DBCLS
Database Center for Life Science

はじめに：
統合データベースプロジェクトとは？

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）
坂東　明日佳

2010年7月6日

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

■ 我が国における生命科学系DBの問題点

統合データベースプロジェクトの立ち上げまで

■ 生命科学系DBの集約・統合のために

統合データベースプロジェクトの体制

■ そのための取り組み

サービスの紹介

前半：全体をながめ、　後半： 実際に使う。

■ 統合データベースプロジェクトの立ち上げまで

我が国における生命科学系DBの問題点

■ 統合データベースプロジェクトの体制

生命科学系DBの集約・統合のために

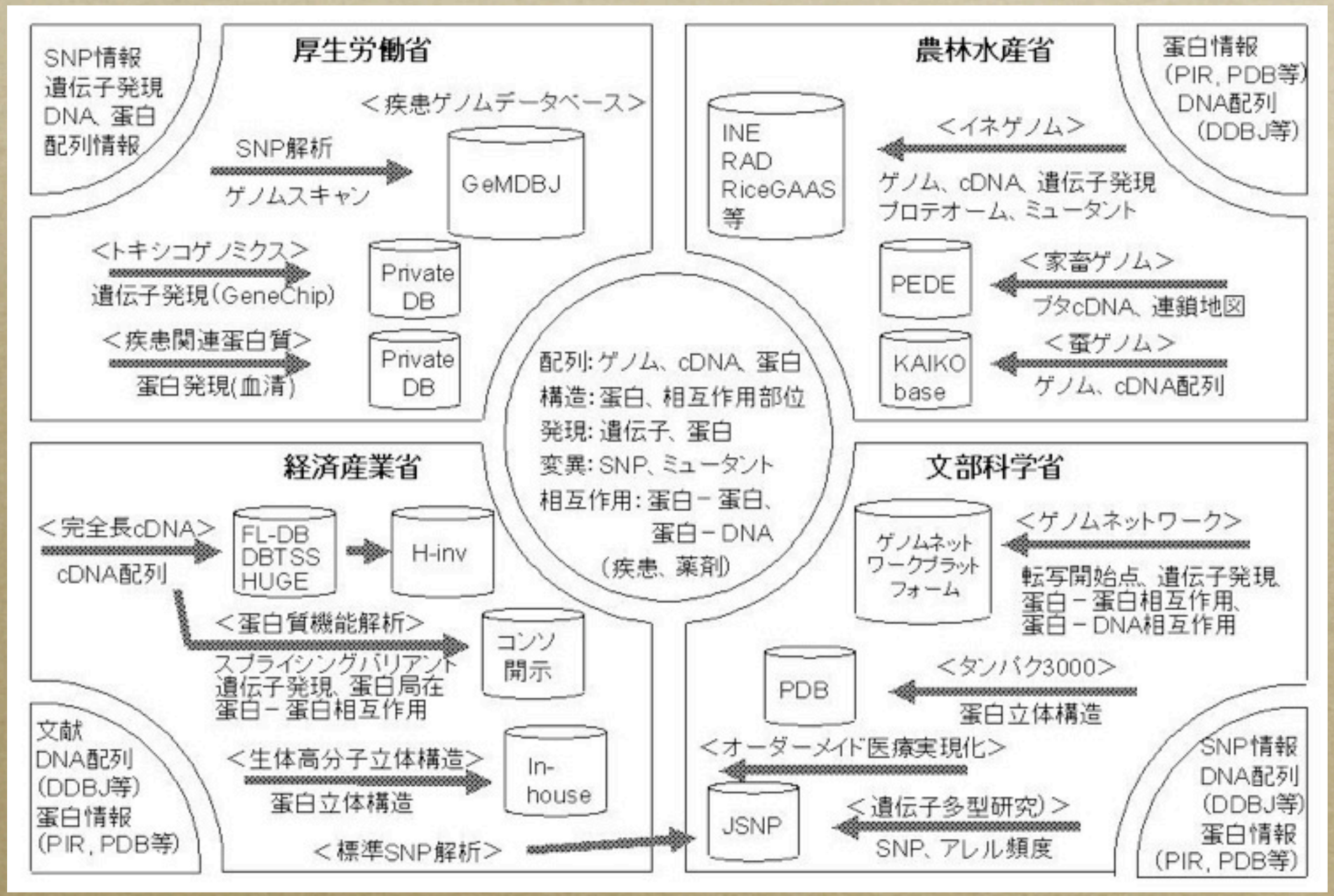
■ サービスの紹介

そのための取り組み

前半：全体をながめ、　後半： 実際に使う。

生命科学系主要国家プロジェクトのデータの管理状況

内閣府科学技術連携施策群の報告書「生命科学データベース統合に関する調査研究」より



多種多様なデータの縦割りによる管理→ 使いにくさ

我が国における生命科学系DBの問題点

1. DBが散在し、所在情報や利用方法が分からない。
2. 大型プロジェクトの成果公開が不十分。
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不足

我が国には、長期的な視野に立った、データベース戦略がない。

統合データベースプロジェクトの発足

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

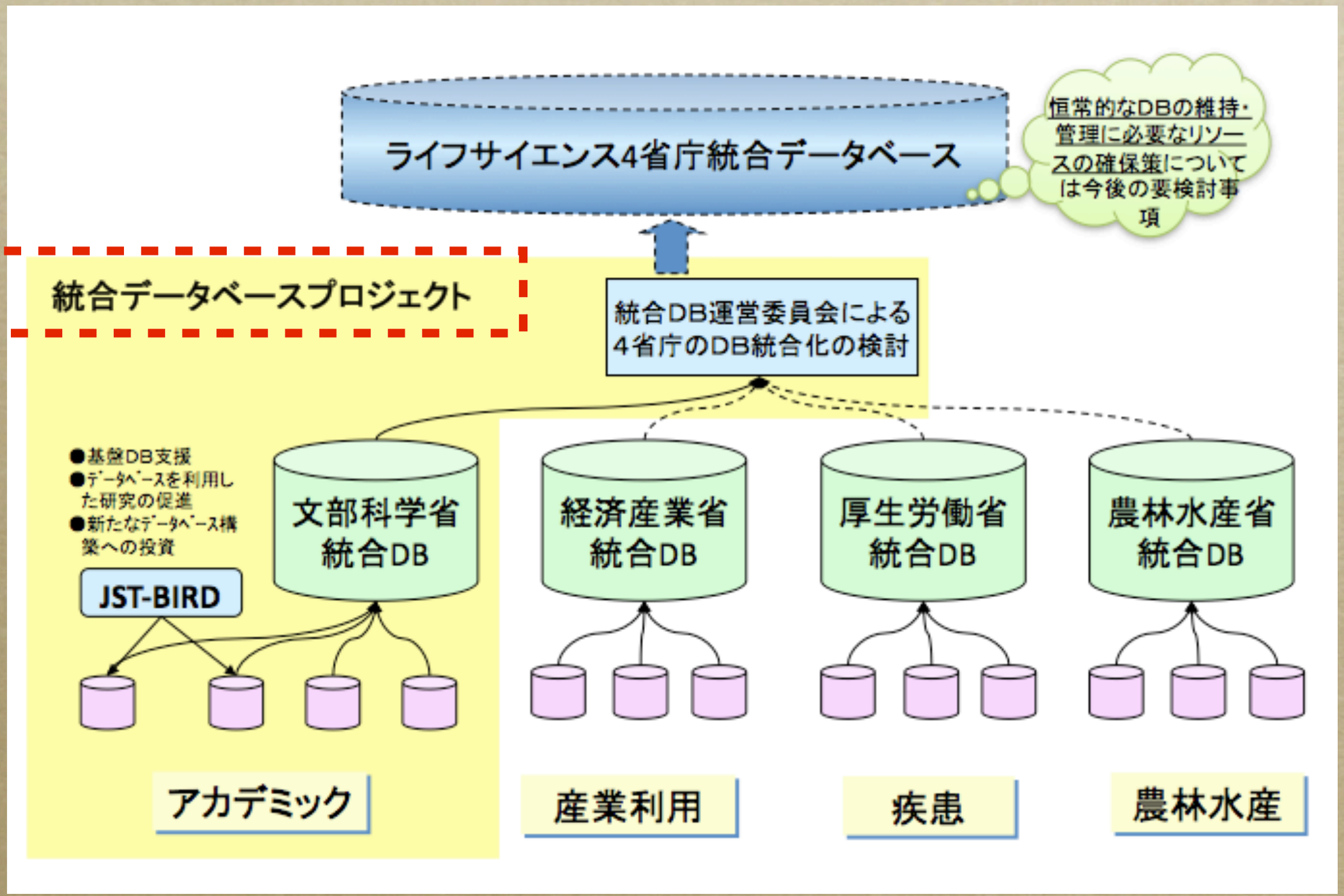
5月17日

「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

4省庁でのDB統合化に向けて



まずは文科省由来の学術的情報を主な対象として、課題を整理

統合DBプロジェクトで、取り組むべき課題

1. 長期的視野に立った、DB整備戦略
2. DBの所在、利用方法
3. 高度な統合DB化の技術
4. プロジェクト終了後のDBの受け入れ
5. 実験データと文献からの知識との対応づけ
6. 高精度注釈（アノテーション）付け
7. DB構築、アノテーションの人材育成
8. 基盤的DBの安定的な支援体制
9. 新しい発想によるDB開発の支援体制
10. バイオインフォマティクスの研究開発の促進方策

そのための体制作りへ

■ 統合データベースプロジェクトの立ち上げまで

我が国における生命科学系DBの問題点

■ 統合データベースプロジェクトの体制

生命科学系DBの集約・統合のために

■ サービスの紹介

そのための取り組み

前半：全体をながめ、　後半： 実際に使う。

中核機関の設置：
ライフサイエンス統合データベースセンター

大学共同利用機関法人
情報・システム研究機構
Research Organization of Information and Systems

大学共同利用機関法人 情報・システム研究機構
国立遺伝学研究所

DDBJ
DNA Data Bank of Japan

国立極地研究所
大学共同利用機関法人 情報・システム研究機構

大学共同利用機関法人 情報・システム研究機構
国立情報学研究所
National Institute of Informatics

大学共同利用機関法人 情報・システム研究機構
統計数理研究所
The Institute of Statistical Mathematics

新領域融合研究センター
Transdisciplinary Research Integration Center

DBCLS
Database Center for Life Science
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

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

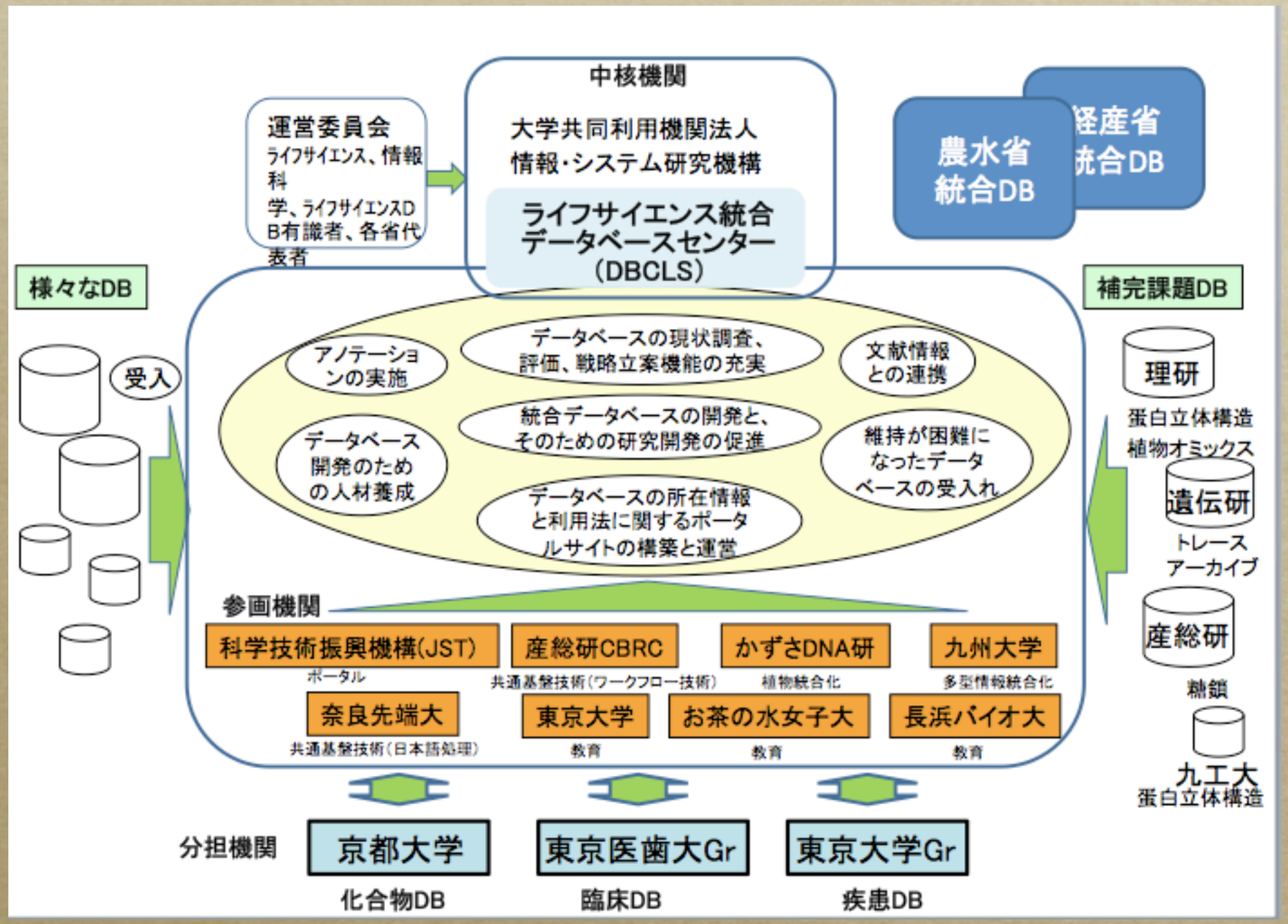
センター長：高木利久

スタッフ：約40名

＊東京大学 弥生キャンパス内にありますが、
東大の所属ではありません。

情報・システム研究機構の中に新センターとして、設立。

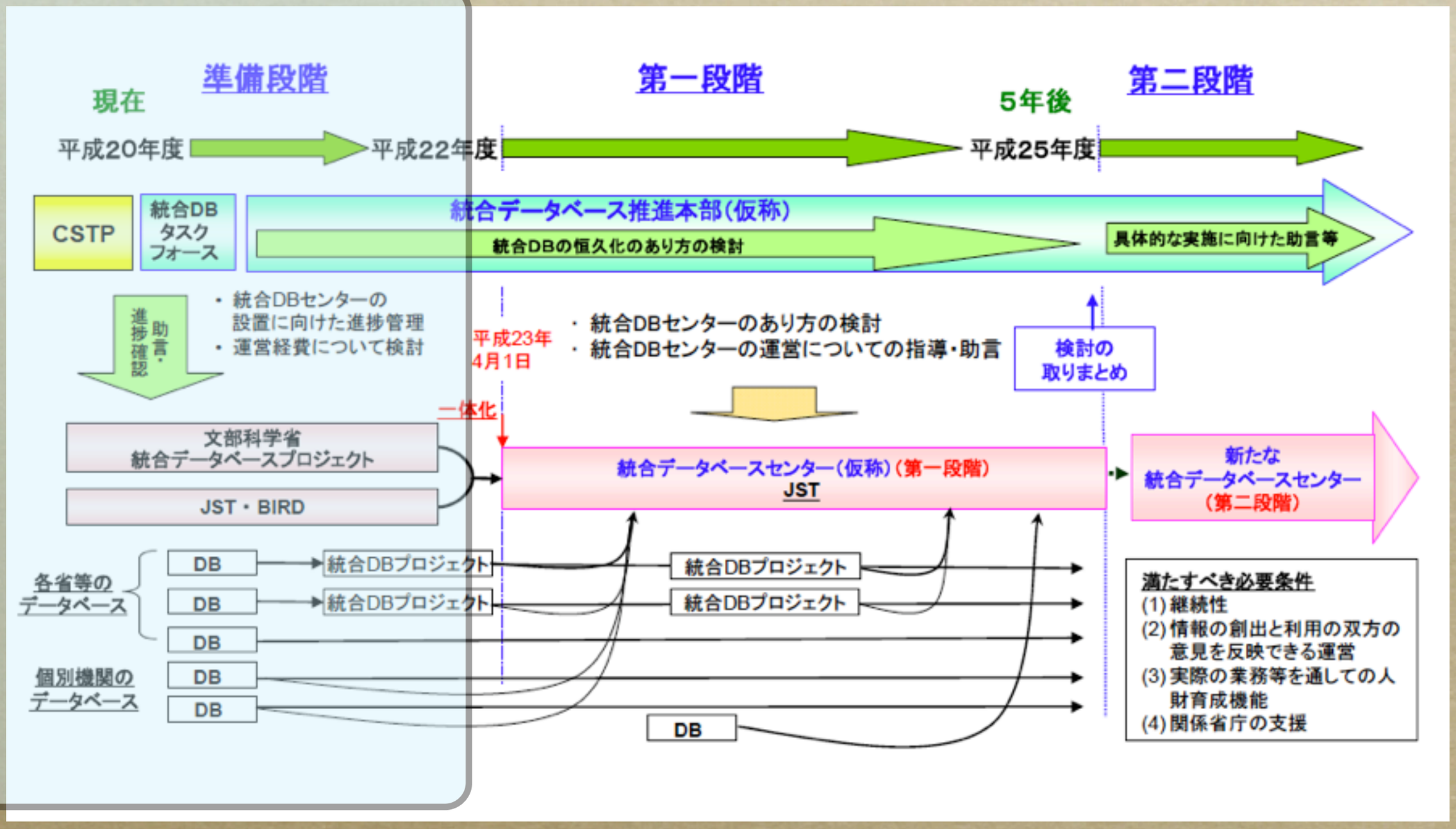
統合データベースプロジェクトの体制図



国内15の大学・研究機関と連携して、課題に取り組む。

今後の体制

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　より



今年度が最終年度、今後の組織のあり方を検討中

■ 統合データベースプロジェクトの立ち上げまで

我が国における生命科学系DBの問題点

■ 統合データベースプロジェクトの体制

生命科学系DBの集約・統合のために

■ サービスの紹介

そのための取り組み

前半：全体をながめ、　後半： 実際に使う。

取り組むべき課題

1. 長期的視野に立った、DB整備戦略
2. DBの所在、利用方法
3. 高度な統合DB化とそのための技術
4. プロジェクト終了後のDBの受け入れ
5. 実験データと文献からの知識との対応づけ
6. 高精度注釈（アノテーション）付け
7. DB構築、アノテーションの人材育成

それぞれの課題に応じて開発しているサービスの例を紹介

統合ホームページからみる

http://lifesciencedb.jp

open ID

各種サービス

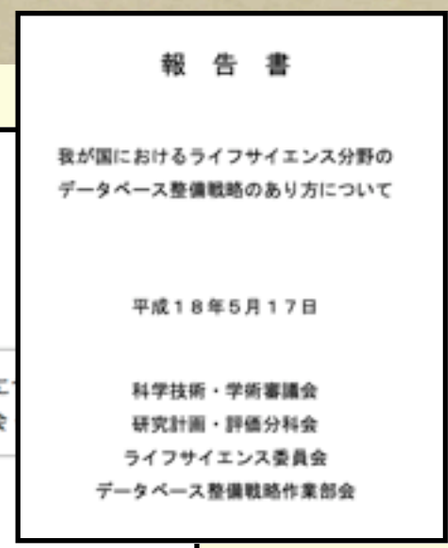
新着情報

ブログ

すべてのサービスはこちらからご覧いただけます。

1. 長期的視野に立った、DB整備戦略

■統合DB事業



運営委員会、シンポジウムでの討論
生命科学DB政策関連の資料の置き場

■LSDB Social Action

データのオープンと共有
についての議論の場

2. DBの所在、利用方法 〜一覧から探す〜

■データベースカタログ

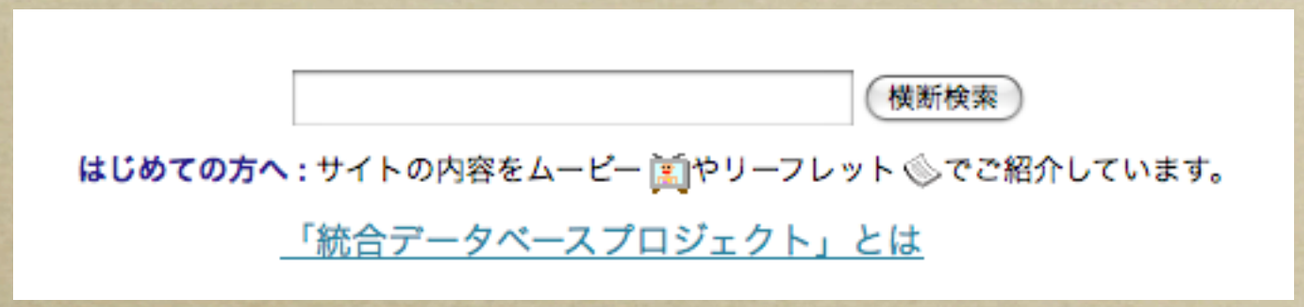
■主要プロジェクト一覧

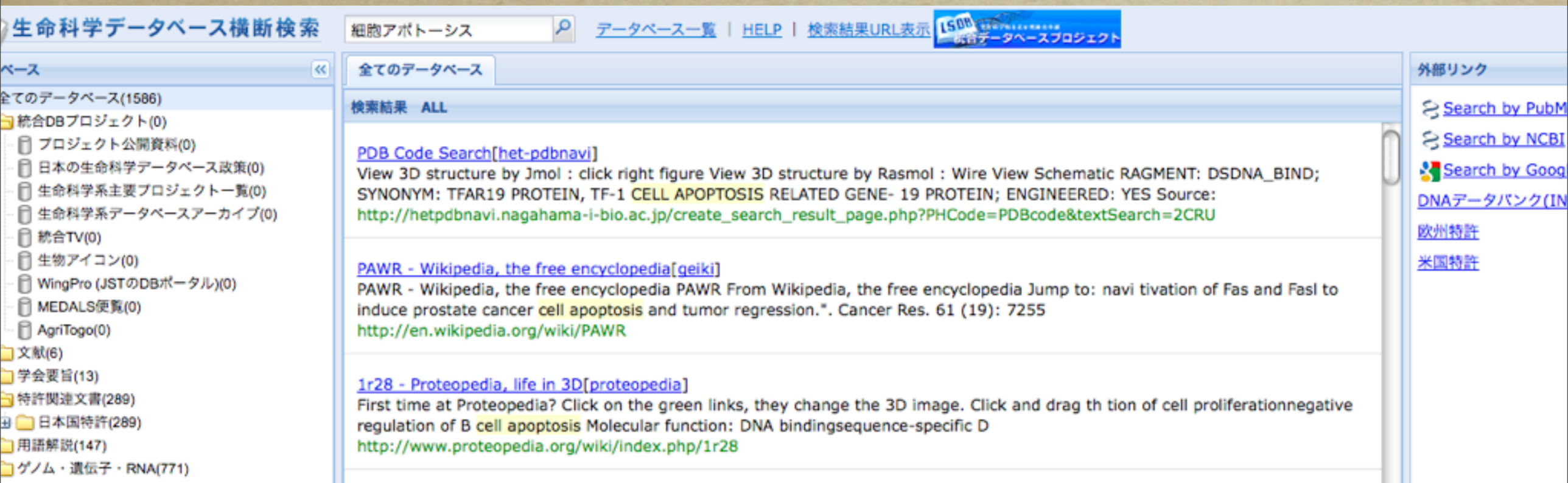
69 Found

プロジェクト名	開始年度	終了年度	概要	データベース公開状況の分類	公開データベース
ゲノムネットワークプロジェクト	2004	2008	遺伝子の発現調節機能に関わる網羅的な解析	コンソ内部公開, 一部共有	ゲノムネットワークプラットフォーム統合データベース(http://genomenetwork.nig.ac.jp/)
農林水産生物ゲノム情報統合DB	2006	2010	農林水産生物ゲノム統合DBの整備	共有	AgriTOGOデータベース(http://togo.dna.affrc.go.jp/)
カイコゲノム解析プログラム	1994	-	蚕ゲノム、cDNA配列、連鎖地図情報の取得と解析	公開	・カイコcDNAデータベース(http://kaikocdna.dna.affrc.go.jp/) ・カイコプロテオームデータベース(http://kaiko2ddb.dna.affrc.go.jp/cgi-bin/search_2DDB.cgi) ・カイコゲノムアノテーションデータベース(http://kaikogaas.dna.affrc.go.jp/)

2. DBの所在、利用方法 〜検索して探す〜

■横断検索

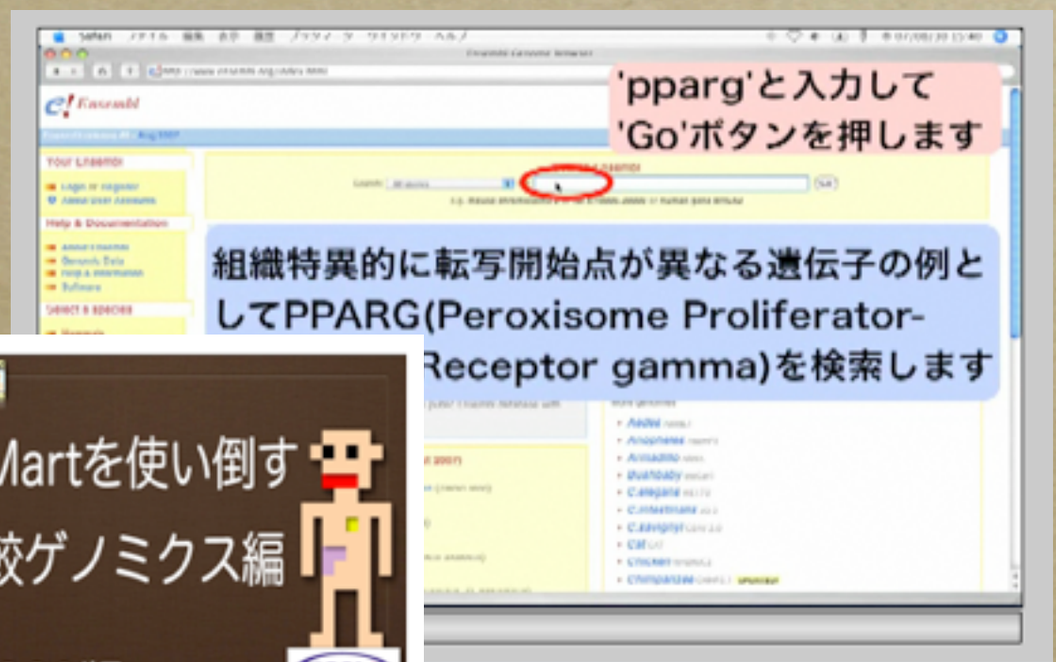




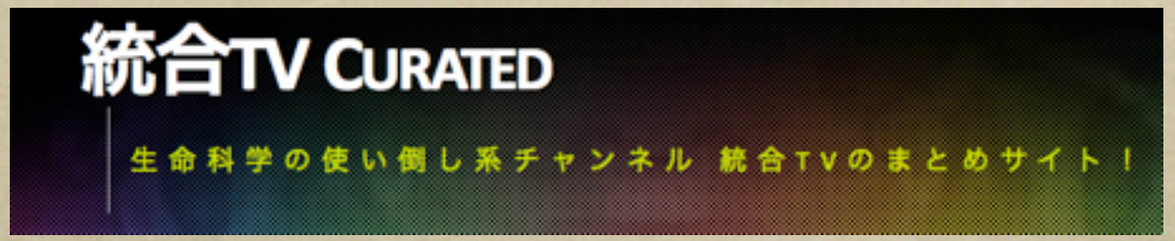
DBCLS内外の、各種生命科学系DBに対する一括キーワード検索

2. DBの所在、利用方法 〜利用法を観る〜

■統合TV



DBの使い方を動画で説明
(本日の講習会の復習にも)



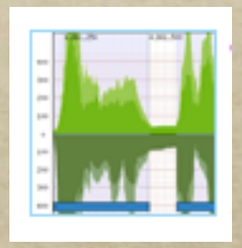
検索:

番組ID	作成日	番組名	番組の説明文
320	2010-07-03	文献執筆支援ツールの紹介	本日の統合TVは、2010年6月23, 24日にライフサイエンス統合データベースセンター（DBCLS）にて開催 回DDBJing 講習会 in 東京より、DBCLS 山本泰智 特任研究員による「文献執筆支援ツールの紹介」をお
319	2010-07-02	統合データベースのサービス紹介〜カタログ・アーカイブ・横断検索を中心に〜	本日の統合TVは、2010年6月23, 24日にライフサイエンス統合データベースセンター（DBCLS）にて開催 回DDBJing 講習会 in 東京より、DBCLS 川本祥子 特任准教授による「統合データベースのサービス紹介 イブ・横断検索を中心に〜」をお送りします。
318	2010-07-01	統合データベースプロジェクトの紹介	本日の統合TVは、2010年6月23, 24日にDBCLSにて開催されたAJACS & 第22回DDBJing 講習会 in 東京より 任研究員による「統合データベースプロジェクトの紹介」をお送りします。
317	2010-06-29	分子可視化ソフト『Chimera』の使い方 2010 応用編	ChimeraはPDBファイルなどのデータを元に生体高分子（主にタンパク質）の立体構造を3D画像で表示 トです。今回の番組では、基本編の内容を踏まえ、より高度な操作について解説しています。
316	2010-06-24	MiGAPの使い方〜導入と基本操作〜	MiGAP(Microbial Genome Annotation Pipeline)はライフサイエンス統合データベースセンター(DBCLS)が提 サービスです。MiGAPの特徴は原核生物ゲノム塩基配列に対しアノテーション（注釈）をつけること サービスは、塩基配列解析にMetaGeneAnnotatorとGlimmerとRNAmmerを、配列相同性解析にNCBI BLA す。
315	2010-06-22	JCGGDBの使い方	JCGGDB は質量分析による糖鎖構造解析のスペクトルや、レクチン-糖鎖相互作用のプロファイリング ど、糖鎖に関係するさまざまなデータベースを集めたウェブページです。 今回は横断検索にかけた結果を通して、データベースの種類と特徴について説明します。
314	2010-06-18	はじめてのRefEx(Reference Expression dataset)	RefEx(Reference Expression dataset) は、ライフサイエンス統合データベースセンター(DBCLS)によるヒト 別学的な発現パターンデータの統合サイトです。4種類の発現データ (EST, GeneChip, iAFLP, CAGE)に対し ふしーく)でデータを整理し、遺伝子発現データ解析のリファレンス（参照）データセットを維持すべ ます。対象生物種はヒトとマウスのみですが、複数の手法による客観的な遺伝子発現データの比較が が特徴です。 今回は、「検索」ボタンから「Descriptionから遺伝子ファミリー(Interpro ID)を検索」する機能を用い ミリーを付与されたInterProのアノテーションの中から検索し（IPR005817 Wnt superfamily）、Wnt フ 子群をその発現パターンと共に表示し可視化する方法を説明しています。なお、マップされる染色体 替えたり、ウェブブラウザの「文字を縮小」する機能を利用して発現パターンを俯瞰することができ 次に、囊胞性線維症（「のうほうせいせんいしょう」と読みます）の責任領域としてしられている7q を発現の高い順に表示する方法を紹介します。さらに、その中のNM_000245 met proto-oncogeneに関 詳細（絶対発現量、相対発現量、3D人体マップ上での発現量のヒートマップ表示）をブラウズする方 す。
313	2010-06-17	今日からはじめるDDBJ Read Annotation	DDBJ Read Annotation Pipelineは国立遺伝学研究所のDDBJが開発・提供している新型シーケンサの出力

番組のまとめサイトもopen！

3. 高度な統合DB化とそのための技術
～テーマごとに統合する～

■タンパク質統合：TogoProt
蛋白質の統合検索システム TogoProt

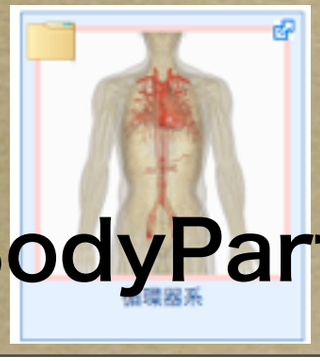


■発現統合：GEO目次、togoexp
発現データの整理、解釈

LSDB Lab. !

■植物統合：KazusaMart, etc.
植物のアノテーション整備

Data management system KazusaMart beta



■ヒト統合：アナトモグラフィー/BodyParts3D

3. 高度な統合DB化とそのための技術
～技術開発～

Clone Id	Length	Definition	Accession
YMU02A01NGRL0003_A01	485	Adiantum capillus-veneris mRNA for unknown protein	DK943491
YMU02A01NGRL0003_A02	342	Adiantum capillus-veneris mRNA, clone: YMU02A01NGRL0003_A02, 5' end sequence.	DK943492
YMU02A01NGRL0003_A03	174	Adiantum capillus-veneris mRNA, clone: YMU02A01NGRL0003_A03, 5' end sequence.	DK943493
YMU02A01NGRL0003_A04	351	Adiantum capillus-veneris mRNA, clone: YMU02A01NGRL0003_A04, 5' end sequence.	DK943494
YMU02A01NGRL0003_A05	518	Adiantum capillus-veneris mRNA, clone: YMU02A01NGRL0003_A05, 5' end sequence.	DK943495
YMU02A01NGRL0003_A07	252	Adiantum capillus-veneris mRNA, clone: YMU02A01NGRL0003_A07, 5' end sequence.	DK943496
YMU02A01NGRL0003_A08	546	Adiantum capillus-veneris mRNA similar to 60S ribosomal protein L13a	DK943497
YMU02A01NGRL0003_A09	496	Adiantum capillus-veneris mRNA, clone: YMU02A01NGRL0003_A09, 5' end sequence.	DK943498

■統合DB

Web操作で容易にデータベースを構築・公開

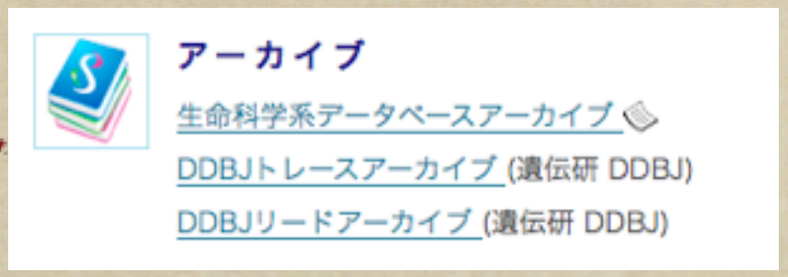
■TogoWS

国内外の主要DBを統一的に扱える

インターフェイス

4. プロジェクト終了後のDBの受け入れ

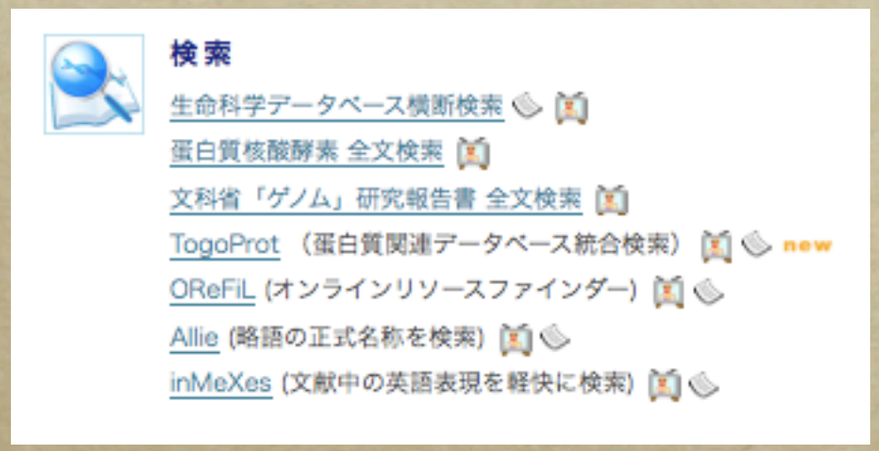
■生命科学系データベースアーカイブ



データベース		データベース運用場所	代表者	データベースカテゴリ	生物種	要約（キーワードを青字表示）	利用許諾
	RA target genes in the Ciona embryo ダウンロード | 簡易検索 オリジナルサイト	高知大学 理学部 応用理学科	藤原 滋樹	発現		カタユウレイボヤ胚におけるレチノイン酸標的遺伝子候補のマイク	
	DIAM - 安全性に関する文献 ダウンロード | 簡易検索 オリジナルサイト	バイオインダストリー協会	–	文献			
	DIAM - バイオテクノロジーの基礎知識 ダウンロード オリジナルサイト	バイオインダストリー協会	–	文献			
	DIAM - バイオセイフティ関連の法規制 ダウンロード | 簡易検索 オリジナルサイト	バイオインダストリー協会	–	文献			
	The NAISTrap database ダウンロード | 簡易検索 オリジナルサイト	奈良先端科学技術大学院大学 バイオサイエンス研究科 動物遺伝子機能学講座	石田 靖雅	変異体			
	DGSP Database - Drosophila Gene Search Project Database ダウンロード | 簡易検索 オリジナルサイト	首都大学東京 大学院理工学研究科 生命科学専攻 統合遺伝学研究室	相垣 敏郎	変異体	ウバエ	入系統のデータベース	詳細
	EGTC - Database for the Exchangeable Gene Trap Clones ダウンロード | 簡易検索 オリジナルサイト	熊本大学 生命資源研究・支援センター	荒木 正健	変異体	マウス	マウスの 可変型遺伝子トラップクローンのデータベース	CC 表示-継承 詳細

プロジェクト終了後のデータの維持管理

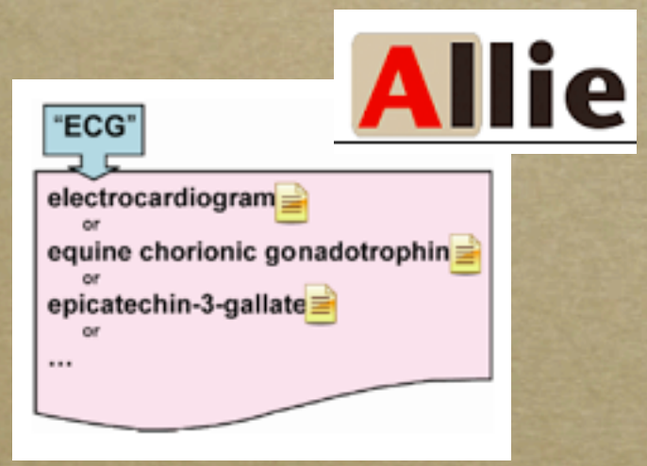
5. 実験データと文献からの知識との対応づけ



■蛋白質核酸酵素（PNE）

■ゲノム特定報告書

日本語リソースの共有化

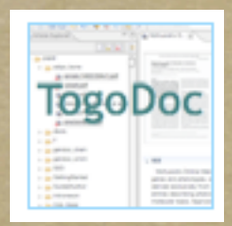


■Allie

略語の名称検索などのテキストマイニングを利用したツール

LSDB Lab. !

■TogoDoc

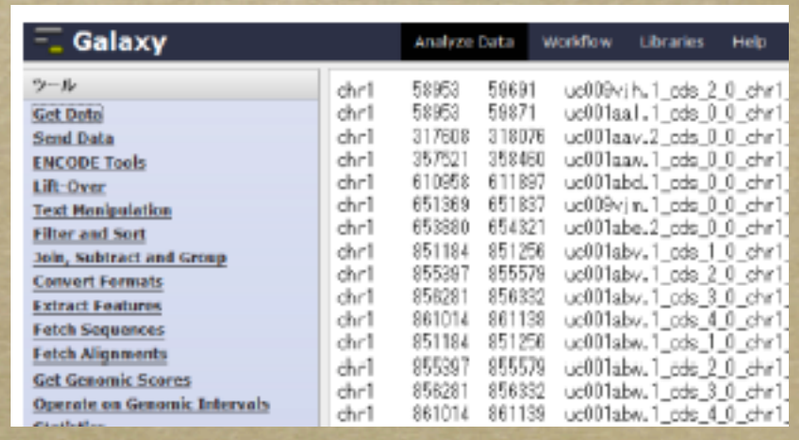


文献の管理支援

6. 高精度注釈（アノテーション）付け

■MiGAP

微生物へのアノテーションパイプライン



■DBCLS Galaxy

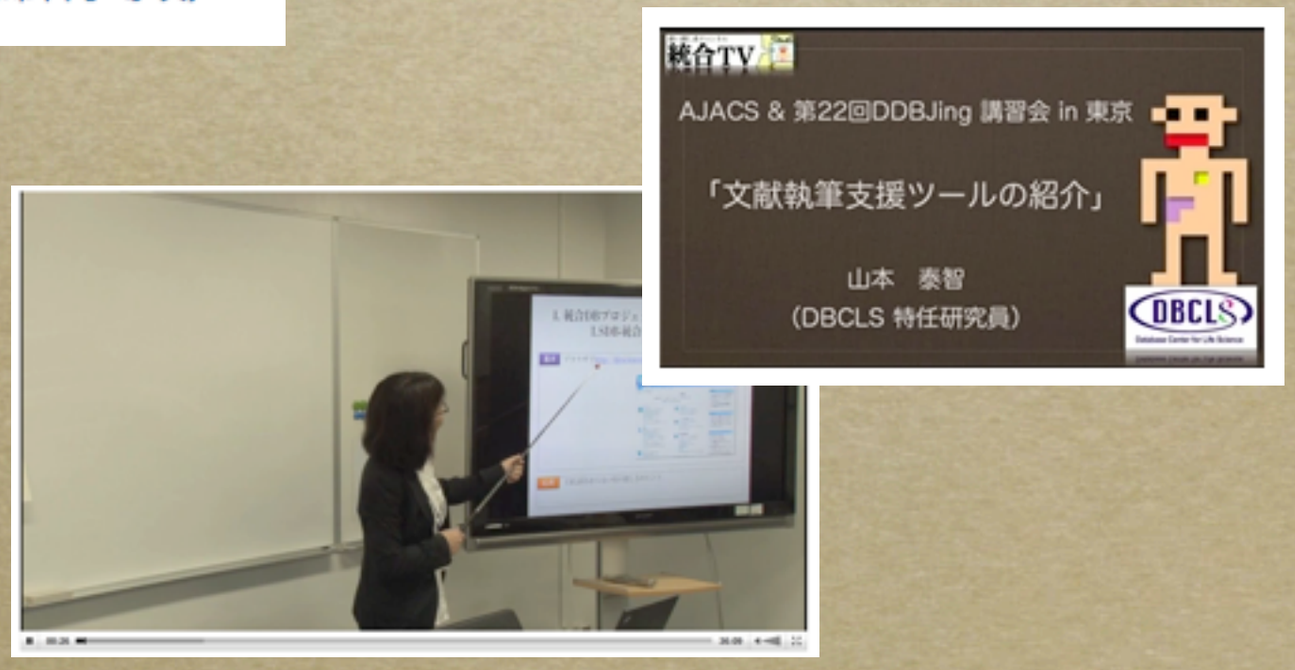
ゲノム解析のツール組み合わせインターフェイス

7. DB構築、アノテーションの人材育成

■MotDB

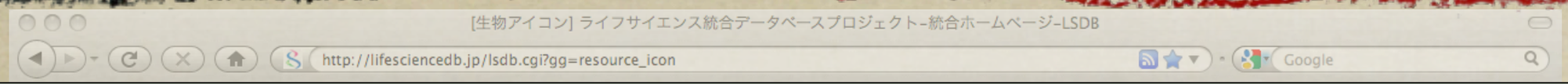
- アノテーター養成 （長浜バイオ大学）
- DB高度利用者養成 （お茶の水女子大学）
- DB構築者養成（東京大学 大学院新領域創成科学研究科 情報生命科学専攻）

■統合TV
では講習の動画配信



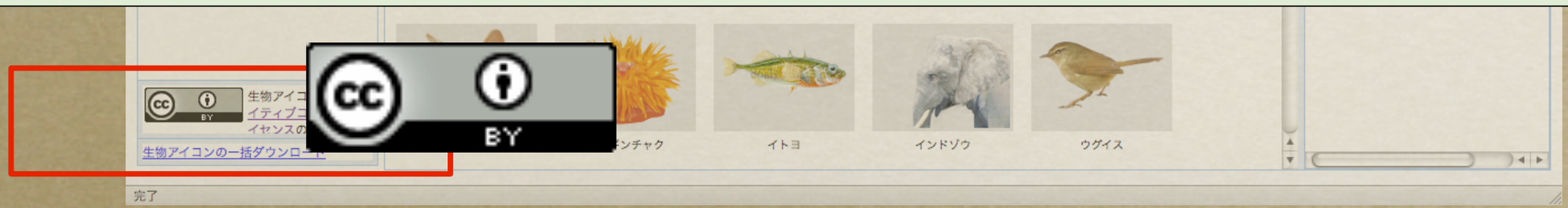
AJACS講習会情報、教材の配信

提供するサービスの利用許諾を明示するために



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。

クリエイティブ・コモンズ FAQより　licensed by CC BY



クリエイティブコモンズ・ライセンスを確認し、
安心してサービスを利用してください。

「実際に使う」へ

「はじめに：統合データベースプロジェクトとは？」(坂東 明日佳)

「生命科学横断検索の利用法」　(川本祥子)

■横断検索

「タンパク質データベースの活用法」(畠中秀樹)

■TogoProt, データベースアーカイブ

11:45 - 13:00　昼休憩

「遺伝子発現データの活用法／次世代シーケンサの活用法」(河野信)

■発現統合、MiGAP

「文献情報を効率的に活用する」(仲里猛留)

■日本語の文献情報、Allie、TogoDoc

本日の流れは、MotDBをご覧ください

http://motdb.dbcls.jp/?MotDB

対象

学部3、4年生や大学院生、教員を始めとしたライフサイエンス分野のデータベースを効率よく利用することに興味のある方や統合データベースが実際にどのように作成、維持管理されているかに興味のある方

日時

2010年7月6日（火）午前9時30分〜午後4時

定員

約90名（応募者多数の場合は大学生、大学院生を優先させていただきます）

会場

金沢大学　角間キャンパス　自然研本館２階　計算機実習室１および２

費用

無料

申込

以下のフォームからお申し込みください。

- 統合データベース講習会：AJACS加賀　お申し込みフォーム