

ライフサイエンス分野の統合データベース整備事業

統合医科学データベース構築方式の開発
20年度 委託業務研究成果報告書

平成21年3月

国立大学法人東京医科歯科大学

田中 博

国立大学法人大阪大学

佐古田 三郎

本報告書は、文部科学省の科学技術試験研究委託事業による委託業務として、東京医科歯科大学、及び大阪大学が実施した、平成20年度の「統合医科学データベース構築方式の開発」の成果を取りまとめたものです。

1 委託業務の目的

ライフサイエンス関係データベースに散在する臨床・疾患データベース（DB）を分子から個体にいたる多階層の疾患横断的な統合医科学DBへ構築するための高度化・統合化方式を開発することを目的とする。がん疾患、神経疾患を対象に実証する。

このため、国立大学法人東京医科歯科大学及び国立大学法人大阪大学は共同で業務を行う。

国立大学法人東京医科歯科大学では、がん疾患に関わる研究開発を中心に、国立大学法人大阪大学では神経疾患に関わる研究開発を中心に実施する。

2 平成20年度（報告年度）の実施内容

2.1 実施計画

2.1.1 東京医科歯科大学

① 疾患・臨床医科学データベースに特化した統合技術開発に関する研究

平成19年度に実施した多階層の統合医科学データベースを統合構築するために必要な統合技術（疾患DB情報モデル、疾患DBオントロジー、セマンティクスなど）の調査・検討結果に基づき、要件定義を行う。

国立大学法人東京医科歯科大学はがん疾患を中心とし、国立大学法人大阪大学は神経疾患について研究を実施する。

② 疾患データベースの高度化と実証的統合に関する研究

平成19年度に実施したがん疾患DB、神経疾患DBの高度化によって得られた、統合技術開発における具体的要求と課題の検討を進め、要件定義を行う。

国立大学法人東京医科歯科大学は、がん疾患（肝臓がん、大腸がん）を中心として、国立大学法人大阪大学は神経疾患（パーキンソン病）について研究を実施する。

③ プロジェクトの総合的推進

分担機関である東京医科歯科大学は、中核機関である情報・システム研究機構の全体戦略に従い連携して本事業を推進する。

プロジェクト全体の連携を密としつつ円滑に運営していくため、技術検討会等を開催し、分担機関グループ内の連携・調整及び進捗状況の確認を行う。

プロジェクトで得られた成果については、倫理委員会が承諾する範囲内で、積極的に公表し、今後の展開に資する。

2.1.2 大阪大学

① 疾患・臨床医科学データベースに特化した統合技術開発に関する研究

平成19年度に実施した多階層の統合医科学データベースを統合構築するために必要な統合技術（疾患DB情報モデル、疾患DBオントロジー、セマンティクスなど）の調査・検討結果に基づき、要件定義を行う。

国立大学法人大阪大学は神経疾患について研究を実施する。

② 疾患データベースの高度化と実証的統合に関する研究

平成19年度に実施したがん疾患DB、神経疾患DBの高度化によって得られた、統合技術開発における具体的要求と課題の検討を進め、要件定義を行う。

国立大学法人大阪大学は神経疾患（パーキンソン病）について研究を実施する。

2.2 実施内容（結果）


2.2.1 東京医科歯科大学


① 疾患・臨床医科学データベースに特化した統合技術開発に関する研究

平成19年度の調査および検討結果に基づき、多階層の統合医科学データベースを統合構築するために必要な統合技術、標準化、公開倫理の要件定義を行った。

まず、本統合医科学データベースの要件定義の礎となる倫理規定草案の作成を行った。作成にあたっては具体的には以下の四段階を経た。

(1) 一般論としてデータベース統合についてその影響や特質について検討。(2) 具体的な医科学データベースの統合を進める場合に想定される問題を、オミックスデータベースと臨床データベースについて検討。(3) 法的側面を整備するため、罰則規定や訴訟の影響について、事例等を踏まえつつ検討。(4) 社会的な適合性も損なわないよう、成立／改定や施行のプロセスについても策定。

上記の検討の結果、1のように倫理規定草案の概要をまとめた。DB統合の是非、統合の程度を最重要項目としてそれぞれの項目について検討を行った。

今後行うべき項目としては、正確性、安全性、透明性（3要素）の確保、また、匿名性が損なわれる可能性を構造的になくすことなどが考えられる（2）。

次に、上記倫理規定案を元に疾患・臨床医科学データベースの統合を行うための疾患DB情報モデルの検討、疾患DBオントロジー構造の検討、セマンティクス検索のアルゴリズムの検討を行い、システムの外部設計書を作成した。

本外部設計書を元に内部設計として、リレーショナルデータベース物理設計、プログラムモジュール設計、外部インタフェース構造を定義し、内部設計書を作成した。

以上の内容を合わせてシステム要件定義書を作成した。

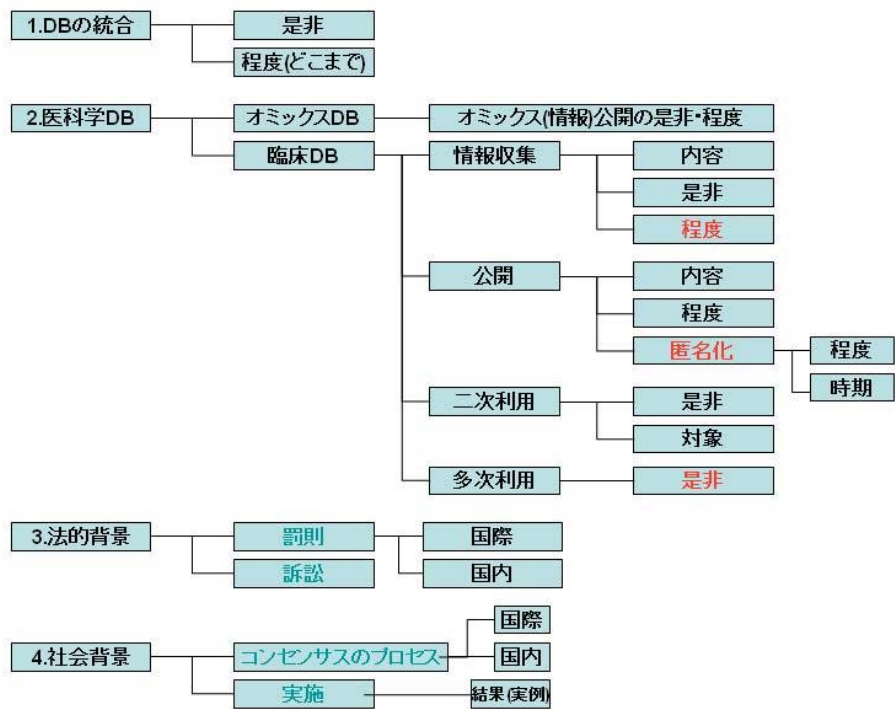


図 1 倫理規定草案概要

・確保すべき3要素

- 正確性 ⇒ 正確かつ最新の内容に保つ義務
- 安全性 ⇒ 患者等個人の人権の安全管理のための措置
取扱い者に対して監督する義務
- 透明性 ⇒ 取得時の利用目的の通知 本人からの開示・訂正請求に応じる義務

・情報を連結する事で匿名性が損なわれる可能性を構造的にゼロにする

⇒連結不可能匿名化

・利用目的による利用者の制限

⇒医療従事者でも利用者登録を必要とする、監査証跡(アクセス記録等)のとれる公開体制

・個人を特定して悪用した場合に関する罰則

⇒安全性の基準としての匿名性の定量的評価

・情報取扱いについての対象者への説明と、これに基づく同意の取得について

⇒現実問題として全ての場合で同意を得る事ができるかの検討

・倫理審査委員会との連携

⇒DB公開前の段階から連携し、倫理上の問題やプライバシー権の侵害の可能性について審査

図 2 倫理規定として予定される項目

② 疾患データベースの高度化と実証的統合に関する研究

平成19年度より先行的に実施している網羅的疾患分子病態DBとパーキンソン病DBを統合的に検索可能なプロトタイプシステムの試験公開によって得られた要求および課題に基づき、高度化と実証的統合にむけた要件定義を行った。さらに要件定義を元に以下の検討、開発を行った。

ユーザインタフェース要件高度化検討の結果、従来の検索方式に加え図 3のようなパネル方式を開発した。本方式により、ユーザは検索カテゴリーの階層構造を意識しながら検索条件を入力することが可能となった。

オントロジー情報の高度化として、大腸癌、口腔癌、肝臓癌、パーキンソン病（パーキンソン症候群）、頭頸部癌において、語数 247語、関連数 190のオントロジー情報を整備した。その結果、例えば、腹痛→(大腸癌)→血便/腹部膨満/下痢というセマンティック変換が可能となり、一つの症状から複数の疾患に検索を広げることが可能となった。また一方で、検索時間の増加という課題が見つかった。要素データベースとのインタフェース部分の高度化に関して、複数の要素データベースを統一して検索することが可能なインタフェースを検討した。結果として、図 4に示すように、平成19年度成果では各要素データベース内の症例検索は統合

医科学データベースの一部として行われていたのに対し、本年度では各要素データベースを統合医科学データベースから切り離し、インタフェースプログラムを介して各要素データベースとやり取りすることで統合医科学データベース仕様および各要素データベースの仕様に関係ない独立したシステム系として設計した。この仕組みは、インタフェースプログラムの変更のみで要素データベースの追加が可能となるため、将来的な要素データベース拡充の業務効率化に大いに寄与する。

また、実施計画に先行して、臨床情報の統合に加え、分子情報解析を統合したシステムの構築を目指し、遺伝子発現量を元に症例情報をクラスタリングし検索結果と共に一覧表示する機能（図5）を開発した。今後ともデータベース統合に関する技術的知識蓄積のための機能の追加、拡張を検討したい。

要件定義を元に平成19年度のプロトタイプシステムを修正し、対象データベースとして、網羅的疾患分子病態DB(447症例)、パーキンソン病DB(400症例)の症例数追加に加え、遺伝子発現の分子情報を持つ疾患ゲノムDB (GeMDBJ;Genome Medicine Database of Japan) の内440症例を統合した。

これを統合医科学データベースのポータルサイトから試験公開を開始した。

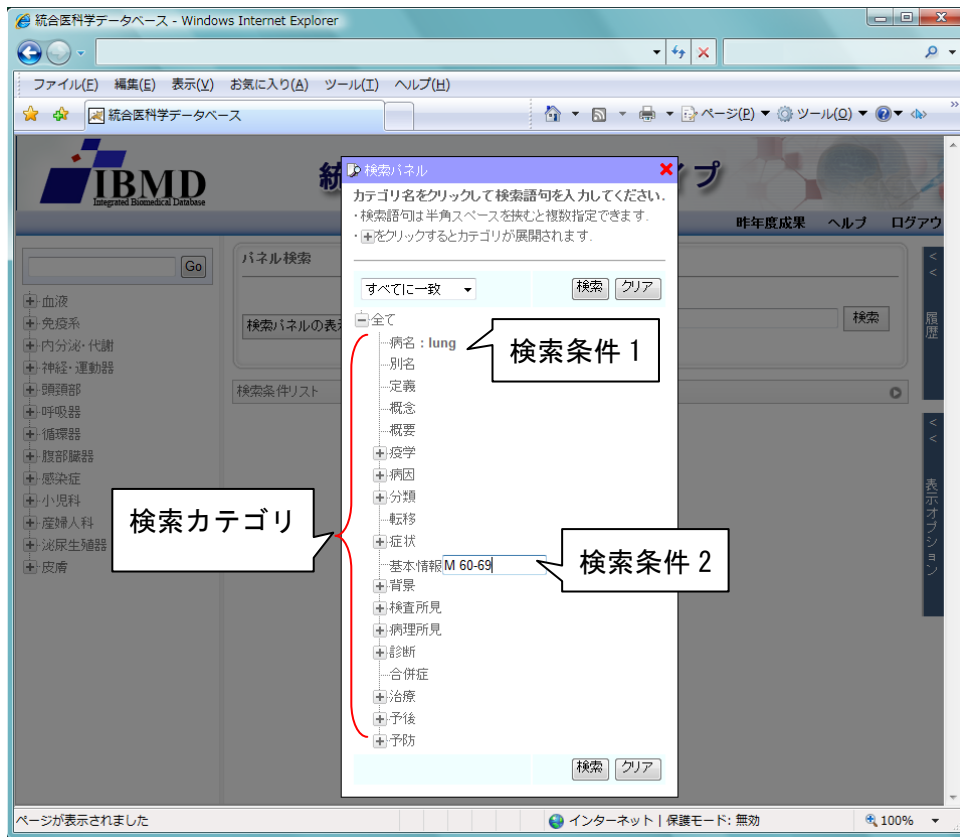


図 3 パネル方式検索

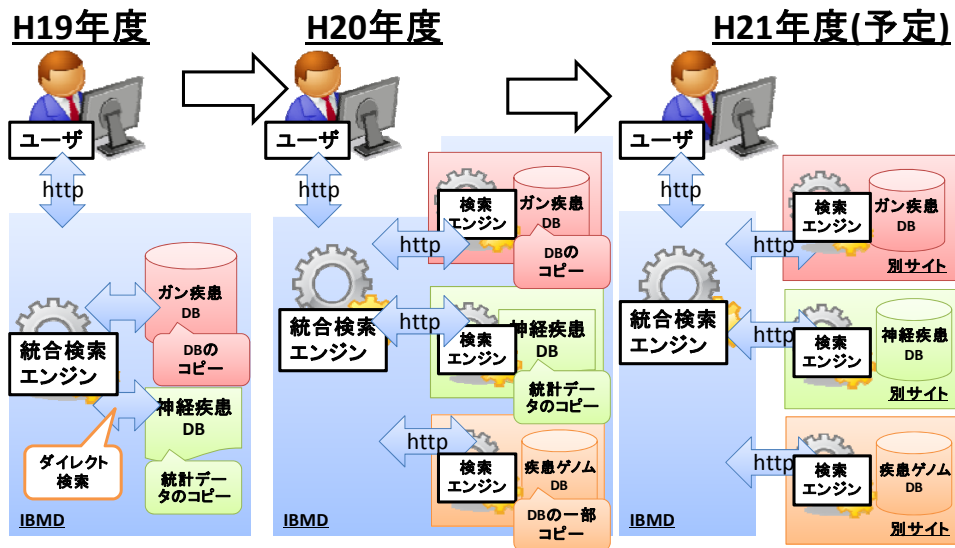


図 4 統合検索インターフェースプログラム

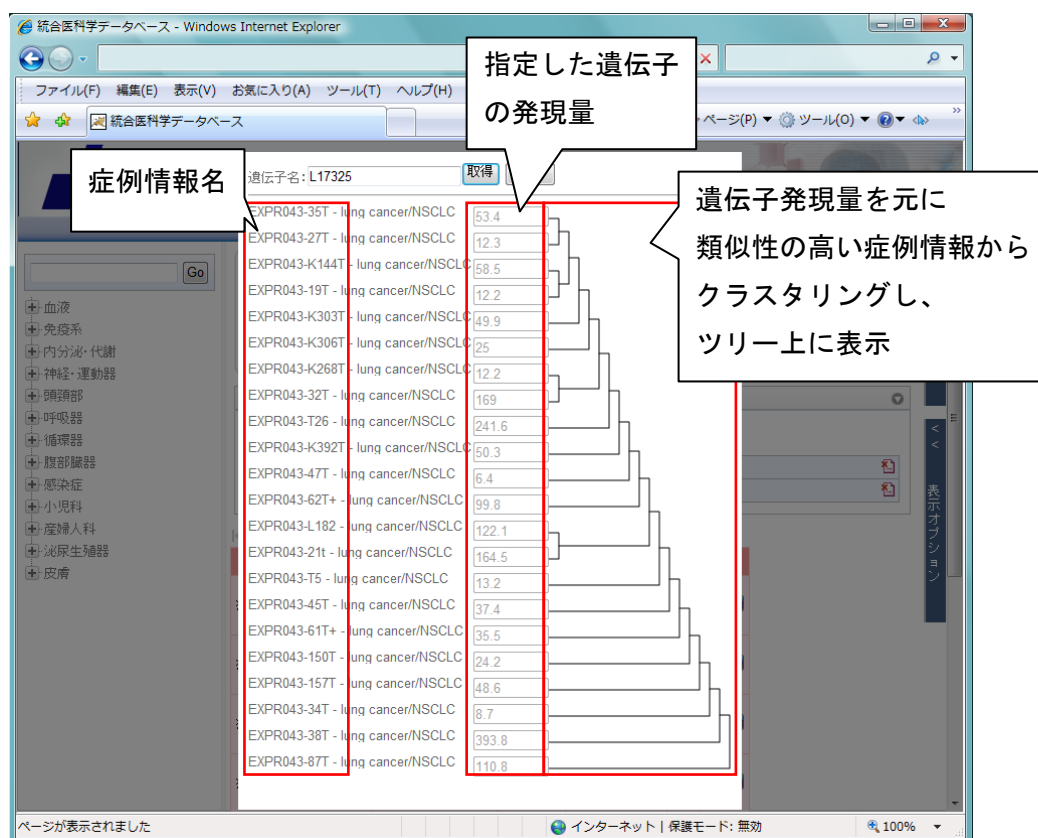


図 5 分子情報クラスタリング

③ プロジェクトの総合的推進

中間評価結果を踏まえ、疾患をがん（特に肝臓がん、大腸がん）およびパーキンソン病に限定したプロトタイプシステムを構築し、医科学データベース統合化のロールモデルを提示し、本モデルを国内に広めるための施策を策定することを、本プロジェクトの目標として再設定した。

がんデータベースの統合化モデルを効果的に提示するため、国立がんセンター研究所のGeMDBJとの連携を推進し、データの連携方法、インタフェースの実装方法について決定し（要件定義書参照）、プロトタイプとして統合を行った。

このGeMDBJや東京医科歯科大学および大阪大学の実証的なデータベース連携の過程において、分担機関内統合化要素技術検討会「2. 4 活動（運営委員会等の活動）参照（20ページ）」を開催し、検索機能やインタフェース設計に関する課題を抽出し、解決策を検討し、当初の計画通り、要件定義フェーズを完了した。次に、モデルを国内外に広めるための施策として、WHO-ICD、ISO、HL7といった国

際主要機関との連携、国際標準化を進めた。2008年5月のHL7、Phoenix会議では、HL7が手がけるFamily History Model、HL7 V3 RIMへのClinical Genomics Dataの反映の仕方、RCRIM Vocabulary専門委員会におけるオントロジーに関する検討内容、EHRと連携するClinical Research Function Profileについての検討内容など主要分野における非公開の検討内容を入手しモデルの国際的適正化のための検討を行った。また、統合データベースとGSVML、ジェノミクスデータとの連携に関する発表を行った。ポルトガル（ポルト）で行われたISMH 2008では、予防医学、東洋医学、温泉医療や統合医療を西洋医学と共存させるための主要な要因として、統合データベースを応用した社会インフラが必須である旨を発表し一定の理解と賛同を得ることができた。バンクーバーで行われた2008年度HL7総会では、セキュリティに関するいくつかの有用な非公開文献を入手することができ、臨床セキュリティ先進国である米国の実態と課題を検討することができた。そのうち、Recent Developments in Patient Safety and Scope of Patient Safety Special Interest Groupは、セキュリティに対するHL7 Patient Safety SIGとしての考え方の方向性がわかる文献であり、Secure HL7 Transactions using Internet Mailは、HL7のセキュリティに関する基本とルーツの理解に有用な文献である。HL7 Role-Based Access Control (RBAC) Role Engineering Processは、セキュリティアクセスコントロールについての新しい方向性を示唆した文献である。チュートリアルに参加して入手したHL7 V3 RIMにおけるRDBのデザインチュートリアルは、HL7 V3 RIMにRDBを連結させる場合のデータベース設計方針についてのチュートリアルであり、今後のEHRの主要な方向性の一つであるHL7 V3と統合データベースにおける主要なデータベース方式の一つであるRDBの連携を考察する上で大変有用な内容であった。クレタ島で行われたIHIC2008では、統合データベースそのものについての発表を行った。その有用性と可能性についてはドイツなどから高い評価が与えられた。しかしながら、国際的な認知度はまだまだ低く、さまざまな国際舞台での宣伝の必要性を感じさせられた。その後、イスタンブールで行われたISO TC215 WG2における検討会議で、臨床ゲノム分野で世界初、医療情報分野で日本初の国際標準（IS）としてGSVMLが認められることが正式に決まった。結果として、臨床オミックス領域世界初の国際標準（IS）としてISOにおいて認定されたGenomic Sequence Variation Markup Language (GSVML)の構造の一部に本モデルが採用される形となっているので、国際標準（IS）の一部となった。また、オントロジー、用語、疾患分類の分野では、次世代医学用語および分類の世界標準となるWHO-ICD11でも、TAG-HIM (Topic Advisory Group for Health Informatics and Modelling)、および内科専門TAG、筋骨格系専門TAGにおいて検討中のContent

Modelの参照・連携対象として、本モデルが具体的に検討されることとなった。また、倫理規定が整備されていないものの、各データベースの倫理規定の範囲内で公開可能な情報をより広く公開し、利用者からのフィードバックを得るため、ホームページ上にて研究者および医療従事者を基本対象者とし、利用希望者へのログインアカウントの発行を開始した。さらに、学会での講演（特に医療情報学会での統合医科学データベースのセッションの共同企画）や雑誌等への論文掲載を推進するとともに、昨年度の調査結果を中核機関と共有し生命科学系データベースカタログの拡充に寄与する等、本プロジェクトの認知を高める活動を実施した。

以上のように、目標に沿ってプロジェクトが進行するように、計画の策定、進捗管理および必要に応じた計画の見直しを随時行い、今年度の業務を効率的に遂行することができた。

2.2.2 大阪大学

① 疾患・臨床医科学データベースに特化した統合技術開発に関する研究

平成 21 年度に予定している疾患・臨床医科学データベースの本格統合に向けて、昨年度に試作した統合方式の暫定的なプロトタイプの開発・運用過程に得られた検討結果と、統合技術の調査結果に基づき、東京医科歯科大学と統合の要件定義に関して意見交換を実施した。図 6 に昨年度に試作した暫定的な統合方式（データ移動型統合方式）と、本年度に医科歯科大学にて作成した統合の要件定義（インタフェース型統合方式）の概要を示す。

データ移動型統合方式では、大阪大学が開発した神経疾患DBから統合検索に必要なデータを抽出し、医科歯科大学で開発したIBMD（統合医科学データベース）にデータを移動することで統合検索を実現する。この際、倫理面に配慮し、移動するデータは個人情報を含まず連結不可能匿名化が施された検索用の統計データのみとした。これにより、データを他機関に移動する際に発生する、データ提供した患者への再同意などの倫理的問題の一部を回避することができると考えられる。この点が本方式の利点となる。ただし、統合データベースの倫理規定を検討する際には、移動するデータの内容にどこまでの情報が含まれているかを精査し、一定の基準を設ける必要があると思われる。また、システム開発の観点から考えると、データ移動型統合方式では、統合の対象とする神経疾患DBのシステム拡張を行う必要がなく、統合に伴うシステム開発コストを軽減できるという利点がある。なお、移動するデータの抽出や、データを移動先のDBが対応する形式に変換する処理は必要となるが、これらの処理に必要なコストはシステム拡張に比べて低い。ただし、本方式では、元となる神経疾患DBのデータが更新さ

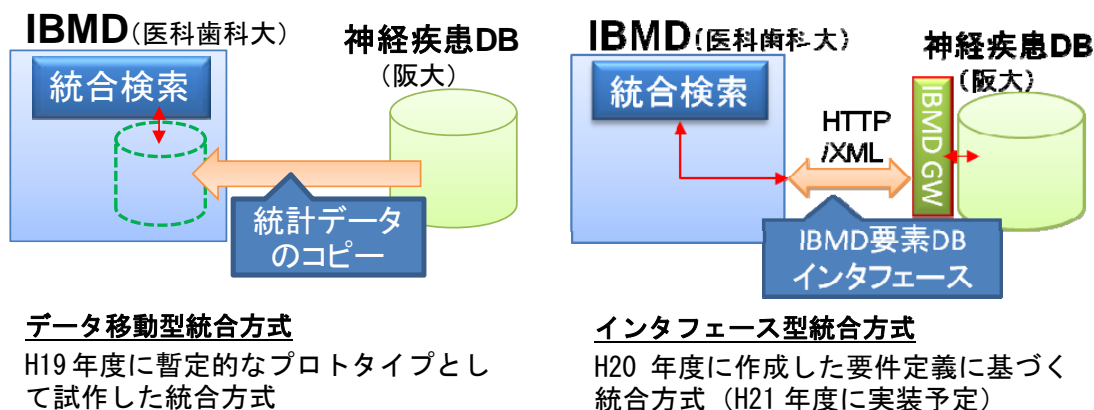


図 6 疾患・臨床医学データベースの統合方式

れる毎に、データの抽出と移動の作業が必要となるため、長期間の運用には、データ移動処理の自動化などの対応が望まれる。

一方、インタフェース型統合方式では、神経疾患DBにIBMDからインタフェ

ースを介して送られてくる統合検索用のクエリを処理する検索エンジン（IBM D GWと呼ぶ）を追加し、IBMD要素DBインタフェースに基づいて検索クエリとそれに対する回答のやり取りを行う。通信にはHTTPプロトコルによるXMLデータの送受信を利用する。この方式は、追加モジュールの開発コストが必要となるが、神経疾患DBのデータを直接検索対象とするので、DBのデータ更新毎にデータを移動させる手間をかけずに済む。ただし、外部のシステムからDBにアクセスすることになるので、アクセスできるデータの範囲やユーザの権限（ロール）管理など、倫理面を十分に考慮しておく必要がある。大阪大学で公開している神経疾患DBにおいては、従来保有していたデータのうち、患者の再同意など倫理面を考慮した上で公開可能なデータのみを登録するという方針で運営しているので、この問題は解決できると考えている。しかし、一般に各機関が保有している疾患・臨床DBとの統合を行う際には、倫理面の問題を十分に検討した上で、統合検索用のインタフェースを設計・実装すべきである。

上述のような統合技術開発に関する検討を行った上で、本年度は、昨年度と同様のデータ移動型統合方式で統合検索可能なデータの更新作業の実施試験を行った。その結果、大阪大学の神経疾患DBから東京医科歯科大学のIBMDに移動させるデータの更新は問題なく行えることが確認できた。しかし、統合データベースの継続的な運用、統合検索の高度化に対応するためのシステム改良の柔軟性などを考慮し、統合される要素データベースに統合検索用のインタフェースを実装することによって統合を実現するインタフェース型統合方式を用いることが望ましいという結論に至った。

よって、東京医科歯科大学が作成した統合医科学データベース統合のための要件定義（IBMD要素DBインタフェース）に沿って、大阪大学では神経疾患（パーキンソン病）DB統合のための要件定義（IBMD GWの仕様設計）を行った。表2に要件定義の一部を示す。この作業により、IBMD要素DBインタフェースが神経疾患DB統合のため要件定義に適用可能であることが確認された。その要件定義に沿ったIBMD GWの実装は平成21年度に実施する。

表 1. 神経疾患 DB の要素 DB インタフェース定義 (一部)

要素DB外接機能一覧								
No	機能名	インタフェース名	区分	データソース	システムフロー	Request 項目数	Result 項目数	概要
1	件数問い合わせ	CountGet()	照会	阪大 DB	External System	N	1	一覧取得resの参照元に記載する項目が検索条件に一致する臨床情報の件数を取得
2	一覧取得	SearchGet()	照会	阪大 DB	External System	N	M	検索条件に一致する臨床ID及び項目内容を取得
3	詳細情報取得	DetailGet()	照会	阪大 DB	External System	N	M	リクエストに記載された臨床IDの詳細項目内容を取得

外部システム間 I / F項目定義							
処理名称		一覧取得_res					
インタフェース名		SearchGet()					
No.	項目名	属性	長さ	必須	参照元		補足説明
					テーブル名	項目名	
1	臨床ID	文字列	8	Yes	MML臨床情報	uid	文書ユニークID
2	病名	文字列	-	Yes	詳細情報取得(病名)参照		
3	別名	文字列	-	No	なし		
4	定義	文字列	-	No	なし		
5	概念	文字列	-	No	なし		
6	概要	文字列	-	No	なし		
7	疫学	文字列	-	No	なし		
8	発生経路	文字列	-	No	なし		
24	感染経路	文字列	-	No	なし		
25	分類	文字列	-	No	MML臨床情報	mmlDh:diseaseClass	診断分類
26	肉眼分類	文字列	-	No	なし		
27	病理分類	文字列	-	No	なし		
28	転移	文字列	-	No	なし		
29	症状	文字列	-	No	MML臨床情報	mmlDh:clinicalConditio	臨床症状
30	時期別	文字列	-	No	MML臨床情報	mmlDh:developmentA	発症年齢と経過
31	重症度別	文字列	-	No	MML臨床情報	mmlDh:parkinsonDegr	パーキンソン病重症度
32	臓器別	文字列	-	No	MML臨床情報	mmlDh:autonomicinfo	自立神経情報
33	発病様式	文字列	-	No	なし		
34	基本情報	文字列	-	No	詳細情報取得(基本情報)参照		
35	背景	文字列	-	No	なし		
36	病歴	文字列	-	No	MML臨床情報	mmlDh:DiagnosticHist	診断歴情報
37	生活歴	文字列	-	No	MML臨床情報	mmlDh:life	生活状況
38	家族歴	文字列	-	No	MML臨床情報	mmlDh:family	家族歴
39	検査所見	文字列	-	No	詳細情報取得(検査所見)参照		
40	血液	文字列	-	No	なし		

外部システム間 I / F項目定義							
処理名称		詳細情報取得_res(血液)					
インタフェース名		DetailGet(血液)					
No.	項目名	属性	必須	参照元		補足説明	
				テーブル名	項目名		
1	検査時間	文字列	No	MML臨床情報	mmlh:date	検査日	
2	検査値情報	文字列	No	MML臨床情報	mmlh:TreatmentHistoryModule	検査値情報	
3	検査分類	文字列	No	MML臨床情報	mmlh:class	検査分類	
4	検査項目	文字列	No	MML臨床情報	mmlh:inspectionItem	検査項目	
5	検査値	文字列	No	MML臨床情報	mmlh:value	検査値情報	
6	単位	文字列	No	MML臨床情報	mmlh:unit	単位	

② 疾患データベースの高度化と実証的統合に関する研究

昨年度に神経疾患DBの高度化の一部として実施した、データベースの統計情報表示、経年情報表示機能などの諸機能について、利用者からの要求に基づいて拡張すべき機能を検討した。その結果、利用者の利便性を高めるために、検索結果の一覧表示や統計情報の表示方法を、より直感的なグラフ表示ができるようにすることが有効であるという結論に至り、表示方法改良の要件定義（プログラム改良の仕様策定）を行った。さらに、その要件定義に基づいて神経疾患DBに改良を加えた。図7に従来の神経疾患DBの検索結果の表示例を示す。従来の統計結

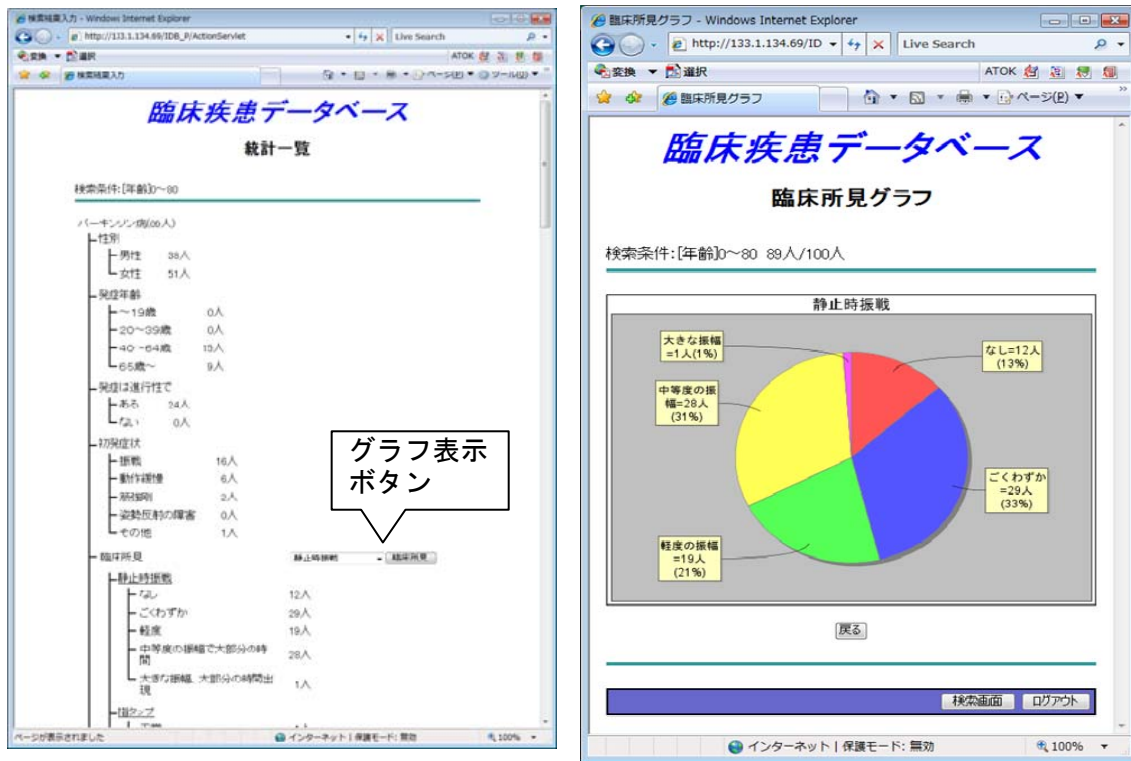


図7 神経疾患DBの検索結果の表示方法の改良

左：従来の統計結果一覧表示 右：臨床所見のグラフ表示例

果一覧表示（図7左）では検索条件を満たした症例データの統計情報を、各データの項目のカテゴリ毎に整理して数字で表示している。今年度の疾患データベースの高度化作業では、それぞれの項目について「グラフ表示ボタン」を押すことでデータの分布をグラフとして表示できるように改良した。図7右は、臨床所見のうちの「静止時振戦」をグラフ表示した例を示している。グラフ表示可能な項目は、臨床所見（静止時振戦、指タップ、筋強剛、椅子からの立ち上がり、歩行、姿勢、姿勢の安定）、臨床重症度分類（Hoehn & Yahrの臨床重症度分類）、薬剤使用割合および薬剤量の経年変化である。

表示するグラフの形式は、臨床所見は円グラフによる分布表示、臨床重症度分類および薬剤使用割合については棒グラフによる比較表示（図8）、薬剤量の経年変化については折れ線グラフによる表示（図9）とした。これにより、神経疾患DBに登録されている主だった項目はグラフで見ることができるようになったが、

未対応の項目（性別、発症年齢、初発症状など）についても、利用者の要望により必要と思われるようであれば、他の項目と同様の方法でグラフ表示できるように拡張を検討する。

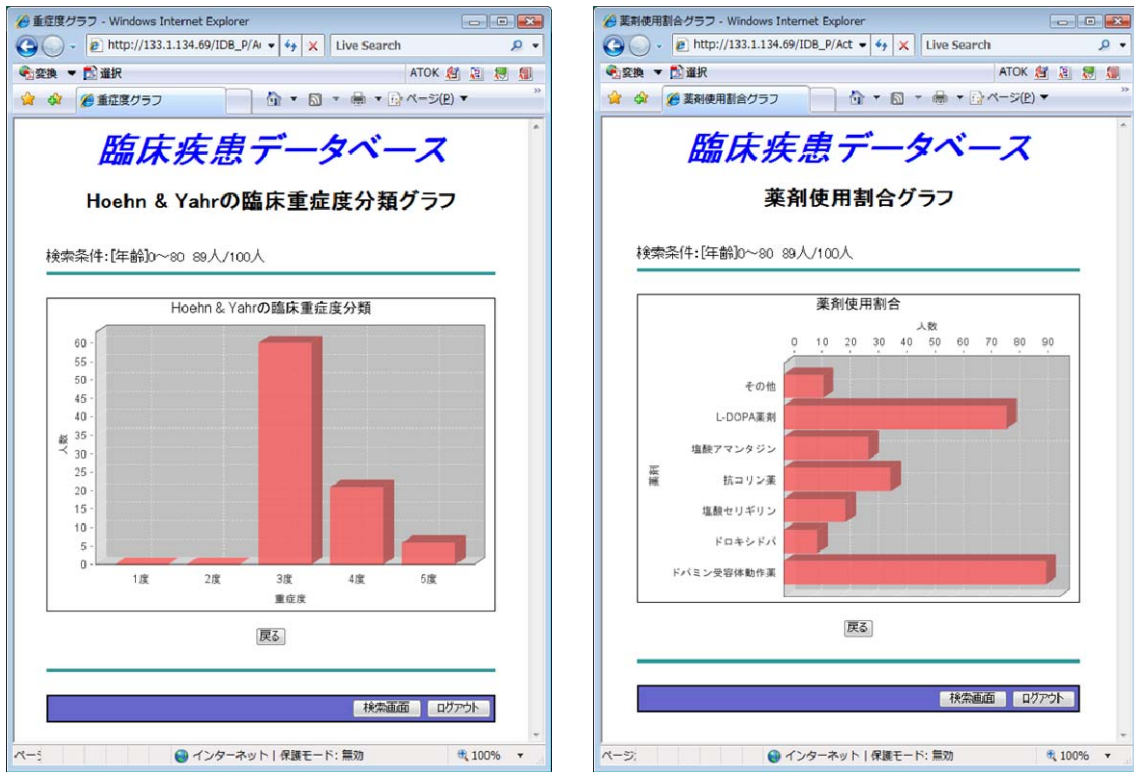


図8 臨床重症度分類（左）および薬剤使用割合（右）の棒グラフ表示



図9 薬剤量の経年変化表示

左：従来の表形式での一覧表示 右：薬剤量経年グラフ表示

またデータベースをより広いユーザに利用できるようにしたいという要望から、パ

パスワード認証なしでのDBの利用の是非についても検討した。本年度末の時点では、倫理規定の策定状況などを勘案し、パスワード認証を設けている。しかし、大阪大学で開発を進めている神経疾患DBに関しては、統計値の公開であるという点と、匿名化により各患者の詳細データにはアクセスに関してはアクセスできない仕組みを取っているという点から、パスワード認証なしで公開しても問題はないという結論に至った。それに伴い、アクセス管理の仕様変更および本プロジェクトの公開Webサイトからのリンク方法の変更などを検討した。これらの仕様変更に対応し、パスワード認証なしで神経疾患DBを公開する時期については平成21年度の上期を予定している。

また、公開データ数については昨年度に公開した100データから400データまで増加させた。なおデータベース公開の倫理規定については、平成19年度に引き続き検討・議論を行い、平成21年度以降の倫理案策定のための準備をおこなった。

2.3 成果の外部への発表

2.3.1 学会等における口頭・ポスター発表

発表した成果（発表題目、口頭・ポスター発表の別）	発表者氏名	発表した場所（学会等名）	発表した時期	国内・外の別
“Current state and future of Omics-Based medicine and systems pathobiology” : BIOMEDICAL ELECTRONICS and BIOMEDICAL INFORMATICS (BEBI'08)、口頭	Tanaka H.	Island, Greece	Aug 20, 2008	国外
「オミックス医療の実現に向けてー統合的医療データベース iCOD-」、口頭	田中 博	オミックスシンポジウム、東京	2008年10月20日	国内
「疾病解明と治療・創薬のためのシステムバイオロジー」、口頭	田中 博	神奈川科学技術アカデミー-KAST システムバイオロジー講座、横浜	2008年10月3日	国内
「生命系の構築原理と進化ー進化する分子ネットワークとしての生命」、口頭	田中 博	物性若手夏の学校、栃木	2008年8月7日~9日	国内
「オミックス医療現実の展望」、口頭	田中 博	「疾患データベースの統合に向けて」オミックス医療が拓く未来 2008、東京	2008年7月3、4日	国内
「病気をシステムで解くーオミックス医療の可能性」、口頭	田中 博	東京医科歯科大学大学院生命情報科学教育部公開シンポジウム「病気に挑む生命科学」、東京	2008年6月3日	国内
臨床データベースについて、口頭	佐古田 三郎	バイオグリッド研究会	2008年5月31日	国内
臨床データベースの問題点、口頭	佐古田 三郎	日本バイオインフォマティクス学会関西地域部会	2008年10月8日	国内

		第6回バイオメディカル研究会		
日本における臨床データベース構築および維持の問題点、口頭	佐古田 三郎	第28回医療情報学連合大会（第9回日本医療情報学会学術大会）	2008年11月24日	国内
Clinical Omics Ontology and National Projects in Japan (toward Social Information Infrastructure)、口頭	Jun Nakaya, Hiroshi Tanaka	InterOntology2008, Tokyo	Feb. 2008.	国外
疾患統合DBとオントロジー、口頭	中谷 純、佐古田 三郎、溝口 理一郎、井戸 敬介、広井 嘉栄、田中 博	第28回医療情報学連合大会、横浜	2008. 11	国内
医療情報におけるオミックス情報のインパクト — 統合データベースを踏まえて —、口頭	中谷 純、田中博	第28回医療情報学連合大会、横浜	2008. 11	国内
医学知識に基づく疾患データベースの統合検索、口頭	井戸 敬介	第4回ナノテクノロジーセンター研究会	2008. 11	国内

2.3.2 学会誌・雑誌等における論文掲載

掲載した論文（発表題目）	発表者氏名	発表した場所（学会誌・雑誌等名）	発表した時期	国内・外の別
Clinical Omics Ontology and National Projects in Japan (toward Social Information Infrastructure)	Nakaya J., Tanaka H.	“Interdisciplinary Ontology Proceedings of the First Interdisciplinary Ontology” Keio University Press.	2008	国外

		61-68		
Loss of Keap1 function activates Nrf2 and provides advantages for lung cancer cell growth.	Ohta T., Iijima K., Miyamoto M., Nakahara I., Tanaka H., Ohtsuji M., Suzuki T., Kobayashi A., Yokota J., Sakiyama T., Shibata T., Yamamoto M., Hirohashi S.	<i>Cancer Research</i> , 68:1303-9	2008	国外
Diversity of preferred nucleotide sequences around the translation initiation codon in eukaryote genomes.	Nakagawa S., Niimura Y., Gojobori T., Tanaka H., Miura K.	<i>Nucleic Acids Research</i> , 36:861-871	2008	国外
Aurora kinase B is a predictive factor for the aggressive recurrence of hepatocellular carcinoma after curative hepatectomy.	Tanaka S., Arii S., Yasen M., Mogushi K., Nguyen T., Zhao C., Imoto I., Eishi Y., Inazawa J., Miki Y., Tanaka H.	<i>British Journal of Surgery</i> , 95:611-619	2008	国外
Prediction of lymphatic	Watanabe	<i>Radiotherapy and</i>	2008	国外

metastasis based on gene expression profile analysis after brachytherapy for early-stage oral tongue carcinoma.	H., Mogushi K., Miura M., Yoshimura R. I., Kurabayashi T., Shibuya H., Tanaka H., Noda S., Iwakawa M., Imai T.	<i>Oncology</i> , 87:2237-42		
A likelihood look at the supermatrix-supertree controversy.	Ren F., Tanaka H., Yang Z.	<i>Gene</i>	2008 (accepted, in press)	国外
Non-uniform survival rate of heterodimerization links in the evolution of the yeast protein-protein interaction network.	Hase T., Niimura Y., Kaminuma T., Tanaka H.	<i>PLoS ONE</i> , 3:e1667	2008	国外
Bioinformatics and Genomics for Opening New Perspective for Personalized Care. “ <i>eHealth</i> : Combining Health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge” (B. Blobel et. al. eds.).	Tanaka H	IOS press, 47-58	2008	国外
Economic Advantage of Pharmacogenomics - Clinical Trials with Genetic Information.	Ohashi W., Mizushima H., Tanaka, H	IOS press, 585-590	2008	国外

” <i>eHealth</i> Beyond the Horizon - Get IT There” (S.K. Andersen, et.al. eds.)				
オミックス医療と Systems Pathology	田中 博	バイオインダストリー、Vol.25、No.12、33-43	2008	国内
網羅的分子病態データベースとシステム病態学—疾患のシステムの理解のための臨床オミックスデータベース	田中 博	医学のあゆみ、vol.225、No.9、765-769	2008	国内
わが国における疾患データベースの統合化	田中 博	蛋白質核酸酵素、vol.53、No.6、774-782	2008	国内

2.4 活動（運営委員会等の活動等）

分担機関内統合化要素技術検討会

平成20年 7月 3日（場所：学術総合センター）

平成20年 9月 4日（場所：大阪大学）

平成20年11月20日（場所：東京医科歯科大学）

平成20年11月27日（場所：大阪大学）

平成20年12月26日（場所：東京医科歯科大学）

平成21年 1月19日（場所：東京医科歯科大学）

2.5 実施体制

2.5.1 東京医科歯科大学

研究項目	担当機関等	研究担当者
疾患・臨床医科学データベースに特化した統合技術開発に関する研究	東京医科歯科大学 情報医科学センター	○中谷 純 野川 裕記 高井 貴子 下川 和郎 荻島 創一 井戸 敬介 茂櫛 薫 広井 嘉栄 福永 恭子
	東京医科歯科大学 難治疾患研究所	三木 義男
	東京医科歯科大学 大学院医歯学総合研究科	新村 芳人
疾患データベースの高度化と実証的統合に関する研究	東京医科歯科大学 情報医科学センター	○中谷 純 水島 洋 下川 和郎 荻島 創一 井戸 敬介 茂櫛 薫 広井 嘉栄
	東京医科歯科大学 大学院医歯学総合研究科	有井 滋樹 天笠 光雄 新村 芳人
	東京医科歯科大学 難治疾患研究所	稲澤 讓治
プロジェクトの総合的推進	東京医科歯科大学 情報医科学センター	◎田中 博 中谷 純 下川 和郎

2.5.2 大阪大学

氏名	所属	役職	役割	業務内容
佐古田三郎	大阪大学 臨床医工学融合研究教育センター (大学院医学系研究科)	教授	業務主任者	研究総括
山本洋一	大阪大学医学部附属病院 臨床試験部 自主臨床研究部門(大学院医学系研究科)	特任准教授 (常勤)	担当責任者	臨床データベース公開
溝口理一郎	大阪大学 産業科学研究所	教授	実施担当	臨床データベース高度化
下條真司	大阪大学 サイバーメディアセンター	教授	実施担当	臨床データベース高度化
濱崎俊光	大阪大学 大学院医学系研究科	准教授	実施担当	臨床データベースバリデーション
古崎晃司	大阪大学 産業科学研究所	准教授	実施担当	臨床データベース高度化
野崎一徳	大阪大学 サイバーメディアセンター	教務職員	実施担当	臨床データベース高度化
安井涼子	大阪大学臨床医工学融合研究教育センター(大学院医学系研究科)	技術補佐員	実施担当	研究補助

2.6 整備実績一覧【代表機関名：東京医科歯科大学】

(1) 保有データ情報

(1-1) データの種類

①生物種	(1) 東京医科歯科大学：ヒト (2) 大阪大学：ヒト
②試料・ライブラリー等の種類、数	(1) 東京医科歯科大学：＜試料＞肝臓癌、大腸癌、口腔癌の手術または生検検体（約350） ＜データ＞肝臓癌、大腸癌、口腔癌の症例情報（約400）、 肝臓癌、大腸癌、口腔癌検体の遺伝子発現解析結果（約200） (2) 大阪大学：＜データ＞神経難病の症例情報（約400）
③測定方法	(1) 東京医科歯科大学：診療情報収集、DNA マイクロアレイによる遺伝子発現解析 (2) 大阪大学：診療情報収集大阪大学：診療情報収集
④データの内容	(1) 東京医科歯科大学：臨床情報（基本情報、病歴・生活歴、臨床検査、画像診断、治療、予後等） 分子情報（マイクロアレイ遺伝子発現情報） (2) 大阪大学：臨床情報（基本情報、症状、治療等）
⑤その他、特記事項	

(1-2) データソース

①現在のデータ量	(1) 東京医科歯科大学：試験公開可能なデータは、約250症例分の症例情報。 非公開データを含めると、約400のがん症例情報を有する。 (2) 大阪大学：試験公開可能なデータは、約400の症例分の統計情報。 非公開データを含めると、約400の神経難病症例情報を有する。
②データ区分	■自前 □第三者 □文献データ ■計算結果等の二次データ □その他（下欄に詳細を記述） 倫理審査委員会が承諾する範囲内で、今後も継続的に症例情報を追加していく。
③将来の増加の見込み	所有者（東京医科歯科大学、大阪大学） 公開（□可 □否 ■その他〔各機関の倫理審査委員会が承諾する範囲での公開〕）
④権利関係	所有者（東京医科歯科大学、大阪大学） 公開（□可 □否 ■その他〔各機関の倫理審査委員会が承諾する範囲での公開〕）

⑤その他、特記事項	個別に試験公開可能なデータの公開をしている。(http://ibmd.tmd.ac.jp) 各データベースの倫理規定に基づいての公開を行うとともに、統合医科学データベースの倫理規定案の策定を推進する。
-----------	---

(1-3) データの管理状況

①更新頻度等の管理状況、体制	症例情報の収集及び検体からの遺伝子発現解析は継続的に実施しており、集積データをクレンジングし、公開用データベースへのデータ更新は年1回～2回を予定している。
②その他、特記事項	

(1-4) データベース関係

①DB管理者数	(1) 東京医科歯科大学：3名 (2) 大阪大学：1名
②キュレーター・アナレータ数	(1) 東京医科歯科大学：3名(CRC他) (2) 大阪大学：1名
③データ構造	各データベースで独自のデータ構造
④DB管理ソフト	PostgreSQL
⑤サーバのOS	Linux
⑥サーバ規模	
⑦DBへのアクセス数	
⑧独立IP数	1
⑨その他、特記事項	DBへの検索メニューは報告書を参照、オントロジーについては東京医科歯科大学が独自に作成したものを使用。

(2) データ(又はDB)の連結、統合化整備(※試験的、限定的公開済みのものも含む。)

通番	データ(又はDB)の名称	公開／未公開	概要(データの種別(生物種)・数量(kB等)、本プロジェクトで実施した特徴点、進捗状況、今後の計画・課題などを簡潔にわかりやすく記述)
1	網羅的疾患分子病態データベース http://ibmd.tmd.ac.jp	試験公開	臨床、病理、分子情報(遺伝子発現情報)を統合化したデータベース。 疾患共通のテンプレートに情報を適応し、250症例のデータを公開した。 最終年度までに300症例のデータ公開を行う予定。
2	パーキンソンデータベース http://ibmd.tmd.ac.jp	試験公開	神経難病特の、臨床情報データベース。 400症例の臨床10項目に対して、統計情報を取りこれを公開した。 最終年度までに500症例のデータ公開を行う予定。

			定。
--	--	--	----

(3) DB基盤システム、ツール等開発成果物の整備（※試験的、限定的公開済みのものも含む。）

通 番	DB基盤システム、ツール 等の名称	公開 ／未 公開	概要（主な機能・特徴点、進捗状況、今後の計画などを簡潔にわかりやすく記述）
1	要件定義書	プロ ジェ クト 内部 公開	統合医科学データベース構築に関わる統合化技術（情報モデル、オントロジー、セマンティクス）、標準化、公開倫理等に関して、全国の全疾患DBを対象とした調査に基づき、要件定義書を完成した。
2	統合検索システムプロトタイプ http://ibmd.tmd.ac.jp/	試験 公開	東京医科歯科大学の網羅的疾患分子病態データベース、大阪大学のパーキンソンデータベース、国立がんセンター研究所のGeMDBJに対しセマンティクス検索技術、ユーザインターフェースを検証するためのプロトタイプシステムを構築した。 機能限定版のプロトタイプシステムを評価用に試験公開する予定。

(4) その他の成果物（(2)、(3)に該当しないもの）

通 番	名称	公開 ／未 公開	概要