

生命科学の基礎・基盤連携施策群 補完課題

生命科学データベース統合に関する 調査研究

情報システム機構・国立遺伝学研究所

大久保公策

データベース統合問題の再定義(H18)

DB = 実世界の写し絵(モデル)

統合=実世界とモデル世界の1:1写像



DB統合と数理モデルがリアルな写像への両道

{ 分割世界を重ね合わせよりリアルに(DB統合):特徴豊富なリアルさ
小さなモデルを数理で表現(システムズバイオ):動くリアルさ

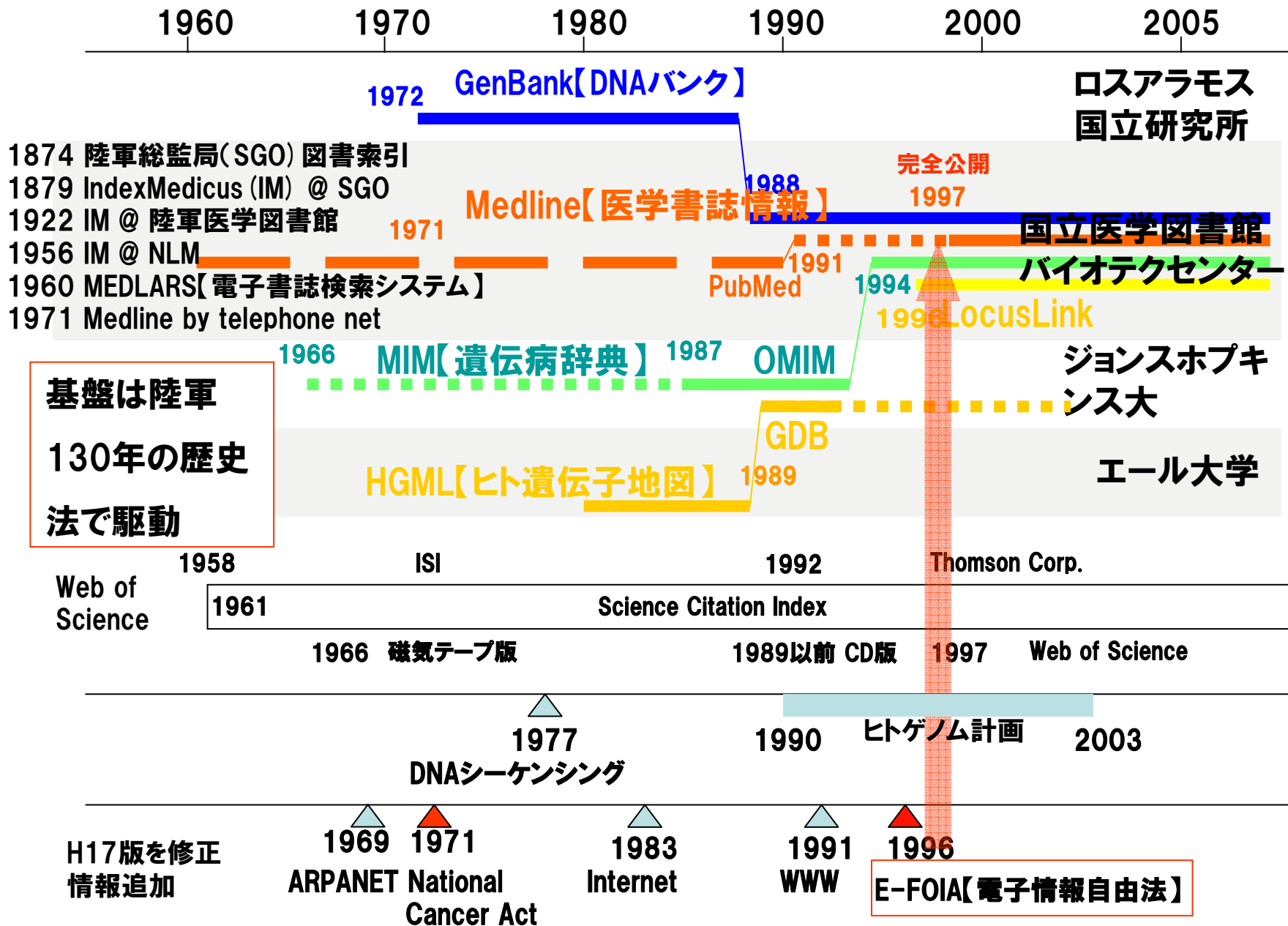
データベース構築法分類 (H17)

型名称	情報源の種類	処理方法	記述者	データ書式	典型	DB数 (国内報告)
バンク型	測定器と 実験者	-	測定器と 多人数	構造化 テキスト	GenBank (1979) PDB (1972)	
プロジェクト型	測定器と 実験者	-	測定器 少人数	量的対応関 係	BodyMap (1992)	
プログラム型	既存データ	機械的処理	計算機	対応関係	UniGene (1996)	
キュレーション型	既存データ	高度情報処理	計算機と 少人数	対応関係	SCOP	
知識モデル型	読み物 (電子・紙)	高度情報処理 (読解)	少人数	対応関係	KEGG (1995)	
総説型	読み物 (電子・紙)	高度情報処理 (読解)	少人数	構造化 テキスト	OMIM (1987)	

それぞれの型に別の課題や共通の課題

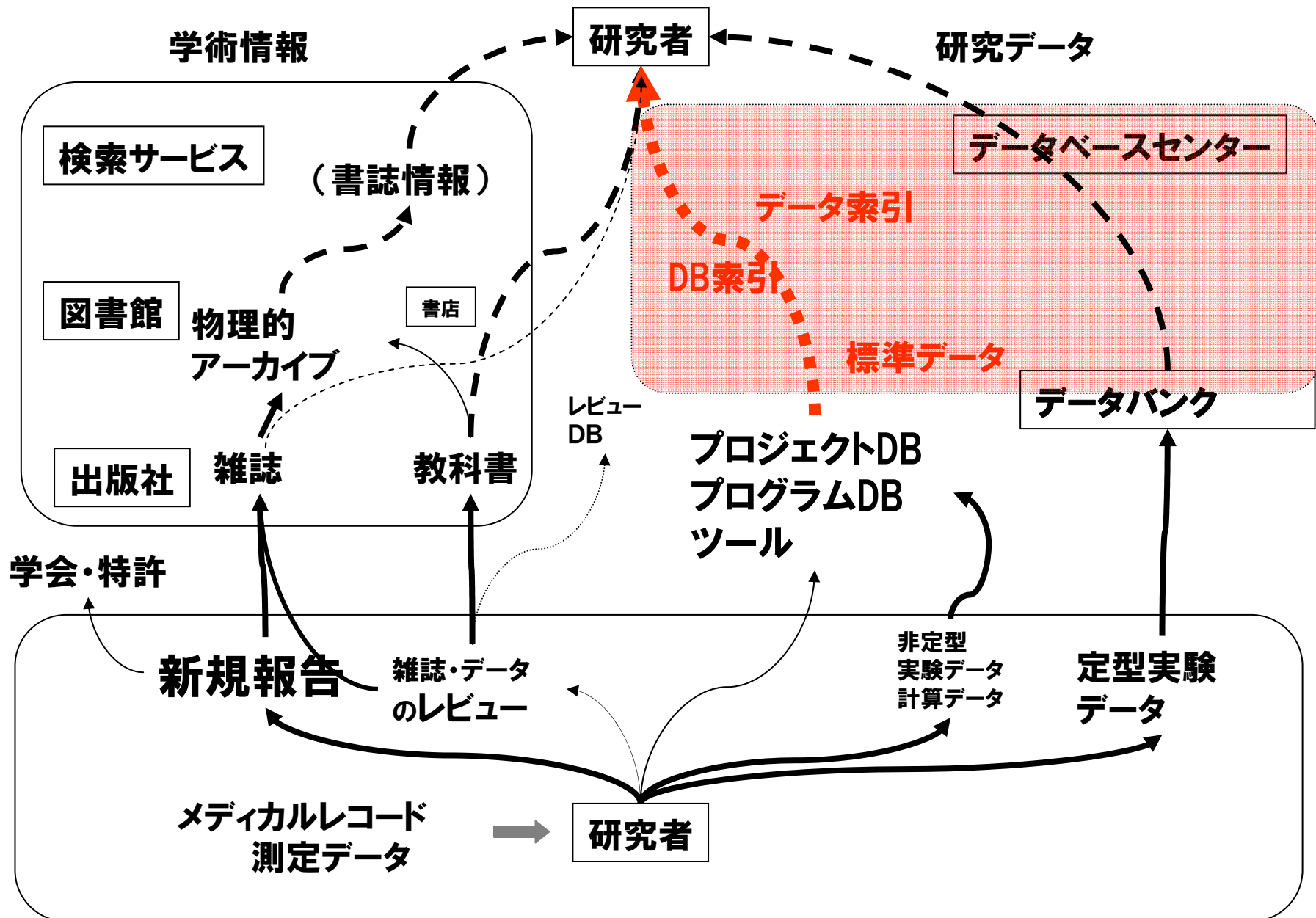
0 70

(参考) 学術情報基盤上での医学情報の統合 (米国の事例)



紙情報にならった解決案 (H17)

媒体の管理による内容の流通(統合は頭の中で)



課題内容

「生命科学データベース統合に関する調査研究」

- ・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

18年度計画と成果



現状調査：情報流通の全貌 ユーザドメインの情報処理のモデル化

- ・研究情報流通のモデル化とオープンアクセスの意味の理解
- ・蛋白構造データベースの調査分類
- ・学協会が編纂した学術基盤情報(学術用語と学会抄録)の流通の現状
- ・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
- ・製薬プロセスモデル上の欠落情報の列挙 制度課題として検討

技術フィジビリティスタディー：省庁連携を生かし技術課題の提示

- ・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(文科省DB統合PJ)
- ・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中
- ・植物解剖および植物分類の整理(文科&農水)→イネアノテーションに反映

制度調査研究：制度上の課題の列挙と克服への立案

- ・有用和文情報死蔵化圧力の調査
- ・プロジェクトデータ死蔵の制度上の欠陥の調査
- ・既存健康情報源の現状調査

DB統合調査要点報告

1. 紙媒体世界の崩壊とDB統合



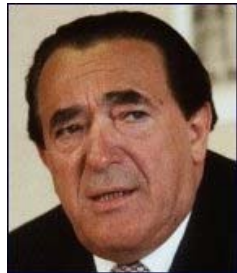
2. 創薬研究者の望むDB統合



3. DB統合と制度



オープンアクセスの背景：寡占による学術雑誌の価格高騰



was forced to sell Pergamon Press and Maxwell Directories to Elsevier for £440 million (1000億円)

Robert Maxwell, Known as “Cap’n Bob” and the “bouncing Czech”, was a British Media proprietor and Member of Parliament.



Crispin Davis Present CEO of Reed Elsevier

2005年 収益1142mil ポンド Reed Elsevier who specialise in publishing and exhibitions amongst other activities recently took over Spearhead, the company organising DSEi (Defence Systems Equipment International ロンドン開催の武器見本市).

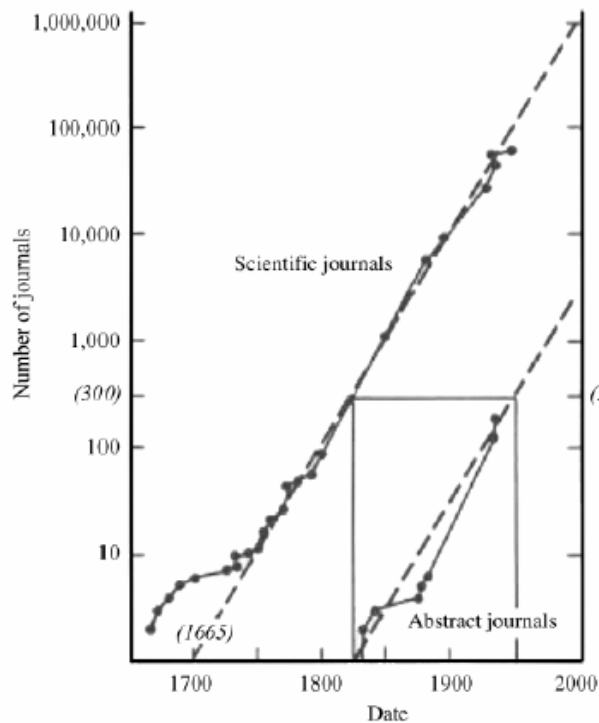
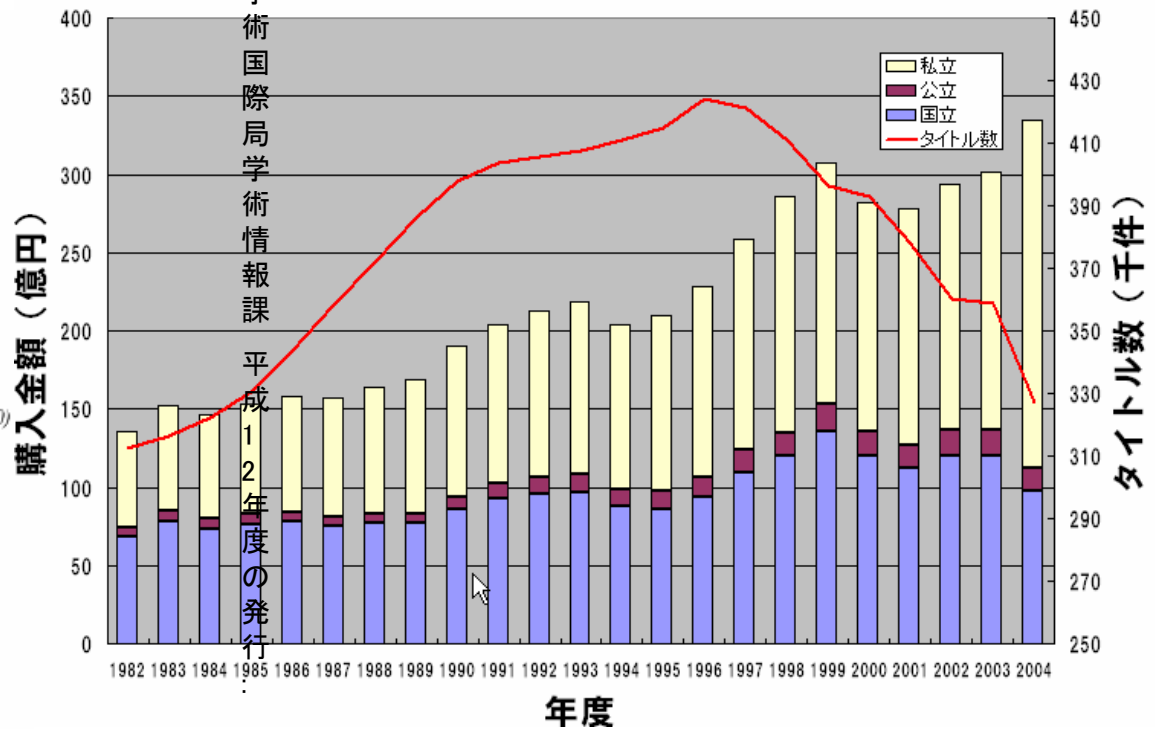


Figure 1 Total number of scientific journals and abstract journals founded, as a fraction of date

日本国内図書館の外国雑誌購入費および受入れタイトル数



文部科学省 研究振興局情報課 大学図書館実態調査結果報告より

文部科学省

学術情報寡占への米立法の挑戦

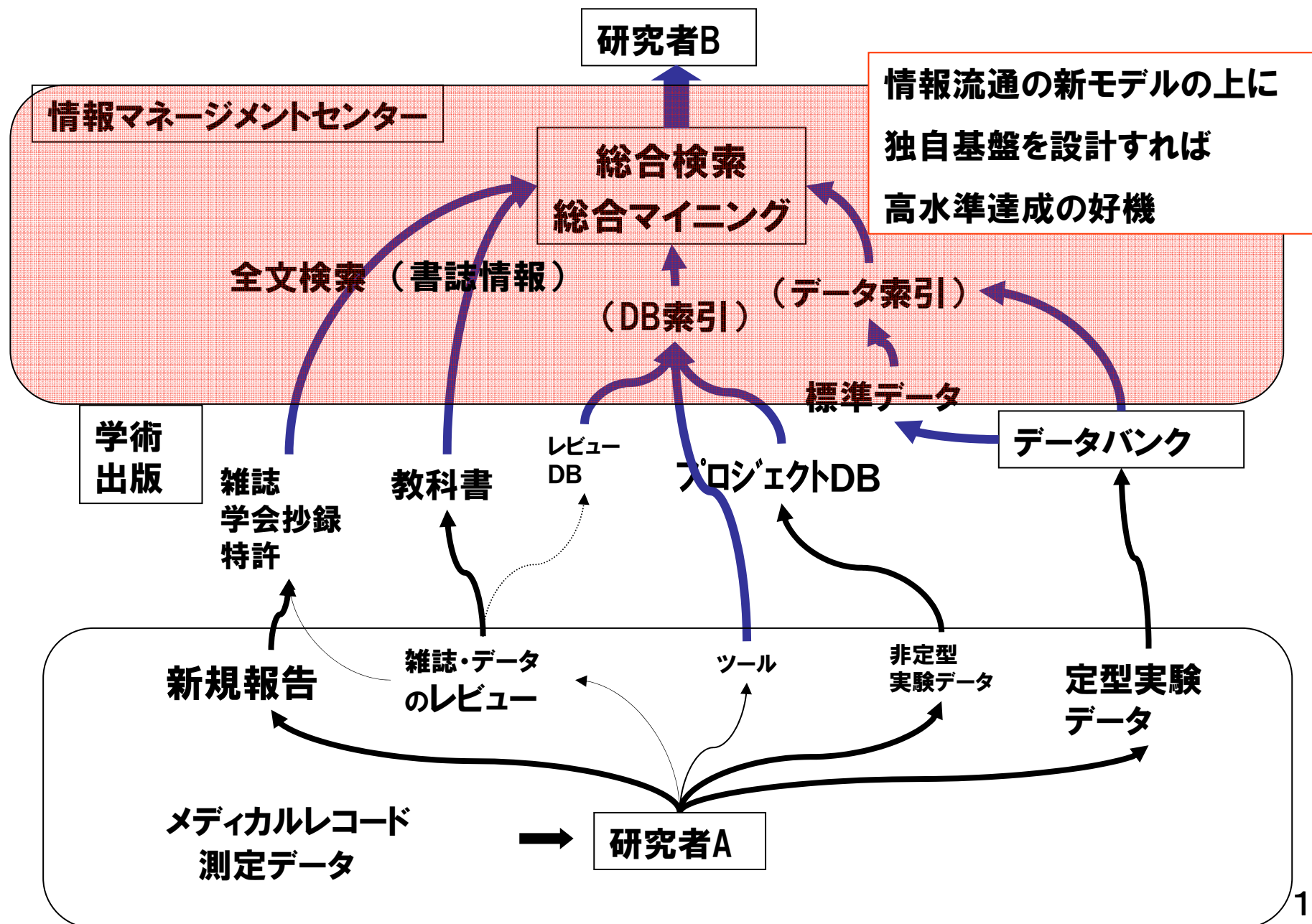
- American Center for Cures Act (CURES法案) 2005 (未通過 2007.2)
 - 保険社会福祉省の援助研究成果論文を6M以内に公開
- NIH自主ガイドライン2005
 - NIH支援の研究成果(論文)は6M以内にPMCに登録
- Federal Research Public Access Act of 2006法案 (未通過 2007.2)
 - 連邦政府が支援した成果論文は6ヶ月商業利用期間のうちに無料公開 PubMed Centralなどの公的機関リポジトリへ

【過去の事例】

- INSDC(DDBJ-GenBank-EMBL), PDBJ-PDB の成功も雑誌(NAR)の申し合わせによる強制
- 電子情報自由法(Electronic Freedom of Information Amendments of 1996)
 - 連邦政府の公開情報に電子化情報を含めた
 - ゴア副大統領が最初の利用者となった「アキレス腱断裂」検索セレモニー MEDLINEの無料化によるPubMedの誕生
 - 社会のオンライン化による紙媒体製造配布コストによる“有償政府出版物の法的根拠消失”

税金研究の報告を米国民に取り戻すための運動オープンアクセスOA

オープンアクセスを前提とする情報基盤設計案



DB統合調査要点報告

1. 紙媒体世界の崩壊とDB統合



2. 創薬研究者の望むDB統合



3. DB統合と制度



平成18年度拡大会議

- 製薬における情報利用基盤のモデルを考える -

参加者 35名			
製薬企業より	13名	大学薬学部(元製薬企業)より	2名
省庁関係より	4名	課題実施者より	3名
事務局より	7名	その他協力者	6名

1月14日(日)

13:05 開会挨拶	研究代表者 大久保公策 (遺伝学研究所)
13:10-13:20 本課題の紹介	ポストゲノム連携施策群プログラムオフィサー
13:20-13:50 「ライフサイエンスの情報基盤の現状と課題」	事務局
13:50-14:20 「文部科学省統合データベースプロジェクト計画」	高木利久(東京大学)
14:20-14:40 「経済産業省の情報基盤取り組みについて」	経産省 製造産業局 生物化学産業課 課長補佐
14:55-15:15 参加者自己紹介	
15:15-15:35 事前調査集計結果紹介	
15:35-16:25 ディスカッション:事前調査結果について	
16:25-16:35 ディスカッション総括	
16:50-17:40 情報基盤について自由意見交換	
17:40-18:30 ディスカッション:創薬研究開発コンセンサスモデル形成(1)	
20:00-22:00 ディスカッション:創薬研究開発コンセンサスモデル形成(2)	

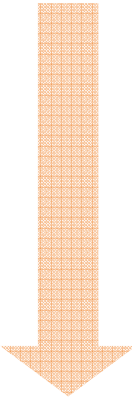
1月15日(月)

9:30-10:00 創薬研究開発コンセンサスモデルまとめ報告
10:00-10:30 コンセンサスモデルに基づいた課題・注力順位付け議論
10:30-11:40 省庁横断DBセンターの役割と形(総合討論)
11:40 閉会挨拶

大久保公策

創薬過程と利用情報

● 開発開始の判断



- 対象疾患と効果(TPP)決定
 - 疫学情報 (罹患・有病率)
 - 臨床(治療)情報
 - 開発・研究情報
- 作用点の決定
 - 病理病態文献DB
 - 特許情報DB
 - 分子生物DB

充実が
課題

海外に対す情報アドバンテージ
(日本の健康情報や日本の研
究の状況)が不在

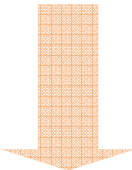
研究開発・前臨床試験

統合が
課題

望まれる対策

- コホート研究充実
- 疾患登録推進
- 行政情報(申請)の利用
- 学会情報の統合利用
- 特許DBの統合
- 分子DBの統合
- 記述研究の推進

● 治験をしない判断



- 動物毒性情報
- 過去の治験結果情報
- 副作用情報

充実が
課題

臨床試験

DB統合調査要点報告

1. 紙媒体世界の崩壊とDB統合



2. 創薬研究者の望むDB統合



3. DB統合と制度



法的動機付けにより政府管轄機関に集積する情報

