



ライフサイエンス統合データベースと国の施策

高木 利久

ライフサイエンス統合データベースセンター(DBCLS) 大学共同利用機関法人 情報・システム研究機構(ROIS)



文科省ライフサイエンス統合DBプロジェクト開始まで

平成12年8月~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

「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17日)

平成13年4月

JST(科学技術振興機構)に BIRD(バイオインフォマティクス推進センター)

平成17年度~平成19年度

内閣府連携施策群 科学技術振興調整費

「生命科学データベース統合に関する調査研究」(平成20年5月)

平成17年8月~平成18年5月

文科省ライフサイエンス委員会DB整備戦略作業部会報告書

「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)

平成18年9月~平成19年3月

文部科学省 ライフサイエンス分野の統合データベース整備事業(公募) (略称 統合データベースプロジェクト)

平成19年4月(~平成23年3月)

文部科学省 ライフサイエンス分野の統合データベース整備事業(公募) ROIS(情報・システム研究機構)にDBCLS(ライフサイエンス統合DBセンター)



生命科学データベース統合に関する調査研究(H17-19)(概要)

D B統合が望まれる背景 ~研究内部の変化と研究環境の変化~

- ・巨大観察データの登場によるライフサイエンス全体のデータ(ベース)依存型への変貌
- ・高額な巨大データに関して大学・企業研究開発の公共データ(ベース)への依存増
- 科学の民営化(データ占有による知財獲得、データベースサービスも競争的)

多面的かつ 複雑な背景

技術的課題は

課題:「生命科学データベース統合に関する調査研究」

実施内容:①国内外のDB調査、産業界等からのヒアリング

②DB統合に向けた技術的FS:省をまたいだDB連携

十分対応可能 ③DB統合に向けた制度調査:DB統合が進む米国等の制度調査

制度上の問題 を明確化

NCBIを代表とする米国での統合DBの成功例とそれを支える社会の仕組み(制度)の変化 OECD等国際的な公的機関によるデータ共有の必要性の認識

DB統合化に向けた提言に資する試案

- 「政府資金によるデータ産生型プロジェクトのデータを我が国の研究社会で 早期に共有するためのルール」が必要
- 2. 公共財としてのデータを保全・管理し、長期にわたるデータの育成と共有を 行う公的機関を設ける(ルール運用の際の受け皿)



表 1 日米欧の主要中核機関の概要

	日本		米国	欧州
	(中核的な機能を果た 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	たしている機関の例) 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NL M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 ビの 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)



ポータル

生命科学系 アータベース カタログ 🐚 🌭

生命科学系 学協会カタログ 🐚 🌭

生命科学系主要プロジェクト一覧 🌭

生物アイコン 🐚 🌭

<u>WingPro</u> (JSTのDBポータル) 📺 🌭

Webリソースポータルサイト (JST解析ツールポータル)



検索

生命科学アータベース横断検索 📺 🌭

蛋白質核酸酵素 全文検索 🗑

文科省「ゲノム」研究報告書 全文検索

TogoProt (蛋白質関連アータベース統合検索) 🕍 🌭 new

OReFiL (オンラインリソースファインダー) 📺 🌭

Allie (略語の正式名称を検索) 🕍 🌭

inMeXes (文献中の英語



データベース

DNAデータベース総覧と検索 😭 🌭

(DDBJ/EMBL/GenBank)

遺伝子発現パンク(GEO)目次 📺 🌭

KazusaAnnotation & Navigation (かずさDNA研究所) 🕍 🌭

KazusaMart (かずさDNA研究所) 📺

ゲノムネット医薬品データベース (京大)

統合医科学アータベース (東京医科歯科大グループ)

疾患解析から医療応用を実現するDB開発 (東大グループ)

持続可能型社会への貢献遺伝子データベース (長浜パイオ) 📉 🖎

tRNADB-CE:エキスパートがキュレートしたtRNADB (長浜パイオ)

(5)

日本糖鎖科学統合データベース (JCGGDBワーキンググループ) 😭

生体分子の熱力学データと構造データの統合 (九工大) 😭

MDeR (ライフサイエンス分野のメタデータ要素レポジトリ) (JST)



アーカイブ

生命科学系データベースアーカイブ 🌭

DDBJトレースアーカイブ (遺伝研 DDBJ)

DDBJリードアーカイブ (遺伝研 DDBJ)



ツール&解析サービス

BodyParts3D/Anatomography 📺 🦠

Wired-Marker 📺 🦠

MiGAP (微生物ゲノムアノテーションパイプライン) 📺 🌭

DBCLS Galaxy 📺 🦠



基盤技術開発

TogoDB (誰でもデータベースが構築できる) 🚳

TogoWS (ウェブサービスの標準化) 📺 🌭

OpenID 認証システム

情報基盤サイト (CBRC) 🌭

築と公開

ab.

BioHackathon(DBCLS パイオハッカソン) 2010, 2009, 2008



http://lifesciencedb.jp/

教材・人材育成

<u>統合TV</u> (DBやツールの動画教材) 🕍 🌭

MotDB (教育・人材育成のサイト)



統合DB事業

<u> バンフレット</u> (PDF)

シンポジウム・講演会

外部評価

文科省 統合アータベース整備事業サイト

運営委員会等の資料

日本の生命科学データベース政策

過去のコンテンプ



日本の生命科学データベース政策

2010年6月17日更新

会議・報告等の流れ

平成12年 (2000)	11月	ゲノム情報科学におけるわが国の戦略について (科学技術会議 ライフサイエンス部会 ゲノム科学委員会)	「人材養成」「研究開発振興」「データベース整備戦略」の3課題をふまえ、2001年度にJST-BIRDが設立されることとなる。 ポストゲノムの戦略的推進に関する懇談会(第5回)によるものか?
平成14年 (2002)	6月5日	ライフサイエンスに関する研究開発の推進方策について (文部科学省 科学技術・学術審議会[第一期] 研究計画・評価分科会) [ダウンロード]	次世代のゲノムインフォマティクス等の新しいゲノム解析技術の開発 の 今後5カ年を見通した 重要な研究課題5つの1つとして 「基盤的な生命情報データベースを整備するとともに、各種の 生命情報を 統合したデータベースの開発」。 また、バイオリソースの整備として、「多数の生 物種についての情報がシームレスに 得られるような統合データベースが必要である。 データ ベース中核機関とバイオリソース構築機関 及びその研究コミュニティが共同してデータベース を作り上げることが必要である。」 産学連携の一環としても、大学での研究成果の内容や権利 関係がデータベースとして 開かれた形態で整備されることが必要、との指摘。 平成18年7月 28日にアップデート版が出ている。
	6月14日	文部科学省 科学技術・学術審議会[第一期]総会(第6回) [議事次第・配布資料] [議事録]	「ライフサイエンスに関する研究開発の推進方策について」(研究計画・評価分科会)の 総会への報告。 委員より「分野別の研究推進方策で触れられているとおり、 各分野の研究成果のデータベースを整理して資産として残し、 広く利用できるものにしていくことは重要である。しかしデータベース作成という作業は論文にはなかなかなりにくい。 そのような仕事に携わる研究者を評価する仕組みも考えてほしい。」との指摘 (議事録より)
	12月6日	バイオテクノロジー戦略大綱 (内閣 BT戦略会議) [ダウンロード]	バイオインフォマティクス分野の基本行動計画として 「統合データベースの構築を進めるとともに、標準化の検討を進める等により、 利用環境や競争環境の整備を図る。(関係府省)」 このころの「統合データベース」は、研究成果としてのデータベース、産学連携の 足がかりとしてのデータベース、「DNAとタンパク質」や「イネゲノムと育種情報」 のような各種情報の統合、という感じか(現在よりもゆるやかな統合?) 第8回 BT戦略会議(平成18年1月26日)には、これを受けて、現在の 統合データベースプロジェクトにつながっている、という書き方がされている(文部科学省の取組資料)。 知財に関しても、日本版バイドール法の適用などについて触れられている。



我が国におけるライフサイエンスDBの問題点

- DBが散在していて所在情報や利用法が分からない
 - 我が国で300、世界で10,000ものDB
 - 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
- DBやDBのエントリに信頼性の高い注釈がついていない
 - DB構築、維持を行える人材不足、DB構築への評価の低さ
- 大型プロジェクトの成果公開が不十分
 - 我が国のライフ関係予算は年間3,000億円以上
 - 公開されているものもプロジェクトが終了すると更新ストップ
- ばらばらに構築・管理されていて検索・解析・応用が困難
 - 現在の統合化は分子レベルで行われていて医療などへの応用困難
 - 日本語化されていないので研究動向や分野の状況の把握困難
- 不可欠な基盤なのに我が国にはDB戦略がない



データベース整備戦略作業部会で提言された推進方策(1)

①現状調査、評価、戦略立案

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。

②基盤データベースの支援

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。

③ポータルサイトの構築と運営

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。

④統合データベース研究開発

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を見出すことが可能になる。

⑤維持困難なデータベース受入れ

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。



データベース整備戦略作業部会で提言された推進方策(2)

⑥文献情報との連携

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。

⑦アノテーションの実施

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。

- ⑧新たなデータベース構築への投資
- 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。
- ⑨データベースを活用した研究の促進

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。

⑩データベース開発のための人材育成

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。

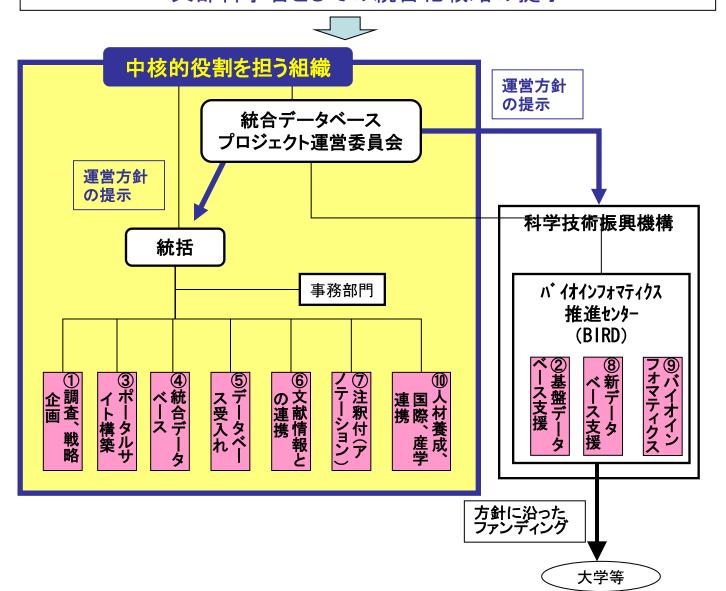


運営体制

文部科学省科学技術·学術審議会 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 文部科学省としての統合化戦略の提示

委託費の範囲

作業部会報告書の 取り組むべき課題





統合データベースプロジェクトの体制図

運営委員会 ライフサイエンス、情報科 学、ライフサイエンスDB有 識者、各省代表者

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

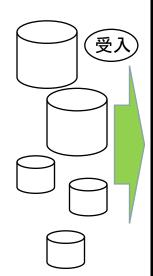
補完課題 (平成19年10月〜)

理研

蛋白立体構造

植物オミックス

既存のDB



アノテーションの実施

データベース 開発のため の人材養成 データベースの現状調査、 評価、戦略立案機能の充実

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 文献情報 との連携

を 維持が困難に なったデータ ベースの受入れ

参画機関

科学技術振興機構(JST)

ポータル

産総研CBRC

共通基盤技術(ワークフロー技術)

かずさDNA研 植物統合化 九州大学

多型情報統合化

奈良先端大

共通基盤技術(日本語処理)

東京大学

お茶の水女子大

教育

長浜バイオ大

教育

分担機関 (平成19年4月〜)

京都大学

化合物DB

東京医歯大Gr

東京大学Gr

臨床DB

疾患DB

九工大 蛋白立体構造



プロジェクト遂行の基本方針

- 文科省だけでなく我が国の中核センターを指向
 - 日本版NCBI、ただし、まねはせず根本から検討
 - 医学図書館やゲノムセンターの一部の機能も
 - 様々な立場、利用者、府省の人からなる委員会
- 研究ではなくサービス事業として推進
 - 我が国の生命研究の効率化と質の向上に寄与
 - それに必要なことは何でもする(技術面以外も)
 - 網羅性、利便性、持続性の追求
 - 自分では知識発見研究しない(縁の下の力持ち)



整備、統合に対する考え方

- 技術面だけではなく制度面での取り組み
 - データの権利関係、個人情報の取り扱い、なども
- 統合は手段であり、目的ではない
 - 物理的に一つのDBが目的ではない
 - 望ましい整備、統合は研究分野や利用者毎に異なる
 - 統合のあり方は研究の進展とともに変化する
- 目的は生命研究の研究開発の効率や質の向上
 - 必要十分なデータにすばやく辿り着ける
 - 公開ではなく<u>共有化</u>を図り、データマイニングを可能に
- 研究開発に関わるデータすべてを対象に
 - データだけでなく、文献(論文、総説)、図表、特許等も



文科ライフ委DB戦略WS報告

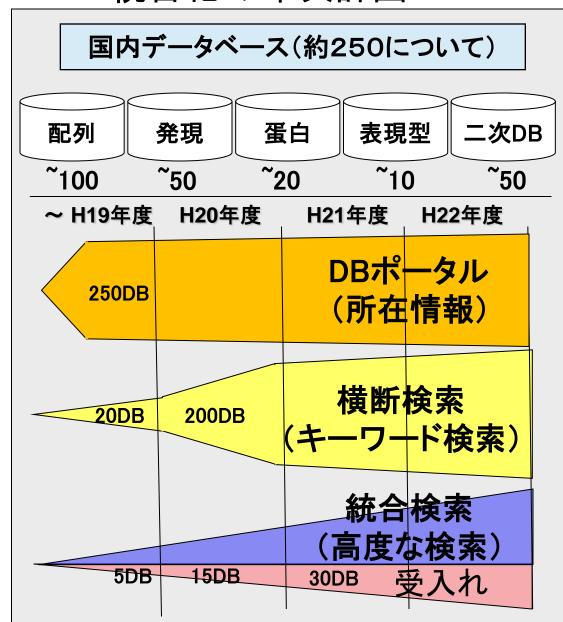
H18年度

戦略立案

ポータル

統合技術

統合化の年次計画



産業・臨床・育種 従事者の課題 解決への展開

4 省 庁 統 合 DB

環境変化と知識 データ増大への 絶え間ない対応



統合化の全体像とステップ(DBCLSの考え方)

データバンク事業 プロジェクトDB 個別DB ツール ①データベース、ツール、プロジェクトのカタログ化、ポータルサイト ②データベースやツールの使い方、使い分けの情報 ③データベースやツールの統一的、シームレスな検索、利用 ④知識発見支援のためのデータベース統合化、解析ワークフロー ⑤目的、用途ごとのデータベース統合化、解析ワークフロー

イノベーション、新たな知識発見、データベース生物学



ライフ課委託業務公募要領のデータ提供条項

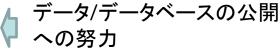
(7) 統合データベースプロジェクトへのデータ提供

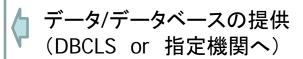
本プログラムはライフサイエンス分野のデータ産生を伴うプロジェクトであるため、本プログラムの実施者は、ライフサイエンス分野における研究に資するよう、本プログラムの実施の結果産生したデータ又はデータベースを、一般に利用可能な形で公開するべく努めるものとします。そのため、本プログラムの実施者は、毎事業年度にまとめる成果報告書の提出に合わせて、論文発表等で公表された成果に関わる生データの複製物、又は本プログラムで構築した公開用データベースの複製物を、ライフサイエンス統合データベースセンター*又は文部科学省の指定する機関に提供するものとします。提供された複製物については、ライフサイエンス分野のデータベース整備(統合データベース構築)に利用するため、これら複製物に関わる知的財産権を非独占的に複製・改変その他必要な形で利用できるものとします。なお、本プログラムの実施者は、複製物の提供を受けた機関の求めに応じ、複製物を利用するに当たって必要となる情報提供に協力するものとします。

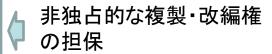
※ライフサイエンス統合データベースセンター(http://dbcls.rois.ac.jp/index.html) 生命情報のデータ又はデータベースはライフサイエンス研究を支える基盤であり、その整備を推進するため、文部科学省では「統合データベースプロジェクト」を実施しています(http://lifesciencedb.mext.go.jp/)。ライフサイエンス統合データベースセンターは大学共同利用機関法人情報・システム研究機構に属し、当該プロジェクトを実施するための中核機関です。本センターでは、我が国のライフサイエンス関係データベースの利便性の向上を図るため、データベース整備戦略の立案・評価、データベースの社合化の基盤技術開発、及び統合化データベース開発等を行い、ライフサイエンスデータベースの統合化を推進しています。

なお、<u>本センターは、ライフサイエンスの基盤となるデータの保存と配布の役割を担うデータデポジトリーのサイトであり、データ提供者と競合するような研究活動を</u>行う機関ではありません。

(平成20年度以降の公募)









DBCLSと統合データベース プロジェクトの説明

┣ 提供者と競合しないこと ■ の説明



データベース統合化に関する農水省の取り組み

プロジェクト名:農林水産生物ゲノム情報統合データベース

期間:2006-2010

予算: 2006 2007 2008 2009 2010

275M¥ 721M¥ 707M¥ 700M¥ 645M¥* *:概算要求額

主要担当機関:農業生物資源研究所

概要:

農水省所管のゲノムデータベースを 統合するAgriTOGO(農林水産生物ゲノム 情報統合データベース)を構築・運営する。 AgriTOGOデータベースの主な内容は 以下の通りである。

- 統合検索(各データベースを横断して検索)
- ・統合データベース(イネ・カイコ)
- データベース便覧(各データベースの カタログ)
- 各種解析ツール





データベース統合化に関する経産省の取り組み

プロジェクト名:統合データベースプロジェクト事業

期間:2008-2011

予算:

2008	2009	2010
70M¥	70M¥	50M¥*

*:概算要求額

主要担当機関:

産総研 バイオメディシナル情報研究センター

概要:

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ついて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

現在までの成果物提供数は以下の通りである。

- データベース情報 60件
- ·解析ツ―ル情報 44件
- ・プロジェクト情報 51件

過去の取り組み状況:

- ・バイオインフォマティクス関連データベース整備 (2000 - 2004; 総額3,026M¥)
- ・ゲノム情報統合(2005 2007; 総額1,581M¥)





厚労省のデータベース関連主要プロジェクト

疾患ゲノムデータベース(GeMDBJ)

疾患ゲノムデータベース(GeMDBJ;Genome Medicine Database of Japan)は、アルツハイマー病、がん、糖尿病、高血圧、喘息および薬理遺伝学(ファーマコジェネティクス)のためのミレニアム・ゲノム・プロジェクト(MGP)研究グループ(SGMGP)により構築された統合データベースで、上記疾患に関わる多型(SNP)、遺伝子発現(GeneChipによる)、タンパク質発現(2D-DIGE, LC-MS/MS)が掲載されている。

トキシコゲノミクスプロジェクト(2002 - 2006)

成果として、150の化合物(医薬品等)をラット個体およびラット・ヒト肝細胞へ暴露した際の毒性情報および遺伝子発現情報などを収載した大規模かつ良質なデータベースおよび解析、毒性予測システム(TG-GATEs: Toxicogenomics Project-Genomics Assisted Toxicity Evaluation system)が構築された。トキシコゲノミクス・インフォマティクスプロジェクトとして継続中。150化合物、7億3000万件のデータを取得。現在は原則としてコンソーシアム内のみデータを公開し、一部の成果を一般に公開している。2010年に一般公開予定。

疾患関連たんぱく質解析研究・創薬プロテオームファクトリープロジェクト(2003 - 2007)

患者と健常者間で発現するたんぱく質の種類・量の違いを比較することにより疾患関連たんぱく質の発見とデータベース化を推進し、国際的に競争力のある医薬品開発のシーズの探索をめざすプロジェクト。具体的には、糖尿病、がん、高血圧、認知症等を対象とした疾患関連たんぱく質の探索、同定、定量を進めることにより、バイオマーカーや創薬ターゲット候補たんぱく質を見つけて創薬のための基盤的なデータベースを構築する。

2010年度まではコンソーシアム・研究機関内のみ公開、それ以降は一般に公開予定。



平成23年度以降の体制についての検討状況

平成20年10月~平成20年12月

文科省ライフサイエンス委員会

ライフサイエンス情報基盤整備作業部会作業部会報告書

「ライフサイエンスデータベースの統合・維持・運用の在り方」(平成21年1月13日)

平成20年12月~平成21年4月

内閣府ライフサイエンスPT統合DBタスクフォース

「統合データベース タスクフォース報告書」(平成22年4月10日)

平成21年9月~平成22年4月

JSTライフサイエンス分野統合データベースセンター設置準備委員会

「ライフサイエンス分野統合データベースセンター設置準備の検討のとりまとめ」

(平成22年6月14日)

平成22年6月~

総合科学技術会議ライフサイエンスPT統合データベース推進タスクフォース



統合データベース事業への取組み



総合科学技術会議における取組み 科学技術連携施策群¹⁾「生命科学基礎・基盤」の開 平成 始 (主監 五條堀 孝) ・補完的研究課題 「生命科学データベース統合に関 17年度 する調査研究」の実施(研究代表者 大久保 公策) 第3期科学技術基本計画²⁾の戦略重点科学技術³⁾ 「世界最高水準のライフサイエンス基盤整備」として統 平成 合データベースを位置づけ。 18年度 科学技術連携施策群の成果及び今後の見通し(中 間報告案)を作成 海外のバイオリソース事業の責任者を招聘し、講 平成 演会を実施 19年度 科学技術連携施策群の最終取りまとめ(終了) 総合科学技術会議ライフサイエンスPTに「統合DB 平成 タスクフォース」の設置 (座長 五條堀 孝) 20年度

関係府省の取組み

- ・関係省庁における検討
- ・文部科学省の統合データベースプロジェ クトの開始
- ・農林水産省の統合データベースプロジェクトの開始

- 経済産業省の統合データベースプロジェクトの開始。
- ・文部科学省「ライフサイエンス情報基盤 整備作業部会」報告書策定。

・「統合DBタスクフォース」の報告書の策定(4月10日)

1) 科学技術連携施策群:各府省の縦割りの施策に横串を刺す取り組み

平成

21年度

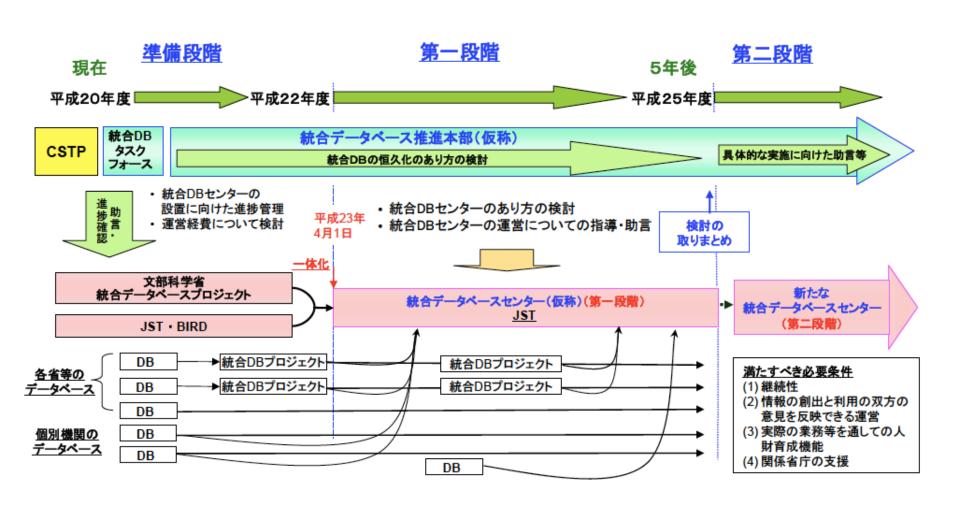
- 2) 第3期科学技術基本計画: 平成18年度から5年間の我が国における科学技術の推進戦略
- 3) 戦略重点科学技術:特に重要な技術を選択し、研究資金を集中させるべき分野

平成21年度概算要求の優先順位付けにおいて、「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(BIRD)との一体化を目指して、加速して事業を実施する必要がある。」と具体的な取組みを評価



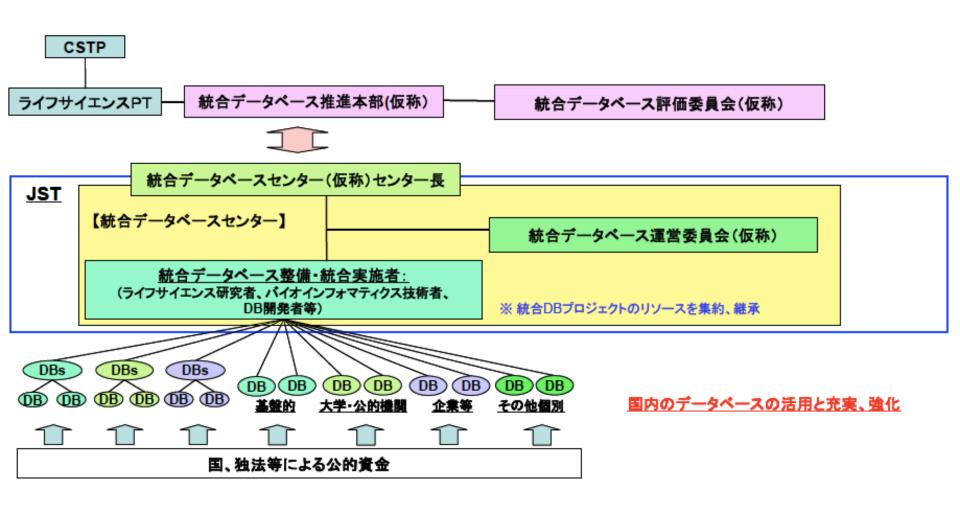
統合データベース整備のロードマップ







新しい統合データベースの組織体制(第一段階)





平成23年度以降の体制と課題

平成23年3月 文科省統合データベースプロジェクト終了予定 平成23年4月 JST新統合DBセンター発足予定

JSTの新統合DBセンター

- ・戦略立案、ガイドライン、ロードマップ策定、国際連携・ポータルの構築維持、統合DB運用
- 統合DB構築支援(ファンディング)

DBCLSは存続

•高度な統合DB基盤技術と人材育成に特化

課題

- データ共有のあり方
- ・ヒト由来データの取り扱い
- 関連府省の連携体制構築