

文部科学省統合データベース プロジェクトについて

東京大学大学院

新領域創成科学研究科 教授

情報・システム研究機構 特任教授

高木利久

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
 - 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
- ◆ 大型プロジェクトの成果公開が不十分
 - 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
 - 統合化されていない、相互運用性低い
 - 統合化もおもに分子レベルで行われていて医療や産業への応用困難
 - 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
- ◆ データベース構築、維持、アノテーションを行える人材不足
 - 我が国のデータに対しきちんとしたアノテーションがされていない
- ◆ わが国には長期的な視点に立ったデータベース戦略がない

文科省におけるこれまでの取り組み

◆ 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
- 平成18年5月17日報告書
- 国際動向を踏まえた問題点整理と推進方策(10項目)・推進体制の提案
 - 戦略委員会、関連機関連携の枠組み、中核的機能を担う機関、などの設置

◆ ライフサイエンス分野の統合データベース整備事業の公募

- 平成18年度(緊急3項目:戦略立案、ポータルサイト整備、統合化技術)
 - 大学共同利用機関法人 情報・システム研究機構を中核としたチームが採択
- 平成19年度(~22年度の4年間) 公募
 - 中核機関1つと分担機関3つ、計4つのチームが採択

◆ 大学共同利用機関法人 情報・システム研究機構での取り組み

- ライフサイエンス統合データベースセンター設置(平成19年4月)

ライフサイエンス統合データベースとは？

- ◆ バラバラに管理運営されている多種多様なDBを一つに
 - 我が国のライフサイエンスDBは約250
- ◆ 仮に統合化が実現されるとすると
 - 利用者にとって: 検索、仮説生成、解析が容易になり生産性向上
 - 管理者にとって: 管理運営が効率的に(トータルで)
- ◆ しかし「一つの統合DB」は現実的に実現困難で有用性も低い、その理由は
 - 分子データの統合化だけでは不十分
 - データの解釈や意味が研究の進展によって変化する
 - ライフサイエンスの最先端はテキスト(論文)で表現される
 - データをどう眺めたいかは研究者によって異なる

データベース統合化に向けた基本的考え方

- ◆ 統合DB構築は完結しないプロセスであると認識
 - 研究の進展に応じて利用者の求めるものが変化することに柔軟に対応
- ◆ 一つの統合DBではなく、研究開発の生産性向上を目指す
 - DB構築者ではなく利用者の思考や意思決定を支援するDBの構築
 - 利用者の興味、知識に応じて必要な情報、判断材料をもれなく提示
 - 複数DBをつなぎ異種データ・知識の関係が俯瞰(仮説生成)できるように
 - いろいろなツールを簡便に組み合わせて解析(知識発見)できるように
 - できれば、上記のことが日本語で行えるように
- ◆ DB化(構造化)されないものもうまく扱えるように
 - 論文(テキスト、ポンチ絵、画像)、特許、教科書、報告書、解説
- ◆ 研究開発は行うが、サービス事業であることを認識

文科省が目指す統合データベースの概要

◆ 想定ユーザ

- ライフサイエンスの研究者および医療、創薬などバイオ産業従事者
- ライフサイエンスプロジェクトの企画立案や評価にかかわる人々
- ライフサイエンスデータベースの構築者

◆ 統合化の対象データ

- ヒト・動植物・微生物の分子データ、文献データ、臨床などの表現型データ

◆ 提供する機能

- DBやツールの所在や利用法を網羅したポータルサイト
- 分子データと文献知識(高次生命機能)を統合したデータベース
- 分野の俯瞰や仮説生成が容易に行える検索機能
- DB構築者へのインデックス、辞書、整理棚、書式、DB構築ツールの提供
- データベース構築者(キュレータ・アナレータなど)の学習用教材
- 上記の日本語による検索と表示

これまでの成果の公開と評価

◆ 成果の公開 (<http://lifesciencedb.mext.go.jp>)

- DB整備戦略のための研究俯瞰提供
 - 学会要旨統合検索、データバンク統合目次、動向調査報告閲覧
- 統合技術開発・提供
 - 遺伝子名称辞書、生物学名日本語一般名対応辞書、施設名辞書
 - 動植物解剖学分類の自動タグ付けプログラム、解剖学3次元辞書
- ポータルサイト構築・提供
 - DB、ツール、ワークフロー
- 分子DB整理統合
 - ヒト遺伝子発現統合、植物ESTボディマップ
- DB構築者人材育成用マニュアル作成支援システム

◆ 評価

- 約100名に評価を依頼、おおむね良好な反応
-