

| 機 関 名 | 本年度成果目標 | 本年度進捗状況(年度末時点) |
|------------------|--|--|
| 中 核 機 関 | <p>DBCLS</p> <p>1. 戦略立案・実行評価 ・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> <p>2. 統合データベース開発 ①共通基盤技術開発 ・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 ②ヒト統合DBの開発・運用 ・文献解説: 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 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) ・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> <p>3. 統合データベース支援 ①ポータル整備・運用、広報、普及啓発(ポータル整備運用) ・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 ②データベースの受入と運用 ・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> | <p>1. 戦略立案・実行評価 ・サイトポリシー見直し完了。英語化も完了。科学データ所有権、知財ポリシーに関する検討を進めた。 ・学会誌1誌公開(ポータルの項参照)。さらにもう1誌、交渉完了。医学系総説誌交渉引き続き検討中。 ・GWASデータ受入れ、運用方針案作成、暫定終了。次年度検討委員会に諮る。</p> <p>2. 統合データベース開発 ①共通基盤技術開発 ・MEDLINE の各タイトルおよびアブストラクトから遺伝子名、タンパク質名、化合物名、病気名、薬剤名、酵素名、症状名を抽出し、その結果を利用して、関連する PNE 記事を検索するシステムを ツール組み合わせインタフェイス Galaxy に組み込み、内部公開(4月)。 ・作成したプロトタイプの性能評価を行い、その結果に基づき 遺伝子名抽出精度を向上させた。 ・Galaxy の調査などで洗い出された問題点について、TogoWS や PNE 関連記事検索がより使いやすいように改良。 ②ヒト統合DBの開発・運用 ・主要3論文サイト(PubMed Central, Blackwell Synergy, Science Direct)及びそこに含まれない36個の個別雑誌書式への対応を達成しソフトウェアを開発終了した。また、2項関係エディタを当初仕様どおりプラグインとして開発終了した。 ・遺伝子名辞書を遺伝子(意味)単位の管理から名称単位の管理法に切り替え、「外部データ(辞書やDB)名称と外部データ中の名称の属性の呼び名(正式名、別名等)」の組を意味として管理するデータ形式を新たに開発し、HGNC、NCBIgene UniProt の3つのDB中での記載名称をデータ化した。管理名称数は635,329件(ユニーク: 168,735件)である。 ・既存の3Dデータを修正し、新たに3Dデータを追加し、フェーズ2(アーティストフェイズ)解剖用語数計642を一般公開配布した。また、データ管理用のBodyParts データベースのデータ編集履歴管理機能を強化した。新規アナトモグラフィビューアーを開発し一般公開した。 ・Wired-marker は21回の機能向上と公開を行った。履歴はhttps://addons.mozilla.org/ja/firefox/addons/versions/6219参照。主な機能改善は表示URLの自動キャッシュとDB化である。またPDFへの対応を開始した。 ・INSDCの4回のmajor update に対応しDB総覧のデータを4回更新した。また、GEOの不定期のupdate に対し GEO目次の1 回のupdate を行った。 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) ・アノテーションレベルを3階層に設定した微生物ゲノムアノテーションパイプライン(Microbial Genome Annotation Pipeline (MiGAP))のうち、配列を全自動で解析してアノテーションを付与するb-MiGAPを開発した。3月にDBCLSで内部公開、センター外でもIPアドレスの登録済みの端末から使用可。</p> <p>3. 統合データベース支援 ①ポータル整備・運用、広報、普及啓発(ポータル整備運用) ・DBカタログは、前年度400DBに196DBの情報を追加し合計596DBの情報を公開した。インタフェースを他のサービスと統一し変更した。 ・横断検索 は、前年度16DBに220DBを追加し計236DBを対象にした横断検索を実現した。必要な検索結果を容易に取り出せるようカテゴリー分類や絞り込み機能を追加した。 ・蛋白質核酸酵素全文検索、ゲノム研究報告書検索、新聞記事検索、横断検索内での特許検索等、日本語文献検索サービスを開始した。(広報) ・7/22ゲノムテクノロジー第164委員会研究会を共催[活動紹介およびデモ]。10/15-17BioJapan[展示]。12/9-BMB2008[展示及びシンポジウム]。 3/28,29農芸化学会[ランチョンセミナー及び展示]。これら展示会等で下記パンフレットを3800部配布し、アンケート等を通じて184名の新たな成果評価者を獲得。 ・LSDB/DBCLS紹介パンフレット(A4サイズ8面)作成。各サービスメニューのリーフレット作成(日英27種類)。ニュース配信(日経BTJ 13件、LSDBウェブサイト[DBCLS, JST 一部ROISと連動] 61件)。英語で公開が可能なサービスや情報について英語版ウェブサイトを整備公開。 ・18年度成果評価結果公開ページを作成。19年度成果に対する評価を92名の評価者に依頼して実施し、結果を公開。寄せられたご意見への対応についてセンター内の検討結果を取りまとめ、12月初旬公開。 (普及啓発) ・統合TVを新たに 96件開発(トータル155件開発)。また、講習会を6回開催。 ②データベースの受入と運用 ・基盤づくり: 受け入れDBの運用システム「生命科学系データベースアーカイブサービスサイト」を公開した。メタデータ仕様と受け入れDB、、TogoDBにより検索して絞り込んだデータをダウンロード可能にした。さらに、DB寄託に関わる利用許諾作成マニュアル、公開に関する覚書書式等を公開した。 ・DBの受入れ: 総計12件のデータベースについてメタデータとデータを受領した(産総研糖鎖医学センター(1件)、九工大(2件)、理研(4件)、JST(1件)、かずさDNA研(4件))。主要プロジェクト一覧の英語化(全35件)、及びデータ内容の追加(10件)、NAR誌収録DBの情報調査(77件)、及び東京医科歯科大学実施アンケートのDB情報からのDBカタログ登録(15件)を実施した。 ・蛋白質統合DB構築: 国内外の蛋白質DB情報を収集し、約40DBの横断検索プロトを構築。また、蛋白ID統合のため、上記DBからのID情報抽出可能性を調査。</p> |
| 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 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 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> | <p>1. 意見集約システムの運用: WINGProの公開継続。新規11DB追加。国内外431DB。 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: 継続運用実施。参画研究者のための情報交換サイト(掲示板)を作成。 3. データベース受入れ Mouse EmbryoのEST メタデータを作成し、データとともにROISへ提出。メタデータエレメントサイトのため、FuGEおよびMAGE-TAB Specificationのデータ作成。今後搭載予定。</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) 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) 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> | <p>1のワークフローは右に記載した通り、8月に完了済。 2のワークフローは開発を完了し、12月26日に一般公開した。 3のワークフローは1月から開発開始、3月末に一般公開する予定。</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 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> | <p>1. 高度情報集積DB インターフェイスに文献書式を摸した、Gene Indexing Veiwを追加するなど使い勝手を改善した。アクティブユーザは現在72人。DBを活用した解析例を国内外学会等で発表し周知をはかり利用を促進している。発表は国際学会3回: Genome Informatics workshop (CSHL/Sanger Inst.), The 5th Rice Annotation Project Meeting (AIST), Biohackathon 2009, 国内9回: 進化学会、植物細胞分子生物学会、明治大学ワークショップ、植物微生物研究会、育種学会、ゲノム微生物学会、統合DB講習会4回の、計12回に達している。 2. ゲノムアノテーション情報蓄積 ゲノム位置情報と論文記載情報の統合のための情報蓄積をすすめている。Gene Indexing型ゲノムアノテーションを106974件蓄積し公開済(2008年3月12日現在)。</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 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 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> | <p>1. 専門用語辞書システム 予定していた項目(同義語に同じ識別子をもたせることにより、表記ゆれや別表記の語の検索を可能にする機能、および、複合語の内部構造のタグ付けと構造の管理機能)の実装に対して、実データでの作業を通じてエラー修正や効率化等の改良を実施。 2. 専門用語解析技術 病名および体の部位に関する専門用語の内部構造解析を文字単位の係り受け解析として実装し、60%程度の精度を確認した。並列構造解析については、設計した解析手法の改良を行い、60%強の解析性能を確認した。 3. 専門用語タグ付け手法 専門用語の意味クラス分類手法の設計とプロトタイプシステムの実装を行った。</p> |

| 機関名 | 本年度成果目標 | 本年度進捗状況(年度末時点) |
|---------|--|--|
| 九州大 | 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。 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 | 1. 現在「応用ゲノム」で行われているAffymetrix SNP Array 6lによるタイピングのquality checkを一貫して行うためのシステムとして、海外先行研究で採用されているGAIN QCソフトウェアを組み込んだパイプラインを作成した。2. 「応用ゲノム」からの解析生データ200件を受入れ、これを材料として上記パイプラインを用いたq.c.及びモデル関連解析を行った。3. これらの過程/結果を共有するためのweb上のデータベースを構築し、限定開示した。 |
| 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 | 延べ30名を対象にDB高度利用者の養成プログラムを実施した。演習においては統合DBプロジェクト内で開発しているTogo Web Serviceの成果を利用し、ネットワークを通じたデータ利用の講習を行うことで、即戦力を育成するカリキュラム開発を行った。講習内容は http://togodb.sel.is.ocha.ac.jp から公開している。受講生の内、2名が講習内容を公開するサーバのコンテンツ作成と管理に、2名がDBCLSの統合TV開発で作成者として活躍している。 |
| 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 | <講義状況>4月～3月の間に25回の講義を行った。4月～7月は9名の受講者に対して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った。8月には東京大学理学部生物情報科学学部教育プログラムでJava言語によるプログラミング演習を行い、9月からは8名の受講者に対してEnsemblゲノムブラウザーをサーバーにインストールする演習を開始した。さらに10月からはEnsemblゲノムブラウザーのインストールや設定を進めながら、6名の受講者に対してUTGBゲノムブラウザーに新しいデータを表示するトラックを作成するためのプログラミング講習も行った。<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までは受講者8名が進み、さらにEnsemblゲノムブラウザーの基本部分のインストールまで進んだ。受講者中2名はEnsemblゲノムブラウザーの基本部分のインストールを終了しトップページの表示を確認するところまで進んだ。また6名の受講者が独自データベース構築に必要なウェブアプリケーションプログラミングを学習した。 |
| 長浜バイオ大 | 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 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 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 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 5. 1名をDBCLSの統合TV開発で活用 | 1. 初級アノテーション教育(260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6月に完了、作成した教材を8月に公開開始。この教材を滋賀医科大学1回生の講義に使用。現在 DB構築中で4月末に公開。 2. 中級アノテーション教育(50名)は新規微生物由来のドラフトゲノム配列を対象に、平成20年11月～平成21年1月に実施。 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名の卒研生が約1,000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のデータベースに約14万件のtRNAを収録し公開を行い、Nucl. Acids Res.のDatabase Issue用の論文が公開された。2008年度更新用データとして、原核生物429種、25,232遺伝子についての精査が完了。 5. 5件の統合TVコンテンツを開発し、公開を開始。 |
| 分担機関 | 京都大学 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 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 | 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: 高速化は一通り完了し、高機能化に向けて機能拡張を実装中。具体的には、部分構造検索プログラムSUBCOMPIにおける、クエリ構造に含まれる部分構造の検索を引き続き実装し、テスト中。また、光学異性体のチェック機能についてもテスト中。 ・化学反応ネットワーク予測: システムの効率化は引き続き検討中であるが、テスト環境にウェブシステムの構築を開始。反応中の基質一生成物ペアの抽出アルゴリズムについては引き続き開発中。 ・酵素番号自動割り当て: 改良版を公開。ゲノム情報とのリンク方法については引き続き開発中。 ②ウェブ技術開発 構造検索とパスウェイ検索、階層分類検索との統合を実現し、公開。 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は引き続き順調に進行中。ライフサイエンス辞書のシノニム情報を用いた同義語・類義語検索を実装し、公開。副作用情報など付属情報については、KEGG BRITEの階層分類情報を使う方式を引き続き開発中。 ・脂質DB等に関してはLipidBank、LIPIDMAPS、KNApSACkとCOMPOUND、DRUGとの対応関係を引き続き更新中。LipidBankに関してはキーワード検索ファイルを提供してもらい実装。KNApSACkについては構造検索の対象に追加。病原性に関しては引き続き関係する生理活性物質のDB化を検討中。 ②LinkDB開発運用 随時追加、変更、更新作業は順調に進んでいる。キーワード検索などの機能拡張については引き続き検討中。 |
| 医科歯科大Gr | 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。 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。 3. 倫理規定草案の作成 | 1. 中間評価結果に従い、対象疾患を絞り肝細胞癌、大腸癌、パーキンソン病、口腔癌、頭頸部癌についてターミノロジー、オントロジーを整備した。 2. プロトタイプシステムの機能向上を行った。 ・癌症例として東京医科歯科大学250症例、大阪大学の神経疾患の症例400症例を統合した。 ・GeMDBJとの連携方式を検討し、癌症例約400症例の統合化を実現した。 ・検索GUIとして、パネル検索方式を実装した。 3. 国際的に広範囲な倫理規定の基本調査を行い、倫理規定草案を策定した。 4. システム化要件定義書を作成した。 |
| 東大医学部Gr | 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。 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 3. リシー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シークエンスデータベースを構築 | 1. 標準SNP DBの構築 Golden gate platformデータに対応した品質管理のための基準作成を行った。 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: 標準CNV databaseの構築を行った。 DB拡張: epistasisの登録表示機能追加、既知疾患関連遺伝子情報の登録表示、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のDB側の対応を行った 3. リシー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シークエンスデータベースを構築した。 |
| 補完課題 | 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。 2. 高等動物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。 3. 微生物由来蛋白質に関する試料調整(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。 4. 変異導入蛋白質に関する実験データ150件、重原子導入蛋白質に関するデータ500件を公開。 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)。 | 1. シロイヌナズナの発現、表現型、リソースに関する計6DBを統合化して公開を開始した。 2. 高等動物植物蛋白質の試料調整に関わる発現検討実験(無細胞蛋白質発現系構築実験3万件、スモールスケール発現実験4万件)に基づき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3千件、精製実験3千4百件)を行い、結晶化実験(結晶化条件: 80万件、結晶観察 881万件)へと進め、タンパク3000プロジェクトで解明された蛋白質構造データのうちH20年度は2万件の回折実験データ(画像枚数)を公開。 3. 微生物由来蛋白質に関する試料調整(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(データセット数)を公開した。 4. 変異導入蛋白質に関する実験データ150件、重原子導入蛋白質に関するデータ500件を公開した。 5. アノテーションシステムの開発運用をすすめ、変換データの中核機関へ提供できるようにした。 |

| 機関名 | 本年度成果目標 | 本年度進捗状況(年度末時点) |
|-------|--|---|
| 産総研糖鎖 | <ol style="list-style-type: none"> 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。 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。 キーワード並びに糖鎖構造による横断検索機能の開発。 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。 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)。 別途開発のAPIを用いた統合検索用プロタイプの構築。 | <ol style="list-style-type: none"> 参加機関が12月には決まり、連携の方法を話し合った。 既に糖鎖統合DBで使用しているHyperEstrailerのインデックスを提供している。 キーワード検索を完了し公開している。糖鎖構造による横断検索機能も完成しているがセキュリティチェック等の作業のためH21年度初めに公開予定。 構造検索の結果にレクチンの認識の有無や抗体の認識の有無などのアノテーションを付けるプログラムを開発した。その結果をどのように表示するかは未定。表示させるための方法を考案中。 名古屋大学の古川氏と協力し合い、登録用のシステムを構築。ノックアウトマウスのデータを入力しているところ(現在10件程度の情報)。平成21年度は登録の作業も続ける。 GlycoEpitopeのAPIを開発・設置した。横断検索から多次元HPLCデータベースの詳細画面にリンクできるように改良した。連携先のDBに検索をかけて結果をXMLで返すAPIを設置した。連携する際に各DBの表記揺れなどを吸収できるように専門用語の整備を行っている(平成21年度も継続)。 |
| 遺伝研 | <ol style="list-style-type: none"> 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。 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。 | <ol style="list-style-type: none"> トレース用FTPサーバを構築し、2008年8月に受け付けた2件のデータのFTP公開を開始した。 トレースデータ検索サイトを構築し、2009年4月より外部公開を開始した。キーワード検索、データダウンロード、統計情報の閲覧が可能になった。 登録処理システムでは、データチェック、DB投入処理システムを開発した。データ受付、投入が可能になった。 波形表示プログラムを開発。検索サイトより波形情報の閲覧が可能になった。 |
| 九工大 | <ol style="list-style-type: none"> 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。 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。 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。 | <ol style="list-style-type: none"> 蛋白質と変異体の熱力学データと構造データの対応表は、これまでに約1,100件(クロスレファレンスの総数は33,000件)のデータについて作成した。 蛋白質と核酸の相互作用の熱力学データの対応表は、これまでに約7000件(クロスレファレンスの総数は6,100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、試験用のデータベースシステムを作成。熱力学データ用XMLフォーマットは、蛋白質と核酸の相互作用の熱力学データについて暫定版を作成し公開した。また、熱力学データのControlled Vocabularyを作成した。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、DBセンターと連携して、文献、蛋白質名、キーワードのリストや、マーキングした論文のサンプルなどの提供、DBセンターが開発したテキストマイニングシステムTogoDoc、Wired-Markerの評価などを行った。 その他、プロジェクトのホームページを作成してWeb上に公開した。 |