

文部科学省委託研究開発事業「統合データベースプロジェクト」
研究運営委員会・作業部会（合同会議） 議事要旨

【日時】 平成23年2月21日（金）14:00～17:30

【会場】 東京ステーションコンファレンス605B+C

（東京都千代田区丸の内1-7-12 サピアタワー6階）

【出席者】

研究運営委員会（*作業部会委員を兼務）：金岡委員、白木澤委員*、豊田委員*、長洲委員、中村桂子委員、中村春木委員、松原委員長、湯元委員、吉田輝彦委員、大久保委員、高木委員*、堀田副委員長
作業部会委員：浅井委員、五斗委員、田中委員、成松委員、森下委員、菅原委員

【陪席】

内閣府：織田上席政策調査員

厚生労働省：高畑専門官

経済産業省：川崎係員

文部科学省：石井課長、田中調査員

情報・システム研究機構：郷理事

【参加機関等】（作業部会委員以外）

内田副調査役（JST）、野口主幹研究員（産総研CBRC）、福井研究チーム長（産総研CBRC）、中村教授（遺伝研兼かずさDNA研）、藤澤研究員（かずさDNA研）、林特任教授（九大）、松本教授（奈良先端大）、中谷特任助教（東大）、瀬々准教授（お茶大）、池村教授（長浜バイオ大）、下川特任講師（医科歯科大）、井戸特任助教（医科歯科大）、山本特任准教授（阪大）小池主任研究員（日立製作所）、土井研究員（理研）、国島副ディレクター（理研）、皿井教授（九工大）

片山助教（東大）、川島助教（東大）、金城准教授（阪大）

永井特任教授、岡本特任准教授、川本特任准教授、金特任准教授、畠中特任准教授、坊農特任准教授、箕輪特任准教授、山口特任准教授、河野特任研究員、高祖特任研究員、仲里特任研究員、三橋特任研究員、山本特任研究員、飯田特任技術専門員、坂東特任技術専門員、吉羽特任研究員、植田事務室長（以上、DBCLS）

【事務局】 呉事務局長、加藤企画課長、逸見係長

【挨拶】 松原委員長より簡単な挨拶があり、開会が宣言された。引き続き、今回の会議は「統合データベースプロジェクト」の最終年度の総仕上げであるため、研究運営委員会と作業部会の合同開催となり、参加人数も多いので議論の時間があまりとれないことをご了承いただきたい旨の発言があった。

【議事】

1. 議事要旨について

前回の研究運営委員会、作業部会についての議事録について、すでに事前に関係者へは配布されているが、改めてご確認いただき、修正あれば別途ご連絡いただきたい。（メール等で別途依頼済み）

2. プロジェクトの成果について（参加機関より報告）

松原委員長よりプロジェクト参加の各機関からの成果発表が求められた。発表件数が多いため、実施項目の多少にかかわらず、発表時間は5分程度、質疑も含めて8分程度。発表は、中核機関、分担機関、補完課題実施機関の順で進める。まずは中核機関から。

（以下、機関名（発表者名））

➤ ライフサイエンス統合データベースセンター（DBCLS）（川本）

◇資料説明◇ 資料2-1

ポータルサイトの概要としては、最終的に全ての参画機関を合わせて50件のサービスやデータベース（以下、DB）が提供されている。そのうち中核機関は31件のサービスを担当しており、それらの月間アクセス数の最大値は5万4,000件（ユニークユーザー）および96万件（ペー

ジビュー)であった。またページビューについては海外からのアクセスが15%あり、英語化されたサービスは海外からも利用されていると思われる。

今年度はライフサイエンス新着論文レビュー、蛋白質統合検索のTogoProt、文献管理システムのTogoDocの3つのサービスを新規にリリースした。

資料「DBCLS成果まとめ」に全31サービスそれぞれの概要とアクセス数をまとめた。

年次計画の達成目標としては、初年度よりDBのポータル、横断検索、DBの受け入れについて具体的な数値目標を掲げ、また統合検索についてはプロトタイプの開発を目標としたが、DBポータルは912DB、横断検索は270DB、受け入れは40DB、統合検索は開発版を公開して、いずれも当初予定を達成できた。このうち横断検索とDBアーカイブについての詳細は以下の通りである。(生命科学DB)横断検索では270件のDB、1,080万レコードを対象として一括検索が可能となっている。また、今年度は、DBアーカイブ(受け入れ)の40DBを追加し、受け入れたDBのメタデータとデータの両方の検索ができるようになった。DBアーカイブでは、公開準備中の5件を合わせ年度内に45DBの受け入れ公開が完了予定である。この中には文科省委託事業の脳科学研究戦略推進プログラム(公募要領に統合DBへのデータ提供が記載されている)やバイオリソースが含まれる。

今年度新規公開サービスである「ライフサイエンス新着論文レビュー」は20のトップ・ジャーナルに掲載された、日本人の著者自身による研究論文レビューのサービスであり、現在トップクラスのアクセスを持つに至った。

次に広報活動として、今年度は皆様のご協力のもと4省合同シンポジウムを、また、分子生物学会においては次期統合に向けたワークショップを開催することができた。講習会活動としては、今年度は年間8回(含む予定)を主催、DDBJ、PDBjとの共催を合わせて10回の講習会を実施した。過去4年間の合計では、単独開催で計21回、延べ700人に参加いただいた。

また、参画機関にご協力いただいたユーザー評価について、2月9日に利用者からのコメントとそれに対する回答を一般公開した。参考までに評価結果のまとめを資料の最後に添付した。

最後に開発されたサービスの平成23年度以降の扱いについて、バイオサイエンスデータベースセンター(以下、NBDC)とDBCLSの協議内容を箇条書きにした。カタログ、横断検索、アーカイブは4月1日よりNBDCドメインで、またそのほかの参画機関のサービスは当面NBDCトップページよりリンクされ従来のドメインで運用することになる。

◆質疑応答◆ 特になし

➤ 科学技術振興機構(JST) (白木澤)

◇資料説明◇ 資料2-2

実施概要は①意見集約システムWINGproの公開・運用、②4つのウェブサイト(「ライフサイエンス分野の統合データベース整備事業」、「ライフサイエンスの広場」、「Webリソースポータル for バイオインフォマティクス」、「遺伝子名称シソーラス」)の運用、③「国際標準メタデータエレメントサイト」、MDeRの構築・運用である。成果概要としては①平成19年3月30日の公開以降、登録のDB数が現在438件、総訪問数が約10万件、②前者2つのウェブサイトについては情報を継続的に更新しており、「Webリソースポータル for バイオインフォマティクス」については、現在、456の解析ツール、29の解析ワークフローが掲載されている。また、「遺伝子名称シソーラス」については、約20万個の遺伝子と1万2,000個の遺伝子ファミリーが掲載されている。③国際標準5種、DB3種のメタデータ要素を掲載している。

本事業成果に対するユーザー評価として、全体については大体65%のユーザーの方に「サービスを継続してもいいのではないか」というご意見をいただいた一方、利便性向上のために統合DBプロジェクトの他サービスとの統合を求める声があった。

最後に、プロジェクト終了後の成果の維持、活用策について、4月に発足するNBDCにおいて継続してサービスを実施する予定である。また、「WINGPro」については、DBCLSのDBカタログと一緒にすべきというご意見もあるので、今後検討を進める。

◆質疑応答◆ 特になし

➤ 産業技術総合研究所情報工学研究センター(CBRC) (浅井)(福井)

◇資料説明◇ 資料2-3

設立して10年になるCBRCでは、計測技術の発展により生み出された大量データからなるDBを生かすバイオインフォマティクスの研究がすすめられ、世界的にトップレベルのものも含め数多くの独自の解析ツールを保有している。CBRCはこれらの技術を用いて、DBに対する高度な解析をダイナミックに行うための開発、ワークフロー（以下、WF）開発を本事業の中で行うことをミッションとし、“DB”の統合ではなく、解析ツールの統合化及びその環境構築を目指した。

開発の内容としては、ユーザーがKNIMEというプラットフォーム上でCBRCのNode Repositoryに公開されている解析ツールを組み合わせることによって自由にWFを構築し、解析できるものである。解析技術の多様化および開発のハイスピード化を目指して、このプラットフォームを利用した。また、この開発WFは、CBRCにある計算資源と連携しており、メモリーを大量に使用する計算などは、我々の計算資源を用いてシームレスに解析が可能である。8ページに示した本プラットフォームを利用したWFの解析例のようにCBRCの計算資源を利用する部分（黄）、DBとの連携部分（赤）、ローカルPCのツール（青）を組み合わせると、自分のPCで結果を可視化することができる。このような複合型WFとして、3タイプのWFを既に公開している。

成果は「統合DB情報基盤サイト」において公開しており、ユーザーはここから解析要素を自由にダウンロード可能である。ここでは14の解析ツール、3つのDB連携ノード、及び3つのウェブ解析サービスを公開している。また、LinuxやMacOSに対応したWFが欲しいというユーザー評価の意見を反映し、今年度はWindows以外に、Linux、MacOSに対応したノードを開発した。

プロジェクト終了後の維持及び活用については、今後もCBRCでは独自に開発したツールを統合化のためにノード化していきたいと考えている。また、計算資源に関しても維持していきたい。

◆質疑応答◆ 特になし

➤ かずさDNA研究所（中村）

◇資料説明◇ 資料2-4

我々の構築したシステムはゲノムDBの陳腐化を防ぐために、あたかも付箋を貼るような手軽さでユーザーサイドからの遺伝子についての参照論文情報等の追加を可能にするものである。

本課題の3月末時点の目標として、19万件以上のアノテーション情報の蓄積、利用者120名、BioMartに植物並びに微生物のデータを入れるという項目が挙げられているが、Gene Indexing型アノテーション18万件以上（2月15日時点）、ゲノムアノテーションジャンボリーでの利用者120名を達成し、BioMartには、植物関連微生物18種類、主要植物8種類を追加して、広報活動を行って利用を拡大している。

プロジェクト終了後は、このツール群と蓄積された情報、並びにCyanoBase、RhizoBaseという植物関連微生物のDBは引き続き維持する予定である。また、関連プロジェクトが採択された場合には、引き続き開発を継続して、主要なモデル微生物全般を対象に拡大していきたい。

成果の概要として、1)「高度情報集積データベースの運用と改良」では、ゲノム上の遺伝子にコメントを付与できるシステムとしてKazusaAnnotationを開発した。これは分子データ上の位置情報＝遺伝子の位置にアノテーション・キュレーションを付与し、高度化できるDBである。平成21年度までに基本機能の開発と改善をほぼ完了し、かなりの利用者、ページビューを得ている。図1に例示したが、植物関連微生物のDBの各遺伝子のページからそれぞれの論文の全リストが見られ、その論文のどこでこの遺伝子が言及されているかがすべて得られるシステムである。

次に2)「ゲノムアノテーション情報の蓄積と高度化」では、このDBを用いて情報を付加する作業を実施し、合計19生物種の7,500報の文献から、遺伝子とタンパク質に関する場所の情報を18万6,000件抽出蓄積して提供している。

最後に3)「植物と関連微生物ゲノム情報の統合」では、生物種横断的なゲノム情報の取得や、理研等の植物オーミクス情報との連携のためにKazusaMartを構築した。これは生物学者によるDBへの問い合わせが複雑であっても簡単にデータを取得できる、BioMartを利用したシステムである。

◆質疑応答◆ 特になし

➤ 九州大学（林）

◇資料説明◇ 資料2-5

ゲノムワイド関連解析（以下、GWAS）では、米国においてはGAIN Project等の先行研究が多くあり、そこでデータの扱いを考慮すると、クオリティー・コントロール（以下、QC）が重

要であると強く認識されていると思うので、それを我が国のGWAS研究にどう適用すべきかを本課題として考えた。

結論として、データを生産する人と、DBは本来分けるべきものだろうと私は考える。そうできない場合には、QCを外部でやるべきで、その結果を公開することにより、メタ解析などが可能になると思う。実際には、国内でGWASを実施している研究グループに申請を出し、データをいただいて、QCを実施してその結果を公開する。しかし、実施するのはQCのみであって、GWAS解析そのものをやるわけではないという立場でやっている。

QC解析の内容としては、標準的なソフトウェアというのは大体固定化されているので、それらを用いるフローをつくり、その解析結果をDBとして公開した。

ゲノム周りの関連解析というのは、データ生産者とDBは分離した組織でやるべきで、公平な利用のためにも中立的なDBは中立な機関で本来やるべきものだと考えている。今後はこのような透明性を確保したQCが公開されていくべきで、やり方としてはデータ生産者とは別の組織として、DB内で独立のものとしてやるべきだと考えている。

◆質疑応答◆ 特になし

➤ 奈良先端科学技術大学院大学 (松本)

◇資料説明◇ 資料2-6

我々のグループは主に日本語の専門用語の辞書システム、およびその拡張について実施した。

実施内容は3つある。1) 辞書を管理するためのツール開発で、京都大学のライフサイエンス辞書(9万5,000語)をベースに複数の辞書を管理して、その内部構造を解析するツールを付与、新規用語を登録できるようにした。内部構造の解析結果は系統樹様の図で示される。2) 内部用語の解析技術の開発で、本事業の中で約3,500語のデータについて手作業で解析し、それをもとに自動解析できるツールを作成した。1) で得られた系統樹のように分解された文字の係り受けの情報が付与される。3) 主にシソーラスを管理するツールの開発で、新規用語のシソーラス上の位置を単語の使い方の類似度に計算し、その配置の支援をするものである。このうち1つは、単語の文脈の一覧が出てくるもので、もう1つは単語と文脈の情報から使われ方の類似度の高い類似語の候補を出してくるので、それらを眺めながら、その新規用語をシソーラスのどこに配置すべきかを検討するためのものである。

本事業終了後の維持・活用について、我々の開発内容はデータの入れ物の開発であったため、ツールの運用は引き続き継続中であり、検索はだれでも、また、内部の情報の操作は登録ユーザーに限定して実行可能としている。登録辞書に関しては、京大ライフサイエンス辞書以外に、病名マスター、DBCLSから提供された辞書などが入っている。用語の類似用例については、現在はデモシステムを運用しているが、実際には文献データなどを保有しているユーザーが使いたいという申し出があれば、他サイトへの移植についても検討できる状況である。

◆質疑応答◆

●アクセス数はどのくらいあるのか。

→まだ運用開始間もないので、あまりアクセス数は上がっていない。

➤ 東京大学 (森下)

◇資料説明◇ 資料2-7

本演習は、バイオDBを構築できる人材を育てるという目標のもと、膨大な外注費を投じずに、組織の中でDBの保守・拡張ができる人材を育てる、やむを得ず外注する場合にも正確な仕様書が書ける力を培うために、1) バイオDBサーバーを構築する演習、2) プログラミングの講習(1と並行して実施)、3) 独創的なサーバーの構築の3つの演習を1サイクル2年間で実施した。平成19年度から3サイクル実施し、1期生5名、2期生10名、3期生5名が修了した。受講者としては、いずれDBCLSなどで活躍できるような外部からの参加者が多数いることが望ましかったが、結果的には外部の受講者は4名で、担当者の専攻の大学院生が中心だった。

サーバー構築実習は具体的には何も入っていない空のマシンに最終的にはEnsemblDBのミラーサイトを自分で構築するまでを10項目の目標を設定して実行していくもので、毎週のように演習を行うが、最後のところまで到達する実習生はなかなかおらず、8項目までたどり着くのがようやくである。独創的なサーバーの構築については、本事業の実施期間が次世代シーケンサーの普及と

重なったので、次世代シーケンサーの出力データの解析を選択する実習生が多かったが、そのためには特に並列計算が必要なので、詳しい内容の教材をつくって指導した。また、ユーザーインターフェースについては我々の開発したtoolkitを中心に説明をした。演習の結果、実習生たちが構築したDB公開の例として、Nucleic Acid ResearchのDB特集号に報告できるものもできた。

作成した演習ノートはウェブ上で公開しており、自習用教材としてアクセスもかなり多い。

最後に、修了生の進路として、まだ在学中のものも多いが、DBCLSの研究者が2名、そのほかの就職先はIBM、JR東日本、日立、野村総研など、一般的なシステムが必要な企業が多い。

◆質疑応答◆ 特になし

➤ お茶の水女子大学 (瀬々)

◇資料説明◇ 資料2-8

我々の目的は、蓄積された生命科学データを活用して、解析や知識発見ができる人材を育成することであった。

本年度の目標は、社会人を2名以上含む7名以上の修了者の輩出であり、実際には社会人8名、学生8名の受講生がいた。また、自習用教材の整備・改訂については、ウェブ上に掲載している演習内容、講義内容について演習時に補足した内容などを加えて再掲載した。また、卒業生の中からDBCLSで文献の情報抽出に2名従事しており、お茶大の教材公開用のサーバーを構築、運営、維持に2名が携わっている。

期間全体での成果として、講義と演習を合計で93名(お茶大の大学院生67名、社会人18名)が受講し、教材の公開、DBCLSでの業務担当、またお茶大のサーバーの維持など、実際に10名程度が統合DBプロジェクトで活躍した。

実習内容は基礎と専門の2つに分かれており、基礎では情報系の受講者に不足している遺伝学、あるいは生物系の受講者に不足しているネットワークやDBの基礎知識をつける演習を行い、専門では、基礎の内容を踏まえて、データの収集の自動化、処理や解析のテクニック、データマイニングの技術に関する講義と演習を行った。(詳細は資料を参照)

授業資料の公開については、一部の大学で講義などに使われているために集中アクセスがみられるなど、毎年、閲覧の人数及びページ数が徐々に高まっている。本事業の実施内容に関する需要は高いなという感じを受けている。

今後の維持について、講義と演習に関しては大学の全学対象の正規授業に組み込まれたので一応継続できることになった。しかし、実施により明らかになった問題点として、ティーチング・アシスタントの費用が大学からは供給が困難だという点がある。教材の公開に関しては、大学でも公開できるが、slideshare社のサービスにスライドを全部登録し、大学とは独立した方法でも公開を継続していく。

最後に育成した受講生の進路について、学生の多くは学んだスキルとは直接関係の無い一般企業に就職した者が多い。この点については本事業の中で、スキルを生かせる仕事の情報提供やあっせんを学生に向けてできればよかったと思う。育成に携わる方としてはスキルを生かしてもらえないのは少し寂しく感じた。一方、企業や研究所から受講した方は、実際にバイオインフォマティクスに関連する業務に携わっているようである。また、子育て等で一時期大学から離れていた方が戻ってくる機会としても有効であったので、大学としても我々としても、非常によかった。

◆質疑応答◆ 特になし

➤ 長浜バイオ大学 (池村)

◇資料説明◇ 資料2-9

我々の大学は、本課題実施の4年の間に「コンピュータバイオサイエンス学科」を新設でき、我々としては満足がいく人材育成のプロトタイプができたと思う。この学科は、ライフサイエンスの研究分野側からの要請のみならず、学生の将来にとって有意義となることが一番大切であるという人材育成の観点から、大量情報が扱えるコンピューターの知識・技術とバイオの知識・技術の両方を備えた人材の育成を目指している。

学科の設立にあたって明確な動機づけをしており、入学時から「π型人間(I型人間、T型人間)に比べて、専門性の深さはそれほどなくても、少なくとも2つの専門性を持つ)のススメ!」という言い方で、学科の趣旨を説明している。学生には新聞記事(4ページ)や文章(5ページ)を示

して、学部1年生の時期から「 π 型人間のすすめ」として、バイオの知識・技術に加えて、コンピューターの知識・技術の重要性や有利性を認識させ、初年次からコンピュータースキルを習得させ、3年間で就職にとっても有利なスキルを得られるようにしている。例えば、大量な情報を取り扱うDBやゲノムアノテーションに関しては、3年の後期に次世代シーケンサーの産出する大量配列情報を実例として扱う。

対象とする学生についての考え方としては、研究者志向型が含まれていてもいいが、それよりも、最初から専門技術者を希望する人材が、我々のカリキュラムに一番マッチしていると考えている。

教員側も、ライフサイエンス分野の研究志向で、研究での必要上からコンピューターも扱うという教員だけでは不十分で、実学的センスを備えた教員の参加が重要になる。幸いコンピュータバイオサイエンス学科にはこういう実学的センスを備えた教員もそろっている。

実習内容の概要としては、1～3年生の間、学部全体の300名へはDBを使う立場から、ポータルサイト活用等の有効な利用法を、50名のコンピュータバイオサイエンス学科の学生には、各種プログラミングの実習を経て、最終的に次世代シーケンサーのゲノムアノテーションを実施している。これらの実習のための教材については、他大学や個人でも自習できる丁寧な教材を作成・公開しており、DBへのアクセスも含めると月に3,000件程度、全体として年間数万件のアクセスがある。

4年生では大学OB/OGのシニアの先生方との共同作業として、世界に発信できる特徴あるDBを構築している。現在、我々は「エキスパートがキュレートしたtRNA遺伝子DB」を構築しているが、これは世界的に最大規模で最高品質でありNucleic Acid Research誌でも発表した。年間30万件程度のアクセス数を持つ。

キュレーターについては、シニア世代の専門家が有望な人材プールであると感じている。そのためシニア世代と学生が共同作業をすることが非常によいと思っている。分野の研究を先導してきたシニアが、DBの中でご自分の名前を明示しながら、DBの構築、キュレーションに当たることが望ましく、国際学術誌や国内外の学会などで発表することで高品質を支えることができていると思う。

今後はこのtRNADBの維持・更新だけでなく、他の分野へも活用を広めていきたい。

◆質疑応答◆ 特になし

引き続き、分担機関から。

➤ 京都大学 (五斗)

◇資料説明◇ 資料2-10

我々は、分子情報を中心とした統合DBの構築を目的とし、既にKEGGに格納された人手で構築された知識の体系とも融合していくことを目的として、DBを構築してきた。具体的な内容としては「共通基盤技術開発」と「統合DB開発・運用」の2つである。資料に各年度の主な成果を挙げたが、今年度に関しては化合物の類似検索と反応の類似検索(共通基盤技術開発)、医薬品データベースに関する種々の拡張(統合DB開発・運用)を行った。並行して、DBの講習会を年1回開催し、医薬品DBとLinkDBは定期的にデータ更新を行っている。

全体の主な成果としては3つあり、1)ゲノムネット統合DBの開発、2)ゲノムネット医薬品DBの開発、3)ケミカル情報解析ツールの開発である。1)ゲノムネット統合DB検索システムについては、初年度に検索システムの拡張を実施し、DBの拡充によりキーワード検索対象DBとしては18、LinkDB登録DBを16追加し、反応DBも整備した。2)ゲノムネット医薬品DBについては、医療用医薬品・一般用医薬品添付文書(JAPICのDB)のキーワード検索、そのDBに対する構造検索、医薬品の相互作用を調べるサイトの作成の3つの開発を行った。ゲノムネット医薬品DBのアクセス状況を見ると、製薬会社、大学、研究所、一般のプロバイダーから幅広く利用されている。3)ケミカル情報解析ツールでは、化合物DBに対し、類似構造検索や反応の経路検索ができる。当初手がけた糖鎖に関するサービスも継続している。また、構造検索結果から機能に関する手がかりを得るために、当面はKEGGのPATHWAYや階層分類に直接リンクするツールも作成した。

DBCLSとの連携不足が指摘されていたゲノムネット統合DBとケミカル情報解析について、今後は連携して技術開発をしていきたい。コンテンツが含まれる医薬品DBに関しては、これから

KEGGで疾患情報の追加を計画しているので、それと統合していくことを考えている。

最後に、我々の京都大学化学研究所にあるスーパーコンピューターを管理するスーパーコンピューターラボラトリーが、今年4月からゲノムネット推進室という名称変更があり、組織としてDBの開発、広報的なことも含めてやっていくことになっている。

◆質疑応答◆ 特になし

➤ 東京医科歯科大学 (田中)

◇資料説明◇ 資料2-11

東京医科歯科大学グループの目的は2つある。1) 全国に分散している疾患のDBを統合し、横断的に検索できるシステムを試行的に開発することで、DB間の用語の差異などを吸収するシステムを考えるものである。事業開始当初、全国的な規模で検討したが、予算規模の関係から不可能と中間的評価で指摘され、プロトタイプの開発ということになった。2) 進展著しい分子情報や遺伝学の情報を疾患の症例DBに取り込むための体系や仕組み、また、これまでの臨床病理情報の配置を検討して、両者の関係を表示できるDBを構築することがもうひとつの目標である。

上の2つのシステムに関して、大阪大学パーキンソンDBとの間での実証実験、あるいはがんセンターとの間で実証実験を行った。

1) の横断検索では特定のキーワードで検索するとそのキーワードが含まれていないと検索結果としては出てこないが、オントロジーツリーから上位あるいは下位の概念を選択して検索することで、網羅的な検索が可能になる。

2) の臨床と分子のDBの縦断的統合に関しては、オミックスデータと臨床所見データを階層的に結合し、個体レベル、組織レベル、細胞内レベルでの関係をつなげた。患者個人のデータはつながっていることはもちろんだが、ある特性を持った患者のグループのなかでどうつながっているかを見ることができるようになっている。そのためにできるだけ多くの症例を、電子カルテ、レポートや臨床病歴などから搭載した。また、分子データに関しては、例えばがんの場合はがんの発現解析のヒートマップが入っており、病理データからCT、薬剤、検査といった臨床表現型もできるだけたくさん掲載した。臨床データと病理データとの関係は12ページにある3層マップで表現した。臨床データを主成分分析してマップに表示し、病理データを二次元に、また、発現データ(分子データ)を3次元目に表すと、例えば分子データのところで特定の条件に入る患者のグループがどのような臨床データ、病理データに対応するのかを見ることができる。グループの選択は対話式なので、直感的に見ることができる。さらに、臨床的な病態と分子名、その両者に正準相関分析を行い、分子と症状を同一平面上にマッピングすると症状と遺伝子の関係が見えてくる。(解析例では肝がんマーカーAFPとサイクリンB1が非常に近接して出現している)

今後の課題としては、ヒートマップとコピーナンバー情報以外の多次元化するオミックス情報への対応、オミックス情報の統合化や、臨床表現型の標準化・カテゴリー化などが必要である。また、オミックスDBの倫理規定ガイドラインの作成も重要である。東京医科歯科大学の難治疾患研究所が共同拠点になったので、その基本的なDBを例に全国的な症例の統合も実施したい。

◆質疑応答◆

●本課題に参加した幾つかの施設のデータが、全て分子から臨床まで統合されているのか。

→iCODとGeMDBJだけがそのように統合されている。パーキンソンDBについては、現在、症状に関してだけ横断検索ができるようになっている。

➤ 大阪大学 (山本)

◇資料説明◇ 資料2-12

我々は、医科歯科グループの課題の一つとして、パーキンソン病の臨床情報(経過、薬剤量や種類の情報、ゲノム情報は含まない)DBのデータを集める仕組みをつくり、一般公開を実施した。

資料に挙げた臨床情報DBの特徴のうち「持続性」が問題であり、長期にわたり地道にデータを収集することが重要である。現在、いろいろな病院や学会が各疾患についてDBを構築しようとしており、DB数は爆発的な増加傾向を示している。例えば治験等で承認された薬でも長期的な予後の観点からの情報はなかなか得られないといったことがあるので、いろいろな医学研究をしていく上で、DBの必要性・重要性は認識されている。そこで我々は、幾つかの医療機関を結んで医療情報、パーキンソン病のデータを集めるSecure Studyシステムを構築してきた。

成果の概要としては、平成19年にデータの公開に向けて再同意をとり、データを匿名化して、公開用DBへ格納する仕組みを開発した。20年には、異なる疾患DB間の統合を実現するための要件定義を行い、双方のDBのスキーマのマッピングや検索処理の手順など、詳細なレベルの設計を行った。21年には、要件定義に沿ってIBMDの要素DBの実装を行い、医科歯科大、阪大のDBの統合検索を実現し、パーキンソン病に特化したシソーラスを阪大のDBで独自に拡張して、検索に利用できる仕組みを開発した。現在、700件のデータを一般公開しており、この公開に用いているプログラムやソース一式も公開している。このことで、1つの公開のロールモデルを達成できたと考えている。

この過程の中に幾つかの問題があった。実はデータ入力に非常に手間がかかったが、医療情報のDBは症例数が多いことが重要なので、今後は電子カルテをうまく利用して臨床情報や薬剤情報を抽出することが必要である。今後、電子カルテからのデータの収集を検討していきたい。また、個人情報管理に関しても、今回は再同意が必要になったが、国全体のことを考えると、今後は個人情報を極力少なくし、同意を省略できる形が必要ではないかと考える。

我々はこの研究の遂行にあたり各施設の異なる考え方から同意取得に非常に手間取った。また日本にはゲノム指針、臨床研究指針、疫学指針など医学研究に関する指針が乱立しているので、研究対象者保護法といった包含するルールを考えていただき、個人情報に関しても、統一した見解を持つことが非常に重要だと思う。また、倫理委員会に関しても、各施設で別個にやるより、どこかで一元的に行える国の施策が必要ではないかと考える。以上を提言としたい。

◆質疑応答◆

●提言については何か具体的なアクションはとられつつあるのか。それともまだ提言の段階か。
→提言の段階である。

➤ 東京大学グループ (小池)

◇資料説明◇ 資料2-13

研究代表の徳永先生に代わりご説明する。

我々のグループでは「疾患解析から医療応用を実現するDB開発」と題し、1) 1検体当たり50万から100万SNPのタイピングをしてゲノム関連解析を行うGWAS用のDBの構築、2) 新たな変異情報と疾患の関係をおさめるDBのロールモデルとして神経変性疾患を主とする変異DBの構築の2つを目的とした。

1) GWASDBについては、①実験結果・解析結果のDB化による半永続的な集約的データ管理、②GWAS以外の研究者にも情報を提供することによる疾患研究の促進、③研究者間の情報共有による遺伝統計研究の促進を具体的な目的とした。米国におけるNCBIのdbGaP、欧州におけるEBIのEGAと同様に、日本人ひいてはアジア人がフェノタイプとSNPの関係を登録できるDBを目指した。標準DBとして構築した健常者のSNPDBには現在は約1,700検体の遺伝子頻度やアレル頻度、遺伝統計値を掲載している。Case-controlDBのほうは、現在19疾患26 studiesを登録しており、そのうち10疾患11 studiesを公開している。17の量的形質も内部用に登録しており、上記未公開分と合わせて、論文公開後に公開される予定である。画面では主な遺伝統計値、SNP間の相互作用、パスウェイ解析などの結果も見られる。GWASデータから同時に解析できるCopy Number Variationの解析もできるので、健常者DB(公開)とcase controlの解析結果(論文未発表のため内部公開)を掲載している。

データの公開に関して、本事業ではまずデータ共有方針を決定し、それから倫理委員会を発足、適切な倫理審査の手続を行った上で研究者からデータを受領するとともにデータ利用者への再配付を行う仕組みを開始して、実際に運営している。研究者から受領したデータは、論文公開前の内部用DBに入っているときは研究者間でのみ使っただけのように、また、倫理審査が終わった後は、一般への公開とデータの再配付ができるようにしている。

2) 神経変性疾患の変異DBでは遺伝子変異と臨床情報の両方を俯瞰でき、疾患の機序の理解と、臨床現場で役立つDBを目指しており、現在はALS、PD、ALD、HSPのデータが対象となっている。データとしては、配列情報、変異と臨床情報、及び配列とorthologous配列のmultiple alignmentsや、構造、ドメイン情報等を格納している。データ量は、疾患ごとに200程度の文献から、1,000~2,000ぐらいの変異データを抽出している。これにより変異と表現型の関係を俯瞰的に見られる。一例を18ページに挙げた。

本サービスは現在中核機関のサーバーを用いているが、人類遺伝学会の学会誌『Journal of Human Genetics』では論文投稿時にこのDBへの登録を推奨することになっており、また、E B I の E G A や欧州の H G V S からの連携の申し出もあるので、今後もしばらく維持したいと考えている。

◆質疑応答◆

●最後に述べられた他DBとの連携について今後検討すべき点はあるのか。

→連携するに当たっては倫理面で解決すべき問題もあるが、現在、中核のサーバー等を借りて運んでいるので、今後サーバーが使えないようになるのなら、何らかの形で同様の計算機環境が使えるようにしていただければありがたい。

引き続き、補完課題担当機関から。

➤ 理化学研究所 (豊田)

◇資料説明◇ 資料2-14

我々の補完課題は「植物オミックス・蛋白質構造統合DB」を核として、理研にあるデータを公開していく流れをつくるというものであった。

まず植物に関しては、現在統合されている29の植物関連のDBのうち17が理研による研究活動の成果に基づくものである。DB全体と各データがそれぞれオントロジーに基づいて分類されている。フェノタイプにもいろいろな種類のDBがあるが、それぞれを共通化したオントロジーとして国際的にはPATOやPOを使うことが主流になっているので、それらを用いて共通したアノテーションを行った。これによって、DB横断的にフェノタイプを検索できるようになった。

タンパク質DBとしては構造DBを中心として、それに関連したタンパク3000の成果に基づく実験データを公開していくものである。これもオントロジーに基づいて、DB及び各データのtableを分類した。例えば構造データについては、PDB全体を、例えばRSGI(理研プロジェクト)などのプロジェクト単位で、部分集合的に分類してみることができる。さらには部分集合として取得したデータをキーワードで絞り込んでテーブルとしてダウンロードすることもできる。テーブルはRDFaにより構造化データ(セマンティックウェブ)にも変換可能である。また、PDBを基準にして、各グループから公開された情報が統合され、ツリー状に表示することもできる。PDBjを基準としているので、PDBのデータとして利用可能である。

さらに上記の植物や結晶、タンパク構造を核に、哺乳類も加えて、理研のDBを含む18DBを分類してオントロジーで統合した。現在のところは、192のDBプロジェクト、table数で664件、750万レコード、2,600万のリンクが公開されている。

まだ未公開、準備中のものが、200近くある。理研には和光、横浜、つくば、播磨、神戸の5拠点に事業所があるが、それぞれにDBの担当者を置き、こちらにデータを公開させるという流れができたので、これからも継続していきたい。

◆質疑応答◆

●その流れはどのようにすると継続できるのか。

→継続のために、現在、委員会をつくり、担当者を配置して、例えばデータをテープに格納して送ってもらうなどして収集する体制をつくっている。維持の費用については、これから工夫が必要。

●理研は非常に膨大な情報を持っているが、公開する傾向になりつつあるということか。

→こういう流れができてきたので、協力関係が得られるようになってきたという状況である。

➤ 産総研糖鎖医工学研究センター (成松)

◇資料説明◇ 資料2-15

本事業開始時にはまだ糖鎖DBが無かったので、補完課題に応募し採用いただいた。応募理由は、産総研で10年にわたってやってきた大型プロジェクトの糖鎖研究の成果の膨大なデータをどうやって世間に公開、公表するかがずっと悩みだったからで、本事業で採用いただいたので、データをまとめて公開することができた。また、高木研究代表者からの要請は、「日本全体の糖鎖研究をまとめ上げる」ことだったので、日本糖鎖科学コンソーシアムのメンバーに全面的に協力願った。

DBの内容としては非常にユニークなものが多い。NEDOプロジェクトの成果(糖鎖関連遺伝子情報DB、糖鎖と結合タンパクの結合・解離に関する実測値情報DB、糖鎖の質量分析測定情報DB、糖結合タンパク質の糖結合位置情報DB)をすべて公開した。そのほか、立命大学、理研、

創価大学に統合用APIを設置して、それぞれの大学等が開発したユニークなDBを公開していただいた。さらに、企業人が糖鎖の研究をするに当たって非常に役に立つ糖鎖関係の知財DB、病原菌が結合する糖鎖のDB、インヒビターのDBなどなどたくさんの小規模DBを新規に構築した。また、産総研の支援により九大では糖鎖遺伝子のノックアウト線虫のフェノタイプDB、名古屋市立大学ではNグリカンの構造解析DB、野口研、名古屋大学はノックアウトマウスのDBが構築されたほか、KEGG、LipidBank、生化学工業のDBにもご協力いただいて、リンクを張っている。

このDBには糖鎖研究をするため、あるいは糖鎖を実用化に向けるための基盤ツールおよび基盤リソースを公開してあるので、このように構築されたDBを利用して、糖鎖研究はまだ難しいと言っている研究者の方々にまだまだ開発要素がある新しい分野に参入していただきたい。

全体の横断検索は糖鎖に関するキーワードあるいは構造によって可能である。構造はアメリカが提唱する糖の図示方法を採用してそのシンボルで検索するか、あるいは化学的な構造式でも一致検索、部分構造検索、Super Structure、類似検索などの検索がかかる。また、単糖のDBや糖鎖合成の支援システムなどがある。さらに専門家による実験プロトコル集も作成した。

最後に、糖鎖研究はまだまだ新しい分野である。アメリカのConsortium for Glycobiologyは9年前からDBを構築しているが、DBの内容は日本とアメリカでは全く異なり、日本では我々が研究してきたユニークな成果をそのままDBにしている。そのDBを日本だけにとどまらず、アジアの中国、韓国、台湾、インド、シンガポールにも協力を呼びかけ、JCGGではなく、Asian Consortium for Glycobiology and GlycotechnologyのDBを構築するべく、現在企画中である。

◆質疑応答◆ 特になし

➤ 国立遺伝学研究所 (菅原)

◇資料説明◇ 資料2-16

遺伝研の補完課題は「塩基配列アーカイブのデータベース構築と統合への貢献」である。

成果物として立ち上げたサイトがDDBJ Sequence Read Archive (URL: trace.ddbj.nig.ac.jp) という新世代シーケンサ由来の配列のアーカイブの正式名で、略称DRAである。

本課題は実際には平成21年からはバイオインフォマティクス推進センター事業のDB高度化、標準化の課題として開発を行ってきた。直近の1月31日の時点で受付件数366件、受付データ量は3.3TBにのぼり、そのうち公開したものが473GBである。

開発内容に対する評価として、我々が作成したメタデータ作成ツールMetaDefineは、NCBIとEBIからも関心を持たれたのでソースコードを提供した。これを用いることで、データ登録ユーザーの負担が大幅に軽減できるので、ユーザーからも好意的なコメントを得た。一方、国内からはアクセス数が少ないことに対して厳しい評価もあるが、(ユーザー) 評価の実施時期以降、メタデータ検索システム、DRAサーチシステムの公開をきっかけにFTPサイトやウェブサイトの利用度が高まっている。

今後に関して、本事業の努力の成果とは全く別のところである種の事件が起きたので、事実関係だけご紹介する。2月16日のNCBIのホームページにSequence Read ArchiveとTrace Archiveのrepositoriesを何か月か後には停止するというアナウンスが出た。EBIは直ちに、EMBL-EBI will continue to support the Sequence Read Archive for raw dataというアナウンスを出した。一方、NCBIの中止のニュースを受けて、「Barcode of LifeというプロジェクトではTrace Archiveは必須なので、そのTrace ArchiveをDDBJで続けてもらえないだろうか」というメールが上記のアナウンスと同じ日付でBarcode of Life担当者からDDBJに届いているという状況である。

◆質疑応答◆

●NCBIのSequence Read Archive中止の件について、DDBJセンター長でもある大久保委員からコメントをおねがいしたい。

→NCBIが資金不足により中止するというお知らせが、EBIにもDDBJにも事前に通知など無く公表された。現在、担当者から直接情報を集めており、できるだけ早期にNCBIへのヒアリングを行い、EBIと2者でどういう方式で継続していくかをこれから探るところである。

もともと生データのアーカイブ事業は米国内で10年ぐらい前に始まったもので、3極共同の事業ではなく、日本の研究者からの要請も特には無かったが、ここ数年の間に信憑性が疑わしいデータが出るので、データを出す者の責任として、また一部の論文誌の要求もあり、DDBJでもアーカイブを持つべきという意見を受けて本事業を機会に始めたものである。現状ではいろいろな面で

NCBIの1%程度の規模である。本事業で整備されたおかげで、ようやく日本以外のオセアニア、アジアの国からもデータ提供を受けられるようになってこれからという段階でもあり、世界的な要請に応えるという側面もあるが、あくまで日本の事業費で実施していることなので、国内の研究者が「生データを保管する必要がある」と強く要望するのであれば、継続できるように検討したい。

▶ 九州工業大学 (皿井)

◇資料説明◇ 資料2-17およびスライド

私の課題は「熱力学データと構造データの統合」だが、まず熱力学データとは何かをご説明する。生体分子は熱力学の法則に従うので、分子の機能を知るためには、熱力学的なデータが必要になる。これらのデータは、分子機能の理解、分子ネットワークの定量的な理解、医薬品の合理的な設計などにも役立つ。

そのために我々は以前からタンパク質の安定性などのデータを集めた、蛋白質の熱力学DBを作成し、既に約2万5,000件のデータを文献から収集した。相互作用に関しては、蛋白質-核酸相互作用を約1万1,000件、蛋白質-リガンドの相互作用を約2万件収集した。

データの内容としては、例えば蛋白質の熱力学データに関しては主として安定性などに関する熱力学的な実験データ、蛋白質-核酸相互作用に関しても結合の定量的な実験データを収集している。この種のDBは世界初であるので、多くの研究者に使われており、これまでに250件以上の論文に引用されている。

本課題では、熱力学データと構造データを結びつけるために必要な技術の開発と、テキストマイニングの技術を用いた熱力学データを抽出する論文の収集およびデータの抽出の自動化を行い、その成果はプロジェクトのウェブサイトにて公開している。

成果の概要は、蛋白質の熱力学データと構造データを統合するために必要なクロスレファレンスの作成、あるいは未整備であった熱力学データのオントロジーを整備するためのControlled Vocabularyの作成、XMLフォーマットへの適用である。また、文献の収集について最も大変なのは関係のあるデータの論文を集め、そこから論文を1つずつ読んでデータを抽出することだが、これをできるだけ自動化するためにDBCLSとの共同研究でTogo Docをカスタマイズして熱力学の論文をリストアップできるようになった。

本事業の終了後は、熱力学データの収集についての資金的な裏づけが現在全くないが、何とか継続するように努力したい。また、未統合の相互作用データ、機能データ、変異や疾病のデータの統合、オントロジーの整備、インターフェースの改良を継続したい。データの収集と入力自動化について共同研究も継続したいと考えている。

◆質疑応答◆ 特になし

各課題からの発表が終了したので、研究代表者である高木委員から全体的な話や今後の問題点について発言するように、松原委員長が要請した。

非常に多数の機関が参加し、内容も非常に多岐にわたって、本事業は推進されてきたが、それでもまだ重要なDBが欠けているのではないかとのご意見もあるかもしれない。

プロジェクトの実施機関が公募により採用され、中核機関と分担機関が同時に採択されたこともあり、プロジェクト全体の推進体制に整合性を考慮することが難しかったように思う。研究代表者としては、知らない分野も多数あり、最初はどうかと大変心配した。

また、プロジェクト推進において何が問題だったかという点、1つは統合すべきデータがそもそもあまりなく、データがあっても権利関係が整理されておらずなかなか提供してもらえないという状況であったことだ。さらには「統合する」とはそもそもどういうことかという概念が委員や参画機関の間で当初なかなか共有されていなかったが、この4年間で目的意識を共有し、実際にデータを出してもらえるところまで何とかこぎつけたと思う。

一見するとインターフェースもまだばらばらで、「これでほんとうに統合か」というご意見もあろうかと思うが、情報技術で何とかできるところまで来たと思っており、その意味ではまだ出発点なのだが、最初はマイナスからの出発であったので、やっと次のプロジェクトにつながる状況まで来たと思う。

このような状況下、JSTのセンターが立ち上がるので、そこでうまく引き継ぎ、これまでの成果を発展させることができる、そういう基盤はできたと思う。今後もしろいろとご意見を頂戴したい。

3. 平成23年度以降の体制について

松原委員長よりJSTへ平成23年度以降の体制やその準備状況についての説明が求められた。

◇資料説明◇ (白木澤) 資料3

JST内に新設されるNBDCの機能として、1) 戦略の立案、2) ポータルサイトの構築・運用、3) DB統合化のための基盤技術の研究開発(基盤技術開発プログラムの実施)、4) バイオ関連のDBの統合化の推進(統合化推進プログラムの実施)の4本の柱がある。

平成23年4月以降、総合科学技術会議(以下、CSTP)の統合DB推進タスクフォース(以下、TF)の助言のもと、NBDCはセンター長に大石先生、副センター長に高木先生という体制で、上記の4つの柱を実施していく。

基盤技術開発プログラムについては研究総括に長洲先生、統合化推進プログラムについては研究総括に高木先生をお願いし、2つのプログラム間でも密接な連携をとりながら事業を推進していく。また、NBDCの運営委員会を設置し、センターの運営に委員のご意見を反映していきたい。

上記2つのプログラムの公募の状況について、基盤技術開発プログラムは12月末、統合化推進プログラムは1月末にそれぞれ募集を締め切り、両方ともに3月下旬に採択課題を正式発表するべく選考を進めている。

前回の研究運営委員会で松原先生から、新体制について広く告知する工夫が必要というご意見をいただいたので、①統合DBプロジェクトのホームページに「統合データベースプロジェクトとBIRDの一本化の検討の経緯」の告知を掲載し、関連資料を紹介するページを設けた。また、昨年12月の分子生物学会において「ナショナル統合データベースの構築に向けて」(オーガナイザー:高木先生、長洲先生)というワークショップが開催された。さらに、統合化推進プログラムの公募説明会を東京、大阪で12月に開催し、認知度を上げるべく告知活動を行った。

松原委員長より各委員へJSTの発表に関連した意見が求められた。

●NBDCの事業が終了する3年後に非常に興味を持っている。本来ならこのような事業はもう少し長い期間を予定して運営されるべきで、省庁を超えた仕組みを3年後につくることがNBDCの本当のミッションだと思うのだが、それについての展望は現在どのようになっているのか。3月に内閣府でのTF会合が予定されているのでそこからの助言もあると思うが、3年後に必ず真のナショナルDBセンターを作らなくてはいけないという強い意志をこの場で皆さんに表明していただきたいと思う。

●統合DBのコンセプトが大事であることは自明だが、その構築方法のプロトタイプを今後どのように運用し、日本のバイオサイエンスに生かしていくのかというコンセプトもはっきりさせておかなければならないと思う。必要な予算や、アカデミアや企業、国際的な役割などの将来像を明確にすべきだが、当面はJSTで実施し、その先はまだ非常に不透明な状況なのだと思う。

●JSTで今後3年間にやるべきこととして、本事業の成果をきちんと引き継いでさらに発展させることに最善を尽くしたい。3年間は橋渡しには非常に短い期間だが、その間に3年後にどこでどのような形で続けていくべきかを皆さんにご議論いただきたい。

●我々バイオインフォマティクスの専門家はDBに関する問題について話し合えばかなり意見が一致するのにその解決が数年間にもわたってなかなか進んでいない。なので、今後3年間の検討に際してもこうあるべしと有識者の意見が一致しただけでは進み切らなかったというこの数年間の原因をよく考えないと、また3年後に同じことが起こる可能性があるかと心配している。

省庁ごとにお金や人があるという現状を認識し、その現実をいかに取り込むか、いかに実際に携わっている研究者に味方になってもらうかを考えることが必要で、単にお金や組織を集中させるだけでは利害が一致しないので、皆さんがついてこない。利害を一致させるために我々研究者ができることは、統合した成果がそこに参加する人たちの利益になることを示すことだと思う。

そこでつくられるものが各省庁の方も含めていろいろな立場の方々に利益になることを統合化するセンターが示していかなければならない。我々もそのように考えていかなければいけない。

●データを生産・解析をしているユーザー側の利益として一番に考えられるのは、幾つかの拠点がフェローシップを提供してバイオインフォマティクス人材のプールになっていただき、そこからそういう人材を必要とする機関に派遣されるといったことで、統合DBには人材のトレーニングもあるし、そこか

ら出て行っているいろいろな現場で働いた人たちが、その経験を中央に持って帰るしくみがあるとうれしい。人材で勝負しないと中国などになかなか勝っていけないのではないか。

大量データの取得についての日本の体力は明らかに足りなくなってくると思うので、データ取得後の頭を使うところやバイオロジーで勝負することは非常に重要だと思う。

もう一点、前回も話題提供したが、内閣官房で進められている医療イノベーションとの関係がどうなるのかが非常に大きな疑問である。“国民総背番号制”や“どこでもマイカルテ”が実現するときに、一方で生物学のデータがきちんと横断検索できる仕組みがあれば、非常に強いのではないかと思う。

●医療のDBの動きに関して言えば、厚生労働省のDBの1つはいろいろな大学病院のDBを統合して、それを医薬品医療機器総合機構に置いて副作用等を統合的に調べるために、1,000万人ぐらいの患者のDBを日本版センチネルプロジェクトという形でつくろうとしているものである。本事業でわれわれが構築したシステムは分子情報などを中心とした研究指向のものなので、上記の副作用等の発見につながるという方向性のDBとどういう形で統合したらいいのかとは思ふ。例えば臨床の先生方は副作用のほうに重点を置かれている。厚労省DBのプロジェクトを推進するために、東大の循環器内科の永井良三先生を中心として1,000万人ぐらいのDBを構築し、医療への応用と同時に製薬会社などに対しても副作用の情報を与えることに関心がある。我々としては本事業についてもそういう先生方に入っていたきたいと思っているが。

●いずれにせよ大きなDBを扱いながらバイオサイエンスをやらなければならない時代になる。今振り返ると4年前は、国産のデータの集積がほとんどない、またそれを扱える人材の養成もないことが問題だった。本事業でこれらの問題の解決の方向が見えたのでこれを継続するということはもちろんだが、その成果をどのように使うのかについて次の目標に向けての意見がそろそろ欲しい。例えば今の医療に関して、日本の医療の将来はどのようになってくるのか、これまでばらばらに集められてきたカルテ情報が国全体として今後どのように生かされるのかという問題に対して、インフラの発展によりどんどん情報が共有化できるのと同時にこれらの問題に対応できるデータの統合・整理がすすみ、どの医者にかかっても共有化された情報に基づいた医療が受けられるといったシステムにしていくための第一歩であるといったことが言えるのではないか。薬にも、基礎的な生命現象の解明にも、あらゆるところへ出口のイメージがあるだろうと思う。ちょうど3年後ぐらいには、本事業の成果がもっと充実し、解析も十分できるし、現在は乖離している情報と生物の基礎知識ももっと融合されるだろう。その上でより発展的なDBをどのようにしていくのかというイメージを作りたいと思う。

●本事業の開始時に、“DB”と“統合”を考えるということが非常に印象的で、大量のデータがある中でバイオサイエンスがどちらへ向かっていくのかを難しいながらも考えていくことができると思った。本日の報告を聞いて着実にできる方向に向いているとは思ふ。ただし、今後は“統合”がより重要になると思う。その点では、今後実施される統合化推進プログラムでは採択件数5から10件と書いてあるが、こういうやり方ではなく、この部分をこそどのように統合化していくのかという議論をして推進することが重要と思う。このプログラムはどのように進められるのか伺いたい。

→5から10件というのは、例えば植物やヒトの多型など分野ごとにまずは統合化するということである。それぞれの統合化にはどういう生物学を目指すのかというメッセージが当然入ってくると思うし、我が国のDBの拠点としつつ、生物学の方向性が問われる課題募集をしているつもりである。幾つかの分野ごとのDBが出てくるとどうつながるかという別の問題は出てくるが、それは目的を限るというよりは横断的に柔軟に行き来できる、いろいろなツールが簡単に利用できるというところをまずは目指したい。新センターではそれ以上の、例えばヒトの疾患への対応といった具体的な目的は限定していない。

●進め方については理解した。ただ、全体として生物学を見ているという意識を持っていただきたい。

●民間企業の立場からは本事業開始当初から権利関係が気になっていた。いろいろな権利関係があるためにそもそもデータが集められないという状況も大分解消はされてきたと思う。国の資金100%で実施されている研究についてはコピーレフトの考え方で進んできていると思うが、特に国の資金と同時に民間企業の資金も投入されているケースではまだまだ統一的な基準などができていないと思うので、今後、そのあたりも課題であると思う。

もう一点。製薬会社で集められている臨床情報や医療の現場から上がってくる情報をもっと活用できたという思いがあるが、分散していたり、個人情報の問題があったりして、なかなかアクセスが難しい。個人情報ということで公開する側も使う側もおっかなびっくりであり、ある程度公開されても利用者が非常に厳しく限定されている。さらにいろいろな機関等のDBごとに公開の仕方がばらばらという状況なので、統合化を進めるにあたり基準の統一も検討してほしい。

→データの共有方法や権利関係の明確化のためのガイドラインの作成と個人情報の管理の2点に関しては、当初から本事業でも相当目的意識を持って取り組んだ。現在は、クリエイティブコモンズの表示・継承ライセンスを採用することで大体決着しており、その先へ踏み込むためには新たな国の方針が必要と考えている。これは、内閣府のTFでぜひ議論していただきたいので、議論のためのたたき台をぜひNBDCから出していききたい。

●研究者の統合だけではなく、「日本の縦割り行政をいかに統合するか」を新事業の一つのモデルにしたらいかがかと提案したい。3年後にまた文科省独自の予算だけでやるのか、それとも国を挙げての予算でDBを考えるか。DBの公開に関して権利化や知財の問題が省庁によって温度差が全然違うので、省間でDB事業を議論すると考え方が全然違うのではないかと。内閣府は省庁を統合する責務があると思うので、研究者の統合だけではなくて、縦割り行政の統合を3年後に目指してほしいというのが、研究者からの意見である。

●いろいろな方が切実に何年も何十年も考えている問題で、一気に解決するのはなかなか難しいかもしれないが、新しいものができるときにはどこの省庁のどこの課がやるというのでない考え方でやったらどうだと具体的に言うと、もう少し現実味が出るかもしれない。

ここで、松原委員長より、CSTPの事務局へ内閣府での取り組みについての発言が求められた。

●内閣府では、統合DBTFから平成21年に報告書を出している。このTFは、ライフサイエンスPTの下で会議体だが、各省（文部科学省、厚生労働省、経済産業省、農林水産省）および内閣府の関係者も出席しているので、その報告書での記載がベースになって今後進めていくことになると思う。そこには、来年度からJSTに統合DBセンターを設置する件と3年後にそれを見直す件が掲載されている。現在の取り組みでは、郷先生を座長として統合DB推進TFを立ち上げ、昨年6月に第1回の会議を開催、3月24日に第2回を開催予定である。その中で、まさに今議論されたようなことも話題になるかと思うが、具体的な議題に関しては、座長の郷先生と本庶委員等で調整している状態である。また、来年度からは内閣府がCSTPで提出している第4次科学技術基本計画が始まり、全体的な上部の会議体の改変等も予定されている。現在、4月以降に関してはその辺の調整をしている段階である。

ここで、松原委員長より、CSTP統合DB推進TF座長の郷先生へ発言が求められた。

●統合DB推進TFにはここにご出席の方も参加いただいているが、これから本格的に動かなければならないと思っている。3年先はあつという間で具体的にはこの1年ぐらいでかなりいろいろなことを、JSTとも一緒に詰めていく必要がある。CSTPもどうなるかわからないが、こういうDBの問題が政治によって変わってしまうのでは困るので、これまでの先生方の努力がちゃんと実を結ぶように、また、これは日本の国の大きな重要なインフラでバイオロジーを進めるための大事な一歩なので、しっかりと検討していきたい。

●難しいのは、許されている条件の中でベストを尽くすという議論から、新しい条件も徐々に作らなくてはならない変化へのボーダーである。4年前には思いもつかなかったことができそうな方向に向かって動いてきた。今後3年でそれをもっと進めて、さらにやらなければならないことも見えてくる。将来的にこのDBを国民のためあるいは科学のためにどう使っていくのかがそろそろ考えられる状況になってきたということをもとにして、CSTPでもこれから大いに議論していただきたい。そのためには、限られた人の中だけでなく、一般の人を巻き込んだ頻繁な議論をぜひお願いしたい。狭いところだけで議論すると、最先端何とか研究資金のように特定なところがちゃがちゃとかき回されて、科学全体が損害をこうむる。パブリックコメントとは言わないが、意見のある人をできるだけ広く、いろいろな機会に求めていただくことを、本会議の意思として科学技術会議にお願いしたい。

4. プロジェクトの事後評価結果の概要について

松原委員長から文科省ライフサイエンス課へプロジェクトの事後評価の概要についての説明が求められた。

◇資料説明◇ 資料4

本日提出した資料は「事後評価報告書（暫定抜粋版）」である。事後評価報告書については本来ライフ

サイエンス委員会に提出されるまでは非公開であるが、今回の研究運営委員会・作業部会が本事業の最終回であるので、評価内容について有効に使っていただきたいという趣旨から主に総評の部分を暫定抜粋版としてライフ課で編集して配付した。取り扱いについてはご注意ください（事務局補足：このうちライフサイエンス委員会が開催され、暫定抜粋版の内容を含め、報告書は確定した）。

《以後、1. 事後評価の概要（資料4、1～3ページ）について抜粋しながら読み上げ》

4ページ以降が事後評価結果概要であるが、4～5ページの“(I) プロジェクト全体について”に評価委員の言いたいことや思いが入っているので、ここを取り上げる。（詳細は資料4を参照のこと、以下ポイントのみ）

- ▶ 総合評価はA（取り組みとしては大変優れている）
- ▶ プロジェクトの意義と達成状況
 - ・本事業の成果は今後のDBの統合に関する一定の指針を提供し、基盤構築にも貢献
 - ・非常に困難な課題を遂行し、期待以上の成果を上げている
 - ・本プロジェクトの果たした役割は大きく、基盤づくりという点では期待以上
- ▶ 今後の課題（国家プロジェクト事業の進め方に関する課題としても）
 - 1) 公募制度、ミッション等の問題
 - ・クリアなビジョンが可視化されていなかった
 - ・「中核と分担を同時に公募した時限プロジェクトの弊害」
 - ・最初の段階で中核機関が方向づけを可能とする施策が必要
 - ・一極集中に近いような体制を組むことが可能となるように、予算／体制を整備すべき
 - 2) 統合DBの在り方（JSTで事業推進の参考に）
 - ・分子レベルのデータと個体の表現型レベルのデータを統合するという姿勢が不足
 - ・「研究階層を超えた連携と統合的理解」をできる人材の育成が問題
 - ・統合DBの在り方を極めて広い立場から展望し、適宜方針を変える姿勢が今後必要
 - ・計算機環境の整備と（統合化が）必要な分野の選定
 - ・継続的に推進されることを期待、多面的な早急な取り組みが必要
 - ・むだや過去の執着を省き、必要な整備を精力的に進めるべき
 - ・今後も事業を推進していくための留意点として、「データの提供者が自由に参加できる」「開発を推進する者がメジャーユーザーである」「継続的な開発と改良」「プロトタイプの開発から公開までの迅速性」などが、評価委員会の議論で挙げられた
- ▶ プロジェクトマネジメント関係
 - ・我が国でのプロジェクトとしては緩いながらも組織的なプロジェクト遂行が評価できる
 - ・中間評価後には中核機関がイニシアチブを取り、比較的短期間で継承し得る成果も上がった

また、代表機関である中核機関（代表機関 情報・システム研究機構）については総合評価としては、得られた成果は大変優れているという高い評価であるが、総評の最終部分にあるように、「プロジェクト遂行の困難度を高め、数々の問題が生じた要因の一つとして、“全てを”プロジェクトベースで行ったことが挙げられる。本来統合DB事業は恒常的なシステムとして維持、整備されるべき部分と、科学の発展に即応して新たに発展させるべき部分があつてしかるべきであるという見解を評価委員会全員の総意としてつけ加えておく」という指摘をいただいた。先ほどのプロジェクト全体についてと中核の代表機関についての指摘は、明らかに文部科学省の仕事のやり方についてのご意見であると受け止め、今後のプロジェクト研究、特にDBの今後の進め方の中でも反映すべき重要な指摘だと考えている。

5. その他

松原委員長より、これ以降、時間の許す限りフリーな議論をするために、各委員へテーマを決めずに意見が求められた。

●今の評価に関連して、重要な分野のこれだけ多くの先生方が入っておられるプロジェクトの最初の公募のやり方、また今回の評価のやり方は非常に難しかったのではないかと思う。参加しているメンバーよりも、かなり少数の方が評価や審査を担当されているが、DBの重要性等、長年議論をやってきてよくわかっている方がほとんどインサイダーになっていて、評価する側が少しアウトサイダー気味なので、よくわかっている部分もあるが、個別の指摘事項については実施者が納得できない部分もたくさんある

と思う。もちろん批判は批判として受け入れるが、政策を進められる担当部署では、新しい政策を進めるときに、評価をやっているメンバーがどれぐらい幅広くその分野をカバーできているかを認識し、より良い意見を聴取できるようにしていただきたい。

→どなたにどのように評価をお願いするかはいつも担当部署の悩みの種である。課題実施者が増えるほど、中身に精通した方に評価者を依頼しにくいので、評価結果の活用の仕方についてはその点を十分考慮し、評価結果を鵜呑みにするのではなく、課題実施者であるか無いかに関わらず専門の方と相談して進めるといふ工夫をしている。一方で、先生方にしてみれば悪い評価をつけられたままの報告書が公開されることについて非常にご批判があるかもしれないが、これは外部評価の仕組みとしてやむを得ないものである。むしろ評価の報告書そのものについてきちっとご批判いただきたい。一部の評価では、結果が出た後、プロジェクトの関係者と評価委員会で意見交換を行った例もある。特に中間評価の場合は評価内容を理解していただく必要がある。そういったご要望があれば対応していくつもりである。

●プロジェクトの中身の評価についてはおおむね妥当だと思っている。しかし今後の進め方や将来の在り方について評価委員の方だけが提言するのは非常に狭いことでそれが問題だと思っている。むしろ、プロジェクトに参加した方々に広く伺うべきではないか。

●非常に困難なときに、本事業に心血を注いで考えてきた方々の意見も聞いてほしいということであるので、ぜひ考慮いただきたい。

●事後評価以外に、ユーザーの意見を聞いた評価もあったが、そちらで非常に評価がよかったものがこちらで評価が悪かったりして、あまり相関していないという印象を受けたので、事後評価結果だけを受け止めるのではなく、ユーザー評価内容とのバランスを見たほうが良い。さきほどご指摘があったように、事後評価は少ないメンバーで決めているため非常に揺らぎもあると思うので、幾つかの評価を多面的に考慮したほうが良いのではないかと思った。

→ご指摘の評価は資料2-1のDBCLSからの報告の最後にあるユーザー評価のことだと思うが、これは数百名のユーザーからの意見なので、この文科省の事後評価委員会だけではなくて、ぜひそちらも参考にいただければと思う。

●人材育成のことが出なかったので意見を述べたい。4年前にできなかったことができるようになった背景には、個別の人材育成による育成もあるが、プロジェクトを実施する中でのOJTとしての人材育成があると思う。“プロジェクト”の繰り返しではそういう人材がなかなか定着せず、次の人材が見つからないという状況になり、今後求められる「研究階層を超えた連携と統合的理解」ができるもう一段高いレベルの人材育成ができないので、人材育成の問題は次の3年間でぜひ検討していただきたい。

●DBをつくるということを研究大学で教育するのは非常に難しい。学生はやはり論文を書きたいし、新しいデータを解析したい。DBづくりを非常に地味なものと考えるので、それをずっと続けていくという動機づけが難しく、本事業の課題を遂行したこの4年間、ジレンマがあった。だから、大学で人材育成を請け負うのではなく、目的を理解できる30代、40代の方が非常に有意義なDBづくりを進めるといった事業を、ある程度予算をつけて展開していくのが適切なのではないかと思う。先ほどご意見のあったバイオインフォマティクスの派遣も研究大学とは違うところでやるべきと思う。

●我々（お茶大）が育成した人材がどこへ行くかと考えたとき、例えば2年間の研究プロジェクトがあるから行ってみようという学生はなかなかいない。やはり企業に就職したほうが収入も安定するのでいい人材が研究者として残らない、あるいは3年間の時限で招くあるいは送り出すのもなかなかできないという状況である。企業が悪いと言っているわけではないが、すごく優秀な学生だけれど多分企業に行くのだろうなと思って育てるといふ悲しい現状である。本事業はDBができあがればいいわけではなく、ライフサイエンス全体の底上げの意味が大きかったと思うので、研究全体のレベルアップ、人材の底上げのために、企業と大学の間、あるいは研究機関の間で、優秀な人材を流動的に扱える仕組みも必要だと思う。

●我々の機関（長浜バイオ大学）では学部や修士レベルで高級技術者を育てるといふミッションがある。学生は、地道で持続的で支援的な要素の高い作業を行う職種、いわゆる技術者になりたいという気持ちで入ってきているので、DBに関係する教育カリキュラムにうまく合っているとされる人材が多く含まれる。DBの構築運営等に関係した情報処理分野には研究者志向とは少し違う人材が必要ではないかと思う。

●人材に関しては、確かに産業側にも問題があると思う。特にライフサイエンスにおいてはバイオインフォマティクスという二足のわらじをはいた人が欲しくても新卒ではなかなか厳しいとか、そのプロジ

エクトが終わってからの雇用などを考えてなかなか採用できないといった状況がある。それを徐々に変えていかななくてはいけないと思っているが、一方で国のライフサイエンスのプロジェクトを見るとウェットな研究の場合にはドライの部分が最初から非常におそろかになっていて、例えばドライには10分の1程度の予算しか使えない。米国などでは人数で言えばドライが半分位入るのが当たり前という状況である。ドライの地位を上げて半分位の人を割いてDBもできるし解析もできるといった国プロを立ち上げていただき、終了時には成果がドライとウェットの研究者に均等に配られるようになると、ドライの研究者も論文は書けるのでインセンティブやモチベーションが上がってくるし、産業側でもドライの重要性が判って採用に結びつくといった動きにもなってくる。そのさきがけとして国プロでもう少しドライを重視していただくことが一計ではないかと思う。

松原委員長より、中核の代表機関で本事業開始にあたりDBCLSを設置した情報・システム研究機構の堀田機構長へ発言が求められた。

皆様のご協力とご努力で達成できた成果は正直言って予想を上回るものであった。法人化に伴い、機構ができた時、機構として独自の事業をしたいと考えていたが、ライフサイエンスの統合DBに踏み出すというのは非常に大きな決断であった。しかし、私自身、以前はライフサイエンスの研究者であり、遺伝研の所長もしていたので、この業界の状況に対していろいろと不満があった。第1の不満は、ライフサイエンスの研究者はそれぞれ細かく分断された狭い専門について一生懸命研究して競争資金を取りに行くだけで、物理の世界で見られるように大きな予算を取りに行くときには普段けんかしている人でも団結するといったことが無い。また、ドライの人たちはウェットの人たちとうまくつながっていない。そういった人たちの団結のために、データ、情報という形でライフ全体をまとめるやり方はないのかと思っていた。何とかしなければという思いはみんな持っているらしいがそういう運動が出てこない状況で、統合DBをやったら多少の貢献ができるかと思い、本事業に応募した。フーズビリティスタディのころ、皆さんに集まっていたときに、非常に大勢の方が集まってくださったことに正直言って私は驚き、やはり需要があるのだと強く思った。その際、ドライの人たちは年中集まる場所があり協力的なのだろうと思って聞いてみたら、「そんなにはない、学会ではただ会うだけ、競争でBIRDに予算を取りに行く」と言われ、競争している場合じゃないと思った。本事業で協力していくことができることは証明されたと思うので、この方向性を保って自分たちで積極的に運動を起こしていくことがぜひとも必要だと思う。

最初に思っていたイメージのなかで未達成だと思うのは、1つは医療の問題である。これは非常に難しいので、今すぐに臨床現場の医療のデータをそのまま取り込むということではなくてもいいと思うが、医学的な判断のためには基礎的な生物学が絶対に必要なので、うまくタイアップしていく道が今後開かれればいいと考えている。もう1つは、ウェットとドライのすり合わせである。今、確かにウェットとドライを両方やる人が目につくようになってきたが、もう少し相互作用が起こるように統合DBが進むといいと思う。

この勢いが今後も続いていくように、JSTにはもちろん、皆さんの協力もぜひお願いしたいと思う。

これが最後の委員会であるので、会の終了にあたり、担当課であるライフサイエンス課長からのあいさつがあった。

平成18年度から5年にわたり実施された本プロジェクトに実際に参画された機関の皆様、研究運営委員会の先生方に多大なご支援をいただき、大変感謝している。

3年間という非常に中途半端な時間であるが、第2期に移行していくためにもこれからの第1期は非常に大事な時期だと思っている。3年後というのは省庁の予算のサイクルで考えると、2年半後の夏には概算要求しなくてはならず、そのためには1年半後には外向けの検討を始めなければいけない。そうすると、実質的に今年の秋から本格的に検討しなくてはならず、その時点で3年後の成果を意識しなくてはならない。つまり、始まったときには成果の見通しを持っていなければいけないという非常に厳しい3年であり、その後に恒久的にどうやっていくのかという議論を同時に進めなくてはならない。縦割りをなくすべきというご意見はそのとおりで、我々も最大限の努力をしたいが、実際に議論をしてみると非常に大きな敵は実は組織の中やコミュニティーの中にいるケースが多々あり、我々の提案に対して、実は研究者や大学の中から反対の意見が出てくることもある。我々もやれることをやるが、先生方には

コミュニティーの中、各組織の中で運動していただき、内閣府、CSTPを中心にした議論、各省での議論とも相まって次の体制ができているので、ぜひ今後ともご協力をお願いしたい。

本日の議題についてはすべて議論が終了したので、松原委員長からの本事業及び委員会、本合同会議への協力に対する感謝の言葉をもって本会は終了した。

以上