

| 機関名 | 本年度成果目標 | 最終成果目標 |
|--------------------------|---|---|
| <p>DBCLS</p> <p>中核機関</p> | <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ースとして、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。胸郭内臓器の臨床レベル化と筋肉系のア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。 ・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。 <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"> ・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、広報、普及啓発</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。(広報) ・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: バンフレット更新、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(普及啓発) ・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"> ・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、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> | |
| 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。 ・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。 | |
| かずさDNA研 | <p>1. 高度情報集積DB</p> <ul style="list-style-type: none"> ・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>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: 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 | |
| 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> | |
| 九州大 | <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。</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>④ DHaploDB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> | |
| 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。 | |
| 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。 | |

| 機関名 | | 本年度成果目標 | 最終成果目標 |
|------|---------|--|---|
| 中核機関 | 長浜バイオ大 | 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」。 | |
| | 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。 ②ウェブ技術開発 ・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。 2. 統合データベース開発・運用 ①医薬品データベース開発・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、副作用情報など付随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。 ③LinkDB開発・運用 ・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。 | 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。 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。 |
| 分担機関 | 医科歯科大Gr | 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 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。 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 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。 3. プロジェクトの統合的推進 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。 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。 | 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。 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。 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。 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。 |
| | 東大医学部Gr | 1. 標準SNP DBの構築 健康者700検体以上(Affy 500K, 6.0, illumina 515Kなど) 2. GWAS(ゲノムワイド関連解析)DBの構築 CNVのケースコントロールDBの構築。Epistasisの解析手法開発。 3. リンクエンスによる臨床情報・ゲノム情報DBの構築 パーキンソン病のリンクエンス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ンクエンスDB構築(2009-2010の2年計画) | 1. 標準SNP DBの構築 健康者700検体以上(Affy 500K, 6.0, illumina 515Kなど) 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。 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。 |
| 理研 | | 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。 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。 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。 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。 | 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。 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。 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。 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)。 |
| | 補完課題 | 産総研糖鎖 | 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 2. 公開している糖鎖データベース(JCGGDB)の普及活動を行う 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 4. オンラインプロトコルの整備を行う 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 6. 糖鎖機能に関する情報を発信する 7. 横断検索用インデックスの更新システムを開発する 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) |
| 遺伝研 | | 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。 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。 | 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 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 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 |
| | 九工大 | 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。 2. 蛋白質・核酸相互作用データをXML化し公開する。 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。 | 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参照を可能とする。 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。 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。 |