

機関名	本年度成果目標	本年度進行状況(9月末時点)
<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"> 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: 次世代シーケンサ、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 <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 生命科学系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"> 公開サイトに関する著作権、個人情報からみのサイトポリシーを実情に合わせて随時再検討している。 医学系総説誌交渉検討中。学会誌2誌公開具体化検討中(ポータル整備運用の項参照)。 医用リシーケンスに関する情報セキュリティの検討を行い、併せて次世代シーケンサの調査を行った。 <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"> MEDLINE 検索システム、PNE検索システムと日英翻訳手法を調査した。 J. Biol Chem の論文から、らん藻に関連する87本を選び、文中に書かれた遺伝子名をタグ付けしたコーパスを作成。 Yahoo pipes!, UCSC Galaxy など、既存のツール組み合わせインタフェイスの調査を行った。 <p>②ヒト統合DBの開発・運用</p> <ul style="list-style-type: none"> 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完了。 種々のソースから得られた遺伝子名辞書を一元管理するための用語とリソースを一行で格納するシステムを構築中。 辞書とデータの管理システム構築。医師との協業でコメント付を進める。一点の特定と情報の埋め込み法を検討中。 キャッシュ機能を開発中。これによりオリジナルリンクが切れても大丈夫になる。累計ダウンロード60,000件。 バンク目次更新継続中。 <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"> 微生物ゲノムのアノテーション用パイプラインの仕様検討を行い、初級、中級に関する仕様を確定した。 <p>3. 統合データベース支援</p> <p>①ポータル整備・運用、広報、普及活動(ポータル整備)</p> <ul style="list-style-type: none"> 生命科学系DBカタログの拡充は、NARのDB issue からデータを作製中。既存レコードの記載内容に関する追加、修正は今後の課題。 生命科学系DB横断検索については順調にDBを追加しており、KEGG DRUG, PIRなど重要なDBの他、微生物DBの検索が可能になった。また特許公報のみであった特許検索に公開・公表特許公報を追加した。現在、タンパク関連DBを中心に索引を作成中である。検索精度の向上について、利用者が望みのDBについて検索結果を取捨選択できるようなインタフェイスを現在設計中である。 コンテンツの拡充について、利用者より要望の高い学会要旨統合検索や、蛋核酵に続く日本語の総説誌の公開を行わなければならないが、今のところ生物物理学会に許諾をいただいたのみで、今後、相当な努力が必要である。(広報) 7/22ゲノムテクノロジー第164委員会第27回研究会を共催し、活動紹介およびデモを実施。10/15-17BioJapanにて展示。12/9-分生にて展示(予定)。 LSDBおよびDBCLSを紹介するパンフレット(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25種類)。ニュース配信(日経BTJ 8件、LSDBウェブサイト[DBCLS, JST 一部ROISと連動] 26件) 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中(10月末公開予定)。(普及啓発) 統合TV: トータル102件開発 講習会: 3回開催(日大、DBCLS/東大、北大; 年度内に更に4回開催予定) <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完了、標準利用許諾作成完了、一括ダウンロードサイト第1バージョン構築完了 補完課題4テーマの各機関、及び経産省統合DBプロジェクト等)の間で、3つの観点(DBカタログ(メタデータ)、横断検索、データダウンロードサイト(利用許諾))での連携方針のもとに具体的な作業計画を策定。産総研からは3つ、九工大からは2つのDBのメタデータを受け入れ済みで、双方ともダウンロードデータを受け入れ準備中。経産省統合DBプロジェクトの間では、メタデータや横断検索、利用許諾、ウェブサービスなどでの共通化をはかることで合意。 蛋白質DB統合化に向けて、国内蛋白質DBのID等の調査を実施し、それをもとに方式を検討し、仕様案を作成した。
<p>JST</p>	<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>	<p>1. 意見集約システムの運用: WINGProの公開継続と新規7DB追加済み</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: 継続運用実施中</p> <p>3. データベース受入れ Mouse EmbryoのEST DB受入れのためROISよりフォーマット入手</p>
<p>産総研CBRC</p>	<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>	<p>1のワークフローはプロトタイプとして8月末に完了し、現在プロジェクト内限定公開(ユーザID、パスワードが必要)を行っている。ホームページはhttp://tgweb01.cbrc.jpで、DBCLSからもリンクされている。</p> <p>2のワークフローは現在開発中であり、進捗としては25%程度である。12月末には一般公開できる予定。</p> <p>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>
<p>かずさDNA研</p>	<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>	<p>1. 高度情報集積DB インターフェイスにタグ入力支援機能を追加し、ユーザの利便性を向上した。利用者は現在64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(計8回。国際: Genome Informatics、国内: 進化学会、植物細胞分子生物学会、明治大学農学部ワークショップ、植物微生物研究会、統合DB講習会3回)</p> <p>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69377万件蓄積し公開済(2008年9月24日現在)。</p>
<p>奈良先端大</p>	<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>	<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表示機能)の実装は完了。</p> <p>2. 専門用語解析技術 用語の内部構造解析データの作成は上記システムを用いて10月から開始の予定。データの作成後、用語の内部構造解析ツールを構築する。並列構造解析については、解析手法の設計がほぼ完了。今後、学習・テストデータの作成を進め、解析精度の測定実験を行う。</p> <p>3. 専門用語タグ付け手法 用語分類手法について検討中。本年度後半にプロトタイプシステムを構築する予定。</p>
<p>九州大</p>	<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>	<p>1. アレルタイプと頻度情報が個人特定に繋がりうるという最近の研究結果から、これを公開しないという米国NIHのガイドライン及び世界的共通理解に従い、本データベースでもこれを非公開とした。2. 現在「応用ゲノム」で行われているAffymetrix SNP Array 61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。3. 「応用ゲノム」からの解析raw data受入れ、これのq.c.及び共有化を行いつつある。</p>

機関名		本年度成果目標	本年度進行状況(9月末時点)
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。本年度はDBCLS2名、自治医科大1名、東大新領域8名	4～9月の間、合計12名の受講者は順調に演習をこなし、スキルを身につけている。
	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名を対象としてDB高度利用者の養成プログラムを実施中。DBCLSの統合TV開発で、受講者2名を活用中。
分担機関	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。 2. 中級アノテーション教育(50名)は新規ゲノム配列を対象に、平成20年11月より平成21年1月の期間を予定。 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600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用の論文がアクセプトされた。
	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: SIMGOMP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: 高速版SIMGOMPを実装し7月に公開。従来版の10倍程度の高速化を実現。 ・化学反応ネットワーク予測: 反応タイプを元にしたシステムを実装中。効率化を重点的に検討中。 ・酵素番号自動割り当て: スコアリング方法を開発し、予測精度の改良を実現した。今年中に公開する予定。 ②ウェブ技術開発 検索エンジンについては従来通りとし、構造検索システムや反応経路検索とキーワード検索との統合を検討中。 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順調に進んでいる。また DailyMed との対応も取り、公開している。副作用情報など付属情報については計算機で利用・解析に使える形での提供を検討中。 ・各種化合物DBのキーワード検索機能を7月に公開。現在、KEGG 以外に7つのデータベースが検索対象。 ・脂質DBに関しては、LipidBank、LIPIDMAPSとKEGG COMPOUNDとの対応を取りLinkDBで統合化した。また2次代謝産物DBのKNAPSAcKもLinkDBで統合化した。病原性に関しては、関係する生理活性物質のデータベース化を検討中。 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検討中。
分担機関	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. 現在肝細胞癌、大腸癌に特化して、使用されているターミノロジー、シソーラスなどを手作業にて収集作成している段階である。同時に自動作業化の試みも行っている。 2. プロトタイプシステムの機能向上に向けて、以下のような作業の途中である。 ・癌、神経疾患の症例を増加中である。 ・GEMDBJとの統合に向けて、具体的な連携方式とハードルの明確化を行っている。また、統合による効果の明確化をするための事例検討を行っている。 ・検索GUIについては、より直感的な操作を可能とする新たなパネル方式を立案し、試作している最中である。 ・検索エンジン設計の要素化、各要素技術のISO、WHOの場での国際標準化などを計画進行している。 3. 国際的で広範囲な倫理規定の基本調査を行っている。草案作成のためのステップ項目を策定し、各項目ごとの必要案を検討している。
	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、及びケーススペ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. 最新のplatformデータに対応した品質管理法を作成し、新しい標準SNPデータを作成済み、近々にDBに追加。 2. CNVの第一次解析を終了し現在品質検計中、またCNV用DBフォーマット案を作成済み、SNP-GWAS-DBの搭載データは着実に増加中、中核機関の計算機資源を利用。 3. パーキンソン病について、症例数増加中、文献情報収集登録中。
補完課題	理研	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。 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。 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と表現型に関するデータベースについて、オントロジー作成とIDの割り当てを行い、XMLやテーブル形式でのダウンロード用データを作成した。 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 高等動植物等由来の蛋白質構造データに付随する実験データ30件分について、各々の位相決定に用いた回折実験データを中心に集積中である。 3. 微生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 試料調整と結晶化実験データについて公開に向けてデータクレンジング作業を行った。重原子導入蛋白質に関するデータについて公開に向けたアノテーション作業を行った。 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群のハブサイトをdatabase.riken.jpとして構築し、理研内部での監査作業が終了したのち、公開できるようにシステムを構築した。また登録したデータをDB群のkeyword検索機能を実装した。
	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、統合化の手順を決定。 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。 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。 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。 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)。 6. 別途開発のAPIを用いた統合検索用プロタイプの構築。	1. 現在、数機関と参加機関との交渉を進めている。 2. 横断検索用index(産総研・立命館大・名古屋市立大・LipidBankの情報)は提供可能。 3. キーワードを利用した横断検索サービスを既に公開した。構造検索に関しては、EuroCarbDBのGlycoCTとCabosMLの融合したXMLを利用し糖鎖の表現の多様化と構造検索の高速化に対応中。 4. CabosMLのには対応できていたが、新しいXMLの変更に伴いツールの改良が必要なので未着手。 5. 登録用エディターと表示用インターフェースのプロトタイプは完了。名古屋大の使い勝手などヒアリングしながら改良を進める予定。 6. 産総研のGGDBは経産省側のプロジェクトで行い、それ以外は産総研で開発を行う予定。参画機関(立命館大学・名古屋市立大学)のDBのデータを引き出すAPIに関しては先方と話し合いながら仕様を決めているところ。決まり次第一般競争入札の手続きを進め、遅くとも2月中旬までに納品の予定。
補完課題	遺伝研	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。 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。	1. FTPサーバ構築済み。8月からddbjに登録された2件をftp公開開始。試験的に受付も開始。キーワード検索、ダウンロード、統計情報表示ページについてはプロトタイプサイトを検証中。 2. 登録処理システムについては、データチェック、DB投入部分について、プロトタイプを検証中。波形情報表示プログラムについてもプロトタイプを検証中。
	九工大	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。 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。 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。	1. 蛋白質と変異体の熱力学データと構造データの対応表については、これまでに約600件のデータについて作成した。 2. 蛋白質と核酸の相互作用の熱力学データの対応表については、これまでに約500件のデータについて作成した。 3. 蛋白質・蛋白質相互作用データについては、データベースのスキーマの設計を行い、データベースシステムとWebページを作成中。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プロトタイプを作成した。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開発をすすめている。