

統合データベースプロジェクト平成21年度進捗状況と平成22年度成果目標

機関名	平成21年度成果目標	平成21年度末進捗状況	最終年度(22年度)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"> ・生命科学分野のデータ共有のあり方を含めた統合データベースの将来像の検討とデータ共有の促進 ・著作権、個人情報に関わる問題点を調査、整理し、具体的事例に適用する。 ・ユーザアンケートとアクセス解析等により提供サービスの評価を行う。 <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"> ・テキストマイニング技術を用いた論文執筆支援のためのシステムを開発し公開する。 ・質問応答システムに向けた、自然文からのクエリ作成システムを開発し公開する。 ・ツールの組み合わせインタフェイスとして、独自機能を付け加えたGalaxyを開発し公開する。 <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。解剖学用語についてはTerminologia AnatomicaとFMAをマップし、日本解剖学会用語との対応付けを行い、解剖学辞書の多言語化及び構造化を行う。BodyPartについては、カノニカルモデルの臨床レベルの正確さを期すために数人の医師グループでのレビューとデータ仕様の明確化を行う。胸部内臓器の臨床レベル化と筋肉系のオートレベル追加を行う。Anatomographyについては、解剖部位の選択の容易化と部品表面の任意の位置へのコメント付け保存システムの開発を行う。 ・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。昨年度までに開発した文献正規化ツールを利用してゲノム報告のある/バクテリアの報告論文から特徴表を試作する。 ・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。DNAデータ総覧については配列検索の高速化 総覧表現の改善を行う。発現データについては、次世代シーケンサーによる配列データとしての発現データの統合サイトを構築し、公開する。 ・(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) ・b-MiGAPを一般公開する。一般公開にあたっては、b-MiGAP担当者を新たに雇用して、利用者対応を担当させるとともに、利用者からのニーズの吸収に努める。 ・解析手法の自由度を高めたb-MiGAPよりは高度なs-MiGAPの開発を進める。 <p>3. 統合データベース支援</p> <p>(1) ポータル整備・運用</p> <ul style="list-style-type: none"> ・ポータルサイトの整備・運用を継続し、今年度成果の公開に対応し必要な部分は変更を加える。 ・DBカタログは1000DBの情報を掲載する。生物アイコンのうち国内で研究が行われている生物に対し文献情報、研究情報、リンク情報などを追加し内容を充実させる。 ・横断検索は更新やバックアップ等の運用管理を中心に行う。横断検索をベースに統合検索に必要な項目や技術について検討調査を行う。 <p>(広報)</p> <ul style="list-style-type: none"> ・学会・展示会展示(6回)、文科省/センター主催シンポジウムの開催(1回) ・広報素材:パンフレット更新、ニュース配信の実施 ・ユーザからのフィードバック:平成20年度成果を中心とした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 <p>(普及啓発)</p> <ul style="list-style-type: none"> ・統合TVをさらに100件追加拡充し、多様なユーザーレベルを想定した実験手法を紹介するサイトとして高機能化を図る。 <p>(2) データベースの受入と運用</p> <ul style="list-style-type: none"> ・我が国のプロジェクトデータベースについて、データベースのタイプに応じた方法で、データベース受け入れを実施する(目標15件)。 ・受け入れたデータベースの公開システムに検索機能・横断検索との連携機能等の拡張を行い、運用を行う。 ・我が国で実施された主要な生命科学系プロジェクトの調査を実施し、その結果を公開する。 	<p>1. 戦略立案・実行評価</p> <ul style="list-style-type: none"> ・JST新センター設置準備委員会の場などを通して引き続き統合化の課題の整理を行うと同時に、将来の医療データ統合化のための調査、検討も行った。 ・データ共有、著作権等に関わるシンポジウム「科学における情報の上手な権利化と共有化」を開催した(10/5)。 ・H20年度成果に関わる統合データベースプロジェクト全体を対象としたユーザ評価結果を公開し、コメントへの回答も併せて掲載した。アクセス解析は継続中。 <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"> ・日本語自然文クエリからのMEDLINE検索システムのプロトタイプを作成し、β公開した。 ・DBCLS GalaxyにTogoWSからのデータ取得などの独自ツールを組み込み公開した。 <p>(2) ヒト統合DBの開発・運用</p> <ul style="list-style-type: none"> ・構築した辞書やデータの管理システムへのコンテンツの充実、キュレーション情報の付加、DNA DB総覧と遺伝子バンク(GEO)の融合化などを行った。 ・ポデバーツ3Dに関しては、全身の筋肉の3次元モデルを新規作成した。また胸部についても解剖学教科書の記述との差異をチェックし、それにできる限り準拠するようにモデルの精密化を行った。 <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"> ・2009年6月にb-MiGAPを一般公開した。当初雇用を計画していたb-MiGAP担当者については募集を行ったが適任者を見出すことができなかったため、利用者対応は業者に委託した。利用者のニーズについては、60名ほどの登録利用者のうち比較的利用頻度が高かった12名からアンケートを回収してその把握に努めた。その結果を受けて、スループット向上に取り組み、また、解析対象をバクテリアとアーケアから菌類へも拡張していくこととした。 ・b-MiGAPに比べて解析手法の自由度を高めたs-MiGAPの開発を進めて、一部利用者には公開した。 ・高い需要に応えるために、国立遺伝学研究所のスーパーコンピュータシステムへのMiGAP移植を進め、3月末までに移植を完了した。 <p>3. 統合データベース支援</p> <p>(1) ポータル整備・運用</p> <ul style="list-style-type: none"> ・プロジェクト成果ポータルサイト「統合ホームページ」の整備・運用は、特に障害無継続中である。 ・DBカタログは3月末808件の情報を掲載したが、目標1000DBの達成は下方修正する。理由はメタデータ、解説の充実化並びに国内DBのリストアップを優先させるため。 ・横断検索は全インデックスを見直し更新を行った。キーワード表示機能、Blast検索連携機能をLSDBLab.にて公開した。 <p>(広報)</p> <ul style="list-style-type: none"> ・学会・展示会展示(7回実施、BioEXPO、脂質生化学会、癌学会、生化学会、生物物理学会、分子生物学会、農芸化学会)、文科省/センター主催シンポジウムの開催実施済(2回、2009年6月12日および10月5日) ・広報素材:パンフレット更新終了、ニュース配信の実施(PJサイトおよびJSTサイトでの配信84件、BTJでの配信9件) ・ユーザからのフィードバック:平成20年度成果を中心としたサービス内容に関するアンケートを実施、集まったコメントを整理して公開し、コメントに対する開発者他の回答も併せて掲載した。 <p>(普及啓発)</p> <ul style="list-style-type: none"> DBCLSで提供しているデータベースの英語版の作成と各種有用データベースの使い方のアップデートを中心に、統合TVをさらに30件(現時点で合計120件)作成した。 <p>(2) データベースの受入と運用</p> <ul style="list-style-type: none"> ・生命科学系データベースアーカイブ(旧アーカイブサービス)において以下を行った。 1) 以下の9件のデータベースについて、アーカイブを公開した。 ①WorTS②GETDB③Atlas (ISH data base)④KTmAbDB⑤HATODAS⑥Transposon insertion site database-germline⑦EGTC⑧DGSP Database⑨The NAISTrap database 2) 11件のデータベースについてアーカイブ公開に向けて準備を進めている。 3) 検索機能に、表示件数や項目の表示順・非表示を可変にする等の機能拡張を行った。 4) アーカイブサイトの構築や更新、運用を支援するシステムを構築した。 5) アーカイブサイトの英語化を行った。 ・国内外の蛋白質関連DBの蛋白IDに基づく統合検索システム(TogoProt)について、インタフェースを改良して、累積400DBの収録を終えた。 ・「生命科学系主要プロジェクト一覧(旧名:ゲノム・ポストゲノム主要プロジェクト一覧)」にプロジェクト情報を追加・公開し、21年度分は22件となった。 	<p>1. 戦略立案・実行評価</p> <ul style="list-style-type: none"> ・生命科学分野のデータ共有のあり方を含めた統合データベースの将来像の検討とデータ共有の促進。 ・著作権、個人情報、倫理に関わる問題点を調査、整理し、具体的事例に適用する。 ・プロジェクト全期間にわたる成果についてのユーザアンケートとアクセス解析等により提供サービスの評価を行う。 <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"> ・現行のテキストマイニングツール群の連携による論文執筆支援環境の充実。 ・データ解析のためのデータ取得解析共有サイバーインフラストラクチャーの統合。 <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、高効率植物、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、新世代シーケンサでコンテイングまたは完全ゲノムを決定した研究グループに、微生物配列に対する標準的アノテーションを提供していく。(補足説明:次世代シーケンサが微生物研究室にも手が届くようになってきたため、決定したゲノム配列からなんとか生物学的意味を取り出したいという要求が近年急速に高まってきたため、MiGAP開発に取り組んだ。2009年6月に公開したMiGAPが事実活用されてきたことから、新規配列に対してMiGAPによって安定したアノテーションを付与することに集中する。) ・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</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(800DB程度)し、利用しやすくするためにメタデータを多く取り入れ、解説を充実させる。 ・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。 <p>(広報)</p> <ul style="list-style-type: none"> ・学会・展示会展示(5回)、文科省/センター主催シンポジウムの開催(1回) ・広報素材:パンフレット更新、ニュース配信の実施 ・英語での情報発信(継続) <p>(普及啓発)</p> <ul style="list-style-type: none"> ・統合TVをさらに100件追加拡充し、これまで作成したコンテンツを整理しまとめるメタ情報を充実させ、より使い勝手をよくする。 (2) データベースの受入と運用 ・我が国のプロジェクトデータベースについて、データベースのタイプに応じた方法で、データベース受け入れを実施する(目標15件)。 ・公開用システム「生命科学系データベースアーカイブ」について、検索機能・他サービスとの連携機能等の拡張を行い、運用を行う。 ・我が国で実施された主要な生命科学系プロジェクトの調査を上げ、その結果を公開する。 ・蛋白質に関する国内外のDB統合検索システム「TogoProt」を拡張するとともに、関連する他機関との連携を進める。 ・これらのコンテンツについて、データ統合に向けた整備を進める。
	JST	<p>1. 意見集約システムを継続運用する。一覧表を自動作成するシステムを開発する。</p> <p>2. 事業サイト等を継続運用する。</p> <p>3. メタデータエレメントサイトを継続運用する。</p>	<p>1. 意見集約システムを継続運用。一覧表自動作成システムも運用。</p> <p>2. 事業サイト等を更新しつつ継続運用。</p> <p>3. メタデータエレメントサイトを継続運用。</p>	<p>1. 意見集約システムを継続運用する。</p> <p>2. 事業サイト等を継続運用する。事業サイトは最終的に「ライフサイエンスの広場」に静的にリンクし保存する。</p> <p>3. メタデータエレメントサイトを継続運用する。</p>
	産総研CBRC	<ul style="list-style-type: none"> ・タンパク質立体構造予測ワークフローの開発(12/E 一般公開):アミノ酸配列を入力とし、タンパク質立体構造データベースから最適なテンプレート構造を検索し、それを基にタンパク質立体構造モデリングするワークフローを開発する。 ・アクティブ・ワークフローに向けた環境開発:ユーザが目的に応じて、独自のアクティブ・ワークフローを構築し実行できる目的指向アーキテクチャーを目指した環境及びツールのプラットフォームを開発する。 	<ul style="list-style-type: none"> ・CBRC情報統合サイトにて11月に公開したタンパク質立体構造予測ワークフローの運用及びプログラム整備を行った。 ・アクティブ・ワークフローとし、遺伝子ネットワーク推定システムであるASIANのプラットフォーム化を実現し、XサーバやWINDOWS上で動作するワークフローを公開した。またその利用マニュアル等のドキュメント作成を行った。 ・CBRCにて独自に開発された解析ツールの標準化/利用推進のため、プラットフォームを用いた解析ツールのSOAP化に着手した。 	<p>ワークフロー技術を用いた統合DB環境構築</p> <ul style="list-style-type: none"> ・利用者の要望を考慮し、利用者の要求に対し、必要なデータ、解析手法などを国内、海外から選び、情報(知識)提供する解析ツールのワークフローをWEBサービスで実現し、公開する。 ・利用者の要望を考慮し、柔軟性・拡張性に優れたプラットフォーム上で動作する解析要素技術を繋げたアクティブ・ワークフローを開発し、公開する。

機関名	本年度成果目標	3月末進捗状況	最終年度(22年度)目標	
中核機関	かずさDNA研	<p>1. 高度情報集積DB</p> <ul style="list-style-type: none"> 登録ユーザ数80人(2010年3月31日現在)。 高度情報集積DBのシステム設計を改善し、検索・表示速度を数倍～100倍程度高速化した。 高度情報集積DBに集積されたデータを用いて、ゲノムDB (CyanoBase, RhizoBase) から新規文献情報が検索・参照可能なシステムを構築した。 ID変換サービスのための、ゲノムDBへの検索システムの設計と、遺伝子名とIDの対応辞書の調査とプロトタイピングを継続中。 植物関連微生物16種類のBiomartシステムへのインストールのためのデータ準備を完了した。 シロイヌナズナ(理研)、イネ(農水)、ミヤコグサ(かずさ DNA研)3種の高等植物ゲノムデータ収集を完了した。 国内学会やDBCLS講習会で高度情報集積DB、Biomart (KazusaMart) の利用例やデータの解析例を発表し利用を促進した。 <p>2. ゲノムアノテーション情報</p> <ul style="list-style-type: none"> 植物関連生物14種の6468報の文献からGene Indexing型ゲノムアノテーションを163173件抽出蓄積し公開済(2010年3月28日現在)。 ゲノム位置情報上に発現情報や論文記載情報など多様なデータを統合するビューアの開発と講習会での広報意見収集を行った。 	<p>1. 高度情報集積DB</p> <ul style="list-style-type: none"> 登録ユーザ100人規模。 高度情報集積DBシステム運用作業工程の改善。 ID変換サービスを統合したゲノムデータベースの利用促進をはかる。 ゲノムデータマネジメントシステムBiomartによる植物と関連微生物ゲノム情報の統合。追加8種の構造注釈つき生物種のデータ入力と公開。 <p>2. ゲノムアノテーション情報</p> <ul style="list-style-type: none"> Gene Indexing型アノテーション情報の2万件追加と既存データの改善。 複合体情報 (Complex Indexing) を1000件追加。 ゲノム位置情報と論文記載情報の統合:ゲノムビューア可視化技術の利用促進をはかる。 	
	奈良先端大	<p>1. 専門用語辞書システムの開発 専門用語の内部構造や同義語識別子を用いた検索や専門用語の意味情報を管理することができる辞書システムの設計を行う。</p> <p>2. 専門用語解析技術の開発 専門用語の内部構造解析済みデータの拡大、および、内部構造解析として80%以上の解析精度を目指す。</p> <p>3. 専門用語抽出ツールの設計と開発 新規の専門用語の意味分類手法の高性能化、および、新規語を用語辞書に登録するためのユーザインタフェースの設計を行う。</p>	<p>1. 専門用語辞書システムの開発 専門用語の内部構造による検索、同義語識別子を用いた同義語の検索機能を実現した。MeSHのシソーラスコードを格納し、シソーラスコードを用いた検索機能を実装した。</p> <p>2. 専門用語解析技術の開発 専門用語の内部構造を解析するための構造解析アルゴリズムを設計・実装し、約85%以上の精度での解析を実現した。</p> <p>3. 専門用語抽出ツールの設計と開発 新規の専門用語と類似する用語を一覧表示し、シソーラス上での位置を表示するインタフェースのプロトタイプシステムを構築した。</p>	<p>1. 専門用語辞書システムの開発 10万語規模の生命科学用語の辞書を格納し、Webブラウザ上で用語の検索や用語のシソーラスコードの表示、および、用語間の上位下位関係を表示・修正することのできる専門用語辞書システムを実現する。</p> <p>2. 専門用語解析技術の開発 用語の内部構造解析の学習データを2000語以上に拡大する。そして、前年度に設計した一般的な統語解析法に基づくアルゴリズムを実装し、90%以上の内部構造解析精度を達成する。</p> <p>3. 専門用語抽出ツールの設計と開発 新規の専門用語にシソーラスコードを付与する支援機能を実装し、専門用語辞書システムの機能として統合する。</p>
	九州大	<p>①JAGQCでのQC機能をPLINKプログラムのモジュールを用いることにより一本化QCパイプラインとする。</p> <p>②集団内階層構造、実験による系統的誤差の検出のために主成分解析(PCA)等を導入し、QC機能を強化する。</p> <p>③ゲノム4領域応用ゲノムの各研究グループからのGWAS1次スクリーニングデータを受け入れ、JAGQCパイプラインによるQCを行い、共有データベースで開示する。</p> <p>④DHaploDBからの日本人ハプロタイプ情報を統合データベースへ移行する方策を検討する。</p>	<p>1. JAGQCでのQC機能をPLINKプログラムのモジュールを用いることにより一本化QCパイプラインとする作業を完了し、webページ表示のための作業を完了した。</p> <p>2. 集団内階層構造、実験による系統的誤差の検出のために主成分解析(PCA)を導入する作業を完了し、webページ表示のための作業を完了した。</p> <p>3. ゲノム4領域応用ゲノムの各研究グループからのGWAS1次スクリーニングデータを受け入れる方法を引き続き検討している。</p> <p>4. DHaploDBに在る日本人標準確定ハプロタイプ約100個の情報を統合データベースへ移行する方策の一環として、同ハプロタイプ中のコピー数多型を決定し、その結果のDHaploDBで表示する方法を引き続き検討している。</p>	<p>・国内各研究グループからのGWAS1次スクリーニングデータを受け入れ、これを順次JAGQC Pipelineを通すことによりQCを行い、これらの結果を開示する。</p> <p>・最近公開されつつあるGWASデータベースと我々がすでに確立している確定ハプロタイプデータを連携させる。</p>
	東京大学	DB構築技術を習得した人材を育成する。本年度の受講者数は1年コース東大新領域3名と医学系研究科1名、1カ月コースは募集中。	<p><講義状況>4月～12月の間に4名の受講者に対して24回の講義を行った。4月～5月はスクリプト言語によるプログラミング・サーバー管理・データベース管理等のバイオデータベース構築に必要な基礎技術の講習を行い、6月～7月はJava言語によるプログラミング講習を行った。7月～10月はEnsemblゲノムブラウザを演習用サーバーにミラーする演習を行った。10月～11月はサーバー障害の対策としてデータのバックアップと復旧について演習した。12月はUTGBゲノムブラウザのインストールとゲノムブラウザ開発のためのウェブプログラミングについての解説を行った。また10月～12月の演習では、昨年度の受講者を講師として、研究で使用しているデータベース・ライブラリ・解析ツールの紹介と演習課題の解説を行った。</p> <p><受講者の進捗状況>OSのインストール・ネットワーク接続・セキュリティアップデートの設定・ウェブサーバーの設置・動的ウェブページの設置・データベースサーバーの設置・データベースの作成・Javaプログラミング講習・Ensemblゲノムブラウザの基本部分のインストール・サーバーバックアップと復旧・UTGBゲノムブラウザのインストール・ゲノムブラウザ開発のためのウェブプログラミングの解説を行った。Ensemblゲノムブラウザのインストールについては、受講者4名中3名が基本部分のインストールまで終わった。</p>	大規模計算のためのクラスター利用技術と独創的バイオDBサーバーの構築技術を指導し、次世代シーケンシング情報のデータベースを構築できる人材を育成する。ウェブで公開中の演習ノートも改善する。本年度の受講者数は東大新領域4名。
	お茶の水女子大	DB高度利用者を養成する。本年度は社会人1名以上を含む7名以上の修了者を輩出する。また統合DB事業へ従事する人材を3名以上輩出する。	社会人(主婦3名)を含む5名及び学生27名が修了。加えて社会人5名が部分的に受講。基礎コースとしてデータベース、ネットワーク、遺伝学、及び、専門コースとしてネットワークによるデータ利用、Java言語によるプログラミング、基礎統計、データマイニング技術を開講。統合DB事業へ従事するものとして、本年度の受講生の内1名がDBCLSでテキストマイニング関連の開発に従事、3名がTA及びサーバ管理者として統合DB事業に従事している。また、昨年までの受講生から、2名が統合TVに従事、講義・演習のTAとして2名が統合DB事業に従事している。	DB高度利用者を養成する。本年度は社会人2名以上を含む7名以上の修了者を輩出する。また統合DB事業へ従事する人材を3名以上輩出する。
	長浜バイオ大	<p>1. 初級アノテーション教育(250名) バイオサイエンス分野におけるデータベースの統合的な利用と教材公開。多様な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更新ならびに公開。</p> <p>2. 中級アノテーション教育(50名) 新規な微生物ゲノム配列を対象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(卒研究生数名) 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(tRNADB-CE)のデータ拡充。</p> <p>5. 1名をDBCLSの統合TV開発で活用。</p> <p>6. ミニシンポジウムの開催。テーマは、「自分が興味を持つ遺伝子がどの範囲の環境微生物に存在するのを探る方法:環境微生物ゲノム配列からのお宝遺伝子発掘の学部や高校教育における活用」。</p> <p>7. 中学生・高校生・高専生を対象にした講習会の開催。テーマは、「生命の謎にコンピュータで挑戦 ～バイオ分野のデータベースの利用～他」。</p>	<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の「健康に貢献する遺伝子発掘」の実習を行い、シニア研究者が精査完了したデータとテキストの更新版を統合DBから公開。</p> <p>2. 2回生のアノテーション教育(270名) 実習授業「バイオ分野のデータベースの統合的な利用」を実施、テキスト最新版のWEB公開。</p> <p>3. 中級アノテーション教育(50名) 次世代シーケンサーで決定した約1000万件の微生物ゲノム由来の大量断片配列データを対象としたアノテーションを実施。次世代シーケンサー由来配列解析用の学内版テキスト作成。同校外版の作成着手。</p> <p>4. 2名の4回生の人材養成として、自己組織化マップによる約2000件の遺伝子の生物系統を推定し、そのアノテーション結果を統合DBセンターより公開。 1. で得られた結果を対象に相同性によらない生物系統の推定を行った。</p> <p>5. 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(tRNADB-CE)に、ウイルスゲノムやイネゲノムを追加し、機能拡充を行った更新版を公開。</p> <p>6. 1名をDBCLSの統合TV開発で活用。</p> <p>7. ミニシンポジウムの開催。テーマは、「自分が興味を持つ遺伝子がどの範囲の環境微生物に存在するのを探る方法:環境微生物ゲノム配列からのお宝遺伝子発掘の学部や高校教育における活用」。</p> <p>8. 近隣の3高等学校で総計180名の高校生を対象とした土曜講座、出前実習の開催。テーマは、「バイオ分野のデータベースの利用～他」。</p>	<p>1. 初級アノテーション教育(270名) バイオサイエンス分野におけるデータベースの統合的な利用の実習と教材の更新版の公開。</p> <p>2. 初級アノテーション教育(270名) 多様な環境由来メタゲノム配列から健康や持続可能型社会に貢献する遺伝子発掘と教材更新ならびに公開。</p> <p>3. 中級アノテーション教育(50名) 次世代シーケンサーで解読した微生物ゲノム配列を対象にしたアノテーションの実施と教材の作成。</p> <p>4. コンピュータとバイオ実験に習熟した人材の育成を目的に、「バイオ技術のデータベース」と教材作成。</p> <p>5. 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(tRNADB-CE)のデータと機能の拡充とキュレータ育成。</p> <p>6. 1名をDBCLSの統合TV開発で活用。</p> <p>7. 環境微生物ゲノム配列からのお宝遺伝子発掘についての、他大学での出前実習の実施。</p> <p>8. 中学生・高校生・高専生を対象にした講習会の開催。テーマは「バイオ分野のデータベースの利用～他」。</p>

機関名	本年度成果目標	3月末進捗状況	最終年度(22年度)目標
京都大学	<p>1. 共通基盤技術開発</p> <p>①知識処理技術開発</p> <ul style="list-style-type: none"> ・化学構造検索の高機能化:類似構造検索、部分構造検索システムを改良する。 ・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インタフェースを改良する。 ・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を公開する。 <p>②ウェブ技術開発</p> <ul style="list-style-type: none"> ・化合物データベースに対する同義語、階層分類情報、LinkDBの等価エントリー情報を考慮した検索機能を開発する。 <p>2. 統合データベース開発・運用</p> <p>①医薬品データベース開発・運用</p> <ul style="list-style-type: none"> ・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、副作用情報など付属情報の検索機能を開発する。 ②化学情報データベース開発・運用 ・各種化合物データベースの統合を引き続き進めるとともに反応データの知識を取り込むことも検討する。 <p>③LinkDB開発・運用</p> <ul style="list-style-type: none"> ・随時追加、変更、更新を行い、併せて、検索機能の拡張を行う。 ・統合データベースプロジェクト内のデータベース間のリンクについて検討する。 	<p>1. 共通基盤技術開発</p> <p>①知識処理技術開発</p> <ul style="list-style-type: none"> ・化学構造検索の高機能化:類似構造検索、部分構造検索システムをユーザーがローカルで使えるように、データベース作成などのプログラムとセットでパッケージ化した。ドキュメントを整備した後に公開する予定である。 ・酵素番号自動割り当て:データの整備とシステムの精度改良に応用するために、化合物ペアのアライメントを改良するためのオプション指定ができるテスト環境を構築した。 ・化学反応ネットワーク予測:生物種ごとに既知の反応データを用いて反応経路を計算する PathComp のインタフェースの改良を進め、テスト中である。 <p>②ウェブ技術開発</p> <ul style="list-style-type: none"> ・LinkDBの等価エントリーの情報を利用し、各種化合物・医薬品データベースのキーワード検索結果から、対応するKEGGのエントリーを取得できるようにした。 <p>2. 統合データベース開発・運用</p> <p>①医薬品データベース開発・運用</p> <ul style="list-style-type: none"> ・JAPICデータの更新は引き続き順調に進めており、3月末で医療用医薬品13,514件、一般用医薬品12,313件の情報がデータベースに登録されている。副作用情報に関連して医薬品相互作用情報を抽出した。 <p>②化学情報データベース開発・運用</p> <ul style="list-style-type: none"> ・反応データの知識として反応パターンの意味付けを行うために、酵素番号自動割り当てシステムの結果と専門家の割り当て結果を比較した。 <p>③LinkDB開発・運用</p> <ul style="list-style-type: none"> ・エントリー表示時にLinkDBの結果リストを表示する仕組みをKEGG以外のデータベースにも拡張した。 ・随時追加、変更、更新作業は順調に進んでいる。 	<p>1. 共通基盤技術</p> <p>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</p> <p>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>
分担機関	医科歯科大Gr	<p>【医科歯科大】</p> <p>1. 疾患・臨床医学データベースに特化した統合技術開発に関する研究:セマンティック検索エンジンの高度化として、外部要素データベースとのHTTP/XMLを利用した統合検索を開始した。また、分子情報に関するデータを取り扱うためのインタフェースに遺伝子発現情報を手がかりとして関連する臨床情報の検索や個別症例を辿ることを可能にするなどの高度化を行なった。</p> <p>2. 疾患データベース(がん、神経難病を中心として)の高度化と実証的統合に関する研究:倫理規定案に準拠の上、東京医科歯科大学の統合検索システムを利用する一般研究者、登録研究者各々のアクセスレベルでの公開を行なった。加えて各ユーザーより本検索システムの評価を得るためのアンケート集計システムを構築した。</p> <p>3. プロジェクトの統合的推進:利用者が必要とする情報への効率的なアクセスに向けた仕組み作りの1つとして日本肝癌研究会、大腸癌研究会への関連情報のリンク付けを行った。</p> <p>【阪大】</p> <p>1. 疾患・臨床医学データベースに特化した統合技術開発に関する研究: 医科歯科大Grが開発した統合検索エンジンと阪大GrのパーキンソンDBの統合検索について、医科歯科-阪大間の打ち合わせで調整した詳細仕様に基づいて、統合検索のためのインタフェースを開発した。</p> <p>2. 疾患データベース(がん、神経難病を中心として)の高度化と実証的統合に関する研究:同意取得作業を継続し、公開データ数を400データから500データまで増加させた。</p> <p>3. プロジェクトの統合的推進: 検索機能の高度化、および関連情報への連携について、利用者の使い勝手向上を中心に次年度以降の機能拡張の方針を検討した。</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</p> <p>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</p> <p>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</p> <p>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</p> <p>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</p> <p>健常者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</p> <p>CNVのケースコントロールDBの構築。Epistasisの解析手法開発。</p> <p>3. リシークエンスによる臨床情報・ゲノム情報DBの構築</p> <p>パーキンソン病のリシークエンスDBの充実化と家族性痙性対麻痺と副腎白質ジストロフィーのリシークエンスDB構築(2009-2010の2年計画)</p>	<p>1. 標準SNP DBの構築</p> <p>健常者約1,300検体のデータを追加登録し、公開を待っている。</p> <p>2. GWAS(ゲノムワイド関連解析)DBの構築</p> <p>CNVの標準DBにおいて、個体ごとのCNV表示機能の追加、クラスタリング有り・無しが表示機能などの機能追加を行った。一方、GWAS DBにおいても、データダウンロード時のオプション追加などの、詳細な機能追加を行った。更に、多系統萎縮症のGWASデータを新たに登録した。また、Epistasis効果・遺伝子間相互作用解析を目的として、変数としてのSNPおよび疾患の関連構造を、グラフィカルモデルを用いてネットワークグラフとして表すツールの精度向上を行った。</p> <p>3. リシークエンスによる臨床情報・ゲノム情報DBの構築</p> <p>新たに、パーキンソン病のリシークエンスDBを公開するとともに、各種会議にて変異データの登録を広く呼びかけた。また、ALS/リシークエンスDBの実験データの追加登録と家族性痙性対麻痺と副腎白質ジストロフィーのリシークエンスDBの文献データの追加登録を行った。</p> <p>4. 中核機関との連携の下、倫理検討委員会で開催・決定したデータの共有に関する方針を広く外部へアナウンスするとともに、実際のデータ配布の運用を開始し、データ利用申請者への対応を審議した。</p>	<p>1. 標準SNP DBの構築</p> <p>健常者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</p> <p>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. リシークエンスによる臨床情報・ゲノム情報DBの構築</p> <p>ALS、パーキンソン病、副腎白質ジストロフィー、家族性痙性対麻痺について、疾患関連・原因遺伝子のmutationデータの網羅的な収集とDB構築。</p>

機関名	本年度成果目標	3月末進捗状況	最終年度(22年度)目標	
補完課題	理研	<p>1. シロイヌナズナオミックス情報の注釈付けと公開 新しいデータベース「シロイヌナズナ変異体情報」を作成し、12月末までに整備した新しい表現型定義情報と、既存のデータベースの対応するレコードへのリンクを設定し、これらを公開した。ここまでで、明確な表現型情報が公開されている既存のレコード全てについて新しいアノテーション情報が関連付けられたことになった。</p> <p>2. 高等動植物等由来蛋白質構造のアノテーション 高等動植物タンパク質の試料調製に関わる発現検討実験(無細胞タンパク質発現用発現系構築実験4万8千件、スモールスケール発現実験8万件)に基づいて X 線結晶構造解析のための大量調製実験(大量合成および大量発現実験4千件、精製実験5千件)を行い、結晶化実験(結晶化条件 138 万件、結晶観察 1517 万件)へと進め、タンパク3000プロジェクトで解明されたタンパク質構造データのうち平成21年度分としては1万1千件の回折実験データ(画像データ枚数)を公開可能とした。</p> <p>3. 微生物由来蛋白質構造のアノテーション 微生物由来蛋白質(変異導入蛋白質を含む)に関わる実験情報および重原子導入蛋白質に関わる情報について、追加データのキュレーション作業を完了した。</p> <p>4. アノテーションシステムの開発運用とデータ変換 検索システムの更なる改良により、より高速かつ精密に表現型情報等をキーとした検索が実行でき、結果もわかりやすく表示されるようになった。</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(豊田) トランスクリプトームデータを集積したリポジトリを公開し、それらの詳細な解析結果に基づいたアノテーションデータを公開する。また、理研で観測されたフェノタイプデータを、理研以外で観測されたフェノタイプデータと統合化し、共通のオントロジーで整理し公開する。</p> <p>2. 高等動植物等由来蛋白質構造のアノテーション(横山) タンパク3000プロジェクトで解明された高等動植物等由来の蛋白質構造データの注釈付け、およびそれに付随する実験データのキュレーションとアノテーションを行う。</p> <p>3. 微生物由来蛋白質構造のアノテーション(国島) 平成21年度に引き続き、これまでに公開したデータについてアノテーション作業を行い、アノテーションデータを公開する。またそれ以外のデータについてもデータ公開を進める。平成22年度末までに、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関し、統合データベースの一部として生命科学分野の幅広い研究者に役立つ体制の確立を完了する。</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(豊田) シロイヌナズナで実施してきた統合化をモデルケースとして、シロイヌナズナ以外の生物種(マウス等)にも適用することで、データベース統合化の範囲を拡大する。また、本事業において公開したデータをクライアントサイドからウェブ経由でプログラマ的に利用できるAPIとツールを構築して公開する。さらに、これまでに構築した、アノテーション作業を行うためのシステムを運用・改善し、アノテーション等のデータ交換やプログラム交換を理研外部者でも実施できるオープンな参加型のシステムとして公開する。</p>	
	産総研糖鎖医学センター	<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、統合化の計画を行う</p> <p>2. 公開している糖鎖データベース(JCGGDB)の普及活動を行う</p> <p>3. 実験データを持っている機関にはDB化し公開するように促し、更に糖鎖統合DBと連携する</p> <p>4. オンラインプロトコルの整備を行う</p> <p>5. 糖鎖関連疾患とその糖鎖関連遺伝子の関連をDB化する</p> <p>6. 糖鎖機能に関する情報を発信する</p> <p>7. 横断検索用インデックスの更新システムを開発する</p> <p>8. 直観的なポータルサイトを旨し統合検索機能を開発し始める(2ヶ年計画)</p>	<p>1. 九州大学の伊東ラボがグリコシダーゼのDBの協力を表明。理研の糖鎖関連遺伝子のマイクロアレイを用いて解析した情報を随時DBに登録して公開したいと要望がある。更に他の機関にも協力を打診中。来年度も立命館大学がオンラインプロトコルの整備に協力することの約束を得る。野口研も合成DBを来年度も継続して協力して頂くことを了解して頂いた。糖鎖合成DBと化合物DBに岐阜大学・木曾研に協力して頂く。</p> <p>2. 糖鎖科学のコミュニケーション(ACGG)で口頭発表。国際複合糖質学会で口頭発表。JCGGで進捗報告と協力の呼びかけ。糖鎖の統合DBとLSDBの普及活動。</p> <p>3. 統合検索の連携のためのAPIを構築した。野口研究所、理化学研究所、創価大学のAPIを完了した。</p> <p>4. 立命館大学と共同で実験オンラインプロトコルの整備に取り掛かった。45種類のプロトコルの原稿をまとめて体裁を整えているところ。</p> <p>5. 糖鎖関連疾患とその糖鎖関連遺伝子の関連した情報を集め(77の病気)、DBに格納した。病態の情報に関する項目に関して著作権者に許可をもらう作業を終え公開した。</p> <p>6. 糖鎖と感染に関する情報を収集。DB化してPACDBとして公開した。構造DBと連携するところは完成。オンラインレポートに関して、糖鎖の合成分解のパスウェイのミニレビューを作成し、DBとのリンクを行った。その他、糖鎖技術を産業に応用している方からのオンラインレポートとして原稿依頼した。</p> <p>7. 横断検索用インデックスの更新システムの構築を完了。</p> <p>8. 専門用語の整備を継続。統合検索技術開発の目処がたった。</p>	<p>1. 平成21年度と平成22年度の2年間で100以上のプロトコルを収集する。年度末までに公開する。</p> <p>2. グリコシダーゼのデータベースを構築する。</p> <p>3. 糖鎖関連遺伝子のDNAマイクロアレイのデータを公開するシステムを構築</p> <p>4. 糖鎖と病原菌のデータベース(PACDB)の更新</p> <p>5. 糖鎖関連のマーカーの情報を収集しDB化する</p> <p>6. 糖鎖関連の特許情報を集め整理しDBとして公開する。</p> <p>7. オンラインレポートの原稿を20前後にする</p> <p>8. 直観的なポータルサイト実現のための統合検索システムの構築を実現する。できるだけDB間の連携を高められるようにする。</p>
	遺伝研	<p>1. トレースアーカイブへの取り組み ・2008年度までに整備したTraceデータの登録、公開システムの運用を継続し、運用中の必要に応じて適宜開発を実施する。</p> <p>2. ショートリードアーカイブへの取り組み ・2009年5月のDDBJ/EMBL/GenBankの国際実務者会議に続いて開催される予定のTA/SRA担当者会議で日米欧3極の協力体制を明確にする。 ・ Short Read Archive登録受け付けの仕組みを確立する。 ・DDBJ独自のID発行開始に向けた環境を整備する。 ・SRAデータの利用者への提供を開始する。</p>	<p>1. トレースアーカイブへの取り組み ・受け付けた1件のデータについて、登録用ファイルを作成中。</p> <p>2. ショートリードアーカイブへの取り組み ・研究手法の進化、次世代シーケンサから出力されるファイル形式の変化に対応すべく、日米欧3極(DRA/ERA/SRA)共同でメタデータのXMLスキーマを、バージョン 1.0 から 1.1 に更新した。また、これに伴い、DRAの全システムをスキーマバージョン 1.1 に対応させた。 ・登録受け付けの仕組みを確立して3月31日までに191件を受け付け、23件を公開して、利用者によるダウンロードを可能とした。 ・ユーザのフィードバックを登録ツールに反映し、登録者がより早く、より正確なメタデータを作成できるように年間を通して改善を続けた。、2010年3月8日にはFlex 技術を利用したウェブベースのメタデータ作成ツール MetaDefine の運用を開始した。MetaDefine はオンラインでのデータチェック、過去の登録データの再利用、既定値の自動埋め込みなどの入力支援機能を提供しており、登録者のデータ作成にかかる負担を軽減している。</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. 蛋白質熱力学データベースおよび蛋白質・核酸相互作用データベースのそれぞれ新規700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する。蛋白質・蛋白質相互作用データは格納用データベースシステムを完成させる。</p> <p>2. 蛋白質・核酸相互作用データをXML化し公開する。</p> <p>3. 統合DBセンターと連携して熱力学データの文献収集とデータ抽出の自動化技術の開発に必要な解析やツールの評価を行う。</p>	<p>1. 蛋白質と変異体の熱力学データと構造データのクロスレファレンスを約1,000件のデータについて作成した。また、蛋白質と核酸の相互作用の熱力学データと構造データのクロスレファレンスを約1,300件のデータについて作成した。蛋白質・蛋白質相互作用データについては、格納用のデータベースのプロトタイプを作成した。</p> <p>2. 蛋白質と核酸の相互作用の熱力学データはXMLフォーマットに変換し、テキストフォーマットでのデータとともに公開した。また、熱力学データのオントロジーについては、海外の研究者と意見交換しながらすすめており、熱力学データのControlled Vocabularyの作成を行った。</p> <p>3. 文献検索と文献からのデータ抽出の自動化方法については、統合DBセンターが開発したTogoDocなどのツールの評価を行った。</p>	<p>1. 蛋白質熱力学データベースおよび蛋白質・核酸相互作用データベースのそれぞれ新規700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する。蛋白質・蛋白質相互作用データは格納用データベースシステムを完成させる。</p> <p>2. 蛋白質・核酸相互作用データをXML化し公開する。</p> <p>3. 統合DBセンターと連携して熱力学データの文献収集とデータ抽出の自動化技術の開発に必要な解析やツールの評価を行う。</p>	