

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
19年度 研究成果報告書

平成20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、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が共同で実施した、平成19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。

従って、本報告書の著作権は、文部科学省に帰属しており、本報告書の全部または一部の無断複製等の行為は、法律で認められたときを除き、著作権の侵害にあたるので、これらの利用行為を行うときは、文部科学省の承認手続きが必要です。

1 委託業務の目的

「統合医科学データベース構築方式の開発」

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19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

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

我が国のライフサイエンス研究を支える社会情報基盤としての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）の要件分析、標準化、公開倫理などの検討に着手し、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を中心として、研究を実施する。

疾患データベースの高度化と実証的統合に関する研究

各疾患DB（がん疾患、神経疾患）の高度化を行いつつ、統合技術開発における具体的要求と課題を検討することに着手し、要件定義を行う。

国立大学法人東京医科歯科大学は、がん疾患を中心に、国立大学法人大阪大学は神経疾患を中心に、実施する。

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

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

我が国のライフサイエンス研究を支える社会情報基盤としての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）の要件分析、標準化、公開倫理などの検討に着手し、要件定義を行う。

国立大学法人大阪大学は神経疾患を中心として、研究を実施する。

疾患データベースの高度化と実証的統合に関する研究

各疾患DB（がん疾患、神経疾患）の高度化を行いつつ、統合技術開発における具体的要求と課題を検討することに着手し、要件定義を行う。

国立大学法人大阪大学は神経疾患を中心に、実施する。

2.2 実施内容（結果）

2.2.1 東京医科歯科大学

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

我が国のライフサイエンス研究を支える社会情報基盤としての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うため、全国に存在し統合化の対象となる疾患データベースの構築・利用状況についての調査を、予備調査と詳細調査の2つのステップに分けて実施した。予備調査では、旧帝国大学および旧制六医科大学の医学部、計13の医学部を対象として、外部公開情報を収集した。収集した情報を分析し、仮説の検証及び次の大規模調査実施に向けた調査方法の設計を実施した。詳細調査では、日本医学会の加盟分科会102学会、及び国公立大学付属・国公立の300病院を調査対象として大きく5つのカテゴリー（表1）に関する調査を実施した。

表1 詳細調査項目のカテゴリー

1	データベース概要 データ件数、内容、方式、データ形式、アクセス件数など
2	公開ポリシー 公開基準、倫理規定、個人情報保護基準など
3	匿名化

	上記個人情報保護基準に準じた具体的方法など
4	拡張性 多施設連携性、標準化基準（ISO,HL7等）など
5	オミックスデータ連携性 オミックスデータの有無、種類など
6	データベース詳細 内容、構造、項目など

調査の実施方法としては、ウェブ公開情報に基づく調査、アンケート票郵送による調査、Web調査票によるオンライン調査を組み合わせ、必要に応じてメール・電話によるより詳細な情報のヒアリング調査を実施し、回答率の向上を図った。

詳細調査の結果、60件（有効回答率15%）の有効回答が得られ、予備調査を含め調査対象機関に572の臨床医学系データベースが存在することが判明した。その中から、症例情報を備え、適切な匿名化の処置によって個人情報が保護され、データの正規化がなされており、かつ調査期間中に運用（公開/非公開は問わず）されていた45のデータベースを統合化の対象候補として選定した。

また、今後検討する統合医科学データベース倫理案策定の下準備として、臨床医学系データベースのポリシー（公開基準、倫理規定、個人情報保護基準）、匿名化（個人情報保護基準に準じた具体的方法）の調査結果を多角的な観点から整理し、ヒト症例データベース公開における社会倫理課題を可能な限り網羅的に抽出・検討し、次年度の実行計画を策定した。

上記調査結果に基づき、医科学データベースを統合するための要求分析を行い、統合医科学DB利用ガイドシステムを構築するための要件定義を行った。機能要件、利便性、実現性、性能、保守性（拡張性）を検討し、統合医科学DB利用ガイドシステムの基本設計（図1）を決定した。

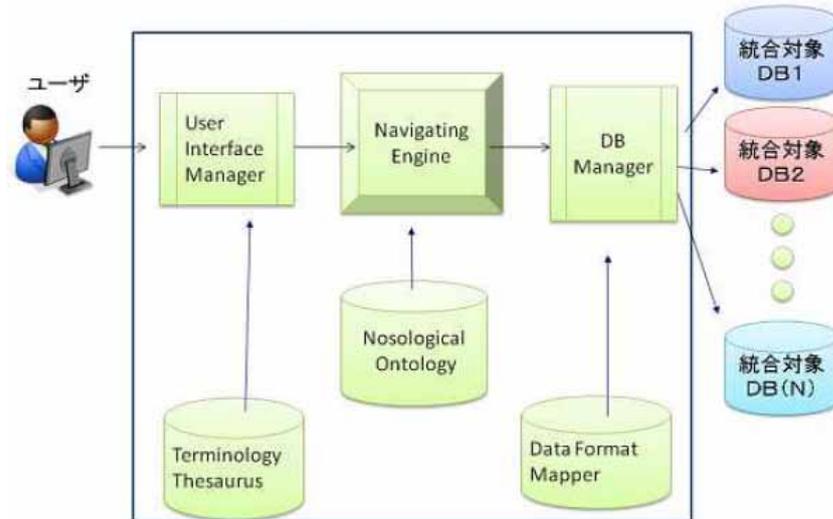


図1 統合医科学DB利用ガイドシステムの基本設計

また本基本設計に基づき、セマンティック検索を実現するためのデータベースモデルとして、データベース3階層モデルを設計した（図2）。

3 Level Method: データベース3階層モデル

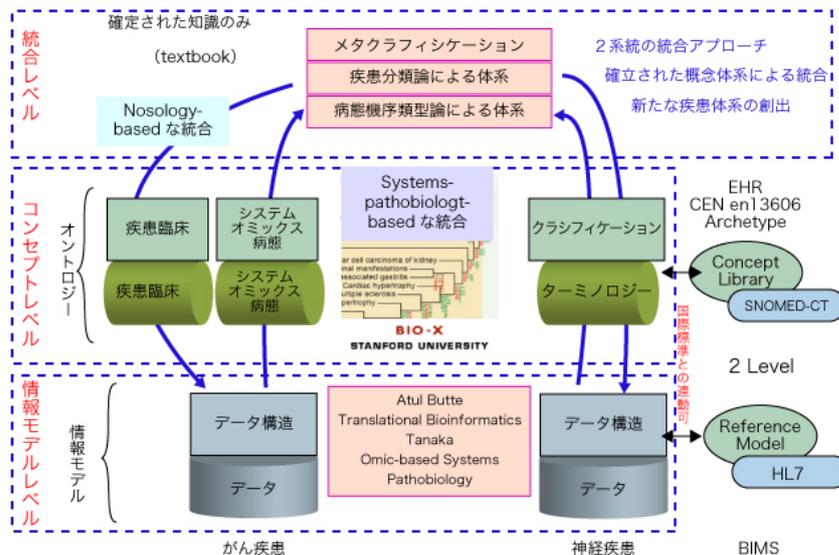


図2 データベース3階層モデル

全疾患に適応可能な疾患DB情報モデル（共通テンプレート）を決定した（表2）。東京医科歯科大学のがん疾患DB、大阪大学の神経疾患DBを高度化し実証的統合を行うために、がん疾患・神経疾患のオントロジーの整備も行った。

表2 全疾患共通テンプレート（一部抜粋）

病名
別名
定義
概念
疫学
病因
分類（肉眼分類、病理分類）
症状（時期別、重症度別、臓器別）
検査所見（血液、尿、ECG、画像診断）
病理所見（肉眼所見、組織所見）
診断
合併症
治療（内科的治療、外科的治療、放射線治療）
予後
予防

疾患データベースの高度化と実証的統合に関する研究

東京医科歯科大学はがん疾患DBとして網羅的疾患分子病態DBに格納されているデータ項目を疾患DB情報モデルに適用させるとともに、試験公開に向けて100症例のデータの整備を実施した。

東京医科歯科大学で高度化した網羅的疾患分子病態DBと大阪大学で高度化したパーキンソン病DBとを各々試験公開するため、一つのサイトから両者に自動ログインできる形の統合医科学データベースのポータルサイトを東京医科歯科大学に設置し試験公開を開始した（図3）。（<http://ibmd.tmd.ac.jp/>）

さらに、統合医科学DB利用ガイドシステムの具体的要求と課題検討を検討するため、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を構築し（図4）、これも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。

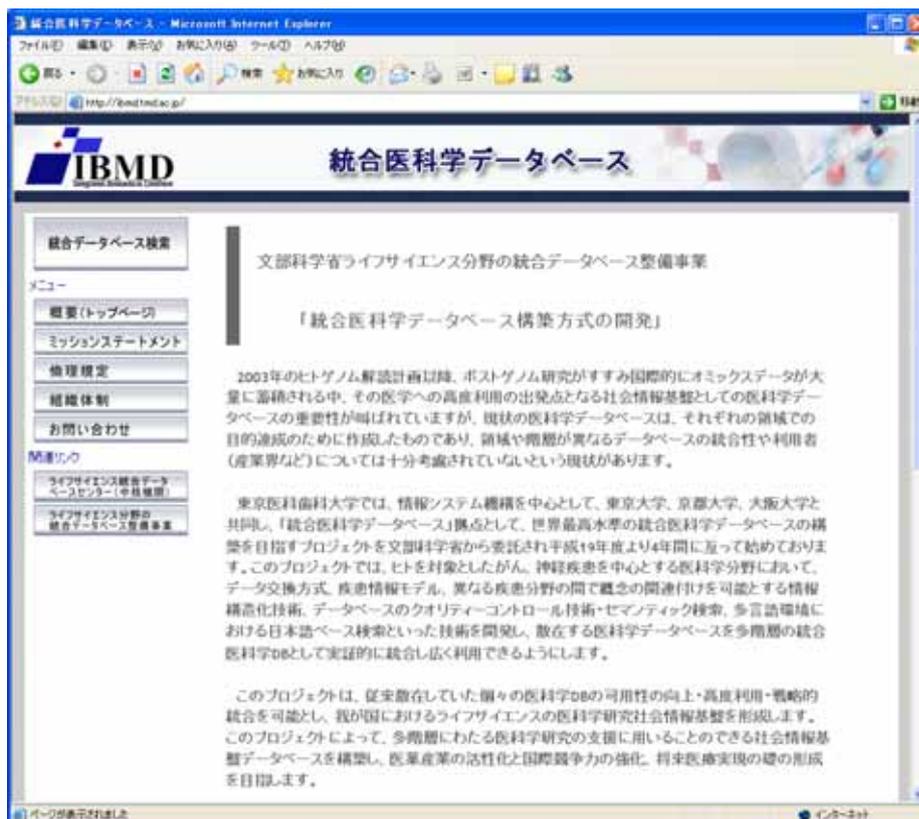


図3 統合医科学DBポータルサイト

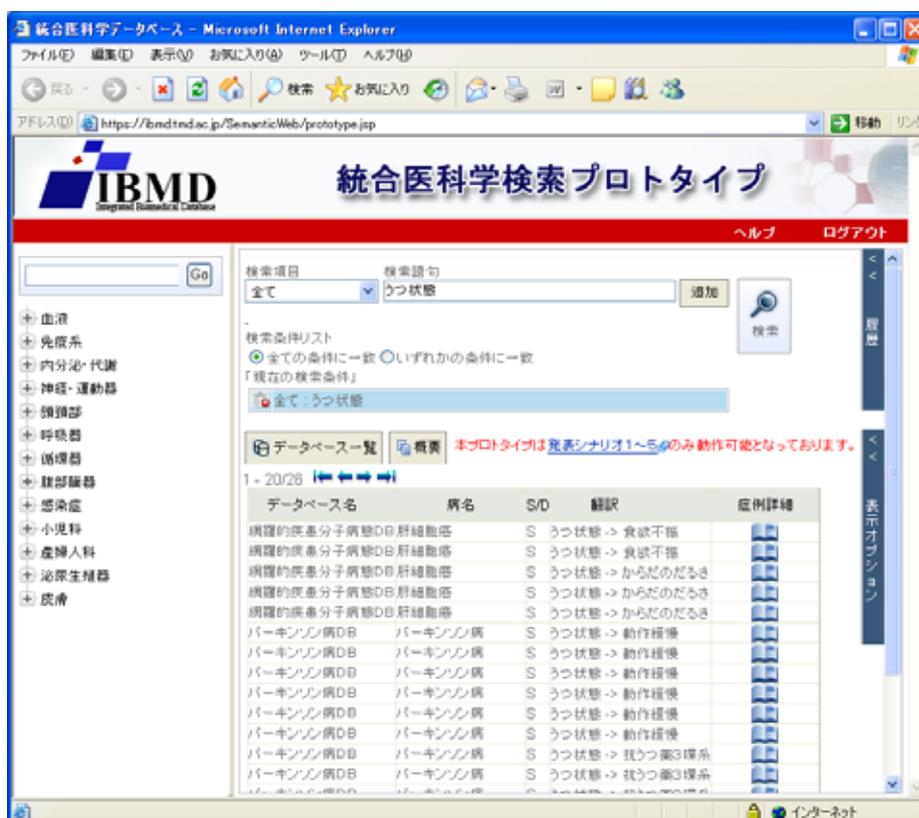


図4 プロトタイプシステム

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、対象を2疾患から全疾患へ拡大するなど、中核機関である情報・システム研究機構の戦略に従い連携して本事業を推進している。

また、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、東京医科歯科大学では月4回程度、大阪大学では6月、8月、11月、1月にマイルストーンを設定し、進捗を管理すると同時に、局面化開発手法に基づいて内容の検証、方向性の確認及び修正、問題点の把握と解決を行った。さらに、統合化要素技術の開発と標準化、及び倫理規定草案の検討会を、9月、10月、1月に東京医科歯科大学と大阪大学と合同で開催し、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った。

2.2.2 大阪大学

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（神経疾患を中心として）

大阪大学が中心となって構築しつつある、神経疾患DB（パーキンソン病DB）は、大阪大学医学部附属病院神経内科脳卒中科ならびに大阪府下の複数の病院の神経内科（表3）に受診している患者データを蓄積している。特定疾患の申請のために保健所などに提出される「パーキンソン病関連疾患臨床個人調査票」の内容に加え、カルテ番号、データとしての利用されることについての同意の有無、服用薬剤の種類と投与量、合併症などを患者データとして入力する。個人を特定できる個人情報と臨床情報とを区別し、臨床情報のみが連結可能匿名化IDを付してデータ管理者である特定非営利活動法人 臨床研究教育支援センター（SCCRE）のサーバコンピュータに保存される。今後、我が国のライフサイエンス研究を支える社会基盤としての多階層の統合医科学DBの構築する一環として、パーキンソン病患者の臨床情報を公開することを当面の目標としている。そのために、さらに連結不可能な匿名化IDを付した臨床情報を公開するための技術的手法を検討する必要がある（図5）。

表3 パーキンソン病DB患者データの収集病院一覧

国立病院機構刀根山病院神経内科、大阪厚生年金病院神経内科、大手前病院神経内科、公立学校共済組合近畿中央病院神経内科、兵庫医科大学内科（神経・脳卒中科）、大阪府立急性期・総合医療センター神経内科、東大阪市立総合病院神経内科、箕面市立病院神経内科、市立豊中病院神経内科、市立堺病院神経内科、多根総合病院神経内科、岸和田徳州会病院神経内科、大道会森之宮病院

統合医科学DBの構築に向けて

(公開用臨床データ変換システムの開発)

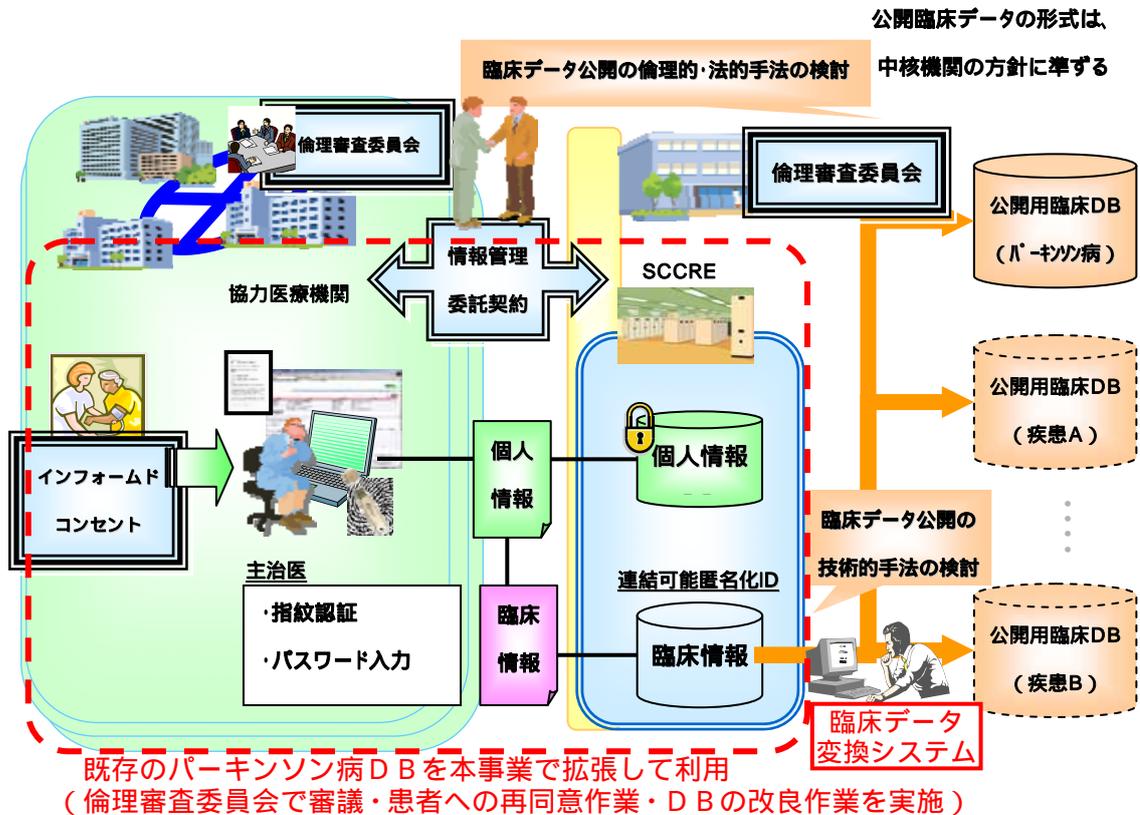


図5 公開用臨床データ変換システム

このパーキンソン病データベースの内容を統合医科学データベースへリンクさせるために、SCCREおよび各病院の倫理委員会での承認が必要であった。提出書類を整備し、SCCRE、大阪大学医学部附属病院でそれぞれ承認を受けた。さらに他の各病院でも現在倫理委員会承認待ちの状態である。パーキンソン病DBに登録予定の症例数は、承認済み・承認待ちの症例全てを含め、現時点で535件である。承認を受けた施設から順次パーキンソン病DBの個人情報を除いた臨床データを公開していく予定である。パーキンソン病患者の臨床データの公開については、SCCREから100症例分のデータの提供を受け、公開した。また今後の統合医科学DBシステムの確立を目指しつつ、パーキンソン病DB患者データの公開に向けて、ソフトウェアおよびサーバハードウェアなどを整備した。

疾患データベースの統合技術を開発するにあたり、パーキンソン病DBに格納されているデータ項目および検索項目を分析した。表4に神経疾患DBの検索項目および各項目に設定可能な検索条件を示す。

表4 神経疾患DBの検索項目

検索項目	設定可能な検索条件
基本情報	
・性別	男性，女性
・年齢	n歳～m歳（n，m：正数）
臨床情報	
・重症度	1度、2度、3度、4度、5度
・発症年齢	～19歳，20～39歳，40～64歳，65歳～
・進行性	発症は進行性である，発症は進行性でない
・初発症状	振戦，動作緩慢，筋強剛，姿勢反射の障害，その他
治療情報	
・薬剤	15種の薬剤各々について使用・未使用・過去に使用
・症状の日内変動	あり，なし，不明
・ジスキネジア	あり，なし，不明
・精神症状	あり，なし，不明

データベースを統合する際には、複数のデータベース間でこのような検索項目の対応関係を統合的に管理する必要がある。そこで本事業では、東京医科歯科大学で検討が進められている疾患DB情報モデルおよび疾患DBオントロジーへの適応可能性を検討した。具体的には、表2の全疾患共通テンプレートと表4の各項目を比較することで、現状の神経疾患DBの検索項目が疾患DB情報モデルへの対応状況と統合技術開発に必要な要件を検討した。その結果、臨床情報（症状）や治療情報など大項目レベルでの対応が取れた。

疾患データベースの高度化と実証的統合に関する研究（神経疾患を中心として）

大阪大学は神経疾患DBとしてパーキンソン病DBについて、公開用臨床データ変換システム（図5）および臨床データ高度検索システム（図6）を開発し、データベースの高度化を行った。本年度の高度化としては、「複合検索条件に対する統計情報提示機能」と「検索結果の時系列変化提示機能」を開発した。

公開用臨床データ変換システムは、データの匿名性と患者の状態変化を示す変更履歴データの整合性を確保した公開用臨床データへ変換するプロトタイプソフトウェアとデータ管理元で公開するデータ項目のフィルタリングを適宜設定できるプロトタイプソフトウェアとで構成される。標準データ形式調査、適用方式検討、データベース設計、データベース実装およびシステム試験・導入を実施した。具体的には、データの匿名性の観点から公開DBに採用すべき標準データ形式

を調査し、SCCREが持つパーキンソン病DBのデータ項目と比較してデータの変換方式を検討した。さらに、検討内容に基づいて公開用のパーキンソン病DBの設計およびプロトタイプの実装を行った。公開用DBには、今後のデータ拡張性を考慮してXMLデータベースを使用した。そして、開発した利用データ変換システムのプロトタイプを用いて変換したデータの一部を公開用DBに試験的に格納し、動作確認を行った。

臨床データ高度検索システムでは、検索ポータルを整備し、疾患臨床データベース検索手順ソフトウェアの開発を行った。検索手順ソフトウェアは、東京医科歯科大学が開発した統合医科学データベースのポータルサイトから、神経疾患データベースにアクセス認証付きで遷移しデータの送受信を行うための手順（プロトコール）を提供する。具体的には下記のような流れでデータの送受信およびページの遷移が行われる。

ポータルサイトから神経疾患 DB へのリンクを選択し、ログイン情報を臨床疾患データベースへ送信する。

神経疾患 DB にて正常にログイン情報を受信した場合、認証処理を行う。

認証処理が正常に行われた場合、ブラウザに検索条件および検索手順の入力画面が表示される。

さらにパーキンソン病DBの検索に特化した検索コアエンジンを実装し、東京医科歯科大学が開発した統合検索プロトタイプシステムとの連携を実施した。

臨床データ高度検索システム

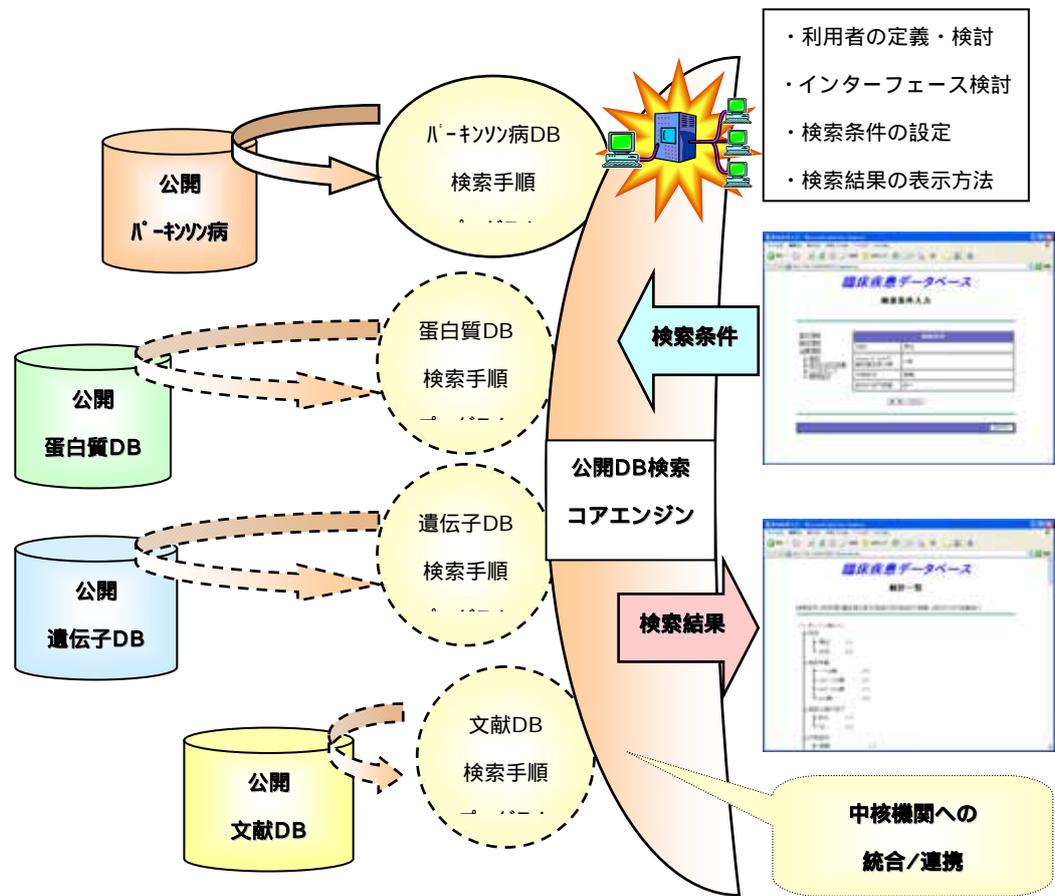


図6 臨床データ高度検索システム

パーキンソン病治療薬は従来のLDOPA製剤などに加え、新規のドパミン受容体アゴニストやCOMT阻害剤など保険診療の範囲内で処方できる薬剤の種類が現在も増えつつある。これに併せて、上記の神経筋疾患DBに入力および検索できる薬物の種類を漸次改変している（図7）。

2.3 成果の外部への発表

2.3.1 論文寄稿

論文タイトル	発表者名	発表誌名	巻	号	ページ	掲載年月
オミックス情報と医療情報の網羅的統合データベースの構築	水島 洋、 田中 博	計測自動制御学会システム情報部門学術講演会 2007 講演論文集			39-40	2007/11
TMDU Clinical Omics Database System - Integrating OMICS data and Clinical Information.	Mizushima H., Arii S., Sugihara K., Miki Y., Inazawa J., Tanaka H.	The 7th International Workshop on Advanced Genomics Abstract book.			82	2007/11
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	53	6	774-82	2008/4
Aurora kinase B is a predictive factor for aggressive recurrence of hepatocellular carcinoma after curative hepatectomy	Tanaka S, Arii S, Yasen M, Mogushi K, Nyuyen ST, Zhao C, Imoto I, Eishi Y, Inazawa J, Miki Y, Tanaka H	Br J Surg.	95	5	611-9	2008/5
Bioinformatics and Genomics for Opening New Perspective for Personalized Care.	Tanaka H.	Stud Health Technol Inform.	134		47-58	2008/5
網羅的分子病態データベースとシステム病態学	田中 博	医学のあゆみ	225	9	7135-39	2008/6

2.3.2 講演

講演タイトル	発表者名	講演会名	発表年月日
文部省ライフサイエンス統合データベースプロジェクトと臨床オントロロジー	田中 博	医療知識基盤データベースと用語・言語・知識処理シンポジウム	2008.3.13
Bioinformatics and genomics for opening new perspectives for personalised care	Hiroshi Tanaka	CeHR 2007	2007.12.3
TMDU Clinical Omics Database Project - Integration of OMICS data and Clinical Information.	Hiroshi Mizushima, Hiroshi Tanaka	The 7th International Workshop on Advanced Genomics	2007.11.27
網羅的臨床情報と網羅的分子情報の統合データベースの構築	水島 洋, 田中 博	第 27 回医療情報連合大会	2007.11.24

2.3.3 データベース

データベース名	内容	URL
統合医科学データベースポータルサイト	試験公開。要件定義書及びセミナーティック検索システムのプロトタイプを公開中。	http://ibmd.tmd.ac.jp/
網羅的疾患分子病態データベース	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、100 症例のデータを統合医科学データベースポータルサイトより公開中。	http://ibmd.tmd.ac.jp/
パーキンソンデータベース	神経難病特の、臨床情報データベース。100 症例の臨床 10 項目に対して、統計情報を取りこれを統合医科学データベースポータルサイトより公開中。	http://ibmd.tmd.ac.jp/

2.4 活動

分担機関内統合化要素技術検討会

平成19年9月19日（場所：大阪大学）

平成19年10月4日（場所：東京医科歯科大学）

平成20年1月25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 高井 貴子 野川 裕記 広井 嘉栄 荻島 創一 井戸 敬介
	東京医科歯科大学 難治疾患研究所	三木 義男
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 広井 嘉栄 荻島 創一 井戸 敬介
	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
	東京医科歯科大学 難治疾患研究所	稲澤 譲治
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
	東京医科歯科大学 大学院医歯学総合研究科	新村 芳人

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
須貝文宣	大阪大学 大学院医学系研究科	特任助教	担当責任者	臨床データベース 公開
戸田達史	大阪大学 大学院医学系研究科	教授	実施担当	パーキンソン SNP データベース
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	助教	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
佐竹 渉	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)	特任研究員	補助者	パーキンソン SNP データベース
桑原世輝	大阪大学 大学院情報科学研究科		補助者	統合医科学データベース構築方式の開発に関するデータ収集、入力作業補助
市川晃平	大阪大学 大学院情報科学研究科		補助者	"
武田伸悟	大阪大学 大学院情報科学研究科		補助者	"
加登聡	大阪大学 大学院歯学研究科		補助者	"
田峰謙一	大阪大学		補助者	"

	大学院歯学研究科			
小松一貴	大阪大学 基礎工学部		補助者	"
田坂優一	大阪大学 大学院基礎工学研究科		補助者	"
岩橋亮平	大阪大学 基礎工学部		補助者	"
竹内康浩	大阪大学 基礎工学部		補助者	"