

## 統合データベースプロジェクト成果目標と進捗状況(平成21年度)

機関名	本年度成果目標	12月末進捗状況
DBCLS	<p>1. 戦略立案・実行評価</p> <ul style="list-style-type: none"> <li>・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</li> <li>・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。</li> <li>・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。</li> </ul> <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"> <li>・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。</li> <li>・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。</li> <li>・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。</li> </ul> <p>(2) ヒト統合DBの開発・運用</p> <ul style="list-style-type: none"> <li>・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アートレベル追加を行う。Anatomographyについては、解剖部品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。</li> <li>・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。</li> <li>・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。</li> </ul> <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"> <li>・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。</li> </ul> <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"> <li>・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。</li> <li>・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。</li> <li>・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。</li> </ul> <p>(広報)</p> <ul style="list-style-type: none"> <li>・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)</li> <li>・広報素材:パンフレット更新、ニュース配信の実施</li> <li>・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</li> </ul> <p>(普及啓発)</p> <ul style="list-style-type: none"> <li>・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。</li> </ul> <p>(2) データベースの受入と運用</p> <ul style="list-style-type: none"> <li>・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)。</li> <li>・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。</li> <li>・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。</li> </ul>	<p>1. 戦略立案・実行評価</p> <ul style="list-style-type: none"> <li>・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。</li> <li>・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)。</li> <li>・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た。アクセス解析は継続中。</li> </ul> <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"> <li>・論文執筆支援の一環として、生命科学者のための文献管理システム TogoDocを開発し公開した。</li> <li>・日本語自然文クエリの分解方法について検討を行った。</li> <li>・DBCLS Galaxyのツール群を公開に向けてテスト・改良し公開した。</li> </ul> <p>(2) ヒト統合DBの開発・運用</p> <ul style="list-style-type: none"> <li>・生物名辞書新規データ入力2万名称。粗データから目視で情報源ごとの表記の不統一を取り除いて整形した、学名::和名::英語名::情報源の対応ファイル(6万種相当)を作成共有。胸部臓器の高度化8割筋肉系の追加8割達成。辞書サービスのドキュメンテーション強化(説明ページ開設)。</li> <li>・GEO目次ウェブサイトの小変更 ドキュメンテーションの強化。</li> </ul> <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"> <li>・システムのデフォルト設定での解析を行うb-MiGAPを6月から公開運用。利用状況は次の通り(集計から開発グループ分は除いてある):登録利用者数59名、実行プロセス数167件、投入塩基長合計184,933,675bp、投入コンティグ数合計15,076本、同定CDS数合計172,707個、処理時間合計2160時間17分4。利用者対応担当者については資格を再検討後再募集することを検討中。</li> <li>・s-MiGAPについては国立遺伝学研究所のPCクラスターへの移植後一般公開予定(年度内)。</li> </ul> <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"> <li>・ポータルサイト運用は継続しているが、サービスメニューの増加に伴い利用者が必要なサービスにたどり着きにくい状況にある。次期インタフェースでは使用感の統一などが課題。</li> <li>・DBカタログ掲載件数は現在712件で目標達成まで約300件を残すが、現在NARのDBIssueの登録を準備中である。カタログにおけるDB分類も課題の一つであり検討を続けている。分類については経産省との連携も課題。両省で統一したものに予定。文科省の特定領域研究報告書から情報を収集し一部を生物アイコンに情報掲載する準備を進めている。</li> <li>・横断検索の運用について、現在インデックスの更新は不定期であるが、各DBの更新頻度の調査をほぼ終了し、度末までに定期更新を目指す。</li> </ul> <p>(広報)</p> <ul style="list-style-type: none"> <li>・学会・展示会展示(6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)、文科省/センター主催シンポジウムの開催実施済(2回、6月12日および10月5日)</li> <li>・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62件、BTJでの配信8件)</li> <li>・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。</li> </ul> <p>(普及啓発)</p> <ul style="list-style-type: none"> <li>・統合TVをさらに31件(現時点で合計90件)作成し、講演の動画に関して使われた各スライドを表示しそれぞれをクリックするとその再生位置にジャンプする機能を追加した。</li> </ul> <p>(2) データベースの受入と運用</p> <ul style="list-style-type: none"> <li>・アーカイブサービスにおいて以下のことを行った。       <ol style="list-style-type: none"> <li>1) 以下の2件のデータベースについて、アーカイブサービスにおいて公開した。           <ol style="list-style-type: none"> <li>①Dcty_cDB - Dictyostelium cDNA Database、②AcEST (Adiantum capillus-veneris EST)</li> </ol> </li> <li>2) 7件のデータベースについてアーカイブテストサイトを構築中である。</li> <li>3) アーカイブサイトトップページにデータベースリストを表示する等の改訂を行い公開した。</li> <li>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の設計を行った。</li> <li>5) アーカイブサイトの英語化の作業を開始した。</li> </ol> </li> <li>・国内外の蛋白質関連DBのUniprotIDに基づく統合検索システムのプロトタイプについて、センター内からの多数の意見を取り入れインターフェースを改善した。</li> <li>・文部科学省の委託研究プロジェクトについて、予算情報の更新を行った。</li> </ul>
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中。一覧表自動作成システムも運用している。</p> <p>2. 事業サイト等を更新しつつ継続運用中。</p> <p>3. メタデータエレメントサイトを継続運用中。</p>
産総研CBRC	<ul style="list-style-type: none"> <li>・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をモデリングするワークフローを開発する。</li> <li>・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。</li> </ul>	<ul style="list-style-type: none"> <li>・タンパク質立体構造予測ワークフローの開発及び動作検証を終え、このワークフローを11月末にCBRC情報統合サイトにて公開した。</li> <li>・アクティブ・ワークフローに向けた開発環境の整備・構築を行っている。このワークフローとして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施し、XサーバやWINDOWS上での動作確認を実施している。またCBRCにて独自に開発された解析ツールのアクティブ・ワークフロー化に向け、プログラムの改良を行っている。</li> <li>・CBRC情報統合サイト(<a href="http://togo.cbrc.jp">http://togo.cbrc.jp</a>)のリニューアルを行い、日本語・英語版ともにWEBにて公開し運用している。</li> </ul>

機関名	本年度成果目標	12月末進捗状況
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普及:登録ユーザ100人規模            ・ID変換サービスを統合したゲノムデータベースの開発と公開            ・Biomartシステムによるゲノム情報統合:シロイヌナズナを含む植物ゲノム3種と植物関連微生物34種類</p> <p>2. ゲノムアノテーション情報            ・Gene Indexing型アノテーション:4万件蓄積追加            ・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開発と公開</p>	<p>1. 高度情報集積DB            ・登録ユーザ数78人(2009年12月15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> <p>2. ゲノムアノテーション情報            ・植物関連生物11種の5613報の文献からGene Indexing型ゲノムアノテーションを145890件抽出蓄積し公開済(2009年12月9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>
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。            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。            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 専門用語辞書システムの表示項目のカスタマイズ機能を実装した。            2. 専門用語解析技術の開発 今年度前半に行った文字単位の決定性係り受け手法に対して、一般的な統語解析アルゴリズムの拡張による全探索解析手法の設計を行った。            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>
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。            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。            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。            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行っている。            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行っている。            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準備をしている。            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での表示法を検討している。</p>
中核機関 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>&lt;講義状況&gt;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演習を行った。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。&lt;受講者の進捗状況&gt;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プログラミング演習・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>
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名が受講。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、基礎統計を開講、データマイニングは、現在開講中。統合DB事業へ従事するものとして、昨年までの受講生から、統合TVに従事するものとして2名、講義・演習のTAとして4名が統合DB事業に従事。本年度の受講生の内4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。
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。            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。            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            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。            5. 1名をDBCLSの統合TV開発で活用。            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かを探索する方法: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」。            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(3回生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」を7月に完了し、シニア研究者の精査完了したデータをデータベースで公開を開始。また2回生270名を対象に実習授業「バイオ分野のデータベースの統合的な利用」を10月～11月に完了。他大学等で利用可能な自習用のテキストとして、最新版をインターネットを介して公開している。            2. 中級アノテーション教育(50名)は、次世代シーケンサーが産出した1000万件に近い、微生物ゲノム由来の大量断片配列データを対象に、配列解析実習を平成21年11月～平成22年1月に実施している。次世代シーケンサー由来配列解析用の学内用テキストを作成した。実習の結果を踏まえ、来年度に自習用の教材を公開するための準備に入っている。            3. 自己組織化マップによる相同性によらない生物系統の推定については、2名の卒研究生が約2,000件の遺伝子の生物系統を推定した。            4. シニア研究者と学部生の共同作業によるtRNAのデータベースに新たにウィルスゲノムのtRNAデータを追加し公開。農業生物資源研究所から依頼を受けて、イネゲノムのtRNAデータの精査も完了した。この更新バージョンを公開した。            5. 1件の統合TVコンテンツを作成し公開。現在1件の統合TVコンテンツを作成中。            6. 9月の日本遺伝学会に於いて、ミニシンポジウム「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かを探索する方法: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」を開催。統合データベースセンターの活動を紹介した。            7. 近隣の2高校で出前実習・授業を実施。ひらめきときめきサイエンスの実施ならびに、120名の高校生を対象に土曜講座を開きバイオ分野の様々なデータベースの実習他を実施。10月以降も高校への出前実習・授業を継続して実施。いずれの実習に於いても本プロジェクトで作成した自習用教材を使用し、あわせて統合データベースセンターの活動を解説している。            8. データベース実習(1回生50名)「バイオ実験技術に関しての統合的なデータベースの作成」用の学内テキストを作成した。平成22年1～2月に実施する実習の結果を踏まえて、自習用テキストを作成し、他大学や一般向けに公開するための準備を開始している。</p>

機関名	本年度成果目標	12月末進捗状況
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"> <li>・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。</li> <li>・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。</li> <li>・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。</li> </ul> <p>②ウェブ技術開発</p> <ul style="list-style-type: none"> <li>・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。</li> </ul> <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"> <li>・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。</li> </ul> <p>②化学情報データベース開発・運用</p> <ul style="list-style-type: none"> <li>・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。</li> </ul> <p>③LinkDB開発・運用</p> <ul style="list-style-type: none"> <li>・随時追加、変更、更新を行い、併せて、検索機能の拡張を行う。</li> <li>・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。</li> </ul>	<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"> <li>・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におけるアライメントをより正確に表示するように改良。</li> <li>・酵素番号自動割り当てシステムの改良は引き続き検討中である。</li> <li>・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PathPredプログラムとして <a href="http://www.genome.jp/tools/pathpred/">http://www.genome.jp/tools/pathpred/</a> から公開した。具体的には環境物質の分解経路と植物の二次代謝経路で使われている反応の変換ルールを使って予測するシステムとして構築した。</li> </ul> <p>②ウェブ技術開発</p> <ul style="list-style-type: none"> <li>・階層分類などを考慮した検索機能については引き続き検討中である。</li> </ul> <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"> <li>・JAPICデータの更新は引き続き順調に進めており、12月末で医療用医薬品13,128件、一般用医薬品12,190件の情報がデータベースに登録されている。副作用情報については米国FDAが提供するAERSとKEGG DRUGとの対応付けを行った。どのように見せるかについては引き続き検討が必要である。</li> </ul> <p>②化学情報データベース開発・運用</p> <ul style="list-style-type: none"> <li>・新たにMassBankとKEGG COMPOUND、HMDBとKEGG COMPOUND/DRUG、SIDERとKEGG DRUGとの対応をとり、リンクに基づく統合化を実現した。12月末でMassBankへは8,400件、HMDBへは約1,500件、SIDERへは約400件のリンク情報を提供している。</li> </ul> <p>③LinkDB開発・運用</p> <ul style="list-style-type: none"> <li>・随時追加、変更、更新作業は順調に進んでいる。検索機能の拡張については、ユーザー定義のリンク情報を使った検索機能をコマンドラインレベルで実装した。</li> <li>・LinkDBのリンク情報を検索しやすく、KEGGの各エントリー表示時にLinkDBの検索結果リストができるようにウェブインタフェースを改良した。</li> <li>・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、ProNIT、ProTherm、長浜バイオ大のデータベースとKEGG GENESとの対応関係を定義し、LinkDBに組み込んだ。遺伝研のPMDへのリンクは引き続き検討中である。</li> </ul>
分担機関  医科歯科大Gr	<p>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</p> <p>プロトタイプの評価を通して、疾患オントロジー及びプロトタイプとして開発したセマンティック検索エンジンの高度化、特に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の高度化を行う。</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</p> <p>症例に基づいた統合検索システムを構築し、東京医科歯科大学の統合検索システムの登録研究者に向けて公開を行う。</p> <p>3. プロジェクトの統合的推進</p> <p>がん疾患とパーキンソン病を中心として、国内の主要な疾患データベースを参照できるような仕組みを検討する。</p> <p>特にパーキンソン病に関しては、パーキンソン病の治療ガイドライン、パーキンソン病患者の会、治療薬提供製薬会社等のホームページで公開されている関連情報へのリンク付け、または中核期間Webサイト内DB横断検索サービス内へのAPIを介した連携など、利用者が必要とする情報に効率的にアクセスできる仕組みを検討する。</p>	<p>【医科歯科大】</p> <p>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:セマンティック検索エンジンの高度化に向けたインターフェースの改修案に優先順位をつけ、それに基づいて改修を開始した。</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:倫理案に準拠したフリーアクセスレベルのデータベースを作成し、その公開を1月初旬より試験的に開始した。また、倫理案については、阪大グループとの打ち合わせを開催し、アクセスレベルの記述等、最終案作成に向けた追記、修正を行っている。</p> <p>3. プロジェクトの統合的推進:阪大グループと打ち合わせを行い、検索機能の高度化のための仕様を決定し、開発を開始した。また、引き続き、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準備を進めている。</p> <p>【阪大】</p> <p>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: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を実施し、詳細仕様の最終調整を行った。その仕様に基づいて、統合検索のためのインタフェース開発作業を進めている。</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、データの公開準備を進めている。</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検討し、機能拡張の準備を進めている。</p>
東大医学部Gr	<p>1. 標準SNP DBの構築</p> <ul style="list-style-type: none"> <li>健常者700検体以上(Affy 500K, 6.0, illumina 515Kなど)</li> </ul> <p>2. GWAS(ゲノムワイド関連解析)DBの構築</p> <ul style="list-style-type: none"> <li>・CNVのケースコントロールDBの構築。Epistasisの解析手法開発。</li> </ul> <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> <ul style="list-style-type: none"> <li>・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)</li> </ul>	<p>1. GWAS(ゲノムワイド関連解析)DBの構築</p> <p>CNVのDBにおいて、個体ごとのCNV情報をクラスタリングして登録する機能を実装した。また、新たに11種類の情緒的気質、及び、自閉的性質の量的形質のGWASデータを登録した。更に、GWAS-DBへの登録の呼びかけを学会と論文を通して行った。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を開発した。</p> <p>2. 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のリシークエンスDBを公開するとともに、厚生省の神経変性疾患の班会議にて変異データの登録を広く呼びかけた。また、パーキンソン病のリシークエンスDBの実験データの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の文献データの登録を行った。</p> <p>3. 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・再配布に関する倫理指針を広く外部へアナウンスするとともに、実際のデータの再配布の運用を開始した。</p>

機関名		本年度成果目標	12月末進捗状況
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 1. シロイヌナズナオミックス情報注釈 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国際的な合意を得つつデータ公開を進める。</p> <p>2. 高等動植物等由来蛋白質構造のアノテーション 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アノテーション、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 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国際的な合意を得つつデータ公開を進める。</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 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うためのシステムを運用・改善し、作成されたアノテーションを公開用データに変換。</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) Plant Ontology (PO), PATOといった公開オントロジーを用いた新しい統一的なアノテーションを実施し、PO/PATOの組み合わせで115種類の表現型を定義した。Fox-hunting, DS transposon, Activation tagging lineの3データベースそれぞれの13142件,221件,1268件のデータに対し、新しいアノテーション情報からのリンクを付加した。また、それ以外のデータとして、マウスの表現型情報についても、国際的な活動体であるInterPhenomeにおいて、同様のオントロジー体系とデータ公開をする国際的な合意を形成し、データ公開を行った。</p> <p>2. 高等動植物等由来蛋白質構造のアノテーション(横山) 高等動植物等由来の蛋白質構造データ約60件分に付随する実験情報について、各々の位相決定に用いた回折実験データに加えて、結晶化実験データ数千件についても集積中である。また、それに伴い、位相決定に用いるか否かの判断のもととした更に広範囲の回折実験データについても付加中である。</p> <p>3. 微生物由来蛋白質構造のアノテーション(国島) 重原子導入蛋白質に関わる情報約2,000件分について、データの集積及びクレンジングが完了した。微生物由来蛋白質並びに変異導入蛋白質に関わる実験情報は、回折実験データを中心に、クレンジング作業中である。プレスリリースの内容及びこれら進捗に関し、日本生化学会年会において発表および意見交換を行った(平成21年10月21-24日)。さらに、公開後の意見交換をもとに、精製実験データ構造などの変更を考えた。</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(豊田) 上記の項目1の成果について、2009年12月に開催された分子生物学会年会においてポスター発表した。所定の表現型について、関連する遺伝子情報を実験手法等の違いに関わらず検索可能となるようにシステムの改良を行った。また、項目1の全データをRDF形式に変換して統合データベースセンター(かずさDNA研究所のグループ)に提供した。</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目指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・木曾研が協力に好感触であり、どこまで協力して頂けるのか検討中。</p> <p>2. アジア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を構築中。野口研究所のAPIは完了。理化学研究所、創価大学のAPIを1月末日に産総研へ納品予定。</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る。プロトコルの原稿を依頼しているところ。3月末日を目処に46種類のプロトコルを整備。糖鎖合成DBに関して今年度開発のAPIと野口研独自に開発しているツールと連携して合成可能な糖鎖化合物のDBをJCGGDB側に構築する。</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(病態の情報に関する項目で著作権者に許可をもらう作業はペンディング)病態を除いて公開する予定。</p> <p>6. 糖鎖と感染に関する情報を収集。DB化してPACDBとして公開間近。構造DBと連携するところは完成。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中。</p> <p>7. 完了。</p> <p>8. 専門用語の整備を継続。センテンス・スプリッターが機能するように前処理が必要でその作業を終えたところ。品詞を判定するEnjuを利用して用語を収集する作業に入る。そのため横断検索の和英・英和に未対応。手動で集めた単語などを利用した統合検索技術開発の目処がたった。</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 ・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 ・登録公開実績(累計)【国内】受付89件、公開17件;【国外】受付2件、公開1件(12月22日現在)。 ・NCBI Short Read Archive (SRA) とメタデータの交換を開始した。 ・簡単な検索機能のついた、DRA 受付データの公開サイトを開設した(10月)。http://trace.ddbj.nig.ac.jp/registered/ ・登録受付ウェブシステム D-way の本運用を開始(11月)。登録者はログイン後、新規登録作成、登録の進捗状況確認、登録履歴の参照やメタデータのアップロードなどができる。 https://trace.ddbj.nig.ac.jp/D-way/ ・高度なメタデータ作成支援、オンラインでのメタデータバリデーションを実現するため、Flash 技術をベースにしたメタデータ作成支援ウェブツールを試作した(12月)。</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750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900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システムを作成し試験を行った。</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ている。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いる。</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継続して行った。</p>	