

生命科学データベース統合に 関する調査研究

情報システム研究機構ROIS

国立遺伝学研究所/DDBJ生命情報研究センター

大久保 公策

平成20年5月7日

課題内容

「生命科学データベース統合に関する調査研究」

- ・DB統合化にむけて関係府省の制度設計やロードマップ作成に資する試案の提示
- ・現状調査による技術的課題並びに解決策の提示
- ・フィージビリティスタディを通じた実現性検証結果の提示
- ・技術的側面、制度的側面からの基礎調査研究の提示

実施内容 H17-19



国内外のDB調査、産業界等からのヒアリング

- ・研究DB実情の調査 研究DBの網羅調査と分類
- ・研究情報流通の現状と周辺動向(オープンアクセス EBM)
- ・研究関連情報流通の現状と動向調査
- ・創薬プロセスの創薬業界参加によるコンセンサスモデルの作成

多面的かつ
複雑な背景

DB統合に向けた技術的FS:省をまたいだDB連携

- ・情報通信研究機構ボクセル人体データの解剖学辞書としての再利用FS
ボクセルモデルの技術的限界を提示→ポリゴンモデルで詳細化成功
- ・産総研(経産)と国立がんセンター(厚労)で分子医学試験統合:ID対応自動生成ロボット

技術的課題は
十分対応可能

DB統合に向けた制度調査:DB統合が進む米国等の制度調査

- ・データ駆動制度および保管流通制度の調査
- ・データ死蔵例の調査及びその原因分析
- ・研究データ 調査データに関する法 制度の調査

制度上の問題
を明確化

研究実施体制

「生命科学の基礎・基盤」連携施策群

指示

報告

研究代表者： 大久保（遺伝研）

3. 研究運営委員会

- ・研究プロジェクトの方針修正
- ・制度設計等の検討

委員)・研究代表者・分担者
・有識者
・関係各省担当者

2. 事務局: (大久保)

- ・研究プロジェクト推進のための調整・支援

遺伝研事務局(プロジェクト運営窓口)

- ・運営委員会との情報・意見の交換
 - プロジェクト運営方針の検討と実施
- ・世界のデータベースの調査・整理と報告
 - ネットワーク経由での情報の収集と整理

1. 分担機関: 基礎調査・検討

- ・各分野のデータベースの調査・検討・報告
 - 事務局との連携による調査と評価
 - 統合化に向けた情報の整理

遺伝研
(大久保先生)
文科省関連DB担当

がんセンター
(吉田先生)
厚労省関連DB担当

生資研
(長村先生)
農水省関連DB担当

産総研
(今西先生)
経産省関連DB担当

東大
(森下先生)
文科省関連DB担当

DB統合が望まれる背景の分析

研究内部の変化と研究環境の変化

- 巨大観察データの登場によるライフサイエンス全体のデータ(ベース)依存型への変貌



- 高額な巨大データに関して大学・企業研究開発の公共データ(ベース)への依存増

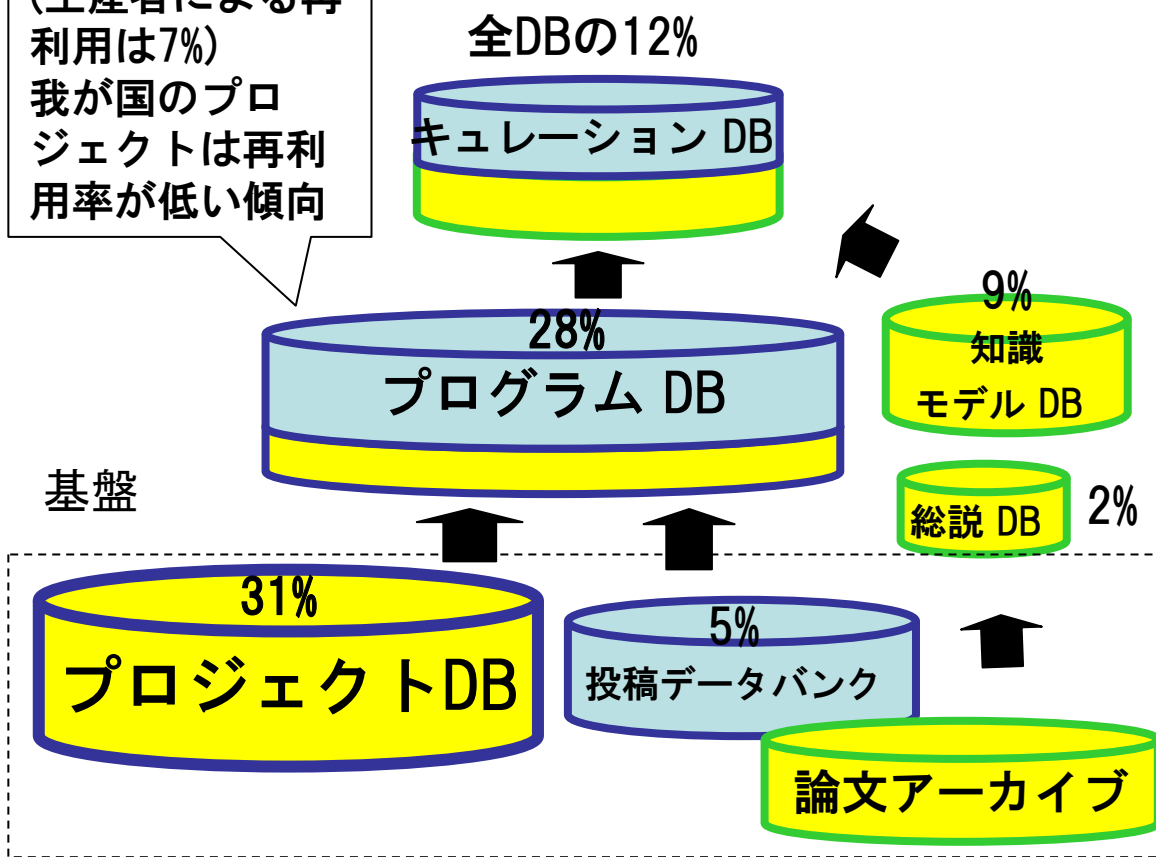


- 大学などの研究機関の自由競争制度
 - データ占有による知財獲得 (選択公開、遅延公開)
 - データベースサービスも競争的 (末端公開、転用不許可)

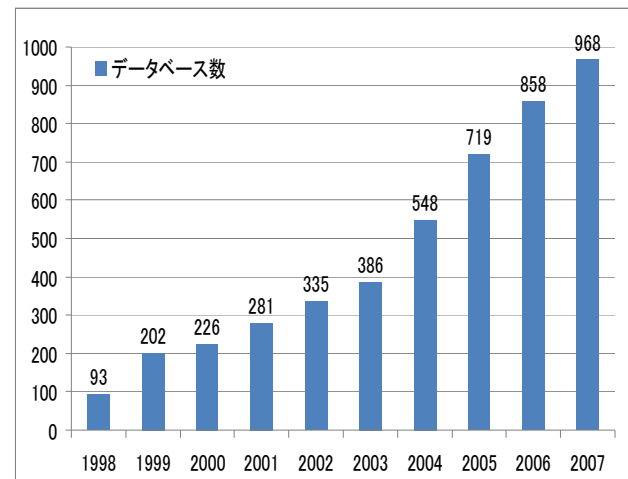
製薬企業を含む63名の有識者と面談、2回の意見交換会議の結論

ライフサイエンス分野のDBの増加と多様化


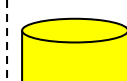


材料国産率3%
(生産者による再
利用は7%)
我が国のプロ
ジェクトは再利
用率が低い傾向



世界的な分子DBの増加



NAR DB issue より

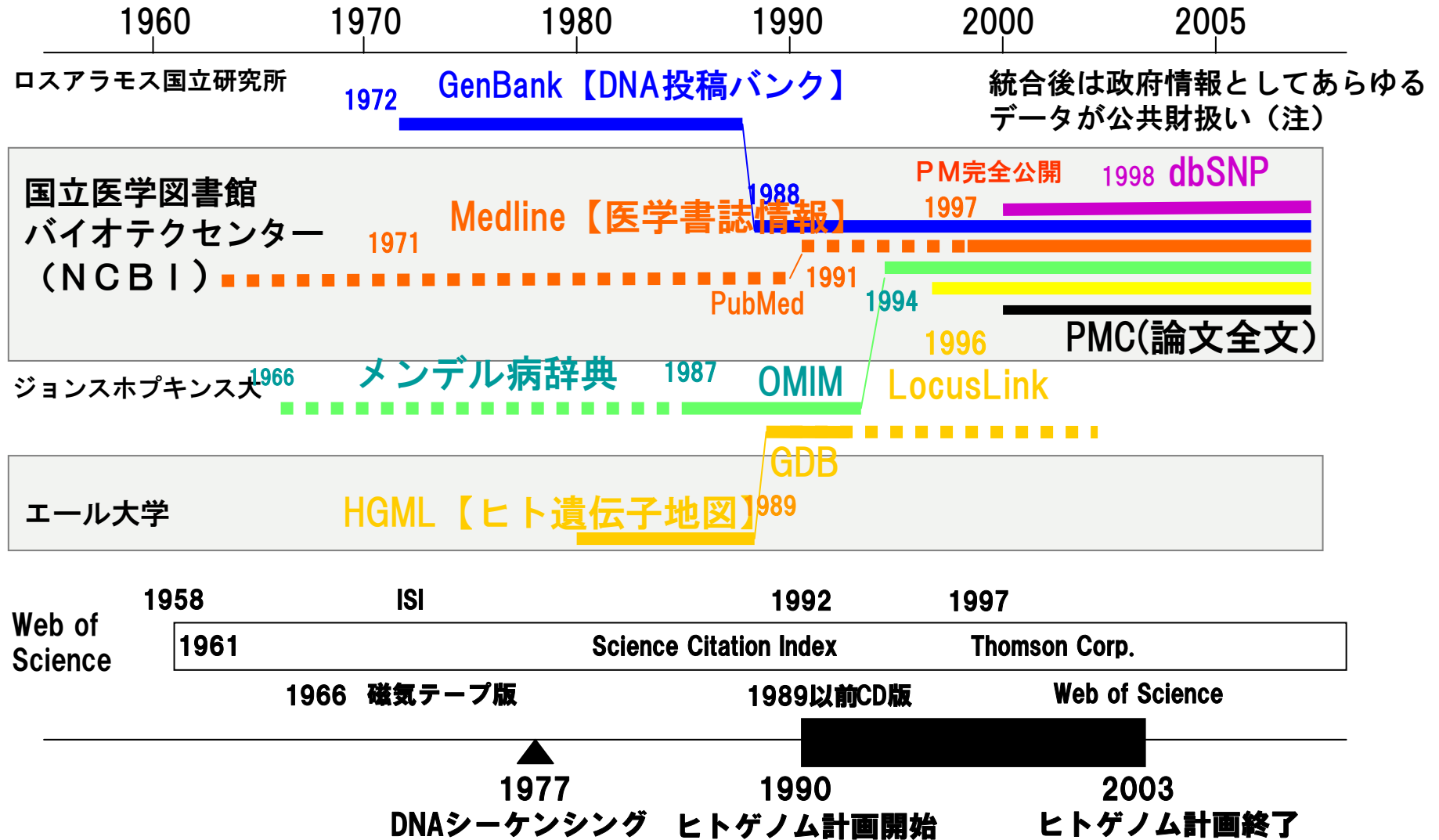
-  利用に制限制約なし
-  末端利用のみ許可
-  分析データ由来のDB
-  論文教科書由来のDB

250を数える我が国の"公開"データベースの内訳
(データベースの網羅的リストアップと独自分類、データ成分への手作業での分解に基づく分析結果)

ゲノム・ポストゲノム 主要プロジェクト名	年 度							プロジェクトの概要
	H12	H13	H14	H15	H16	H17	H18	
文部科学省	プロジェクトデータベースの増大増加							
ゲノムネットワーク								遺伝子の発現調節機能に関わる網羅的な解析
タンパク3000								主要タンパク質約3000種の基本構造及びその機能解明
遺伝子多型研究								ヒトゲノム遺伝子領域中のSNP関連情報の取得と解析
テーラーメイド医療実現化								約30万人のSNPと薬剤の効果、副作用などとの関係解明
理研ゲノム、植物、遺伝子多型								ヒト、マウス、植物のゲノム、cDNA解析、遺伝子多型解析
バイオインフォマティクス研究								生命科学分野の基幹データベースの構築・高度化
統合データベース								生命科学分野DB戦略立案支援、ポータルサイト整備
経済産業省								
データベース結合								国内外の有用なヒトゲノム関連情報、解析ソフトの統合的連携
ゲノム情報統合								
完全長cDNA								約3万のヒトの全長cDNA配列情報の取得と解析
生物システム制御基盤技術								創薬支援のためのゲノム、タンパク、化合物一貫解析技術開発
生体高分子立体構造								膜タンパク質及び関連複合体の立体構造・機能解明
蛋白質機能解析								完全長cDNAの遺伝子発現頻度など多方面からの機能解析
遺伝子多様性モデル解析								ヒトのモデル疾患に関わる遺伝子多型情報の取得と解析
標準SNP解析								日本人集団768人に関するSNP15万種のアレル頻度の解析
厚生労働省								
疾患ゲノムデータベース								がん等5疾患のゲノムワイドなSNP解析などのデータベース構築
トキシコゲノミクス								遺伝子発現解析によるゲノムレベルでの毒性発現機構解明
疾患関連蛋白質								主要疾患を対象とした疾患関連たんぱく質の探索、同定
農林水産省								
イネゲノム								イネゲノム配列の解読および遺伝子の機能解明
家畜ゲノム								ブタのcDNA配列情報、発現頻度、マーカー情報の取得と解析
蚕ゲノム								蚕のゲノム、cDNA配列情報、連鎖地図情報の取得と解析
農林水産生物ゲノム情報統合DB								イネその他農林水産生物統合ゲノムデータベースの整備

実線はデータ生産型

米国NCBIにおけるDB統合の実例

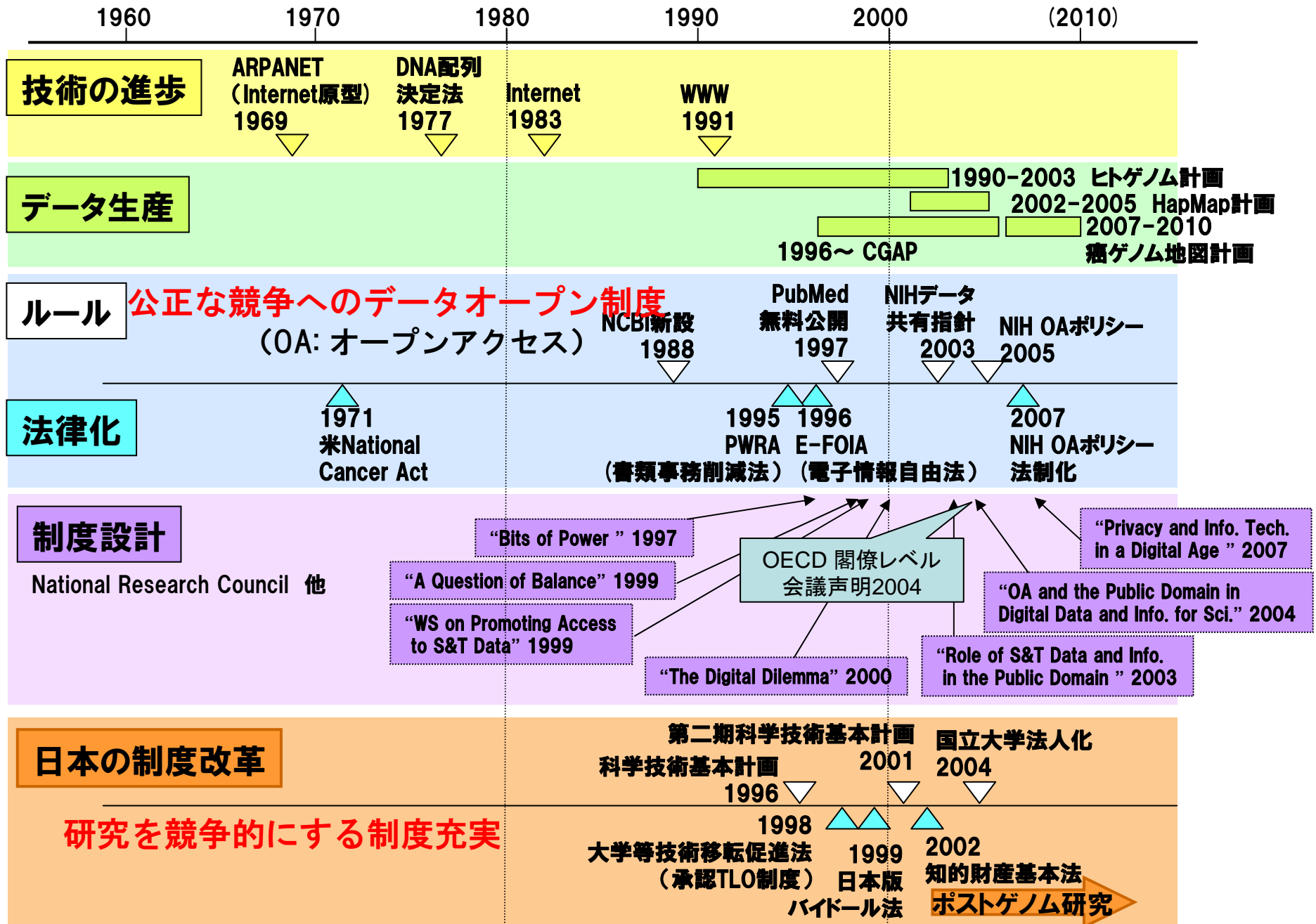


統合後は政府情報としてあらゆるデータが公共財扱い(注)

(注)OMIMや電子教科書は商業転用制約

OMIM著者D.McKusickは2008 日本国際賞

統合を可能にした制度整備（米国）



DB統合化に向けた提言に資する試案

「政府資金によるデータ産生型プロジェクトのデータを我が国の研究社会で早期に共有するためのルール」が必要

- 例えば、政府資金によるデータ産生型のプロジェクトについて、早期にデータ共有する為の計画を提案採択時の判断基準とする。

公共財としてのデータを保全・管理し、長期にわたるデータの育成と共有を行う公的機関を設ける

- データの生産者としての研究プロジェクトは研究計画終了後に解散となるため、データを維持提供し続ける主体とはなりえない。

科学・経済両面から 科学データの公共財化の議論が始まりつつある

- 多くの人々の各種活動に利用される「基盤」とするためには、蓄積された知的資産が体系化され、それが広く「**供用可能**」とされている必要がある。
- （出展：文部科学省 知的基盤整備計画「知的基盤とは」科学技術・学術審議会 技術・研究基盤部会 H19.9）
- 「Open Access to **publicly funded** data provides greater returns from the public investment in research and generates wealth through **downstream commercialization** of outputs」
- （出展：第11回 OECD科学技術閣僚級会議最終コミュニケ@パリ
- （稲葉副大臣(当時) 野依理研理事が参加）
- http://www.oecdtokyo2.org/pdf/theme_pdf/science_pdf/20040130communique.pdf

すべての調査結果は動く報告書の形で公開予定（成果公開DB）

サイト内検索

検索

データベース公開の現状

MAIN MENU :

- ▣ はじめに
- ▣ シンポジウム発表資料
- ▣ 調査レポート
 - ▣ データベース公開の現状
 - ▣ 情報リテラシー向上のために
 - ▣ 統合に向けたフィジビリティスタディ
 - ▣ 著作権・知財権
 - ▣ OpenAccessの動き
 - ▣ 情報公開に影響を及ぼす法体系等
 - ▣ その他
- ▣ 研究運営委員会資料
- ▣ サイトポリシー
- ▣ Contact

調査レポート

調査レポート >> データベース公開の現状

国内

- DB分類資料
- ライフサイエンス関連独立行政機関
- 生物資源寄託機関
- J-STAGE関連調査
- JSTデータとPubMedのリンク
- 科研費DB調査(医学)
- 疫学研究者からのヒアリング
- 臨床疫学研究データベース
- 平山計画調査のデータについて
- EBMのためのDB調査
- 医学文献専門家からのヒアリング
- 厚労省関連情報統合の動き
- 厚労省関係財団法人

国外あるいは国内との比較

- DB歴史年表
- 学術情報と内外の動き
- MedlinePlusと国内の類似サービス
- ゲノムコホート比較
- 各国のデータ公開状況
- DBの利用条件とアクセス数

我が国のDBの説明つき目録

48の日本のコホート研究のカタログ化

出生前の母親の喫煙・食事・環境・職業要因調査、母体血、臍帯血・胎盤、生後1ヶ月時に母乳を採取、生後6ヶ月、1歳6ヵ月後の発達度合い及びアレルギーテスト				
動脈硬化の危険因子を明らかにする多施設コホート統合研究(JALS)				
2001	既存の33コホート対象者	日本動脈硬化予防基金	滋賀医科大学(右記参照)	既存の循環器コホート研究の個人データを統合 委員長は滋賀医科大学 (統合研究のデータ収集・データ解析等の事務局業務は、NPO法人日本臨床研究支援ユニットが実施)
2014(予定)	120,000	滋賀医科大学医学部、東京大学大学院医学系研究科他多数		学会報告のリストはあり
所定の問診項目(家族歴、既往歴、現病歴、嗜好、運動習慣、服用薬など)、身長、体重、血圧、血液検査(総コレステロール、HDLコレステロール、血糖)、心電図などの記録が必要				
シックハウス症候群の疫学研究				
2001	築浅住宅住民		北海道大学	住環境・住まい方と症状の関係
2005	564+96+約1,000件			論文、口頭発表、研究報告書
住宅の状況と家族の健康に関する質問紙調査及び訪問調査				
鶴谷プロジェクト				
2002	仙台市宮城野区鶴ヶ谷地区70歳以上住民		東北大	健康長寿に関する要因についての研究
2003(一部追跡中)	1,179		宮城野区保健福祉センター、東北大11分野、東北文化学園大	
高齢者の機能(運動、うつ、認知機能、歯科、生活習慣、動脈硬化等)調査、死亡・入院・医療費等の状況				
オーダーメイド医療実現化PJ				
2003	プロジェクトの協力医療機関	文科	東京大学医科学研究所	がん、糖尿病、心筋梗塞など47疾患