

# 統合化のビジョン、ステップ、課題(案) (運営委員ご意見のまとめとDBCCLSの取り組み)

---

平成21年4月13日

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  
高木 利久

# 統合化の理念とアプローチ

---

- 公的資金を用いて得られたデータ、DBは広く共有し、その活用を図るべし
- DB統合により、生命科学の新たな展開とバイオ産業の国際的競争力向上に貢献すべし
- 統合DBは重要な基盤であり、省庁に囚われない存在として国が継続的支援すべき



- 公的資金で得られたデータをコミュニティで早期に共有するためのルールを作成する
- 公的資金で得られたデータを保全、管理し、育成するための中立的機関を設置する

# 統合化の目標

---

アカデミアの知識を集約し、日本で生産するデータの価値を最大化することにより、日本のユーザ、更には世界のユーザに貢献する



制約のない共有    操作感の統一    知識発見支援



ブレークスルー、イノベーション、新たな研究スタイル

# 統合データベースの持つべき機能

---

## <サービスの側面>

- 個々のデータベース、解析ツールへの案内機能
- 複数のデータベースの統合化による付加価値の向上
- 標準インタフェース、フォーマット、オントロジーなどの共通基盤の提供
- 維持困難なデータベースの保存によるサービスの継続
- 一次データの登録、編集、標準化、品質管理、閲覧サービス
- 企業アクセスも視野に入れたセキュアな環境の提供

## <基盤的側面>

- 使い勝手およびデータの付加価値向上のための研究開発機能
- データベース統合利用促進のためのファンディング機能
- 世界的情報共有化のための国際的な連携機能
- データベースの構築や統合化のための人材育成機能と教材提供

# 今後の進め方

---

- 文科省統合DBプロジェクトを核として、本格的恒常的な統合データベース事業を構築する
- JST事業の早期具体化と中長期的なプランニングの場を設定する
- 4省連携など全体としての統合を実現するための組織と委員会を設置する
- 統合DB構築に関わる中核拠点の整備と併せて、基盤DBの維持、個別DBの構築への支援にも留意する
- データ所有権、公開原則、DBへの登録義務等への考え方を省庁を越えて統一する
- 個人情報保護、著作権、営利目的利用に関する考え方を整理し、関連法令、規則等の整備を推進する

# 統合化の全体像とステップ (DBCCLS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

~H20

~H22

イノベーション、新たな知識発見、データベース生物学

# DB統合推進の基本的考え方(DBCLSの考え方)

---

- データの共有化、価値の最大化が目的であり、必ずしも一カ所集中型を目指すものではない
- データを統合化し、価値を高めるには、知識の集約、デジタル化が必要であり、これに積極的に取り組む
- 技術的には可能でも、制度的その他の理由により、共有化できていないものは、その障害の解消に積極的に取り組む
- DB利用者だけでなく、DB構築提供者にも喜ばれる仕組みを作る
- サービスの安定的、永続的な維持更新

以下參考資料

# 統合化の全体像とステップ (DBCCLS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

~H20

~H22

イノベーション、新たな知識発見、データベース生物学

## 統合化ステップ①

---

- データベースカタログ
- 国家プロジェクトカタログ
- 学協会カタログ
- 学会要旨、科研費報告書
- 各種報告書、会議録
- 特許
- 有用サイト検索
- データベース、ツールなどのダウンロード
  - データの所有権や共有ルール作成
  - 個人情報保護(個人ゲノムデータ等)
- メタデータや各種辞書整備

赤: ほぼ完了  
青: 進行中  
黒: 未着手

## 統合化ステップ②

---

- 統合TV
- 利用者マニュアル(オンライン)整備
- 人材養成マニュアル(オンライン)整備
- ツール検索システム
- 解説記事、学会要旨、などの全文検索
- 解析ワークフロー
- 利用者の意見集約システム
- DB評価システム

赤: ほぼ完了  
青: 進行中  
黒: 未着手

## 統合化ステップ③

---

- 複数データベース一括横断検索
- 統合検索 (NCBI ENTREZ風)
- 統合検索環境 (統一的な操作感)
- DB受け入れ、DB構築支援、DB公開支援
- リンクやインデックスの自動生成
- 遺伝子辞書や日英対訳辞書などの整備
- オントロジーのための基盤整備
- テキストマイニング
- データベース権処理、著作権処理

赤: ほぼ完了  
青: 進行中  
黒: 未着手

## 統合ステップ④

---

- 個別だけど共通性の高い統合データベース
  - 蛋白質統合データベース
  - 糖鎖統合データベース
  - SNP統合データベース
  - 疾患統合データベース
  - 植物統合データベース
  - 熱力学パラメータ統合データベース
  - プロジェクト単位統合配列データベース、など
- 解析ワークフロー、統合利用環境
- アノテーション支援ツール
- DBやツールの部品化標準化、  
ウェブAPI

赤: ほぼ完了  
青: 進行中  
黒: 未着手

## 統合化の各ステップに必要な基盤技術、仕組み

---

- サービスの安定的な維持更新の仕組み、体制
- データベースの評価や品質保証の仕組み
- データ加工、流通の追跡可能性確保の仕組み
- データベースセキュアアクセス法の開発
- 個人プロフィールによる検索の高度化
- データベースやリンクの自動更新
- アノテーションのための意見集約システム開発
- データベースやリソースの自動検出
- アノテーション、キュレーション体制構築

赤: ほぼ完了  
青: 進行中  
黒: 未着手

## 統合ステップ⑤(今後の検討課題)

---

- 分野特有の統合データベース
  - がん、脳、免疫、進化、発生、再生、育種、発酵、、、
- 分野特有の解析ツール、ワークフロー
  - がん、脳、免疫、進化、発生、再生、育種、発酵、、、
- 分野特有のオントロジー
  - がん、脳、免疫、進化、発生、再生、育種、発酵、、、
- 上記実現のための関連機関・研究者との連携
- 一般市民向け、学生教育用
- 分野特有のDBやキュレーションのための基盤技術

## その他の検討、考慮すべき項目

---

- 個人ゲノム時代への対応
- 次世代シーケンサーへの対応
- オープンアクセス化への対応
- 医療、健康情報への対応
- 医薬品、化合物への対応
- 分子イメージング、画像、動画への対応
- 環境、生物多様性情報への対応
- DBの視点からの国家プロジェクト管理
- 世界的な連携、役割分担への対応
- 上記のためのさまざまなガイドライン、法整備

DDBJその他  
関連機関との  
緊密な連携