

ライフサイエンス統合データベースの目指すもの

- ◆ 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

統合DBプロジェクト推進体制

- ◆ 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DBセンター
- ◆ 文科省統合DB整備事業
 - 中核機関: ライフサイエンス統合データベースセンター
(参画機関) 科学技術振興機構(JST)、産総研CBRC、かずさDNA研、奈良先端大学、九大、東大、お茶大、長浜バイオ大
 - 分担機関(医療にかかわるDBの統合化): 京大、東京医歯大、東大
- ◆ 運営委員会設置
 - 利用者(生命科学の研究者や医療従事者)、情報処理の専門家、DB構築者、府省関係者、産業界の代表、分担機関の代表、等からなる委員会

事業計画の概要

◆ 平成19年度事業概要

- 動向調査と利用者意見の集約に基づく戦略立案、DB知財の調査
- 人体3次元モデル、図表の検索技術、ワークフロー技術、等の開発
- ヒト統合DB、植物・微生物の整理棚・辞書の構築

◆ 平成19年度新規提供サービス

- 臓器レベル細胞レベルのヒト統合DB、辞書、ツール
- 高精度アノテーションを施した植物ゲノム統合DB、微生物ゲノム統合支援DB
- 18年度提供サービス(ポータルサイト等)の量的質的充実

◆ 20年度以降の計画概要

- 日本語の学会要旨や総説記事の統合利用
- プロジェクトDBの受け入れ(そのための審査委員会の設置)
- 分担機関との連携による医療にかかわるDBの統合化
- 専門分野(がん、免疫、など)毎の知識整理とそれによる統合DB開発
- モデル生物の統合化DB開発
- 解析ツールの部品化

中核機関の19年度事業計画と実施体制

委託業務の目的

ライフサイエンスやバイオ産業に従事する研究者や技術者がいわゆるゲノムプロジェクト・ポストゲノムプロジェクトの成果や多様なDBや解析ツールをストレスなく利用し、より高度な研究開発が効率よくできる環境（統合DB）を実現することを目的とする。このため、情報・システム研究機構、科学技術振興機構、産業技術総合研究所生命情報工学研究センター、かずさDNA研究所、九州大学、奈良先端科学技術大学、長浜バイオ大学、東京大学、お茶の水女子大学は共同して、以下の3つの業務を行う。

（1）戦略立案・実行評価

ライフサイエンス（基礎生物学および応用生物学（医・薬・農））、情報処理技術、ライフサイエンスDBの3分野の専門家による研究運営委員会を組織し、各々の分野の動向に即した我が国のDB整備戦略を立案する。また、3分野の動向に関する情報の網羅的収集・分析を日常的に行い、戦略立案や実行評価を支援するチームを組織する。

（2）統合データベース開発

ライフサイエンス、バイオ産業にかかわる情報へのアクセスと利用に関する格段の利便性向上とそれによる研究開発の飛躍的な効率化と質的向上を目指して、ヒト統合化DBおよびモデル生物・産業応用生物統合DBの開発・運用と、それに必要な統合DB構築技術、DBの標準化、文献等からの知識獲得技術等の共通基盤技術開発を実施する。

（3）統合データベース支援

個々の公開DBや利用可能なDB解析ツールなどを網羅的に調査し、提示するポータルサイトを構築する。運用にあたって専門家及び利用者の意見集約システムを開発し、意見をフィードバックして運用する。また日本語による研究情報の流通促進に注力する。プロジェクトデータベースを受入れ、相互運用可能なDBへ変換した上で公開し、運用する。また、広報・普及啓発の活動を行う。人材育成は実績のある機関と連携し階層別に、アノテータ・キュレータ教育、DB構築者養成、及びDB高度利用者養成を行う。

機関名	委託業務の題目	各機関毎の委託業務の目的
ROIS	戦略立案・実行評価 / 統合データベース開発 / 統合データベース支援	(1)戦略立案・実行評価、(2)統合データベース開発、(3)統合データベース支援、及び(4)プロジェクトの総合的推進を実施する。
JST	意見集約システム運用・広報・データベース受入 / 運用	(3)の中で、意見集約システムの運用、広報(の一部)、並びにデータベース受入・運用の一部に関わる研究開発を実施する。
かずさ	植物・植物関連微生物のゲノム情報データベース統合と高度化	(2)の中で、モデル生物・産業応用生物統合DBの開発・運用の一部(植物ならびに植物関連微生物のゲノム情報データベースの統合と高度化)に関わる研究開発を実施する。
CBRC	ワークフロー技術を用いた統合DB環境構築	(2)の中で、共通基盤技術開発の一部(ワークフロー技術を用いた統合DB環境構築)に関わる研究開発を実施する。
長浜バイオ大学	アノテータ・キュレータ養成	(3)の中で、人材育成の一部(アノテータ・キュレータの教育)を担当する。
東大	DB構築者養成	(3)の中で、人材育成の一部(DB構築者の養成)を担当する。
お茶大	DB高度利用者養成	(3)の中で、人材育成の一部(DB高度利用者の養成)を担当する。
奈良先端大	専門用語辞書管理システムと専門用語解析技術の開発	(2)の中で、共通基盤技術開発の一部(専門用語辞書管理システムと専門用語解析技術の開発)に関わる研究開発を実施する。
九大	多型知識表現技術	(2)の中で、ヒト統合DBの開発・運用の一部(多型知識表現技術開発)に関わる研究開発を実施する。