

## 平成22年度委託業務概要一覧

区分	機関名 (代表者名)	業務項目	「委託課題」	9月末時点において望む達成状況(目標)	目標達成に向けて注力していただきたい点	22年度 (千円)
			(注)DBはデータベースの略記			
中核機関	かずさDNA研 (中村 保一)	①高度情報集積DBの運用と改良	「統合DB開発:植物及び植物関連微生物のゲノム情報DB統合と高度化」 ①(9月末)KazusaAnnotationの利用者・運用者の利便性向上のため、DB応答速度、データ入力速度高速化を中心に開発。MicrobeBaseに蓄積したデータを種々の表現方法で閲覧可能。(年度末)遺伝子名等の食い違いを解決する対象を増やす。	・モデル生物・産業応用生物統合DBとして研究者コミュニティにおいて認知され、対象とする生物種、領域において中心的なツールとして機能している。 ・統合DB基盤システムのツールとして組み込まれている。	・多くの研究者に利用されるように進めていただきたい。	24,000
		②ゲノムアノテーション情報の蓄積と高度化	②シニア/バクテリア・根粒菌中心に、論文記載情報・実験情報の蓄積、ゲノム位置情報と論文記載情報の統合、公開。			
	産総研CBRC (浅井 潔)	①ワークフロー技術を用いた統合DB環境の構築	「統合DB開発:ワークフロー技術を用いた統合DB環境構築」 1)アクティブ・ワークフローに向けた要素技術開発 解析ツール(LAST, CentroidFold)、要素技術をアクティブ・ワークフロー向けにノード改良・開発(8月末公開予定)。 2)ワークフローによるプラットフォーム開発 既存ツールを組み合わせて、利用者要望を考慮した解析プラットフォーム開発(9月末公開予定)。 利用促進のための広報活動。	・研究者コミュニティにおいて、標準的なワークフローツールとして位置づけられる。 ・統合DB基盤システムのツールとして組み込まれている。	・どの程度利用され必要なツールとして認知されているのか不明。 ・ある程度開発を絞り、既存の開発ツールの利用促進に注力していただきたい。	19,400
		②ゲノムアノテーション情報の蓄積と高度化	②シニア/バクテリア・根粒菌中心に、論文記載情報・実験情報の蓄積、ゲノム位置情報と論文記載情報の統合、公開。			
	奈良先端科学技術大 (松本 裕治)	(1)専門用語獲得技術 ①専門用語辞書システムの開発	「統合DB開発:専門用語辞書管理システムと専門用語解析技術の開発」 (1)専門用語獲得技術 ①(9月末)10万語規模の生命科学用語の辞書格納。 Webブラウザ上で用語検索、シソーラスコード等の表示・修正可能な専門用語辞書システム実現。 (年度末)用語間意味関係表示機能の実現、システム完成。	・統合DBとして必要な辞書ツール、システムとして基盤システムに組み込まれている。	・本事業期間内に確実に実用レベルの成果が出るのであれば、業務内容を絞って進めていただきたい。 ・見込みがなければ、研究開発は中止し、現時点で整理をお願いします。	10,500
		②専門用語解析技術の開発	②(9月末)内部構造解析済みの学習データを2000語以上に拡大。 (年度末)統語解析法に基づくアルゴリズム実装。90%以上精度達成。			
		(2)専門用語抽出ツールの設計と開発 ①専門用語辞書拡張支援ツールの設計と開発	(2)専門用語抽出ツールの設計と開発 ①(9月末)新規語にシソーラスコード付与機能を追加。			
	九大 (林 健志)	1)JAGQC Pipelineのデータ充実	「統合DB開発:多型知識表現技術開発」 1)(9月末)GWAS1次スクリーニングデータ受入れ、QC解析、結果開示。 ゲノム多型研究者コミュニティに認知、有効なツールとして機能。	・国内のゲノム多型研究者コミュニティにおいて認知され、有効なツールとして機能している。	・研究者コミュニティに認知され、利用されるような取り組みに注力していただきたい。	4,950
2)日本人確定はプロタイプDBとGWAS DBの連携		2)(9月末)九大提供データの整備、相互のデータQC、東大へのハプロタイプデータ提供完了。 (年度末)東大Gと共同で日本人標準SNPはプロタイプDBを公開。				
東大 (森下 真一)	①独自のサーバー構築演習	「統合DB支援:DB構築者の養成」 (10月迄)演習終了。 (年度末)教材等を整理、公開。 H21年度受講修了者(4名)に大規模計算のためのクラスター利用技術を習得。 特に次世代シーケンシング情報のDB構築人材を育成。	・統合DBとして必要とする人材が育ち、本事業内で活躍している。	引き続き実践的な人材が輩出するよう進めていただきたい。	12,000	
長浜バイオ大 (池村 淑道)	①アノテーター・キュレーター教育	「統合DB支援:アノテーター・キュレーターの教育」 ①(9月末)DBの統合的利用、ゲノム配列実データのアノテーション実習、自習用テキスト公開用教材の提供。 大量ゲノム配列を対象とした有用遺伝子配列の探索、アノテーション実習、DB・テキスト教材公開。 (年度末)Short Read大量DNA配列のアノテーション実習。 決定された塩基配列への新規アノテーション付加。 実験技術原理・機器画像等の情報収集、自習用テキスト教材公開。 他大学・社会人・高校・高専生への講習会開催。	・統合DBとして必要とする人材が育ち、本事業内で活躍している。 ・シニア専門家の協力によるキュレーター、アノテーター人材育成モデルとしては、長浜バイオ大のみの取り組みにとどまらず、全国的に波及しつつある。	・引き続き実践的な人材が輩出するよう進めていただきたい。	12,000	
	②自己組織化マップによる大量ゲノム配列データへのアノテーション手法を用いた人材育成	②(9月末)環境微生物類等の大量遺伝子のアノテーション結果可視化、公開。 (通年)インフルエンザウイルスのアノテーション結果の可視化、公開決定された塩基配列のアノテーション結果の可視化。				
	③シニア研究者のチームによる分散型の高度なキュレーションシステムの改良・運用	③(9月末)新規登録の大量塩基配列データを対象にtRNA遺伝子DBの更新、検索機能向上。 (年度末)タンパク質遺伝子の高度キュレーションシステムの構築・運用。 シニア専門家のキュレーターグループ完成。 (通年)他大学・一般企業等人材、市民対象の講習会開催。				
お茶の水女子大 (瀬々 潤)	①DB高度利用者の養成	「統合DB支援:DB高度利用者の養成」 (9月末)H21年度迄構築したDB高度利用者養成カリキュラム、自習用教材整備・改善。 (年度末)カリキュラム実施時のフィードバックの自習用教材反映。 (通年)DB高度利用者養成カリキュラムの実施。 (9月末一部実施)本事業参画機関のデータ解析・DB整備等の業務経験。	・教材が整備されるとともに、統合DBとして必要とする人材が育ち、本事業内で活躍している。	引き続き実践的な人材が輩出するよう進めていただきたい。	11,500	

区分	機関名 (代表者名)	「委託課題」 (注)DBはデータベースの略記		9月末時点において望む達成状況(目標)	目標達成に向けて注力していただきたい点	22年度 (千円)
		業務項目	業務計画概要(主な目標)			
分担機関	京大 (五斗 進)	「ライフサイエンス知識の階層化・統合化事業」		①医薬品・化合物DBについて、統合DBシステムとしての基本的な骨格が完成。想定ユーザー(特に国内製薬企業研究者)が最低限必要とするデータが網羅されている。 ②想定ユーザー層に認知され、必要性が認められている。	・感染症、食品・環境について、どこまで実施すると達成したと言えるか不明であり、H22年度は実施を見送るのが望ましい。 ・さらに、実施内容に優先順位を付け、選択と集中をしていただきたい。	66,400
		(1)共通基盤技術開発 ①知識処理技術開発	(1)共通基盤技術開発 ①(9月末) 複数化合物入力の実現による化学構造検索システムの高機能化。化学反応NW予測システム整備、ゲノム情報関連付け。 (年度末)ゲノムネットケミカル情報解析サービスとして整備。			
		②ウェブ技術開発	②(9月末)化学情報DBのキーワード検索の高機能化。化合物情報、遺伝子・タンパク質情報、パスウェイ情報関連検索の仕組み整備。			
		(2)統合DB開発・運用 ①医薬品DB開発・運用	(2)統合DB開発・運用 ①(9月末)医薬品化学構造、相互作用、ターゲット、関連パスウェイの検索仕組み実装。 (年度末)知識処理技術開発で実装した技術を応用する仕組み整備。			
		②化学情報DB開発・運用	②各種化合物DBの統合。			
	③LinkDB開発・運用	③(9月末)LinkDB自動更新の仕組み構築。 ユーザー定義のリンク情報をLinkDBと組合せ検索する仕組み構築。				
		(3)プロジェクトの総合的推進	(3)プロジェクトの総合的推進			
	東京医科歯科大 (田中 博)	「統合医科学DB構築方式の開発」(がん疾患を中心として)		統合医科学DBのロールモデルが一疾患だけでも完成している。 ①統合医科学DBとしての基本構造(骨格)が9月迄に完成。 ②対象疾患について、HUBとして当該分野・領域で認知・位置づけられること。 ③対象疾患データの提供/リンクでの協力体制が構築できていること。	・倫理基準の確定がボトルネックであれば、これを最優先にし(9月までに完了)、この目処が立つまでは開発は控えるべき。 ・対象とするがんデータについては、オールジャパンの統合DBであることから、国内にデータを保有する機関との協力体制を構築すること。 ・さらに実施内容に優先順位を付け、選択と集中をすべき。 ・倫理基準を策定するにあたっては、DBCLS関係者も参加した協力体制を進めること。	18,300
		①疾患・臨床医科学DBに特化した統合技術開発に関する研究(がん疾患を中心として)	①疾患オントロジーの整備及びヒマンティック検索エンジン、インターフェースの高度化、基本構造完成。			
		②疾患DBの高度化と実証的統合に関する研究(がん疾患を中心として)	②(9月末)倫理規定に基づきアクセスレベル設定。統合検索システムのアクセス制御機能高度化、公開。			
	③プロジェクトの総合的推進	③(年度末)倫理規定を大阪大学と連携し最終案作成。 中核機関を含め検討、作成。国内の主要疾患DBとの緊密な連携を図る。利用者が効率的にアクセスできる仕組み構築。 連携希望機関に対して、必要な情報を提供する仕組み構築。 肝臓関連学会、大腸(消化器)関連学会等へ、HUBとして認知・位置づけられるよう働きかけ。 倫理委員会が承認する範囲内で、成果をHP、学会・講演会等にて情報発信。				
阪大 (山本 洋一)	「統合医科学DB構築方式の開発」(神経疾患について)		①(9月末)統合検索システムを介したパーキンソン病DBの検索機能の開発・改良作業完了。 (年度末)これまでに実施した開発作業及びDBの運用過程から得た知見を、統合医科学DBのロールモデルとしてまとめる。 ②(9月末)パーキンソン病の治療ガイドライン等関連情報とのリンク付け、関連機関への情報提供を完了。 (年度末)日本神経学会、大阪府の神経難病センター、NPO法人SCORE等を介して、広く認知されるための情報発信。 データ提供等の連携を希望する機関に対して統合を実現するために必要な情報提供の仕組み構築。 神経疾患(パーキンソン病)について公開患者データ数目標:約700。	・システムの改良等の開発は9月までとする。 ・さらに実施内容に優先順位を付け、選択と集中をすべき。 ・神経疾患(パーキンソン病)のデータについては、オールジャパンの統合DBであることから、国内にデータを保有する機関との協力体制を構築すること。	12,300	
	①疾患・臨床医科学DBに特化した統合技術開発に関する研究(神経疾患について)	①(9月末)統合検索システムを介したパーキンソン病DBの検索機能の開発・改良作業完了。 (年度末)これまでに実施した開発作業及びDBの運用過程から得た知見を、統合医科学DBのロールモデルとしてまとめる。				
	②疾患DBの高度化と実証的統合に関する研究(神経疾患について)	②(9月末)パーキンソン病の治療ガイドライン等関連情報とのリンク付け、関連機関への情報提供を完了。 (年度末)日本神経学会、大阪府の神経難病センター、NPO法人SCORE等を介して、広く認知されるための情報発信。 データ提供等の連携を希望する機関に対して統合を実現するために必要な情報提供の仕組み構築。 神経疾患(パーキンソン病)について公開患者データ数目標:約700。				
東大・医 (徳永 勝士)	「疾患解析から医療応用を実現するDB開発」(ゲノムワイド関連解析のデータベース開発)			・研究開発はほぼ問題なく着実に進んでいると判断する。 ・今後は、データ提供の仕組み作り、利用・普及に注力した取り組みをお願いしたい。	34,030	
	(1)標準SNP DBの構築	(1)標準SNP DBの構築 解析データの追加登録、提供用のデータの作成。 データの利用申請者へ所定の手続き・審査を経た上でデータ提供。				
	(2)GWAS DBの構築 ①GWAS 第1ステージDBの構築	(2)GWAS DBの構築 ①(9月末)登録データの拡充。CNVのGWAS結果登録・表示機能の拡張、及び既存GWASの量的形質データ登録への対応とインターフェース拡充。 データの利用申請者へ所定の手続き・審査を経た上でデータ提供。 (年度末)GWAS-DB、CNV-DBの改良、データの収集、提供について整備した方針の改良。				
	(3)リシークエンスによる臨床情報・ゲノム情報DBの構築 ①リシークエンスDBへの外部DBからのデータインポート	(3)リシークエンスによる臨床情報・ゲノム情報DBの構築 ①(9月末)家族性痙攣性対麻痺、副腎白質ジストロフィーのリシークエンスDBの第2次DBの残り部分であるデータ登録部の拡張実装。 (年度末)ユーザの意見を参考にしつつリシークエンスDBの改良。				
	(4)プロジェクトの総合的推進	(4)プロジェクトの総合的推進				
東大・病院 (辻 省次)	「疾患解析から医療応用を実現するDB開発」(リシークエンスDBの開発)		・ヒトゲノム多型・変異データについて我が国のRepository DBとなること。 ・できれば関係する研究者コミュニティにおいて認知され、データ提供の恒常的仕組みが形成されることが望ましい。 (例えば、日本人類遺伝学会の論文誌 Journal of Human Geneticsへの投稿に本DBへデータ登録義務を課す。)	特になし。	8,660	
	(1)リシークエンスによる臨床情報・ゲノム情報DBの構築 ①リシークエンスDBの構築	(1)リシークエンスによる臨床情報・ゲノム情報DBの構築 ①(9月末)家族性痙攣性対麻痺、副腎白質ジストロフィーのゲノム変異情報と臨床情報の拡充。 インターフェースの拡充。同疾患に関するゲノム変異情報、臨床情報の外部データ受付・登録。 (年度末)ユーザの意見を参考にしつつDBの改良。				
東海大 (井ノ上 逸朗)	「疾患解析から医療応用を実現するDB開発」(ゲノムワイドSNPの統計遺伝学的解析手法の開発)			特になし。	8,070	
	(1)標準SNP DBの構築 ①標準SNP DBの構築のための統計遺伝学手法の開発	(1)標準SNP DBの構築 ①健康日本人の品質管理に必要となる基準値検討。日本のゲノム医学分野における代表的なDBとして広く認知され、データ提供が恒常的になされるよう、日本人類遺伝学会学会誌の投稿に際し、本DBへのデータ登録を義務付けるよう働きかける。				
	(2)GWAS(ゲノムワイド関連解析)DBの構築 ①GWAS DBの統計遺伝学手法および解析ツールの開発	(2)GWAS(ゲノムワイド関連解析)DBの構築 ①(9月末)SNP間相互作用解析ツールを大規模なGWASデータセット等にも対応できる、より一般性の高いツールに開発。 (年度末)GWAS DB内のデータを対象とした解析を、これまで開発したツールを用いて行う。				
日立製作所 (小池 麻子)	「疾患解析から医療応用を実現するDB開発」(ゲノムワイドSNPの疾患関連解析手法の開発)			特になし。	10,617	
	(1)GWAS DBの構築 ①GWAS 第一ステージ疾患関連SNP探索手法の研究開発	(1)GWAS DBの構築 ①(9月末)Bioinformatics的候補SNP絞りこみ手法開発。CNVの検出力向上、CNVのGWAS精度向上。量的形質についての統計遺伝学的解析を行うパイプライン構築。 (年度末)パイプラインの操作性向上、CNVの検出力等の計算精度向上。				
	(2)リシークエンスによる臨床情報・ゲノム情報DBの構築 ①リシークエンスDBの解析手法の開発	(2)リシークエンスによる臨床情報・ゲノム情報DBの構築 ①(9月末)家族性痙攣性対麻痺、副腎白質ジストロフィーについて文献からの変異情報、患臨床情報抽出。対象疾患遺伝子の転写産物の立体構造予測等。 (年度末)ユーザーの意見を取り入れながらデータの追加等。				

区分	機関名 (代表者名)	「委託課題」 (注)DBはデータベースの略記		9月末時点において望む達成状況(目標)	目標達成に向けて注力していただきたい点	22年度 (千円)
		業務項目	業務計画概要(主な目標)			
補完課題	理研 (豊田 哲郎)	「植物オミックス情報および蛋白質構造情報」		・理研外の研究者から見て、理研のDBが統合されたミニマムモデルの成功事例として、研究者コミュニティにおいて認知される。 ・理研内においては、ミニマムモデルを展開することにより、理研内で保有するDBが順次、整備・統合される動きにある。	・理研のDB統合の成功モデルとして、研究者コミュニティに理解され、認知されるような取り組みにも注力して進めていただきたい。	63,000
		① シロイヌナズナオミックス情報注釈	① シロイヌナズナオミックス情報注釈 i. (9月末)トランスクリプトームデータを集積したりポジトリ公開。 (年度末)アノテーションデータ公開。 ii. (9月末)理研内外で観測されたフェノタイプデータの統合化、共通のオントロジーで整理・公開。 iii. (9月末)シロイヌナズナによる統合化モデルケースを、他の生物種(マウス等)にも適用し、DB統合化の範囲を拡大。			
		② 高等動物等由来蛋白質構造のアノテーション	② 高等動物等由来蛋白質構造のアノテーション (9月末)タンハク30007プロジェクトの高等動物等由来蛋白質構造データのアノテーション、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実施・公開。DBCLSIに成果データ提供。			
		③ 微生物由来蛋白質構造のアノテーション	③ 微生物由来蛋白質構造のアノテーション (9月末)公開済みデータのアノテーションデータ公開。 (年度末)理研播磨研究所における微生物由来蛋白質構造データに付随する実験データが、幅広い研究者に役立つための体制確立。			
		④ アノテーションシステムの開発運用とデータ変換	④ アノテーションシステムの開発運用とデータ変換 i. (9月末)公開データをクライアントサイトからウェブ経由でプログラマ的に利用できるAPI・ツールの構築、公開。 ii. (9月末)アノテーション作業を行うためのシステムを、データ投稿・プログラム投稿を理研内外から実施できるオープン参加型のシステムとして公開。 iii. (9月末)理研外の研究者から見て、理研のDBが統合されたミニマムモデルの成功事例として、研究者コミュニティから認知されるための統合検索サイトを公開。			
	産総研RCMG (成松 久)	「糖鎖修飾情報とその構造解析データの統合」		・糖鎖に関する国内のHUBデータベース、世界に誇れるデータベースとして認知されること。	特になし。	52,000
		① 運営と開発体制の準備	① 運営と開発体制の準備 糖鎖関連研究者への検索システムの活用方法を広める。糖鎖の統合検索システムを実現のため、開発体制、情報基盤を整え、統合検索システム完成。			
		② データ提供機関との交渉	② データ提供機関との交渉 糖鎖関連データを保有する研究機関と会合を繰り返し、円滑にデータを統合できるよう、データ提供の際に発生する作業、費用負担等にて、良い手段で解決し業務遂行。			
		③-1 研究を支援するDB構築 i オンラインプロトコルの統合検索システムのコンテンツ増強	③ 統合データベース構築 糖鎖科学統合DBのコンテンツ、検索機能の拡充(コンテンツを増やし統合しやすく。統合検索開発)。 ③-1 研究を支援するデータベース構築 i 糖鎖標準品の合成を支援するシステム完成。 外部研究者が反応経路と合成に関わるノウハウ登録システム開発。糖鎖標準品の合成法が検索可能。 本合成以外のプロトコルを、初心者でも糖鎖解析可能にする各種様々な解析に関するオンラインプロトコル整備。累計100以上プロトコル収集。年度末までに公開。			
		ii 新規DBを構築	ii (9月末)DB間をつなぐためのデータ収集。グリコシダーゼのDB構築。 解析後のDNAマイクロアレイデータの公開システム構築。他の研究領域と連携できるように「糖鎖と病原菌のDB(継続)」、「糖鎖のマーカー(新規)」、「糖鎖関連の特許情報(継続)」等様々なデータを収集。			
		iii 糖鎖の機能に関する情報のDB(オンラインレポート)	iii 一般から専門家まで理解しやすい易しい読み物の作成、DB化。 糖鎖専門用語と遺伝子名等を結びつける情報抽出技術の確立。 糖鎖専門用語集作成等の整備。各種ドキュメント、専門用語とDBとの相互リンク開発。			
	③-2 統合DB構築に向けた開発と中核機関との連携 i 検索用インデックスの自動更新システムの運用	③-2 統合データベース構築に向けた開発と中核機関との連携 i 横断検索のインデックスの自動更新。				
	ii 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	ii 統合検索システムの開発。APIの統合検索システムへの組み込み。 キーワード・専門用語の情報と糖鎖統合DBへのリンク先情報の中核機関へ提供。				
九工大 (皿井 明倫)	「生体分子の熱力学データと構造データの統合」		・本DBが当該分野の中心的なDBとして研究者に認められること。	・事業内で完成(達成)すべき目標を明確にしていきたい。 ・開発は極力絞り込み、既にあるDBが多くの研究者に利用され、利用者が増えるような取り組みに注力していただきたい。	8,000	
	① 蛋白質と変異体の熱力学DBの構築と統合	① 蛋白質と変異体の熱力学データベースの構築と統合 平成22年度収集の蛋白質、その変異体の構造安定性に関する熱力学データについて、熱力学データと構造データに対応づけるクロスレファレンステーブルを作成。統合検索と連携するため、DBCLSIにデータ提供、インデックス作成。 テキストマイニングによるデータの自動収集、自動抽出法の開発。				
	② 蛋白質・核酸相互作用の熱力学DBの構築と統合	② 蛋白質・核酸相互作用の熱力学データベースの構築と統合 平成22年度収集の蛋白質、核酸の相互作用の定量的な熱力学実験データについて、蛋白質・核酸複合体の構造データがあるものについて熱力学データと構造データに対応づけるクロスレファレンステーブルを作成。統合検索と連携するため、DBCLSIにデータ提供、インデックス作成。テキストマイニングによるデータの自動収集、自動抽出法の開発。				
	③ 蛋白質・蛋白質相互作用データの生成と統合	③ 蛋白質・蛋白質相互作用データの生成と統合 蛋白質・蛋白質相互作用データを格納するDB作成、試験運用。				
	④ XMLデータフォーマットやオントロジーなどの統合化技術の開発	④ XMLデータフォーマットやオントロジーなどの統合化技術の開発 オントロジーの整備。熱力学データの新規発生データについてのXMLフォーマットの作成、公開。  文献の自動収集のプロトコルを作成。データ更新は最低月1回、インターフェイス機能を更新。広報活動(9月末)。 自動化した文献収集を実際のDB構築で運用。統合化に関わる作業をルーチン的に実行するプロトコルの作成(年度末)。				