

# いかにして「使える」データベースを維持し続けるか？ ライフサイエンス統合データベースプロジェクトから 学ぶこと

東京大学大学院 理学系研究科 生物科学専攻

科学技術振興機構(JST)  
バイオサイエンスデータベースセンター(NBDC)

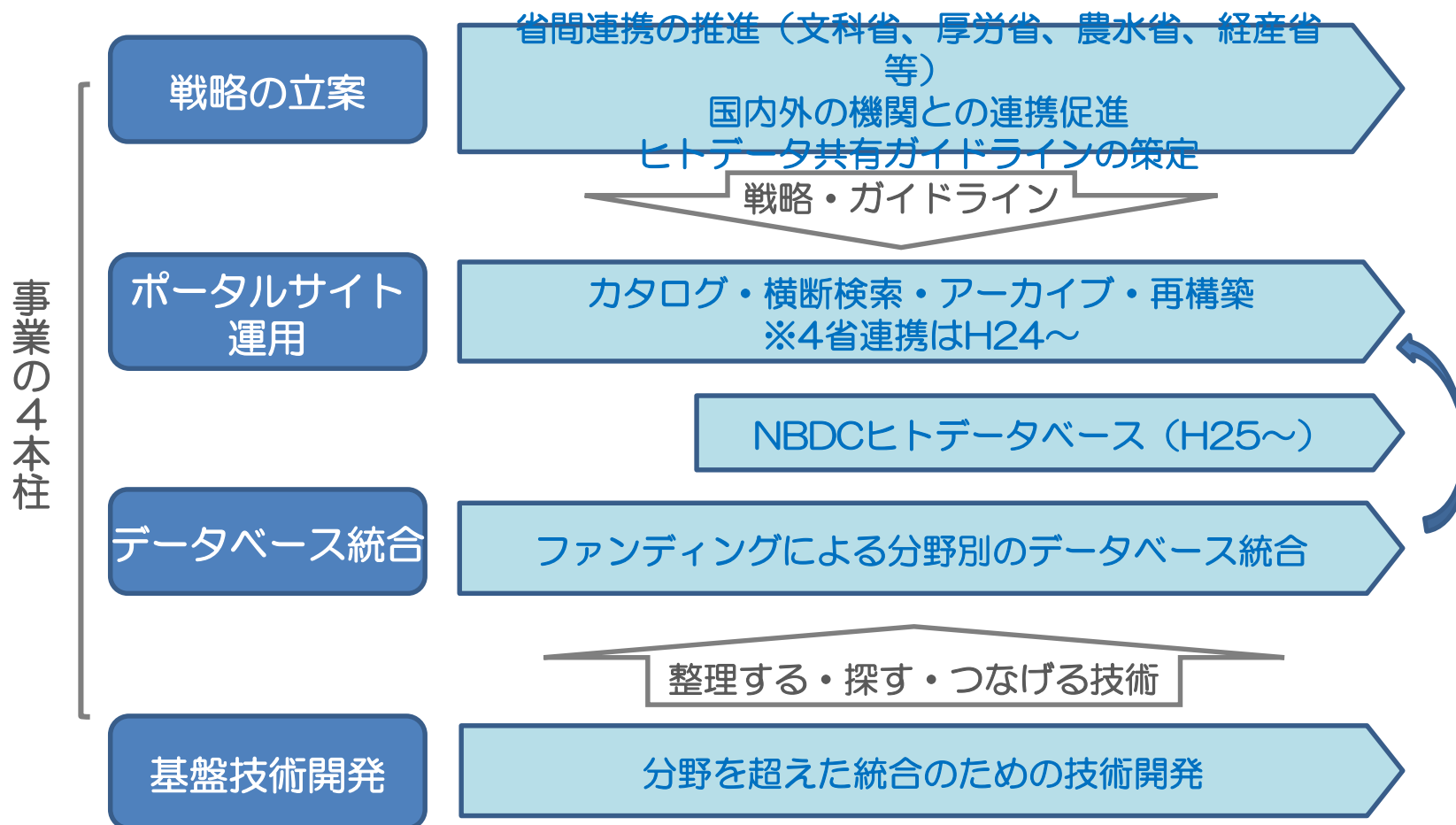
情報・システム研究機構(ROIS)  
国立遺伝学研究所(NIG)  
DDBJセンター(DDBJ)

高木利久

- 1987 NIG DDBJ(DNA Data Bank of Japan)
- 2001 JSTバイオインフォマティクス推進事業(BIRD)
- 2006 文科省統合データベースプロジェクト  
(前後して、他の省庁でも類似プロジェクト発足)
- 2007 ROISライフサイエンス統合データベースセンター(DBCLS)
- 2011 JSTバイオサイエンスデータベースセンター(NBDC)

データを公共財化・統合化して、価値を最大化する

# NBDCの主要な取組み：4つ柱



## NBDCポータルサービス

# 生命科学系データベースカタログ

## 1,643のDBをカタログ化

生命科学系データベースアーカイブ

## 134のDBをアーカイブ

生命科学データベース横断検索

## 621のDBを横断検索可能

## NBDCヒトデータベース

公開(制限公開含む)65件(Studies)

## RDFポータル

## 20件のRDFデータセットを公開

## 4省連携ポータル

4巻の生命科学データベースの統合を目指して

このサイトは、生活の質、社会の豊かさを測る指標として、経済産業省により、全国消費実況調査の一環として調査されたものである。

[illegible]

100

19-21

7-1400

Figure 1. The effect of the concentration of the polymer on the gelation time.

1. 1998年，中国人口出生率为16.5‰，死亡率为6.5‰，自然增长率为10.0‰。

このように、 $\mathcal{H}_1$ と $\mathcal{H}_2$ は、 $\mathcal{H}$ の直交補空間である。

100

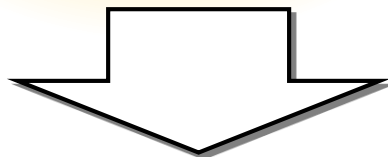
年1

41

\_\_\_\_\_

# セマンティックWEB技術の 研究開発による 分野を超えた DB間統合・活用への 挑戦

- RDF化等による高度な検索の実現
- データベース関連ツール開発
- データベース統合化の国際的標準化など



## 分野別統合データベース



※各サービスのDB数等は平成29年11月現在の数値。

- ✓ 従来型の研究では困難な知見の発見
- ✓ 効果的なイノベーションの創出

# 統合化推進プログラム(TPP)

研究代表者	所属・役職	研究開発課題(採択年度)
石濱 泰	京都大学 大学院薬学研究科・教授	プロテオーム統合データベースの構築(H27-H29)
大浪 修一	理化学研究所 生命システム研究センター・チームリーダー	生命動態システム科学のデータベースの統合化(H24-H26)/ 生命動態情報と細胞・発生画像情報の統合データベース(H27-H29)
沖 真弥	九州大学 大学院医学研究院・助教	エビゲノミクス統合データベースの開発と機能拡充(H29-H33)
金久 實	京都大学 化学研究所・特任教授	ゲノム情報に基づく疾患・医薬品・環境物質データの統合(H23-H25)/ ゲノムとフェノタイプ・疾患・医薬品の統合データベース(H26-H28)/ ゲノム・疾患・医薬品のネットワークデータベース(H29-H33)
黒川 顕	国立遺伝学研究所 生命情報研究センター・教授	ゲノム・メタゲノム情報を基盤とした微生物DBの統合(H23-H25)/ ゲノム・メタゲノム情報統合による微生物DBの超高度化推進(H26-H28)/ データサイエンスを加速させる微生物統合データベースの高度実用化開発(H29-H33)
菅野 純夫	東京大学 大学院新領域創成科学研究科・教授	疾患ヒトゲノム変異の生物学的機能注釈を目指した多階層オミクスデータの統合(H26-H28)/ 疾患ヒトゲノム変異の生物学的機能注釈を目指した多階層オーミクスデータの統合(H29-H33)
田畑 哲之	かずさDNA研究所・所長	ゲノム情報に基づく植物データベースの統合(H23-H25)/ 植物ゲノム情報活用のための統合研究基盤の構築(H26-H28)/ 個体ゲノム時代に向けた植物ゲノム情報解析基盤の構築(H29-H33)
中村 春木 栗栖 源嗣	大阪大学 蛋白質研究所・所長／教授 大阪大学 蛋白質研究所・教授	蛋白質構造データバンクの国際的な構築と統合化(H23-H25)/ 蛋白質構造データバンクの高度化と統合的運用(H26-H28) 蛋白質構造データバンクのデータ検証高度化と統合化(H29-H33)
成松 久 木下 聖子	産業技術総合研究所 糖鎖医工学研究センター・センター長 創価大学 理工学部・教授	糖鎖統合データベースと研究支援ツールの開発(H23-H25)/ 糖鎖統合データベースおよび国際糖鎖構造リポジトリの開発(H26-H28)/ 糖鎖科学ポータル構築(H29-H33)
岩坪 威	東京大学 大学院医学系研究科・教授	ヒト脳疾患画像データベース統合化研究(H23-H25)
金谷 重彦 有田 正規	奈良先端科学技術大学院大学 情報科学研究科・教授 理化学研究所 環境資源科学研究センター・チームリーダー	メタボローム・データベースの開発(H23-H25)/ 生物種メタボロームモデル・データベースの構築(H26-H28)
徳永 勝士	東京大学 大学院医学系研究科・教授	ヒトゲノムバリエーションデータベースの開発(H23-H25)/ 個別化医療に向けたヒトゲノムバリエーションデータベース(H26-H28)
豊田 哲郎 榎屋 啓志	理化学研究所 情報基盤センター・統合データベース特別ユニット リーダー 理化学研究所 バイオリソースセンター・ユニットリーダー	生命と環境のフェノーム統合データベース(H23-H25)/ 生命と環境のフェノーム統合データベース(H26-H28)
松田 文彦	京都大学 大学院医学研究科 附属ゲノム医学センター センター長・ 教授	大規模ゲノム疫学研究の統合情報基盤の構築(H23-H25)

## BioHackathon

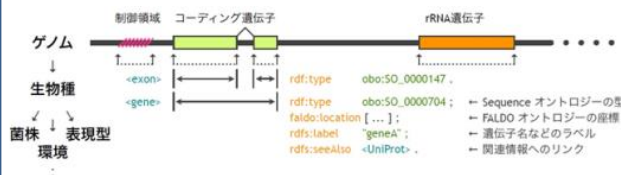
- ✓ 国際的な標準化の取り組み
- ✓ RDFモデルとオントロジーの整備

## RDFサミット

- ✓ ゲノムRDF規格化

## TogoGenome

- RDFによるゲノム情報の統合DB



## RDF化ガイドライン/RDFポータル

- ✓ RDF化ガイドラインを策定
- ✓ それに準拠したRDF化データベースを集約
  - RefEx FANTOM5 RDF (発現)
  - NBDC NikkajiRDF (化合物)
  - wwPDB/RDF (タンパク質立体構造)
  - MBGD RDF (微生物ゲノム)
  - GlycoEpitope (糖鎖)
  - Linked ICGC (がん変異データ) など



## GA4GH - Beacon Project

ヒトゲノムの変異情報検索

例: 12番染色体の34567番目の塩基が  
Aのデータはあるか → yes or no



Global Alliance  
for Genomics & Health



公共オミックスデータ検索・解析技術、教育教材  
テキストマイニング技術、質問応答システム



# 4省及びAMEDとの連携

## 4省連携

## 4省連携によるDB統合

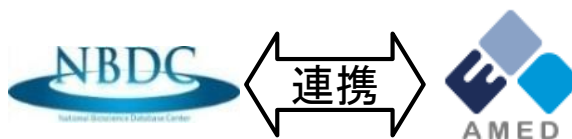


所管省庁	実施機関
厚生労働省	国立研究開発法人 医薬基盤・健康・栄養研究所
農林水産省	国立研究開発法人 農業・食品産業技術総合研究機構
経済産業省	国立研究開発法人 産業技術総合研究所
文部科学省	国立研究開発法人 科学技術振興機構

## 各機関での実施事項

- ・データベースカタログに搭載するデータベースの探索
- ・横断検索を実施するためのインデックスの整備
- ・データベースアーカイブ搭載データの整備 等

## AMEDとの連携



ゲノム医療の実現に資するデータシェアリングに係る基盤整備及び利活用について、AMEDと基本連携協定を締結(H28.9)し、ゲノム医療関連事業コンソーシアム内でのヒトゲノムデータ共有基盤を運用中。

## 産総研 統合データベースポータルサイト

# 4省連携の4ステップとFAIR principles

データ共有を推進する国際コミュニティFORCE11が提唱

## カタログ

データベースの所在情報を提供  
国内の公開データベースは、ほぼ網羅。

Findable!

## 横断検索

複数データベースを横断的に検索  
Google検索での埋没リスクを回避。  
国内の実施可能分は、ほぼ網羅。

Accessible!

アーカイブ 「統一フォーマット」でのダウンロードの実現  
各省データベースをガイドラインに沿って  
ライセンスを明確に整理しアーカイブ化。

Interoperable!

データベース再構築 意味づけされた用語を整備し  
より高度な検索を実現

Re-usable!

FAIR principlesの考えを先取りする形で早い時期からDB統合を実施

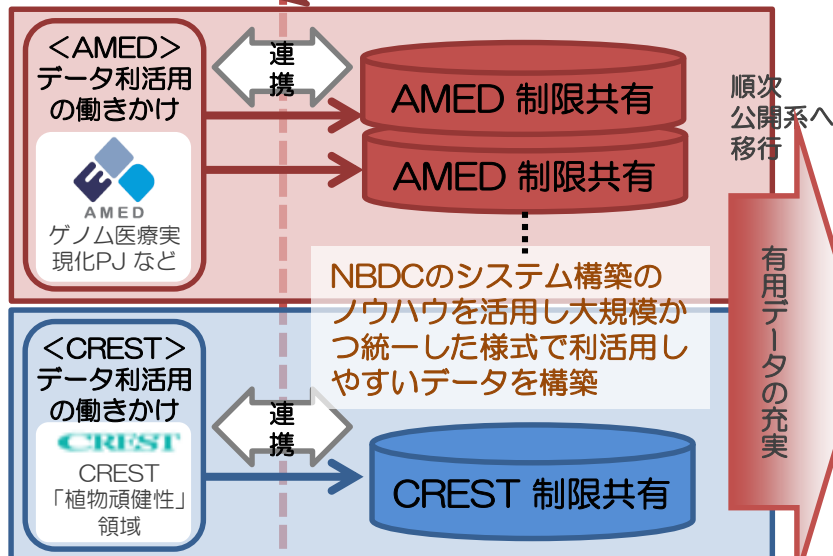


# 新たな取り組み

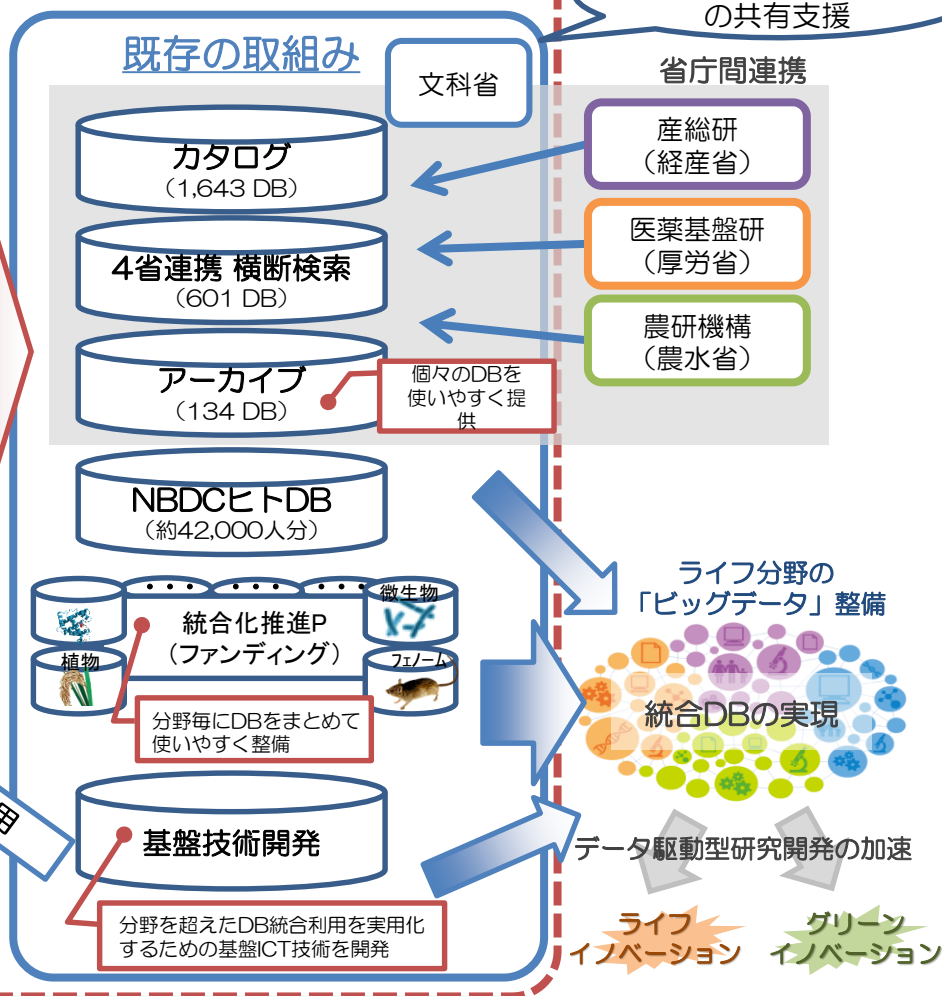
公開前データの  
プロジェクト内共有  
も支援

将来のデータ公開・データ共有が期待できる、研究プロジェクトにおけるプロジェクト内データベース構築をNBDCが支援し、将来のデータ公開やデータ共有を促進する。

## 新規の取組み



## 既存の取組み



グループ共有DBの取り組みを、他の大型プロジェクト研究に拡大していく

※資金・人的リソースは当該PJと相談



National Bioscience Database Center

－数あるデータベースを、まとめて、使い易く－

## バイオサイエンスデータベースセンター

English   サイトマップ



国立研究開発法人  
科学技術振興機構  
Japan Science and Technology Agency

文字サイズ変更   大   中   小

Search for...   Search

[ホーム](#)
[NBDCについて](#)
[研究開発](#)
[公募情報](#)
[採用情報](#)
[イベント](#)
[人材支援](#)
[アクセス](#)
[リンク](#)

NBDCは、日本の生命科学研究を推進するために、データベースをつなげて使い易くします。  
 そのためにNBDCや協力機関は、以下のようなサービスやウェブサイトを作成・提供しています。

【メンテナンス】2017年5月12日(金) 16:30～5月15日(月)12:30(最長)の間、NBDCの一部サービスが停止します。【詳細】

 生命科学全体のデータベース統合

[Integbioデータベースカタログ](#)

データベース横断検索   国内外DBを一括検索   

[生命科学系データベースアーカイブ](#)

[NBDC RDFポータル](#)

 統合のための連携

[integbio.jp: 4 省合同ポータルサイト](#)

[NBDCグループ共有データベース](#)

[BioHackathon](#)

 日本語や動画でわかりやすく

[新着論文レビュー / 領域融合レビュー](#)

[統合TV](#)

 論文をもっと読みやすく、書きやすく

[Allie / inMeXes / TogoDoc](#)

 大量の配列データを扱いやすく

[DBCLS SRA](#)

[RefEx / 統合遺伝子検索 GGRNA](#)

 分野ごとのデータベース統合

 ヒトと医・薬

[NBDCヒトデータベース](#)

[ヒトゲノムバリエーションデータベース](#)

[KERO: 疾患マルチオミクスデータベース](#)

[KEGG MEDICUS: 疾患・医薬品統合リソース](#)

 生命を支える分子

[DDBJ: 日本DNAデータバンク](#)

サービスを活用して得られた研究成果発表に関する情報提供をお待ちしております。

 NBDCパンフレット  
 (PDF: 3.17MB /  
 2016/06/30更新)

お問い合わせ・ご意見・ご要望

 サービスや事業に関するご意見等をお寄せください。

新着情報      

2017/04/24

「NBDCヒトデータベース」東京大学大学院 医学系研究科 アレルギー・リウマチ内科 からの制限公開データ (Type I) および非制限公開データを公開しました (hum0099)

2017/04/13

平成29年5月12日(金)16:30～5月15日(月)12:30までの間、『RDFポータルサイト』と『HOWDY-R』サービスが停止します。

<http://biosciencedbc.jp/>

# ConBio2017 NBDC関連イベント紹介

## ① 特別企画

12月6日（水）～8日（金） 10:00-17:00  
神戸国際展示場 2号館1階 BioDBコーナー

使ってみようバイオデータベース  
— つながるデータ、広がる世界 —

## ② フォーラム

12月9日（土） 11:45-13:15  
神戸ポートピアホテル本館B1布引（第7会場）

生命科学のデータベース活用法

ConBio2017 特別企画

使ってみよう  
バイオデータベース  
— つながるデータ、広がる世界 —

生命科学のDBをポスター・配  
布資料やPCデモで紹介

NBDC事業の取り組み：デー  
タベース統合化の紹介

日時  
2017年12月6日（水）～12月8日（金）  
10:00～17:00

会場  
神戸国際展示場 2号館 1階  
バイオデータベース（BioDB）コーナー

ブース名一覧

- BioDB 1 使ってみようデータベースのカテゴリ、検索結果、アーカイブ
- BioDB 2 知識発見につながるデータベース検索へDBCLSからの提案
- BioDB 3 経産省関連ライフサイエンスサイト：MEDALS
- BioDB 4 創薬・健康・食糧研究を支えるNBDCのデータベース
- BioDB 5 ChIP-Atlas：公的 ChIP-seq データをフル活用できる
- BioDB 6 統合ゲノム情報 DB「ヒト統合ゲノムデータベース」
- BioDB 7 KEGG NETWORK：代謝経路の可視化
- BioDB 8 SSBD：生命科学研究のためのデータベース
- BioDB 9 NBDC：生命科学のデータベースセンター

NBDC関係14機関による  
集合ブース展示

ConBio2017 フォーラム

生命科学の  
データベース活用法

「統合化推進プログラム」のデー  
タベースの活用方法紹介

日時 2017年12月9日（土） 11:45～13:15

会場 第7会場（神戸ポートピアホテル本館B1布引）

NBDCでは、わが国のライフサイエンス研究を推進するために、データベース（DB）をつなげて使いやすくする事業「ライフサイエンスデータベース統合推進事業」を実施し、その一環として、ライフサイエンス DB に関するサービス提供と研究開発を行っています。特に研究開発においては、1) DB 統合化の実現に向けて基盤となる技術開発、2) 分野ごとに関連する情報を網羅的に収集して整理する「統合化推進プログラム」を支援しています。

本企画では、2) 「統合化推進プログラム」にスポットを当て「ゲノム、エピゲノム、タンパク質の構造、プロテオーム、糖質、微生物、生命情報など生命科学における様々な種類のデータについて「どんな DB があるのかわかりたい」、「入手したデータを十二分に活用したい」といったニーズを満たすための DB やツールを紹介し、皆様のご参加をお待ちしています。

紹介する DB と講演者

- KERO (Kashiwa Encyclopedia of Research Objects)
- ChIP-Atlas (ChIP-seq Data Analysis Platform)
- Protein Atlas (Protein Data Bank)
- JPO (Japan Protein Open Access Project)
- GlycoAtlas (Glycan Data Analysis Platform)
- Microarray Atlas (Microarray Data Analysis Platform)
- SSBD (Systematic Sequencing and Bioinformatics Database)
- KEGG (Kyoto Encyclopedia of Genes and Genomes)
- KEGG NETWORK (KEGG Pathway Database)
- KEGG GENES (KEGG Gene Database)
- KEGG PATHWAY (KEGG Pathway Database)
- KEGG DRUG (KEGG Drug Database)
- KEGG LIGAND (KEGG Ligand Database)
- KEGG BRITE (KEGG Brite Database)
- KEGG COMPOUND (KEGG Compound Database)
- KEGG ENZYME (KEGG Enzyme Database)
- KEGG GLOSSARY (KEGG Glossary Database)
- KEGG NAMESPACE (KEGG Namespace Database)
- KEGG REFERENCE (KEGG Reference Database)
- KEGG SOURCE (KEGG Source Database)
- KEGG TAXONOMY (KEGG Taxonomy Database)
- KEGG VECTORS (KEGG Vectors Database)
- KEGG IMAGES (KEGG Images Database)
- KEGG AUDIO (KEGG Audio Database)
- KEGG VIDEO (KEGG Video Database)
- KEGG TEXT (KEGG Text Database)
- KEGG TABLE (KEGG Table Database)
- KEGG FORM (KEGG Form Database)
- KEGG SCRIPT (KEGG Script Database)
- KEGG STYLE (KEGG Style Database)
- KEGG FONT (KEGG Font Database)
- KEGG COLOR (KEGG Color Database)
- KEGG SOUND (KEGG Sound Database)
- KEGG SMELL (KEGG Smell Database)
- KEGG TASTE (KEGG Taste Database)
- KEGG TOUCH (KEGG Touch Database)
- KEGG FEEL (KEGG Feel Database)
- KEGG THINK (KEGG Think Database)
- KEGG EMOTION (KEGG Emotion Database)
- KEGG BEHAVIOR (KEGG Behavior Database)
- KEGG PERSONALITY (KEGG Personality Database)
- KEGG INTELLIGENCE (KEGG Intelligence Database)
- KEGG CREATIVITY (KEGG Creativity Database)
- KEGG IMAGINATION (KEGG Imagination Database)
- KEGG DREAM (KEGG Dream Database)
- KEGG MIND (KEGG Mind Database)
- KEGG SOUL (KEGG Soul Database)
- KEGG SPIRIT (KEGG Spirit Database)
- KEGG BODY (KEGG Body Database)
- KEGG MIND (KEGG Mind Database)
- KEGG SOUL (KEGG Soul Database)
- KEGG SPIRIT (KEGG Spirit Database)
- KEGG BODY (KEGG Body Database)

7名の研究者による  
統合データベース紹介

データ解析のよろず相談を実施!!

是非、お越し下さい!!

いかにして「使える」データベースを維持し続けるか？

# 「使える」データベースとは？

- 誰にとって？
  - 実験研究者（非常に多様）
  - バイオインフォマティシャン
  - データサイエンティスト
- 最大公約数的なDBとは？
- 研究のどの場面で使う？
- いつを目指して開発する？
- どれくらいの期間使うことを想定するのか？

# DBCLS データコア構想

想定ユーザ

想定ユーザに応じた開発の4階層

エンドユーザ  
実験科学者  
医療従事者

アプリケーション開発

ゲノムデータ・オミックスデータ解析支援および  
そのための基盤技術開発

アプリケーション開発者  
TPP  
AMED  
ToMMo

ミドルウェア開発

リンクトデータへの効率的、安定的な  
アクセスのための基盤技術開発

データ科学者  
国際連携  
人工知能  
機械学習

データベース整備

リンクトデータ・コンテンツの生成および  
そのための基盤技術開発

データベース構築者  
TPP  
AMED  
ToMMo

データソース整備

既存データのリンクトデータ化支援および  
そのための基盤技術開発

国際/国内  
協力

広報



# いかにして「維持し続ける」か？

- 資金

- どう調達？誰が負担？
- 国？DB利用者？データ生産者？

- DB開発人員

- 評価の仕組み
- インセンティブ、ポジション

- どの分野のどのDBを？

- DB数は世界で20,000程度、日本だけでも1,000超える
- 優先順位？
- DBの価値をどう測るか？（アクセス数？引用？）
  - データ、一次DB、統合DB

いかにして「使える」データベースを維持し続けるか？

- DB利用者からの厳しい注文と暖かい応援

- まずは使ってみる
- 注文をつける
- 使えるデータベースを育てる
- 使ったデータ、DBを引用する
- 維持すること、維持する人を褒めてあげる
- 色々なメディアで、様々な機会にDBの重要性を訴える

- 長い目で見て利用者の利益になる