

ライフサイエンスにおけるデータの共有について －国際的な動向および我が国の取り組みと課題－

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・システム研究機構国立遺伝学研究所DDBJセンター(DDBJ)
高木利久

研究データ共有の歴史と動向

- 文献データは1960年代より
- 研究データは1970年代より
- 塩基配列データバンク(INSDC)は日米欧の3極体制で
- タンパク質データバンク(wwPDB)は日米欧の4センターで
- 他の種類のデータにも共有の枠組みが広がりつつある
- 分野別目的別単位でのデータ共有の枠組みも多々ある
 - 国際がんゲノムコンソーシアム(ICGC)
 - Alzheimer's Disease Neuroimaging Initiative (ADNI)
 - Global Alliance for Genomics & Health (GA4GH)

データ共有がなぜ進んでいるのか？

- 少数の数式や法則で表現できない
- DBは研究のインフラでありフロンティア
- ファンディング機関からの義務付け
- 論文投稿時における出版社からの義務付け
- 研究コミュニティにおける合意
 - ヒトゲノムの公開に関するバミューダ原則
- 受け皿としてDBセンターの整備
- 民間企業も共有への理解
- DB構築が評価の対象という面も
- データ共有による研究の促進
 - データ駆動型研究へシフト

データの利用に関する障害

- 自分の専門外のDBを使う必要性あり
 - DBや解析ツールの数が多すぎて使い方分からない
 - 生体内相互作用DBだけでも500以上のDB
 - 注釈が信頼性のあるものとないものが混在
 - フォーマットや用語がバラバラ
 - 遺伝子の概念さえDBによって違う
 - 同じ遺伝子にも多数の名前あり
 - データの文脈依存性、曖昧性、冗長性、複雑性、誤差
-
- データの囲い込み(公開されないデータも多数)
 - データの権利関係不明
 - データ解析できる人材(バイオインフォマティクス)の不足
 - 受け皿となる中核DBセンターがない
 - プロジェクト終了するとデータの維持管理更新されない

統合データベース事業への取組み



総合科学技術会議における取組み

関係府省の取組み

平成
17年度

- ・科学技術連携施策群¹⁾「生命科学基礎・基盤」の開始（主監 五條堀 孝）
- ・補完的研究課題「生命科学データベース統合に関する調査研究」の実施（研究代表者 大久保 公策）

平成
18年度

- ・第3期科学技術基本計画²⁾の戦略重点科学技術³⁾「世界最高水準のライフサイエンス基盤整備」として統合データベースを位置づけ。
- ・科学技術連携施策群の成果及び今後の見通し（中間報告案）を作成

平成
19年度

- ・海外のバイオリソース事業の責任者を招聘し、講演会を実施
- ・科学技術連携施策群の最終取りまとめ（終了）

平成
20年度

- ・総合科学技術会議ライフサイエンスPTに「統合DBタスクフォース」の設置（座長 五條堀 孝）

平成
21年度

- ・「統合DBタスクフォース」の報告書の策定（4月10日）

- ・関係省庁における検討

- ・文部科学省の統合データベースプロジェクトの開始
- ・農林水産省の統合データベースプロジェクトの開始

- ・経済産業省の統合データベースプロジェクトの開始。
- ・文部科学省「ライフサイエンス情報基盤整備作業部会」報告書策定。

平成21年度概算要求の優先順位付けにおいて、「文部科学省統合データベースプロジェクトとJSTバイオインフォマティクス推進センター（BIRD）との一体化を目指して、加速して事業を実施する必要がある。」と具体的な取組みを評価

1) 科学技術連携施策群：各府省の縦割りの施策に横串を刺す取り組み
2) 第3期科学技術基本計画：平成18年度から5年間の我が国における科学技術の推進戦略
3) 戦略重点科学技術：特に重要な技術を選択し、研究資金を集中させるべき分野