

機関名	本年度成果目標	最終成果目標	
中核機関	DBCLS	<p>1. 戦略立案・実行評価</p> <ul style="list-style-type: none"> ・知財・法律関係: サイトポリシー・著作権方針を策定。科学データ所有権、知財ポリシーに対応する見解提示 ・日本語文献対応: 総説誌全文検索の対象拡大(1誌追加)。学協会著作物の扱いに関する方針策定 ・調査関係: 新型シーケンサ、医用リシーケンスへの対応、人材養成ニーズに関する調査報告資料を作成 <p>2. 統合データベース開発</p> <p>①共通基盤技術開発</p> <ul style="list-style-type: none"> ・遺伝子名、蛋白質名をPNEの記事に対応付ける技術を開発し、検索システムに組み込み、公開 ・遺伝子名、蛋白質名でfull paperに対応付ける技術を開発し、特定生物を対象としたプロトシステムを完成 ・TogoWSの操作性向上のためのツール組合せインターフェイス設計のための調査を行い、報告資料を作成 <p>②ヒト統合DBの開発・運用</p> <ul style="list-style-type: none"> ・文献解読: 論文正規化情報の取り出しと2項関係エディターの開発 ・遺伝子名辞書の高度化と辞書のDB化 ・アナトモグラフィ・ボディパーツ3Dの高度化 ・Wired-PDFの開発と普及化 ・バンク目次: DNA DB総覧と遺伝子発現バンク(GEO)の維持更新(3か月毎)と次世代シーケンサデータ受入れ体制整備 <p>③モデル生物・産業応用生物統合DBの開発・運用(微生物関係)</p> <ul style="list-style-type: none"> ・オートアノテーション用パイプラインの構築: 標準タイプ決定(9月)後、仕様を決めて開発 <p>3. 統合データベース支援</p> <p>①ポータル整備・運用、広報、普及啓発(ポータル整備運用)</p> <ul style="list-style-type: none"> ・生命科学系DBカタログの拡充(150DB追加)と記載内容の充実 ・生命科学DB横断検索の拡充(目標200DB達成)と検索精度の向上 ・コンテンツの拡充: 新聞記事、特定研究報告書、総説誌などがターゲット(広報) ・学会・展示会展示、シンポジウムの開催(年間3回) ・広報素材: パンフレット作成、ニュース配信の実施 ・ユーザからのフィードバック: サービス内容に関するアンケートの実施と結果のサービス内容への反映(普及啓発) ・統合TV: 100件開発(トータル150件) ・講習会: 6回開催 <p>②データベースの受入と運用</p> <ul style="list-style-type: none"> ・基盤づくり: メタデータ仕様作成、標準使用許諾作成、一括ダウンロードサイト構築、運用 ・補完課題、関連機関及び個別プロジェクトDBを対象に年間10DBを受入れ、公開 ・蛋白質に関する国内外のDB統合化のための内容検討と仕様書の作成 	<p>1. 戦略立案・実行評価</p> <ul style="list-style-type: none"> ・生命科学分野におけるデータ共有のあり方に関する考え方を提示する。 ・データ提供を促進する方策の立案し、実行する。 ・データベース統合に関わる著作権、個人情報に関わる問題点の整理と解決策の立案し、実施する。 ・統合データベースに収載すべきコンテンツの検討と収載策の立案し、実行する。 <p>2. 統合データベース開発</p> <p>(1)共通基盤技術開発</p> <ul style="list-style-type: none"> ・同義語対応や検索結果の重みづけ等を付加した高度な検索技術を実現し、サービスに適用する。 ・統合データベースの利用およびアノテーションに関わる解析手順のワークフロー化を実現し、公開する。 ・辞書、オントロジーの構築、キュレーション支援のための専門用語解析・獲得技術を実現し、辞書整備に活用する。 <p>(2)ヒト統合DBの開発・運用</p> <ul style="list-style-type: none"> ・ヒト統合DB構築に必要な辞書、オントロジーを整備し、公開する。 ・キュレーションに必要な文献解読システム及びツール群を整備し、DB構築に活用する。 ・配列データ、発現データ、多型データ、構造データ等の統合サイトを構築し、公開する。 <p>(3)モデル生物・産業応用生物統合DBの開発・運用</p> <ul style="list-style-type: none"> ・植物及び植物関連微生物については、コミュニティ参加型のゲノムアノテーションサイトを構築し、高等動物、植物関連微生物を網羅した文献からのキュレート情報を集積し、公開する。 ・産業応用微生物については、オートアノテーション用のパイプラインを構築し、ゲノム解読された全ての微生物に標準的なアノテーションを実施し、公開する ・アノテータ・キュレータ教育で得られた情報を上記のアノテーション情報に追加し、公開する。 <p>3. 統合データベース支援</p> <p>(1)ポータル整備・運用、広報、普及啓発</p> <ul style="list-style-type: none"> ・生命科学系DBカタログについては、国内DB並びに国外主要DBを網羅(600DB程度)する。 ・生命科学DB横断検索については、国内外主要DBを網羅(200DB)し、検索の高度化も実現する。 ・普及啓発については、統合TV拡充(250サイト)、高機能化と開発教材とのドッキングサイトを構築する。 <p>(2)データベースの受入れと運用</p> <ul style="list-style-type: none"> ・プロジェクト関連機関の開発DBとの連携サイト構築と横断検索への編入 ・個別プロジェクトDBの受入れとアーカイブ化(30DB) <p>(3)人材育成</p> <ul style="list-style-type: none"> ・DB構築者養成については、計20名に教育を実施し、うちDBCLSで計6名活用 ・DB高度利用者養成については、計80名に教育を実施し、うちDBCLSで計6名活用 ・アノテータ、キュレータ養成につ初級中級延べ1,000名に教育を実施し、うちDBCLSで計6名活用
	JST	<p>1. 意見集約システムの運用 WINGProの公開継続と新規10DB追加</p> <p>2. 広報 本事業全体の広報活動とプロジェクト内サイトの構築・運用</p> <p>3. データベース受入れ メタデータサイトの2件追加とMouse EmbryoのEST DB受入れ</p>	
	産総研CBRC	<p>1. アミノ酸配列から広範囲な立体構造に関する予測を行うワークフロー(8/E 限定公開)</p> <p>2. 予測とデータ取得により蛋白質に関する網羅的な情報を得るワークフロー(12/E 一般公開)</p> <p>3. 蛋白質の比較情報を提示し、保存部位、変異部位を推定するワークフロー(3/E 一般公開)</p>	
	かずさDNA研	<p>1. 高度情報集積DB インターフェイスの改善と利用者の拡大(100人規模) イネを対象にネームサービスの充実</p> <p>2. ゲノムアノテーション情報 GeneIndexing型アノテーション4万件蓄積 ゲノム位置情報と論文記載情報の統合</p>	
	奈良先端大	<p>1. 専門用語辞書システム 同義語に同じ識別子を持たせ、表記ゆれや別表記を検索し表示する機能実装 複数語から成る用語のタグ付け機能と内部構造の表示機能実装</p> <p>2. 専門用語解析技術 用語の内部構造を90%以上の精度で解析、500語の内部構造解析データ作成 70%以上の精度で複数の用語の並列構造を解析</p> <p>3. 専門用語タグ付け手法 MeSHオントロジーに従った用語分類システムのプロト構築</p>	
	九州大	<p>データ生産者とは独立して各データソースの倫理規定にしたがった3段階の共有形態による多型情報提供を行う。</p> <p>1. アレルタイプと頻度情報 2. 国際水準に則ったQuality control解析結果 3. 多型タイピング一次データ</p>	
	東京大学	DB構築技術を習得した人材を育成する。本年度はDBCLS2名、自治医科大1名、東大新領域8名	
	お茶の水女子大	DB高度利用者の養成。本年度は20名を対象。2名をDBCLSの統合TV開発で活用	
	長浜バイオ大	<p>1. 初級アノテーション教育(250名) 環境由来メタゲノム配列から健康に貢献する遺伝子発掘と教材公開</p> <p>2. 中級アノテーション教育(50名) 新規ゲノム配列を対象に実際にアノテーションを実施</p> <p>3. 自己組織化マップによる養成(卒研生数名) 1. の結果をもとに相同性によらない生物系統の推定</p> <p>4. シニア研究者と学部生の共同作業によるtRNAのデータベース化</p> <p>5. 1名をDBCLSの統合TV開発で活用</p>	

機関名		本年度成果目標	最終成果目標
分担機関	京都大学	<p>1. 共通基盤技術開発 ①知識処理技術開発 ・化学構造比較の高速化: SIMCOPMベースに高速化を実装し公開(7/E) ・化学反応ネットワーク予測: 2つの化合物構造を入力し、その間の反応経路を予測するシステムプロト ・酵素番号自動割り当て: 化合物ペアからその間の反応を触媒する酵素番号を割り当てるシステムの改良 ②ウェブ技術開発 高速・高機能な検索エンジンを開発し、構造検索システムと統合</p> <p>2. 統合データベース開発・運用 ①医薬品・化合物データベース開発・運用 ・JAPIC添付文書の更新、DB間のリンク情報の更新作業を行う。併せて、付属情報の検索機能も検討 ・各種化合物DBのキーワード検索機能の公開(7/E) ・脂質DB、病原性に関わるDBを統合し、ゲノムネット化合物DBとして提供 ②LinkDB開発運用 随時追加、変更、更新を行い、併せて、キーワード検索などの機能拡張を行う</p>	<p>1. 共通基盤技術 化合物、医薬品データベースに対してキーワード検索から構造検索まで様々な方法で検索するための技術を開発する。また、化合物や医薬品が生体内外でどのように使われ、変換されているかを調べることができる技術も開発し、検索から解析までウェブ上でできるデータベース基盤技術を確立する。</p> <p>2. 統合データベース開発・運用 各種化合物、医薬品データベースの統合化を、データ間のリンク情報を中心として実現し運用する。さらに各化合物の生体における作用や病原性など付随する情報を統合化することにより、研究者から一般のユーザーまでが有効利用できるデータベースをウェブを通じて広く公開する。</p>
	医科歯科大Gr	<p>1. ターミノロジー、シソーラスを肝細胞癌、パーキンソン病から、大腸癌、舌癌及びGeMDBJに搭載されている癌疾患を対象に拡充する。</p> <p>2. プロトシステムの機能向上 ・平成19年度分を含め、癌200例、神経疾患400例を公開予定 ・GeMDBJとの統合: 平成20年度は、特にがん症例を追加することによって、癌と神経疾患では分かりにくかったセマンティック検索の効果をより明確にユーザーに認識させることを期待する。 ・検索GUIの高度化、セマンティック検索エンジンの高度化: 検索GUIの高度化では、直感的(分かり易く)で、複数の臨床・病理所見から類似症例を検索可能とする機能を追加する。 ・最終年度のシステム化に向け、より効率的に他のデータベースとの連携を可能とするため、検索エンジン等の標準化をプロトタイプレベルで先行して検証を行う。</p> <p>3. 倫理規定草案の作成</p>	<p>1. がん疾患、神経疾患を対象にした臨床・疾患データベースの統合化技術開発 疾患DBの統合化に必要な技術要件をまとめ、疾患DBを統合するための標準様式と標準APIのモデルを開発・整備する。</p> <p>2. 実証的統合 東京医科歯科大学のがんDB300症例、大阪大学の神経疾患DB500症例の実証的統合に加え、GeMDBJなどがん、神経疾患に特化してDBの統合を進め、統合データベース検索システムをプロトタイプ的に構築し、公開する。</p> <p>3. オントロジー、ターミノロジー、シソーラスの整備 がん、神経疾患を典型例として、オントロジー、ターミノロジー、シソーラスを整備する。</p> <p>4. 倫理面の検討結果 臨床疾患データベースを統合公開するにあたり必須となる倫理規定案を作成し、中核機関へ提案する。</p>
	東大医学部Gr	<p>1. 標準SNP DBの構築 ← 08/07から一部のデータで公開 新たなplatformデータに対応した品質管理のための基準作成。</p> <p>2. GWAS(ゲノムワイド関連解析)DBの構築 ← 08/08から一部のデータで公開 CNV対応: 検出、標準、及びケースコントロールのデータベース化、可視化 DB拡張: 新規データ受入れ用の計算のパイプライン化、第2ステージデータ受入れ対応</p> <p>3. リンクエンスによる臨床情報・ゲノム情報DBの構築 パーキンソン病について、臨床情報と変異情報を搭載したリンクエンスデータベースを構築</p>	<p>1. 標準SNP DBの構築 健康者700検体以上(Affy 500K, 6.0, illumina 515Kなど)</p> <p>2. GWAS(ゲノムワイド関連解析)DBの構築 ゲノム4領域、厚労省、CRESTなどのデータ登録。Epistasisの解析手法開発。</p> <p>3. リンクエンスによる臨床情報・ゲノム情報DBの構築 5疾患程度について、疾患関連、原因遺伝子のmutationデータの網羅的な収集とDB化。</p>
	理研	<p>1. シロイヌナズナの発現、表現型、リソースに関する計6DBを統合化して公開。</p> <p>2. 高等動植物等由来の蛋白質構造データを付随する実験データを含めて30件公開。</p> <p>3. 微生物由来蛋白質に関わる試料調整(発現プラスミド構築実験 1万、培養実験 5千、精製実験 3千)、結晶化実験データ(結晶化条件: 90万件、観察1000万件)と200件の回折実験データ公開。</p> <p>4. 変異導入蛋白質に関わる実験データ150件、重原子導入蛋白質に関わるデータ500件を公開。</p> <p>5. アノテーションシステムの開発運用と変換データの中核機関への提供(契約締結後)。</p>	<p>1. シロイヌナズナオミックス情報の注釈付けと公開 シロイヌナズナの発現、表現型、リソースに関するデータベースを標準的なオントロジーやIDに基づくアノテーションをつけて公開し、XMLやテーブル形式でのダウンロードを可能にする。</p> <p>2. 動植物由来蛋白質構造データと実験データの注釈付けと公開 タンパク3000プロジェクトで解明された高等動植物由来の蛋白質構造データに付随する実験データに標準的なオントロジーやIDに基づくアノテーションをつけてXMLやテーブル形式でのダウンロードを可能にする。</p> <p>3. 動植物由来蛋白質構造データと実験データの注釈付けと公開 理研播磨研究所における微生物由来の蛋白質構造データに付随する実験データに標準的なオントロジーやIDに基づくアノテーションをつけてXMLやテーブル形式でのダウンロードを可能にする。</p> <p>4. 理化学研究所のデータベース統合化のためのモデルケース構築 理化学研究所のDB群を対象とする横断検索を可能にするための理研側のハブサイトを構築し、シロイヌナズナと蛋白質立体構造情報をモデルケースにしなが、他のDBにも横断検索の対象を広げていく(10~100 DB)。</p>
補完課題	産総研糖鎖	<p>1. 糖鎖データ統合への参加機関を確定し(12/E)、各機関と相談の上、統合化の手順を決定。</p> <p>2. 検索アイテム(糖鎖構造、遺伝子名など)の統一と中核機関へのデータ等の受け渡し。</p> <p>3. キーワード並びに糖鎖構造による横断検索機能の開発。</p> <p>4. 糖鎖科学統合DBの検索機能追加: 糖鎖構造全体の推定や抗体、レクチン等の部分構造推定情報表示。</p> <p>5. ノックアウトマウスを題材にモデル生物の情報を集約するカテゴリーの構築を行う(名古屋大と連携)。</p> <p>6. 別途開発のAPIを用いた統合検索用プロトタイプの構築。</p>	<p>1. 糖鎖業界に散在するデータベース集約し、糖鎖科学統合データベースを構築 糖鎖に関連するDBであるGlycoEpitopeとGALAXYとLipidBankの3件のDBを統合。リソースを共有できるように名古屋大学と共同でDBを構築する。</p> <p>2. 糖鎖関連データベースを保有している機関との連携を支援する標準化技術開発 DB間でデータのやり取りをするために糖鎖構造番号、遺伝子名など共通項目を標準化。</p> <p>3. 専門的な情報を直感的に理解できるインターフェースを開発 初心者・専門の用に別々のインターフェースを開発する。可能な限り入力を控えるインターフェースであったり、選択しながら情報までとどり着けるインターフェースを開発する。</p> <p>4. 中核機関の検索方法との連携を可能にするための開発 HyperEstrailerで検索indexを中核機関と共有できるようにする。</p>
	遺伝研	<p>1. トレースデータ用FTPサーバ構築。キーワード検索、ダウンロード、統計情報閲覧サイト作成。</p> <p>2. トレースデータ登録処理システム、波形表示システム開発。</p>	<p>1. トレースデータ用データベースの開発 ID発行とアーカイブデータベースの作成、および外部公開データベース作成</p> <p>2. トレースデータ用登録システムの開発 アノテーションプロトコルの確立と一般からのデータ受付の開始</p> <p>3. トレースデータ活用のためのソフトウェア開発 公開用FTPサイトの整備とwwwサイトの公開</p>
	九工大	<p>1. 蛋白質と変異体の熱力学データ1000件と構造データの対応表作成。</p> <p>2. 蛋白質と核酸の相互作用の定量的な熱力学実験データ1,300件を対象に構造データの対応表作成。</p> <p>3. 蛋白質・蛋白質相互作用データ格納用、熱力学データ用XMLフォーマット、データ抽出プロトを作成。</p>	<p>1. 蛋白質の安定性や相互作用の網羅的な熱力学データを構造データと統合 年間1000~1500件程度の熱力学データと構造データのクロスレファレンスを作成し双方向からデータを参照できるようにする。</p> <p>2. XMLなどのデータ交換フォーマットの整備、オントロジーなどの統合化技術の開発 熱力学データをXML化し公開する。また、熱力学のオントロジーを整備する。</p> <p>3. 統合検索との連携を可能にするための開発 統合DBセンターと協力して熱力学データの文献収集とデータ抽出を自動化するためのテキストマイニング技術を開発。これにより研究室レベルで行っているデータベース開発の負担をできるだけ軽減することを目指す。</p>