

ライフサイエンスデータベース 統合推進事業の取り組み

坂東 明日佳
JSTバイオサイエンスデータベースセンター (NBDC)

日付 AJACS20 2012/2/10

事業のミッションと体制

ライフサイエンス分野のデータベースを統合し、データの価値を最大化することにより、ユーザに貢献できる日本が誇るべきデータベースセンターとなること (H23~)

基盤技術開発プログラム
データベース統合化の基盤となる
技術開発

DBCLS (ライフサイエンス統合
データベースセンター)

統合化推進プログラム
分野ごとのデータベース統合化

脳疾患画像G フェノームG
メタボロームG 蛋白質構造G
医薬・疾患G 糖鎖G
微生物G ゲノム疫学G
植物G
ゲノム多型G

JST-NBDC (バイオサイエンスデータベースセンター)

- カタログ、横断検索、アーカイブの運用
- 戦略の立案 (基盤技術と統合化推進のコーディネート、統合化ガイドライン)
- ファンディング●広報

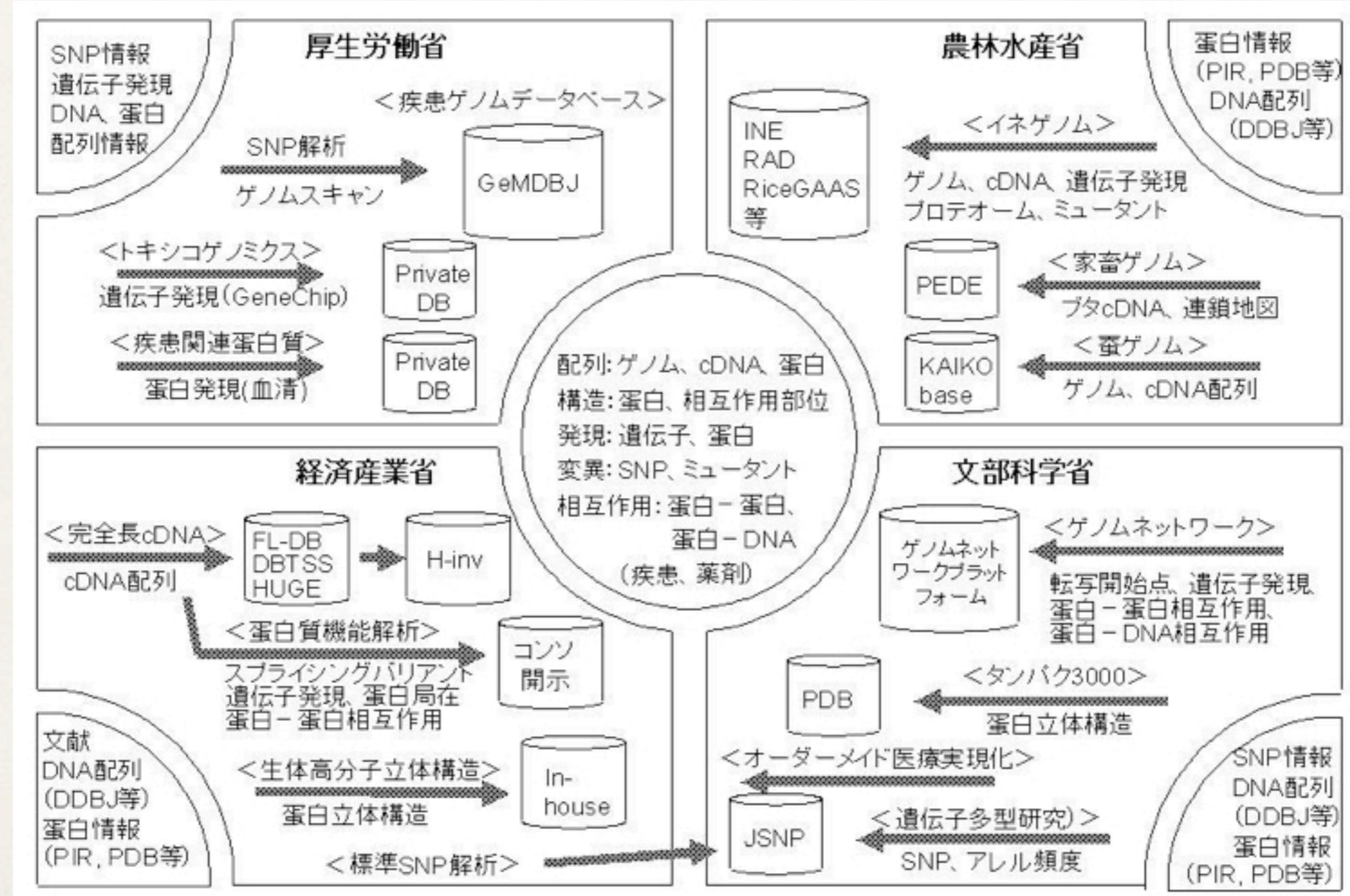
目次

- ＊ 前半：国内のDB基盤構築の現状と問題
- ＊ 後半：そのための取り組み～NBDCでの取りくみを中心に

目次

- * 前半：国内のDB基盤構築の現状と問題
- * 後半：そのための取り組み～NBDCでの取り組みを中心に

国内のDB生産の状況



・多種多様なデータの蓄積、・データの縦割りによる管理、・巨額な予算投資

我が国におけるDB基盤の問題

- * 1. DBの散在
- * 2. 成果公開が不十分
- * 3. ばらばらに構築、管理されているため、検索、解析、応用が困難
- * 4. DBの構築、維持、アノテーションを行う人材の重要性が浸透しない

我が国には長期的な視野にたったデータベース戦略がない

統合されるとは？論文の場合

- 書誌情報の整備 PubMedの存在
- 検索のための語彙やシソーラスの整備
- 論文の構造がジャーナル間で共通 (AbstractからReferenceまで)
- 英語で記述する文化
- 電子化の普及
- 評価のしくみ

「目的の論文を探し、入手し、効率よく読む、論文同士を比較する」が通常的に可能。

…論文に付随するデータについては？

統合データベースプロジェクト発足

H17年度～H19年度 内閣府連携施策群での取り組み
振興調整費「LSDB統合に関する調査研究」
(代表：大久保公策)

文部科学省 ライフサイエンス委員会での検討

データベース整備戦略作業部会

H18年度 準備期間

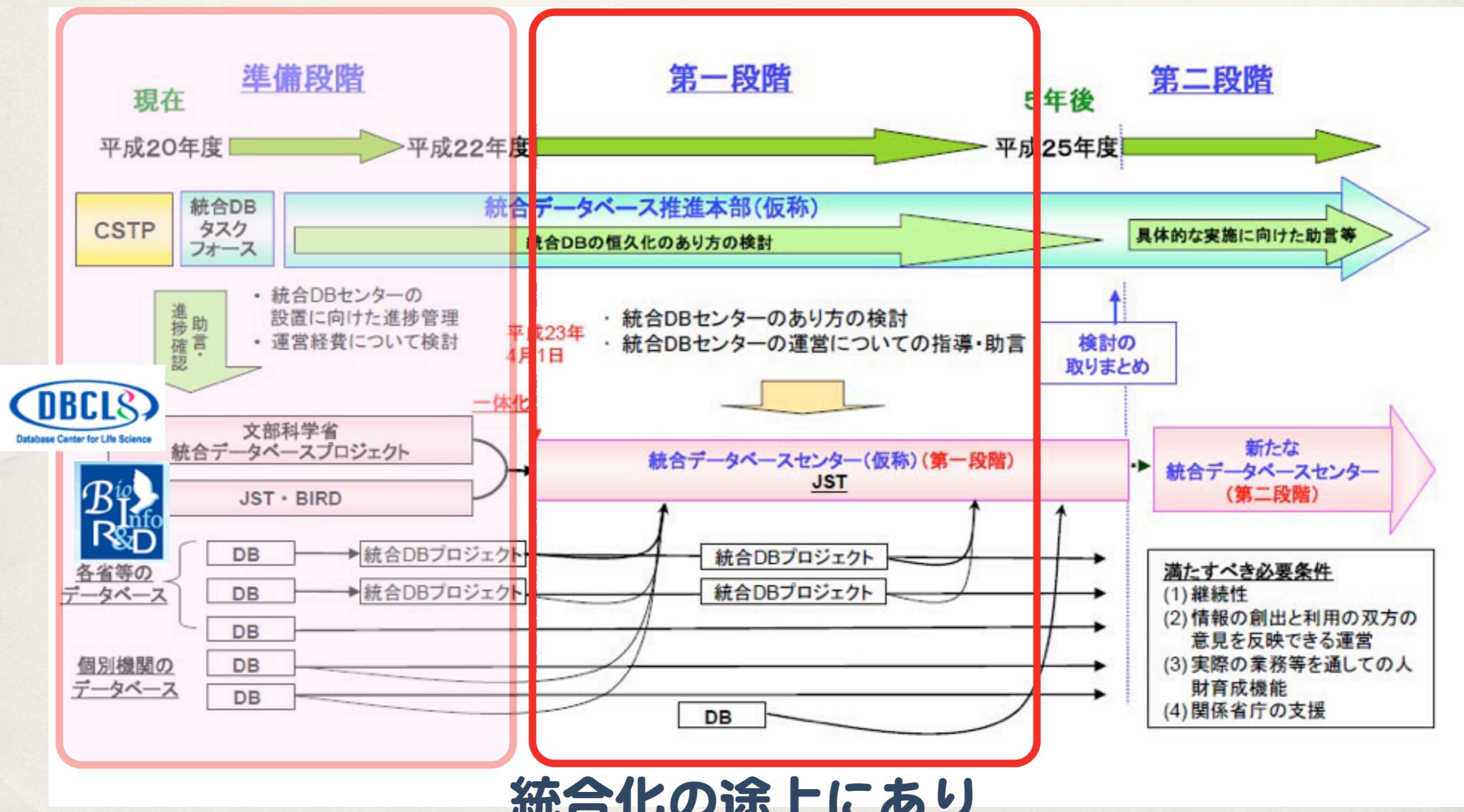
5月17日

「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」

→動向を踏まえた課題の整理と推進方策

H17年度～H22年度 統合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の取り組みを引き継ぎ発展させる



目次

- * 前半：国内のDB基盤構築の現状と問題
- * 後半：そのための取り組み～NBDCでの取りくみを中心に

事業の体制

基盤技術開発プログラム
データベース統合化の基盤となる
技術開発

DBCLS (ライフサイエンス統合
データベースセンター)

→ 「DBCLSの紹介 (河野) 」

統合化推進プログラム
分野ごとのデータベース統合化

脳疾患画像G	フェノームG
メタボロームG	蛋白質構造G
医薬・疾患G	糖鎖G
微生物G	ゲノム疫学G
植物G	
ゲノム多型G	

JST- NBDC (バイオサイエンスデータベースセンター)

- カタログ、横断検索、アーカイブの運用
- 戦略の立案 (基盤技術と統合化推進のコーディネート、統合化ガイドライン)
- ファンディング ● 広報

我が国におけるDB基盤の問題に対するアプローチとは？

The screenshot shows a search results page for 'Medaka full length cDNA'. The left sidebar has sections for 'System' (System, Image Gallery, Name Registration, Submission), 'cDNA/BAC/Fosmid' (Search, Submission, Download, Library Information), and 'Information' (Funding, Examples). The main content area displays a table with columns: Library Name, Organism, Stage, Entries (5', 3', Total), and Clusters (5', 3', Total). The table includes rows for olea, oleb, olec, olvi, olli, olki, olsp, olte, and others. A 'Request Help' button is at the top right.

ライブラリ名	組織名	発生段階	Entries			クラスター		
			5'	3'	Total	5'	3'	Total
olea	胚	ST22	22258	22215	44473	6077	5554	11631
oleb	胚	ST35	23134	23127	46261	7426	6876	14302
olec	稚魚	ST40	23000	23138	46138	6532	6211	12743
olvi	雄肝臓	成魚雄	18867	18978	37845	2497	2376	4873
olli	雌肝臓	成魚雌	23311	23534	46849	2678	2558	5236
olki	腎臓	成魚腎臓	22022	22442	45024	7763	6984	14747
olsp	脾臓	成魚脾臓	23402	23433	46835	5060	4792	9852
olte	精巢	成魚雄	24301	24342	48643	7078	6852	13930
olte	卵巣	成魚雌	23183	23184	46118	5686	5521	11207
			299944	299945	51581	21581	21582	47902
					15381			15381

* 1. DBの散在

* 2. 成果公開が不十分

The screenshot shows a 'Not Found' error page with the message 'The requested URL /cgi-bin/wndb/content.cgi was not found on this server.' Below it is a file manager interface with a tree view showing categories like 'Top(800)', 'Biology(94)', 'Chemistry(5)', 'Physics(304)', 'Mathematics(201)', 'Astronomy(108)', 'Engineering(9)', and 'Medicine(16)'. The right panel shows thumbnail previews of various files and documents.



3. ばらばらに構築、管理されているため、検索、解析、応用が困難



統合化のステップ

1. カタログ構築 DBを探す
リンクが行われる状態
2. 横断検索 データを探す
相互に一括横断検索できる状態
3. アーカイブ構築 データを使う
統一形式でのデータダウンロードが可能となる状態
4. データベース再構築連携
統一化されたフォーマットで再構築された状態

3つのサービスとして公開中

1. カタログ構築

Integbioデータベースカタログ
(生命科学系DBカタログ、WINGpro)

2. 横断検索

生命科学横断検索

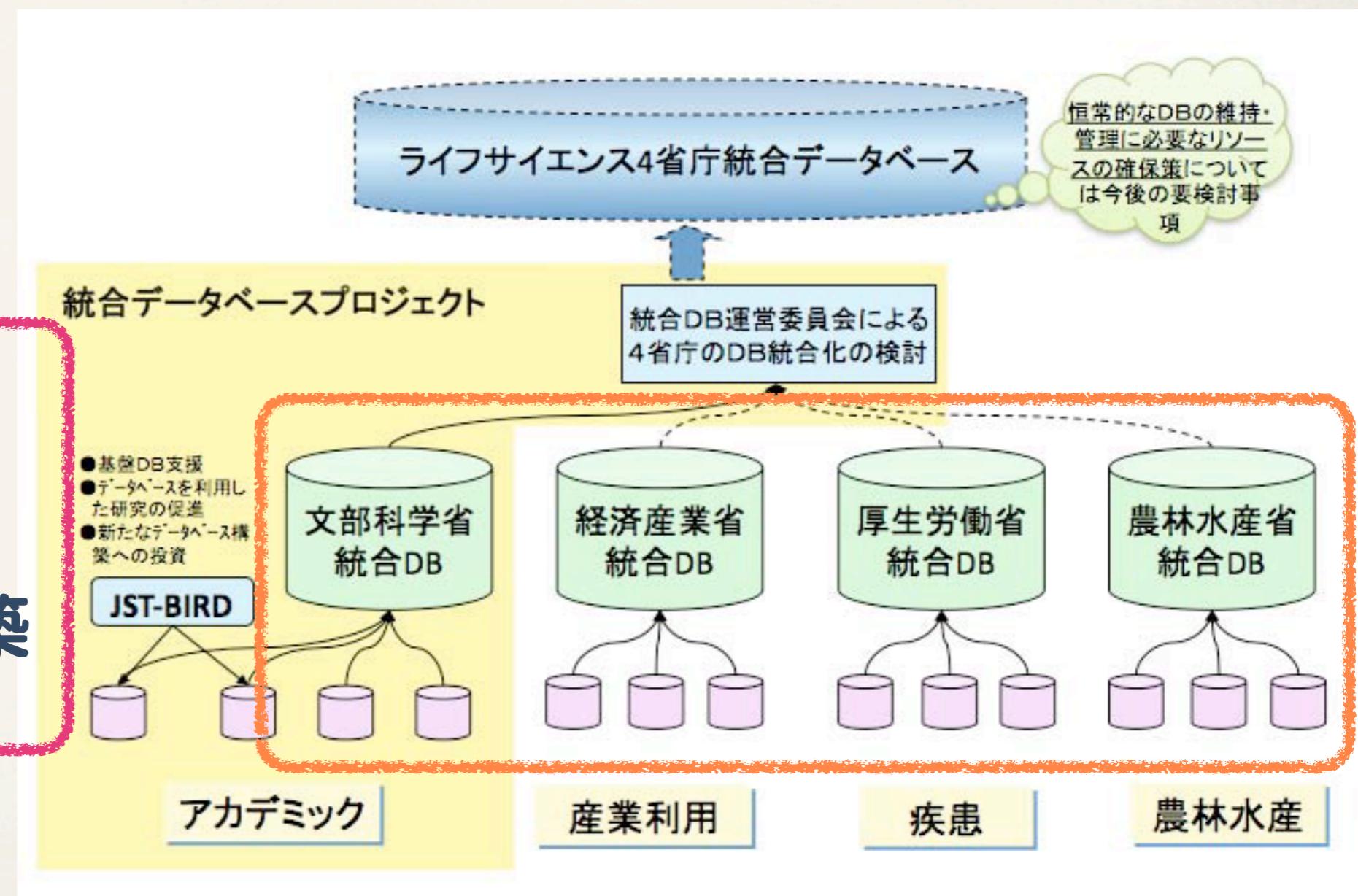
3. アーカイブ構築

生命科学系データベースアーカイブ

→ 「NBDCの三大サービスの紹介 櫛田研究員」にて

省間連携でのデータベース統合

1. カタログ構築
2. 横断検索
3. アーカイブ構築



縦割り問題の解決へ

integbio.jpの公開

文字サイズ - +

 **integbio.jp** 生命科学系データベース統合のための合同ポータルサイト

ホーム 統合へのステップ 技術交流 イベント

4省の生命科学系データベースの統合を目指して

このサイトは、文部科学省、厚生労働省、農林水産省、経済産業省による、生命科学系データベース統合のための合同ポータルサイトです。

内閣府 総合科学技術会議ライフサイエンスプロジェクトチーム 統合データベース（推進）タスクフォースの示す方針を踏まえて、平成23年度より、4省の間で生命科学系データベースの統合が推進されています。その統合のためのステップとして、データベースのカタログ、横断検索、アーカイブ構築などの連携を4省合同で進めています(図1)。

(1) カタログ連携：データベースカタログへの記載、リンクの実現
各者のデータベースに関する情報を記載・蓄積し、各者のデータベースに対して、リンクが行われる状態を実現する。

(2) 横断検索連携：「横断検索サーバー」の設置、一括横断検索の実現
各者における適切なサイトに、横断検索サーバーを設置するとともに検索インデックスを統一し、各者のデータベースに対して、相互に一括横断検索が可能となる状態を実現する。

(3) アーカイブ構築連携：「統一フォーマット」でのダウンロードの実現
各者のデータベースをガイドラインに沿ってアーカイブ化して収集し、各者のデータベースに対して、統一形式でのデータダウンロードが可能となる状態を実現する。

(4) データベース再構築連携：「データベースの再構築」による統合の実現
各者から譲り受けたデータベースを再構築して収容し、各者のデータベースが、統一化したフォーマットで再構築された状態を実現する。

図1. 4省データベース統合の段階的進展

以下は、各省で統合を進めている4つの機関それぞれのポータルサイトです。


NBDC National Bioscience Database Center


NIBIO 農林水産省生物資源データベースセンター


AgriTOGO 農林水産省生物資源データベース


MEDAIS METI database portal for life science



NBDCは、データベース整備・統合の様々な実務ティングを介して、統合のための基盤技術開発やデータベースの統合を進めています。平成22年度までの「データベース統合プロジェクト」と平成23年度までの「バイオデータベース事業」の流れを継承しています。

URL: <http://biosciencedbc.jp/>







学会の合間に省間連携

組織に依存しない場で統合のあるべき姿を模索する

統合化推進と基盤技術の連携 1



基盤技術開発プログラム
データベース統合化の基盤となる
技術開発

DBCLS (ライフサイエンス統合
データベースセンター)

統合化推進プログラム
分野ごとのデータベース統合化

脳疾患画像G
メタボロームG
医薬・疾患G
微生物G
植物G
ゲノム多型G

フェノームG
蛋白質構造G
糖鎖G
ゲノム疫学G

JST- NBDC (バイオサイエンスデータベースセンター)

- 戦略の立案 (基盤技術と統合化推進のコーディネート、統合化ガイドライン)
- カタログ、横断検索、アーカイブの運用
- ファンディング●広報

統合化推進と基盤技術の連携2

担当者編集項目

オントロジーおよびデータフォーマットの活用（実績、予定）

- ・このプログラムで活用している（する予定）既存のオントロジー/データフォーマット形式

- ゲノム関連
GO, Sequence Ontologyなど
- メタゲノム関連
Environment Ontology (EnvO)

- ・このプログラムで開発する（あるいは予定）新規のオントロジー/データフォーマット形式

- メタゲノムのメタデータについてのオントロジー

メタゲノムのメタデータは微生物群集が生息する環境を記述したものであり、本DBにおける3つの柱の一つ、環境についての情報を担っているため非常に重要となり
インは一応存在しますが、あくまでMinimum Informationであるため、各研究者がそれぞれ重要と考えたメタデータを比較的自由に公共データベースに登録している。

- 課題で必要なオントロジー/データフォーマット
- 注目技術
- 国際標準化
- 進行状況

注目技術

- ・データベースの統合化、高度な検索機能を実現するために活用している（する予定）、あるいは注目している技術

- RDF、RDFDBとRDBとの連携、オントロジー、自然言語処理

国際標準化に関する情報

- ・各分野で行われてる国際的なプロジェクトがある場合、そのプロジェクト名、URL

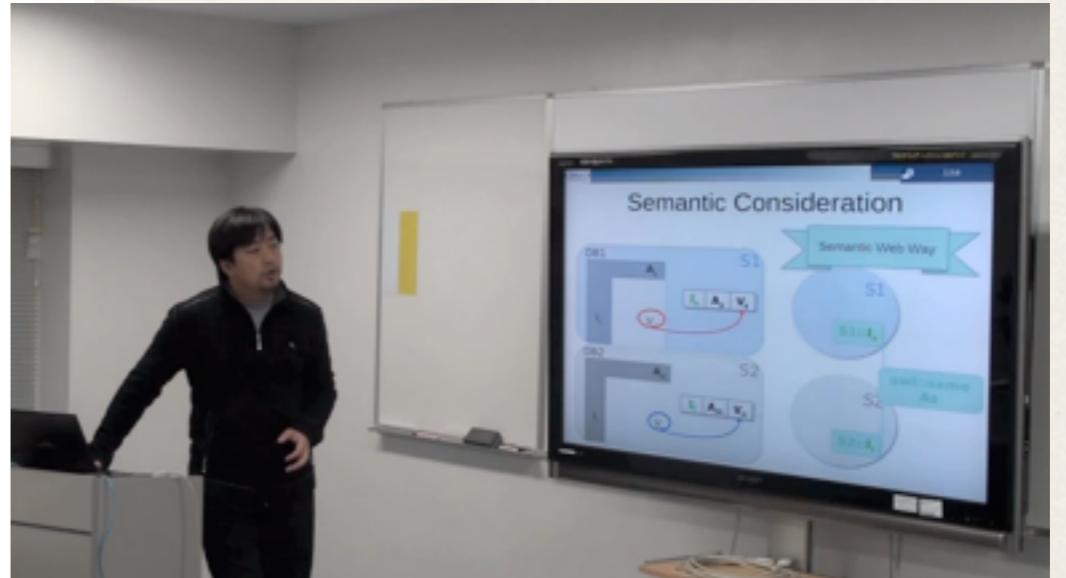
メタゲノム関連

- Genomic Standards Consortium (GSC) http://gensc.org/gc_wiki/index.php/Main_Page
- Minimum Information about a (Meta)Genome Sequence (MIGS & MIMS) http://gensc.org/gc_wiki/index.php/MIGS/MIMS
- Minimum Information about a MARKer gene Sequence (MIMARKS) http://gensc.org/gc_wiki/index.php/MIMARKS
- Human Microbiome Project (HMP) <http://www.hmpdacc.org/>
- Earth Microbiome Project (EMP) <http://www.earthmicrobiome.org/>

- ・各分野で、デファクトスタンダードとなっているデータベースや、オーソリティーが提供するデータベース

未だスタンダードと言えるものは存在しませんが、以下の2つはメタゲノム解析の分野では比較的主流です。

- Genomes OnLine Database (GOLD) <http://genomesonline.org/cgi-bin/GOLD/bin/gold.cgi>
- MG-RAST <http://metagenomics.anl.gov/>



テーマ別の勉強会

Wikiを利用したオープンな技術情報の交換

権利関係、倫理、設備整備

●データ共有について方針

クリエイティブ・コモンズ・ライセンス（ccライセンス）とはなんですか？

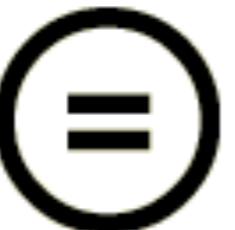
法律や技術に関する専門的な知識がなくても、簡単な4つのアイコンの組み合わせを選択するだけで、誰でも自分の生み出した作品を、自分の好きな条件で、インターネットを通じて世界に発信することができる画期的なライセンスシステムです。



表示



非営利



改変禁止



継承

クリエイティブ・コモンズ FAQより [licensed by CC BY](#)

●ヒト由来データの取り扱いについての方針

脳疾患画像G
ゲノム多型G
ゲノム疫学G

●大規模データの受け入れ、サーバー資源や運用

データ生産者と利用者両方の視点から検討する

展示、講習会、講演で広める、 交流する



NBDC の広報サイト

バイオサイエンス ×DB=∞

検索



○ Web ○ events.biosciencedbc.jp/

Home シンポジウム 講習会 展示会 連載

ようこそ！

バイオサイエンスデータベースセンター
(NBDC) が企画するライフサイエンス分野
データベース関連のイベント情報をお知らせし
ます。

連載のお知らせ

2011年11月より、生命科学系データベースの
現状と問題点、さらには将来のデータベースの
理想像についてご紹介しています。

第3回は 田畠哲之 先生
(かずさDNA研究所)です。



[詳細はこちら](#)



研究成果を聞く (シンポジウム)

2011年11月14日(月)に「バイオインフォマティクス推進センター事業(BIRD) 第8回研究開発成果報告会」を日本科学未来館みらいCANホールで開催します。たくさんの参加をお待ちしています。



使い方を学ぶ (講習会)

2012年3月2日(金)に統合データベース講習会(AJACS本郷11)を開催します。本講習会では、NBDCで提供しているサービスやDBCLSで開発しているサービス、日本の代表的なデータベースであるDDBJ、PDBJ、KEGGの使い方を実習を交え



DBを使ってみる (展示会)

日本農芸化学会2012年度大会の附設展示会(2012年3月23日~25日@京都女子大学)に出展します。ライフサイエンス分野のDBやDB周りの情報技術を紹介します。

細胞工学の連載公開中！

脳疾患画像G
メタボロームG
医薬・疾患G
微生物G
植物G
ゲノム多型G
フェノームG
蛋白質構造G
糖鎖G
ゲノム疫学G

NBDC, 基盤技術開発

講習会の募集を まもなくはじめます。

本日の進行 : motdb.dbcls.jp/? AJACS28

13:00～13:20 ライフサイエンスデータベース統合推進事業の取り組み

13:20～14:20 NBDCの三大サービスの紹介 櫛田 (NBDC)

14:20～14:30 休憩

14:30～15:20 DBCLSの活動の紹介 河野 (DBCLS)

15:20～16:10 BodyParts3D/アナトモグラフィーの使い方

藤枝 (DBCLS)

16:10～16:20 休憩

16:20～17:50 DBCLS Galaxy/TogoDBの使い方 山口 (DBCLS)

開発サービスのリンク

櫛田 (NBDC)

 データベースのカタログ
[Integbioデータベースカタログβ版](#)
(生命科学系 データベース カタログ とWINGproが含まれています)

 コンテンツ
[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](#)
[Webリソースポータルサイト](#)
[ゲノム解析ツール リンク集](#)
[ライフサイエンス 新着論文レビュー](#)
[統合TV](#)

 検索
[生命科学データベース横断検索](#)
[TogoProt](#)
[GenLibi](#)
[Allie](#)
[inMeXes](#)
[SRAss](#)
[鎖錠](#)
[統合遺伝子検索 GGRNA](#)

 広報
[NBDC広報サイト](#)
[パンフレット\(PDF:1.98MB\)](#)

 アーカイブ
[生命科学系データベースアーカイブ](#)

 開発ツール
[TogoDB](#)
[TogoWS](#)
[DBCLS Galaxy](#)
[アナトモグラフィ-/BodyParts3D](#)
[TogoDoc](#)

 データベース
[HOWDY](#)
[MDeR](#)
[RefEx](#)

 連携
[integbio.jp](#)
[BioHackathon 2011、2010、2009、2008](#)

山口 (DBCLS)
藤枝 (DBCLS)

本ポータルサイトで提供しているサービスについて

<http://biosciencedbc.jp/>

終わり