

統合データベース タスクフォース報告書

平成21年4月10日
ライフサイエンスPT
統合DBタスクフォース

1. はじめに

これまで、ライフサイエンス分野においては、ヒトを含めた動植物のゲノム解読を行うプロジェクトや、3000種類のタンパク質の基本構造型の立体構造の決定を目指したタンパク3000プロジェクトに代表される大量のデータが産出される研究事業が多く行われてきた。今後、我が国のライフサイエンスを更に発展させていくためには、これら産出されたデータを利用者の視点に立って統合化し、効率よく研究者、産業界、さらには国民に還元し、新たな知見を得たり、新たな産業を創出していくことが不可欠となっている。こうしたことから、平成17年度から、各府省の縦割りの施策に横串を通すことを目的に設置された科学技術連携施策群「生命科学の基礎・基盤」の中で、内閣府、文部科学省、厚生労働省、農林水産省、経済産業省が一体となったライフサイエンス分野のデータベースの統合化に向けた取り組みが開始された。その成果として、いくつかの統合データベース事業が新たに立ち上がり、また、将来の省庁横断的なデータベースの構築に向けた検討も始められた。

しかし、この連携施策群も、平成19年度に終了したことから、平成20年5月に、総合科学技術会議のライフサイエンスPTに本「統合データベース タスクフォース」を設置し、1年間を目処として、関係府省の役割分担を明確にしつつ一層の連携強化を図り、早急にデータベースの統合化に向けた具体的な制度設計や行動計画を策定することとした。

本タスクフォースとしては、平成20年12月25日に第1回の会議を開催し、5回にわたり検討を深め、ライフサイエンス分野における我が国の恒久的かつ一元的な統合データベースの具体的な仕組みと、その整備に向けたロードマップについて、以下のように取りまとめた。今後、関係者において、この考え方にに基づき、着実な統合データベースの構築がなされるよう、望むものである。

2. 現状及び問題点

個人のゲノムデータやタンパク質の立体構造データ、遺伝子発現データ、細胞レベルでの発現情報など、ライフサイエンス研究にデータベースの活用は必須であり、我が国のライフサイエンス研究全体を推進・加速させていくためには、産出された大量のデータを生かすためのデータベースの整備が不可欠であり、また、データベースが効率的に活用されるための仕組みも必要である。

我が国においては、国立遺伝学研究所のDDBJ(DNA Data Bank of Japan)、大阪大学のPDBj(Protein Data Bank Japan)、京都大学のKEGG(Kyoto Encyclopedia of Genes and Genomes)、東京大学・独立行政法人 科学技術振興機構(JST)のJSNP(A Database of Japanese Single Nucleotide Polymorphisms)、独立行政法人 農業生物資源研究所のRAP-DB(Rice Annotation Project-Database)、社団法人 バイオ産業情報化コンソーシアム(JBiC)のH-InvDB(ヒト遺伝子アノテーションデータベース)、独立行政法人 医薬基盤研究所のGeMDBJ(疾患ゲノムデータベース)、独立行政法人 理化学研究所のFANTOM DB(Functional Annotation of Mouse

Database)といった世界的にも評価されているデータベースの構築等が進められている。また、文部科学省の統合データベースプロジェクトにおける大学共同利用機関法人 情報・システム研究機構(ROIS)の取組をはじめとして、各省庁で統合データベースの構築に向けた取組が行われている。しかし、これらの取り組みは、欧米と比較すると規模や体制、活用状況等、まだ十分とは言えず、更なる取組みの強化が必要である。

また、データベース維持のための恒久的な財政措置がとられていないため、研究プロジェクト等で整備されたデータベースが、プロジェクトの終了とともに更新停止、提供終了となることがあり、研究成果の継続的な利用が保障されていない。

また、個人情報保護等の観点や、公的資金により取得されたデータの所有権に関する様々な考え方や状況が存在する上に、さらにはデータベースの種類によっては、その統合の有用性が研究者等に十分に認識されていない点等もあり、研究の結果生じたデータベースの共有化に向けた研究者のインセンティブは必ずしも高くないのが現状である。

その他、国家プロジェクトの成果として産出されたデータベースは、公的な面を持つという考え方が我が国の研究者に必ずしも十分には普及していないという課題や、公的研究資金を活用したデータ蓄積型の研究から産出されたデータベースを登録し、広くアクセスを可能とする制度も整備されていないという状況がある。

3. 目的

本報告書は、上記現状及び問題点を解決し、我が国におけるライフサイエンス研究の成果が、広く研究者コミュニティに共有かつ活用されることにより、基礎研究や産業応用研究につながる研究開発を含むライフサイエンス研究全体が活性化されることを目指し、データベースの統合とそれに関連したバイオインフォマティクス研究の育成に向け、下記の機能を有する組織の設立を提言することを目的とする。

4. 体制整備

統合データベースの構築について、我が国においては、端緒に就いたばかりであり、時限的な競争的資金を活用してデータベースを管理・運営しているが、欧米のように、一元的で継続的な体制を整備することが必要である。

(1) オールジャパンとしての統合データベース構築

ライフサイエンスの研究から発生する大量のデータは、研究プロジェクトの成果として、その研究主体によってデータベース化され、それぞれの研究機関等に維持・管理されていることが多い。これらのデータベースの統合化に向けては、情報通信技術の進歩により、物理的に一ヶ所に集約する必要はないが、それぞれのデータベースを維持・管理している研究機関等が、統合しやすい形でデータを入力・管理し、またデータの品質を十分確保した上で連携していく事が必要となる。

将来的な我が国における統合データベース構築のための体制としては、まず、基本方針や推進方策の策定、関係省庁の調整など、司令塔機能を担う機関として、総合科学技術会議に「統合データベース推進本部(仮称)」を設置することが求められる。

関係省庁も参加した「統合データベース推進本部(仮称)」で策定された方針に従い、我が国と

して一元的な統合データベースの事業を着実に実施していくためには、システム設計やポータルサイトの構築など、具体的な実務に必要な事項を決定・実行するとともに、統合データベースに係る研究開発を担うセンター機能として「統合データベースセンター(仮称)」を整備・運営する必要がある。また上記の統合データベース事業の推進・協議・調整のための会議体として、「統合データベース運営委員会(仮称)」を設置することが必要である。

また、現在、各省庁や研究機関(事業)等で公的資金を用いて整備を図っている専門的なデータベースについては、その充実強化を図りつつ、それぞれが整備しているデータベースの品質や内容を十分確保した上で、上記の「統合データベースセンター(仮称)」と連携を図り、ネットワークを構築することにより、我が国としての一元的な統合データベースを整備することが必要である。

(2)「統合データベース推進本部(仮称)」の整備

「統合データベース推進本部(仮称)」は、データベース統合を我が国の科学技術施策の重要課題の一つとして、省庁を超えたオールジャパンの視点で、ライフサイエンスの分野を広く俯瞰し、データベース統合に関わる基本方針の決定、統合データベース関連事業に関する方針の策定、「統合データベースセンター(仮称)」の運営全般に対する指導・助言等、統合データベース構築の司令塔機能を担うことが求められる。

なお、この本部の「司令塔機能」は、データベース統合に関する範囲を所掌し、その推進に向けた施策を策定するものであり、ライフサイエンスの各個別分野において、データベース構築を内容として含む個々の研究までには及ばないことが重要である。その趣旨を広く関係者に周知し、個々の研究者の創意工夫に基づく研究を尊重して、開発研究を加速していくような配慮が必要と考えられる。

本部の具体的設置については、総合科学技術会議のライフサイエンスPTに附属して設置することが現実的である。

(3)「統合データベースセンター(仮称)」の整備

研究資金を活用して先行的に取り組んでいる文部科学省の統合データベースプロジェクトが平成22年度で終了する。恒久的な研究基盤として、統合データベースを整備していくためには、この取り組みの成果をどのように引き継ぎ、新たな恒常的な体制として再構築するのかという喫緊の課題の現実的な解決が、まず、必要となる。

したがって、我が国における恒久的な統合データベース整備に向けてのロードマップとしては、平成23年度以降の現実的体制を第一段階として、「統合データベースセンター(仮称)」の整備を行い、その後、第二段階として、整備した体制の強化を図りつつ、我が国として目指すべき統合データベースに相応しいセンター機能が発揮できる体制を構築していくこととする。

① 当初の設置場所

これまでの我が国における統合データベース関連の事業実績や規模等を考慮すると、恒久的な統合データベースの整備に向けて、ROISが設置したライフサイエンス統合データベースセンターを中心として実施してきた文部科学省統合データベースプロジェクトとJSTのバイオフィーマティクス推進センター事業とを一本化し、それを母体として、我が国のデータベース統合にかかる実務や研究開発の中核機能を担うものとして「統合データベースセンター(仮称)」を整備することが現実的である。

その「統合データベースセンター(仮称)」は、これまで我が国のデータベース開発者に対する支援を行ってきたファンディング機能としてのノウハウと、文献情報等のデータベース利用者に対する支援を行ってきたノウハウの両方を兼ね備えているJSTの新たな組織として整備し、設置場所は、実際のデータベースの構築や研究を行っている現場と一体となって活動できる場所が望まれる。

② 求められる機能

「統合データベースセンター(仮称)」の機能としては、文部科学省の統合データベース事業でROISが中心に取り組んできたライフサイエンスデータベースの整備(統合・維持・運用)、JSTが取り組んできたライフサイエンス分野における基盤的データベースの高度化・維持・運用を一体化して引き継ぎ、更なる発展を目指していくとともに、利用者が、各省等におけるデータベースも含めて我が国全体に散在するデータベースを有効に使うことのできる運用を目指した一元的なネットワーク化の構築を主なものとする。

具体的には、データベース統合に必要な調査、データベースの統合に必要な標準化、システムの構築・維持・管理、ポータルサイトの構築、データベースの受入れ・管理・更新、データベースの品質管理、各省等のデータベースとのネットワークの構築、海外との連携等の実務機能に加えて、データベースの統合化や高度な検索等、統合的利用のための技術開発(インデックス、辞書、データフォーマットなどの構築)の機能とする。

特に、各省や研究機関(事業)等で公的資金によって整備を進めているデータベース、さらには企業等で整備が進められているデータベースを含め、将来的な一元的なネットワーク化を目指すにあたっては、事業開始段階からの関連事業間の調整機能が重要であり、ポータルサイト等の共通化のみならず、必要に応じてデータベースの構築を行う機関との事前の連携についても重点的に取り組んでいくことが必要である。

③ 組織体制

「統合データベースセンター(仮称)」は、センター長の強いリーダーシップの下、上記の機能が十分発揮される組織とすべきである。また、統合データベースの整備・統合をおこなう実務者には、広く有能な人材を集めることが必要である。

また、システム設計やポータルサイトの構築など、具体的なセンターの企画・運営については、情報を創出する側と統合データベースのユーザーとなる側の産業界や研究者コミュニティの意見を反映することが可能となるよう、「統合データベース運営委員会(仮称)」を設置し、協議・調整を行いながら推進していくことが必要である。

特に、この委員会には、関係省庁や産業界に加えて、ネットワークに参加するデータベース関係者の参加が必須であり、データベースを構築する側と利用する側の双方の視点から広く研究者コミュニティの意見を十分反映できるような委員構成とし、透明性、公平性、客観性に十分配慮した運営を行うことが求められる。

④ 運営費

恒久的なデータベースの統合・維持・運用を行う統合データベースセンターの基盤的運営経費(共通に必要な実務の企画、そのために必要となる調査、システムの構築・維持・管理、ポータルサイトの構築・維持・管理、データベースの受入れ・更新・品質管理、ネットワークの構築等に必要経費)は、センターを担う機関の運営費交付金のみで賄うのではなく、オールジャパンの

体制に相応しい関連府省等による新たな支援の仕組みを検討することが必要である。

その際、新たな体制の発足が求められている平成23年度を鑑み、平成23年度の予算要求時まで、センターの基盤的運営経費については、内閣府を中心に関係府省間で検討しておくこととする。

また、新たな仕組みによって各省庁及び研究機関における先進的な研究が阻害されることのないよう配慮するとともに、個々のデータベースの開発・高度化等の研究については、各省庁からの研究資金の活用も検討すべきである。その他、民間企業やファンディング機関等からの支援を受けることも検討する必要がある。

(4)「統合データベース評価委員会(仮称)」の整備

我が国における統合データベースを着実に整備していくためには、その取組み状況をチェックする機能が求められる。具体的には「統合データベース推進本部(仮称)」の諮問機関として「統合データベース評価委員会(仮称)」を設置し、「統合データベースセンター(仮称)」の運営全般について、評価を行い、その結果を「統合データベース推進本部(仮称)」へ報告することが必要である。

(5)国内のデータベース等の整備

ライフサイエンス分野における我が国としての一元的な統合データベースの整備のためには、これまで関係省庁や研究機関(事業)、企業等で取り組まれてきたデータベースの整備や、統合データベース化への取組の成果を更に発展させ、それぞれ整備を図っているデータベースの品質と内容を十分に確保した上で、単独のデータベースとして、あるいは分野のいくつかのデータベースを束ねたものとして、「統合データベースセンター(仮称)」と連携することによって、ネットワーク化を図り、オールジャパンの体制として編みこんでいくことが最も適切である。

具体的には、統合データベースに参加する、既存の各省庁や研究機関(事業)、企業等におけるデータベース事業等(既存の各省における統合データベース事業を含む)について、「統合データベース運営委員会(仮称)」において、データベースの品質管理や更新のあり方等について、関係者間で協議・調整しつつ整備を図っていくことが求められる。

更に、ネットワーク化に必要となるような共通基盤の整備にあたっては、先行して統合化が進められている文部科学省の統合データベースプロジェクトを基礎として、可能な範囲で活用していくことが効率的である。

なお、今後、データ蓄積型の研究事業を実施する中で開発されていくデータベースについても、整備を図り、ネットワーク化を進めていくことが必要である。

(6)国際連携

「統合データベースセンター(仮称)」は、わが国のライフサイエンスの統合データベースを整備していく上で、わが国の特徴あるコンテンツや先進的な情報技術を活かした取り組みが重要であるが、これらをもって国際的にも貢献できるように、他国との緊密なる連携を行うことが必要である。このような国際連携を行う際、欧米諸国はもちろんのこと、今後の関連分野において大きな発展が見込まれるアジア諸国との連携が重要となろう。

(7)人材育成

これまで実施されてきた統合データベース事業等で育成されてきた人材を活かしつつ、高度な

専門性を備えた人材の継続的な育成や、大学、各独立行政法人研究機関等、産業界とも連携したキャリアパスの構築を進めていくべきである。

(8) 人体に由来するデータ等の取り扱い

人体に由来するデータ等については、収集、保存、公開の方針が、個人情報保護等の観点から、人以外の動物や物質等由来の情報とは異なり、慎重な対応が不可欠である。そのため、人を対象とする研究に関連するデータベースのネットワーク化等の整備等に向けては、「統合データベースセンター(仮称)」の整備に関する議論と並行して、「統合データベース推進本部(仮称)」に部会を設置して、できるだけ速やかに方針を検討すべきである。

5. データベース登録の推進(インセンティブ)

データ蓄積型の研究の成果として産出したデータベースのうち、統合データベースに登録できる内容のものについては、その登録が進むよう、例えば、総合科学技術会議が産出されたデータベースの登録状況を評価する仕組みが必要である。なお、民間が参加するような研究成果の登録については、一定の配慮が必要である。例えば、公的研究資金と企業とのマッチングファンドを活用したような研究や、知的財産権の確保を目指した研究の場合には、特許出願が公開された以降にデータベースの登録ができるよう、一定の期間、登録を猶予するような仕組みとするべきである。

また、統合データベースに登録できるデータベースが成果として予測される研究事業については、予め当該課題を採択する際の条件として、研究成果として産出されたデータベースの登録について条件づけるようにすることも求められる。

さらに、根本的なインセンティブの向上のためには、統合データベースを活用すれば、孤立したデータベースを用いた研究からは達成できない成果があがるという実績を発信し、研究者自らがデータベースの連携を推進するような状況を作り出すことが必要であり、そうした経験や実績が積み重なるような共同研究等を推進・支援していくことも重要である。

6. 目指すべき統合データベースの整備に向けたロードマップ

本報告書においては、まず、平成22年度末までを「準備段階」として、文部科学省統合データベースプロジェクトとJST バイオインフォマティクス推進センター事業を一体化した新たな「統合データベースセンター(仮称)」の設立に向けた準備を行い、平成23年度から平成25年度末までを「第一段階」として、新たに設置したセンターを運用していく中で体制整備を図りつつ、恒久的な統合データベース構築のあり方について検討し、平成26年度以降を「第二段階」として、恒久的な統合データベースを構築することを目指し、それに向けたロードマップについて取りまとめた。

(1) 準備段階の整備のあり方

平成22年度末までの準備段階における整備について、

- ①総合科学技術会議に司令塔機能として「統合データベース推進本部(仮称)」を設置し、新たな「統合データベースセンター(仮称)」を整備していくための助言・指導や、進捗の確認等を行うこと、
- ②具体的には、本部で策定した方針の下、ROISが中心に行ってきた文部科学省統合データベ

ープロジェクトとJSTのバイオフィォマティクス推進センター事業とを一本化して、平成22年度末までにJSTの新たな組織として「統合データベースセンター(仮称)」を整備すること、に取り組んでいくこととする。

(2) 第一段階の整備のあり方

平成23年度から平成25年度末までの第一段階の整備については、

- ①「統合データベース推進本部(仮称)」において、「統合データベースセンター(仮称)」の運営等について指導・助言を行っていくとともに、平成26年度以降の恒久的な統合データベースのあり方について検討していくこと、
 - ②本部で策定した方針の下、具体的な実行機関としてJSTに整備された「統合データベースセンター(仮称)」を着実に運営していくこと、また、ROISにおいては、人材育成機能等を引き続き行うとともに、統合データベース構築の基盤技術の開発を行うこと、
 - ③「統合データベースセンター(仮称)」は、センター長のリーダーシップの下、データベース統合化に向けた活動を行うこと、
 - ④「システム整備等、具体的な企画・運営に関しては協議・調整を行う「統合データベース運営委員会(仮称)」を「統合データベースセンター(仮称)」に設置して推進すること、
 - ⑤「統合データベースセンター(仮称)」の運営全般について評価を行う機関として、「統合データベース評価委員会(仮称)」を「統合データベース推進本部(仮称)」の諮問機関として設置すること、
 - ⑥これまで各省や研究機関(事業)、企業等で取り組んできたデータベース(既存の各省の統合データベースを含む)についても、引き続き適切な整備を図ること、
- などに取り組んでいくこととしている。

(3) 第二段階の整備のあり方

平成26年度以降、第二段階として目指すべき「統合データベースセンター(仮称)」のあり方については、第一段階で整備した成果を踏まえつつ、同センターの更なる機能強化を図るため、「統合データベース運営委員会(仮称)」等の議論も踏まえ、引き続き「統合データベース推進本部(仮称)」において検討していくことが必要である。

なお、ライフサイエンス分野だけでなく、環境分野やナノ分野など、科学技術に関するデータベースを広く包括的に整備していくという方向性が考えられ、それに相応しい体制の構築を目指して、科学技術全体の観点からデータベースの統合のあり方を検討していくことも求められる。

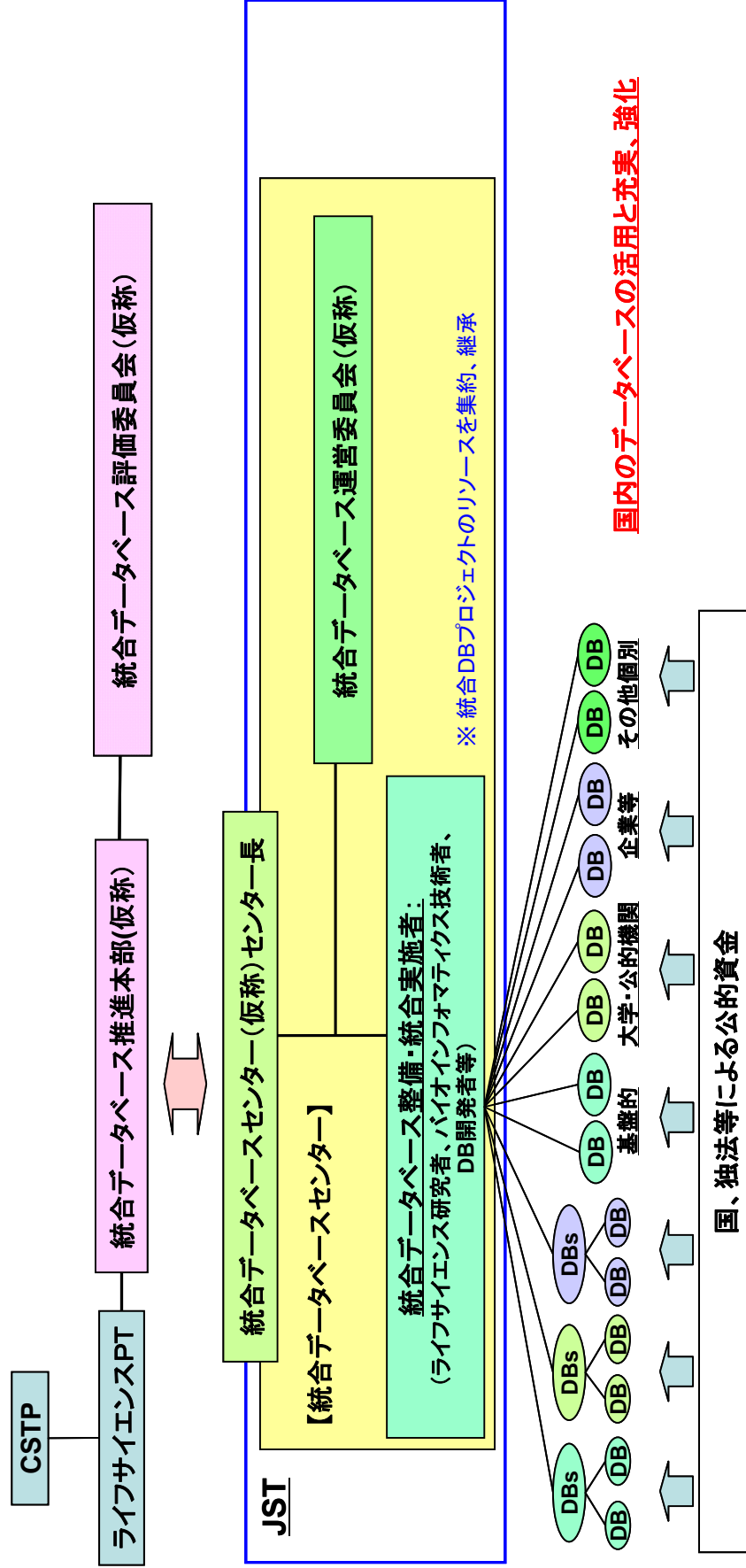
また、将来的にもライフサイエンス分野に特化した統合データベースの整備機関として機能強化を図っていく場合には、第一段階で整備した体制の取組を踏まえた上で、我が国のライフサイエンスの先導的な研究基盤整備という観点から、研究開発要素も含めた総合的な機関として「統合データベースセンター(仮称)」を更に発展させる方向で検討していくことが必要である。

この際、第二段階の「統合データベースセンター(仮称)」が満たすべき必要条件は、①継続性、②情報の創出と利用双方の関係者の意見を反映できる運営、③実際の業務を通して人材を養成する機能を持つこと、④関係省庁による支援と考えられる。

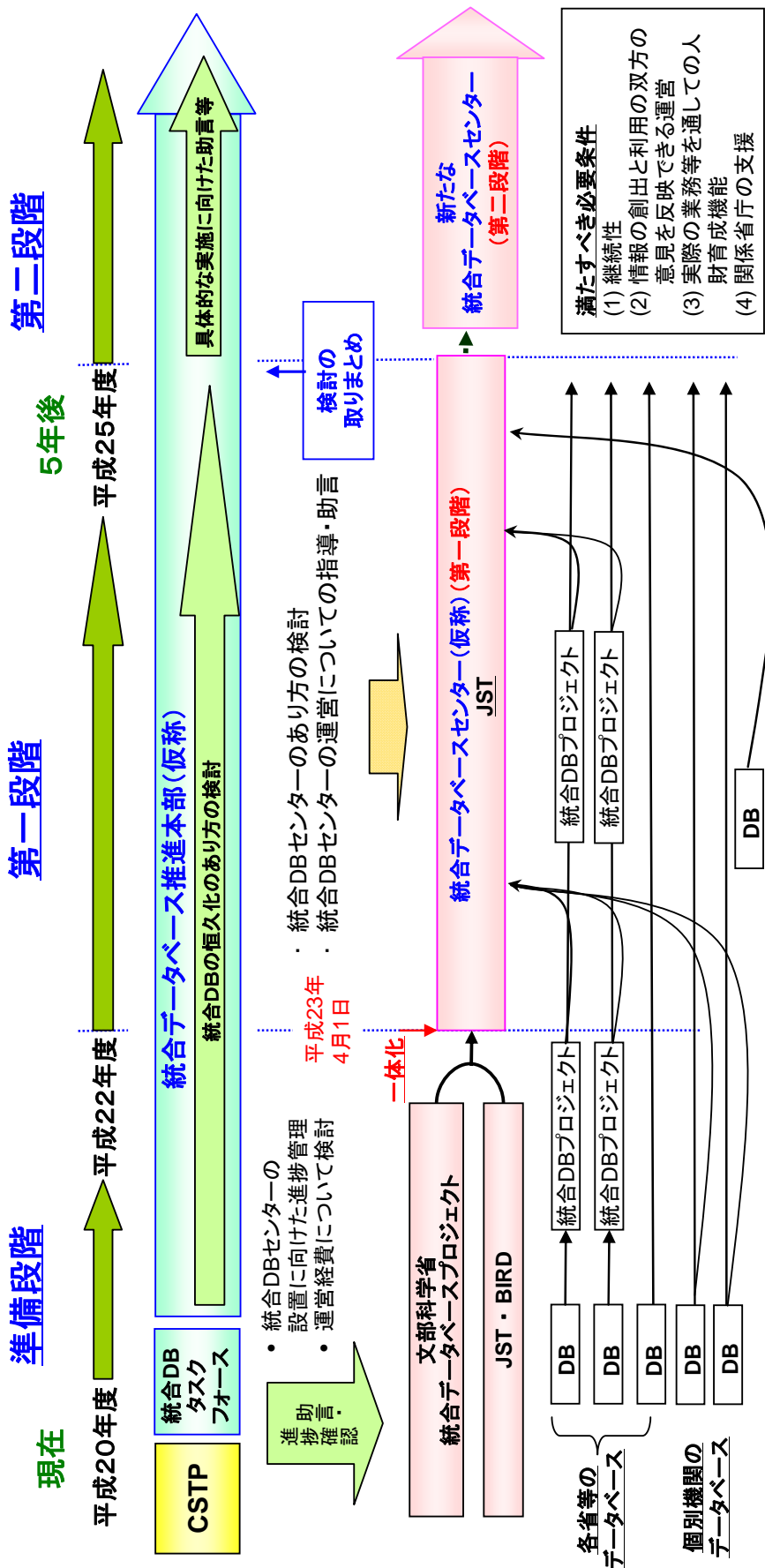
これら第二段階としての整備の検討は、第一段階として平成23年度より整備される「統合データベースセンター(仮称)」の成果等を踏まえながら、総合科学技術会議の「統合データベース推進本部(仮称)」において、平成26年度以降の恒久的な統合データベースの組織のあり方として、例えば、①引き続きJST内の組織として更に充実強化を図り組織を編成していくこと、あるいは、

②分野的に近い機関に移行すること、③新たな機関を設置すること、などの実現可能性等について検討していくこととし、必要であれば法制化も視野にいれつつ、平成25年度末を目処に、その方向性を取りまとめることとする。

新しい統合データベースの組織体制(第一段階)



統合データベース整備のロードマップ



【 統合データベース タスクフォース委員 】

五條堀 孝 (座長)	国立遺伝学研究所 副所長・教授
浅島 誠	(独)産業技術総合研究所 器官発生工学研究 ラボ長
大石 道夫	(財)かずさDNA 研究所 所長
小幡 裕一	(独)理化学研究所バイオリソースセンター センター長
小山 博史	東京大学大学院医学系研究科 教授
勝木 元也	(独)日本学術振興会 学術システム研究センター 副所長
小原 雄治	大学共同利用機関法人 情報・システム研究機構 理事
榊 佳之	豊橋技術科学大学 学長
佐々木 卓治	(独)農業生物資源研究所 理事
諏訪 牧子	(独)産業技術総合研究 所生命情報工学研究センター主幹研究員
祖父江 友孝	国立がんセンター がん対策情報センター がん情報・統計部長
高木 利久	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教授兼センター長
豊田 哲郎	(独)理化学研究所 生命情報基盤研究部門長
長洲 毅志	エーザイ株式会社 理事・研究開発担当付 担当部長
中村 春木	大阪大学蛋白質研究所附属プロテオミクス総合研究センター センター長・教授
増井 徹	(独)医薬基盤研究所 生物資源研究部 部長
水上 政之	(独)科学技術振興機構 理事
横山 茂之	(独)理化学研究所 生命分子システム基盤研究領域 領域長
吉田 輝彦	国立がんセンター研究所 腫瘍ゲノム解析・情報研究部 部長

【 検討経過 】

- ・ 第1回会合開催 (平成20年12月25日): 統合データベースに関するヒアリング
- ・ 第2回会合開催 (平成21年 1月21日): 論点整理
- ・ 第3回会合開催 (平成21年 2月17日): とりまとめ骨子案検討
- ・ 第4回会合開催 (平成21年 3月18日): 報告書案検討(1)
- ・ 第5回会合開催 (平成21年 4月 2日): 報告書案検討(2)