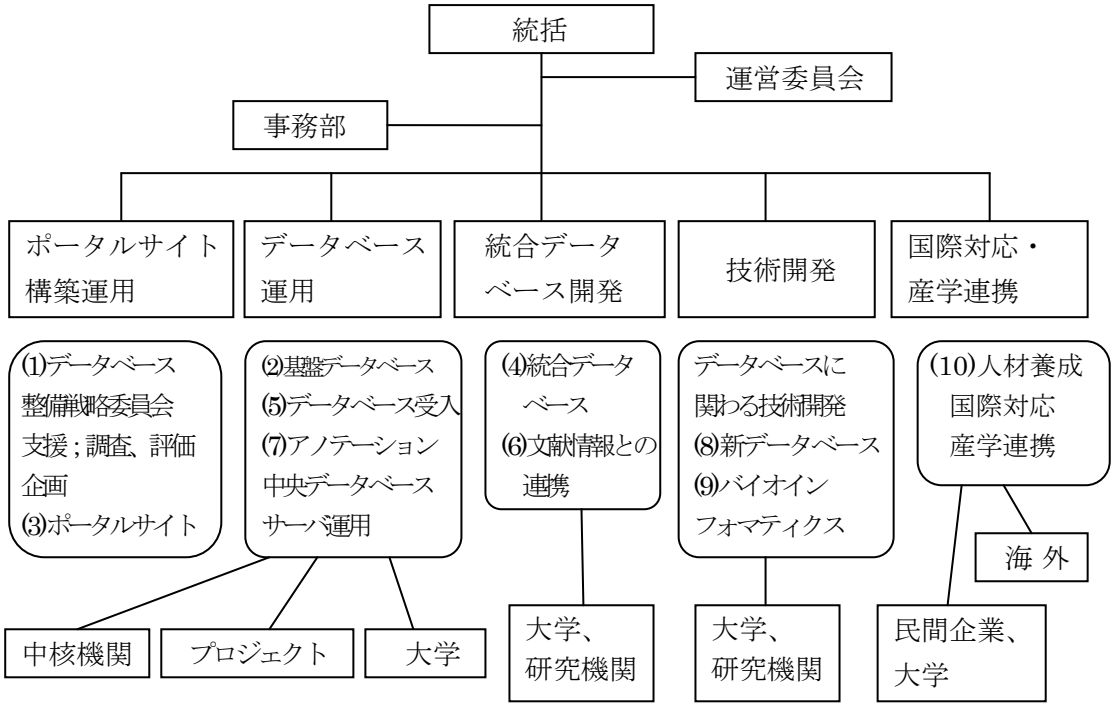

図 1　中核的機能を担うための体制案

7. 緊急に取り組むべき課題

ライフサイエンス研究は進展が速い。その成果であるデータベースも例外ではない。1年、2年の遅れが取り返しのつかない事態を招く可能性もある。そのような事態が起きないようにするには、6節「推進方策とそれを実現するための体制」を一日でも早く実行に移すことが肝要である。しかしながら、予算の確保や体制の整備には、早くても1年ほどの時間を要する。そこで、それほどの予算や体制を必要としないもののので、かつ、緊急性を有するものを6節で示した10個の課題、あるいは(A)から(C)の3つの機能・制度の中から優先度の高いものを選び出し、実行に移すことが望まれる。

また、すでに指摘したとおり我が国において整備されているデータベースの数は把握されているだけでも二百弱を数える。また、それらのデータベースは様々な背景に基づき多様な主体・資金によって作成されたものである。従って今後の整備にあたっては、始めは国家プロジェクトの成果としてのデータベースを対象とし、その後順次対象を広げていくなど、段階的な整備の計画を立案することが望ましい。

まず、1番目に(A)連携のための戦略委員会の設置は、すべての推進方策の基盤となるもので、これを最も急がなければならない。戦略委員会の設置を支援するために十分な透明性や客観性を確保した上で早急に作業グループを整備し、 6-1の(1)「データベースの現状調査、評価、戦略立案機能の充実」を図るべきである。

2番目は、6-1節の(3)「ポータルサイトの構築・運用」に向けたポータルサイトの設計と試作品の作成である。試作品を早急に構築し、利用者に試験的に使ってもらい、その意見を反映させて本運用のサイトを設計することが重要であろう。また、可能なら、我が国で作られた主なデータベースに関してはその試作品に盛り込み、産業界などのからの要望に直ちに応えるべきであろう。各データベースの使いやすさの指標や利用法による分類手法の開発などデータベースの評価法・分類法の開発にも急いで取り組むべきであろう。

3番目は、6-1節の(4)「統合データベース開発」に向けた、用語の統一、記述形式の標準化、データの共有化およびそのための技術開発である。これも、今後開発するデータベースやポータルサイトの基盤になるものであり、これの開発を急がなければならない。

4番目は、何といっても(10)の人材養成である。これにはある程度長い時間がかかるのでできるだけ前倒しで実施することが望ましい。具体的には、アノテータやキュレータの確保、養成に着手すべきである。大学や研究機関や学会と連携して、また、民間企業の力も借りて、集中的な講習会を開くなどして、潜在的なアノテータやキュレータの掘り起こしや教育を実施すべきであろう。

なお、上記以外の(2)、(5)、(6)、(7)、(8)、(9)に関しては、JSTバイオインフォマティクス推進センターや科研費をはじめとした既存の取組みはあるものの、可能な限り急いで取り組むことが重要である。

8. おわりに

ライフサイエンスにおけるデータベースの位置づけは、研究成果の整理・編集だけを目的とする所から、様々な階層に対する様々な種類のデータの統合化により、新たな発見や俯瞰的な理解を与えるだけでなく、個々のデータが持つ生命科学全体における意味を鮮明にし、対応する個々の研究自体の意味とその将来への方向付けを明確にするという極めて重要な役割を担うようになってきた。データベースの整備が覚束なければ、ライフサイエンスやバイオ産業の未来も危ういと言えるほどの存在にまでなってきた。

もちろん、これらデータベースに蓄えられた文献データをはじめ、ゲノム、プロテオーム、生体分子構造、遺伝子やタンパク質の発現や分子間相互作用、ネットワークやパスウェイ、臨床疾病などのデータそのものは、日本はもとより世界中のライフサイエンス研究者の努力と社会からの請託の結晶であり、それゆえ、データベースは、長い歴史における人類の叡智をまとめた宝とも言えよう。この宝物をどのように活かし、さらに発展させられるかが現在、問われているのである。この活用とは、直接的なライフサイエンス研究に対してだけでなく、産業界、一般社会への活用も含まれ、特に多数のデータベースの連携や統合による高度化によって、データベース構築時には考えもしなかった異分野を横断する新たな応用が実現し、生命科学が全く新たな次元での発展を迎えられるものと確信している。

本報告書では、このような現状の認識と将来への展望をもとに、これまで我が国においてあまり議論されることがなかったライフサイエンス分野におけるデータベース、その中でもとくにライフサイエンスの基盤となるデータベースの整備に対して、問題意識を明確にした専門家による議論を通じて様々な問題点を鮮明にし、それを解決するための戦略と具体的な施策のあり方を論じた。本報告書の結論を一言で要約すると、ライフサイエンスのデータベース整備に関して、省庁の枠を超えた権限と責任をもった、国家の司令塔的な役割を担う連携のための戦略委員会の設置、そこで立案された計画を実行する関係機関による連携、調整のための枠組みおよびその中核機能を担う体制の整備が不可欠であるということである。

なお、本報告書の冒頭に述べたように、ライフサイエンス分野には本報告書ではあまり触れなかったデータ（生物資源等の研究用材料に関するもの、医療現場で用いられる臨床情報や医薬品情報、化合物の構造や毒性情報、食品の成分や安全性に関するもの、作物や家畜の育種に関するもの、産業上有用な微生物の情報、など）が数多く存在する。これらに関しては、現在内閣府科学技術連携施策群（ポストゲノム）で進められているような各省庁連携の体制と連携をとって取組を進めることが望ましい。また、緊急に取り組むべき課題については、研究の進展が速いことから、その決定過程における透明性には留意しつつ、遅滞のない施策の立案と実施が望まれる。

データベース整備戦略作業部会委員名簿

（委員）

	秋山　泰	（独）産業技術総合研究所生命情報科学研究センター長
	江口　至洋	三井情報開発（株）フェロー
	宇高　恵子	高知大学医学部教授
	金久　實	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長
	鎌谷　直之	東京女子医科大学附属膠原病リウマチ痛風センター所長
◎	郷　通子	お茶の水女子大学学長
○	五條堀　孝	国立遺伝学研究所生命情報・DDBJ 研究センター長
	佐藤　清	（社）バイオ産業情報化コンソーシアム事務局長
	菅原　秀明	国立遺伝学研究所生命情報・DDBJ 研究センター教授
	高木　利久	東京大学大学院新領域創成科学研究科教授
	田中　博	東京医科歯科大学情報医科学センター長
	中村　春木	大阪大学蛋白質研究所教授
	姫野　龍太郎	（独）理化学研究所情報基盤センター長
	深海　薫	（独）理化学研究所バイオリソースセンター　情報解析技術室長
	細江　孝雄	（独）科学技術振興機構理事
	宮野　悟	東京大学医科学研究所教授
	山本 博一	アステラス製薬㈱研究本部研究企画部日本橋部長

◎主査、○主査代理

（参考人）

大久保　公策	国立遺伝学研究所生命情報・DDBJ 研究センター教授
楠木　正巳	大阪大学蛋白質研究所助教授
藤山　秋佐夫	国立情報学研究所教授

データベース整備戦略作業部会における審議の過程

○第１回　平成１７年８月１２日

（１）データベース整備戦略作業部会について
（２）ライフサイエンス研究に関するデータベースに係る施策について
（３）データベース整備の現況について
（４）課題整理と今後の展開について

○第２回　平成１７年１１月１０日

（１）内閣府総合科学技術会議の科学技術連携施策群における統合データベースの動きについて
（２）科学技術振興機構　バイオインフォマティクス推進事業における「生命情報データベース高度化・標準化」研究開発課題の公募について
（３）データベース整備に係る作業部会委員の意見概要について

○第３回　平成１８年１月１６日

（１）データベース整備に係る作業部会委員の意見概要について
（２）データベース整備戦略として必要な機能について
（３）データベース整備戦略として必要な組織体制について

○平成１８年１月１９日
ライフサイエンス委員会に対してデータベース整備戦略作業部会における議論の状況を報告

○第４回　平成１８年２月２８日

（１）データベース整備戦略作業部会報告書骨子（案）について
（２）平成１８年度「統合データベースプロジェクト」について

○第５回　平成１８年３月２４日

（１）データベース整備戦略作業部会報告書（案）について
（２）平成１８年度「統合データベースプロジェクト」について

○第６回　平成１８年５月１１日

（１）データベース整備戦略作業部会報告書（案）について

付録：用語解説

ヒト完全長 cDNA(p.6 下から 8 行目)：cDNA とは、個々の遺伝子が、ゲノム DNA から読み取られてタンパク質を作る際の鋳型となる塩基配列情報を有する DNA であり、その配列が完全な状態で取得できたものを完全長 cDNA という。これを得るために日本独自の技術が使われている。

オントロジー(p.8 上から 9 行目)：知識、語彙、概念などと、それらの間の関係を明確にした辞書。生物学では異なる分野で同じ用語を異なる意味に用いたり、異なる用語で同じ意味を表したりすることがあり、これを明確化することを目的に遺伝子オントロジー (GO)プロジェクトなどが進められている。

パスウエイ(p.8 上から 12 行目)：生体分子間の相互作用の一連の流れのことである。生命科学の進展により、個々の生体分子の役割の解析から多様な生体分子の一連の相互作用がおりなすシステムに解析の興味が移り、代謝パスウェイや信号伝達パスウェイといった高次の生命現象に関わるパウスェイの解析が精力的に進められている。

SNP 解析(p.11 上から 13 行目)：SNP とは、一塩基多型と呼ばれるヒト一人一人の遺伝子中に見られる DNA 配列の違いのことである。これを解析することで、疾患感受性や薬剤応答性の個人差が分かり、疾患関連遺伝子の同定や薬剤による副作用を避ける診断方法の確立が期待できる。

ファーマコジェネティックス(p.11 上から 14 行目)：医薬品の効果や毒性に及ぼす遺伝的特性の影響を調べる学問のことで、特に遺伝的多型が薬理作用に及ぼす影響をゲノムワイドに調べた情報をベースにしたテイラーメイド医療が注目されている。

アクセッション(p.24 下から 1 行目)：データベースに格納される個々のデータであるエントリー単位に割り振られた固有のＩＤ。DDBJ に代表される国際塩基配列データベースなどでは、決まった文字数の英数字が一定の規則のもとに割り当てられる。