

【日 時】 平成22年6月3日(水) 15:30~18:40

【場 所】 ライフサイエンス統合データベースセンター大会議室

【出席者】 林健志(九大)、時松敏明(京大)、田中博(東京医科歯科大)、下川和郎(東京医科歯科大)、井戸敬介(東京医科歯科大)、山本洋一(阪大)、古崎晃司(阪大)、徳永勝士(東大)、小池麻子(日立製作所)、成松久(産総研)、新聞陽一(産総研)、鹿内俊秀(産総研)、田中康博(文科省)、大久保公策、高木利久(作業部会主査) 永井啓一、川本祥子、坊農秀雅、畠中秀樹、吉羽洋周、河野信、高祖歩美、坂東明日佳、箕輪真理(以上、DBCLS)

(敬称略・順不同)

【議 事】

1. 議事要旨について

事前送付済み。修正は2-3日中に連絡していただきたい。

2. 評価結果と進捗状況・成果目標に関する事前配布資料について

事前送付済み。各機関からの発表の際に各コメントなどに対応した内容をご発表いただきたい。

(高木主査より) 今回の会議の趣旨は、9月末をめどとした成果の取りまとめについてを中心に、作業部会委員による評価への対応などご発表いただきたいというもの。まずは、中核から昨年度の対外的活動内容の概要と、今年度の予定をご報告する。

3. 中核機関より(箕輪)

前回の作業部会では、中核機関の開発内容について紹介したので、今回は対外的活動を中心に紹介する。

平成22年度のスケジュール(研究運営委員会、作業部会、行事[講習会、展示会、学会対応、シンポジウム]、進捗報告とユーザ評価)についての説明を行い、特にシンポジウムへの当PJへの参加機関からの協力、9月の成果取りまとめを受けて10月に実施するユーザ評価への協力の依頼があった。続いて、21年度の対外的活動の概要(研究運営委員会・作業部会、シンポジウム、広報活動、BioHackathon、AJACS、ユーザ評価)について報告された。

<補足>スケジュールでは今年度の分しか示していないが、皆さんも気になる来年度に関しては、正直なところまだ何も決まっていない。10月には何かお話できるかと思う。

○遅れている理由は、政策立案が進んでいないからか。

→4省の話は内閣府では話題に上っている。JSTでの体制についても話は進んでいる。事業仕分けの影響もあるかもしれないが、確定していることは無い状態。

4. プロジェクトの平成21年度進捗状況及び最終目標について

➤ 東京医科歯科大学グループ

◇資料説明◇

【医科歯科大】(田中)

21年度は統合技術開発としてプロトタイプの開発とDBの高度化・実証的統合については処理の分散化、ユーザ登録の自動化などを行った。また、前回高木主査から質問のあった[どういうデータをフリーにしてよいか]については、今回資料に内容をまとめてある。22年度は検索エンジンの高度化・インターフェースの高度化について9月には基本構造を完成させ、アクセス制御機能を高度化したうえで利用者に向けて9月にDBを完全公開にしたい。またPJの総合的推進として、倫理規定の作成、関連情報へのリンク付けな

どを実施する。

【阪大】(古崎)

21年度は統合技術開発としてプロトタイプの開発とDBの高度化・実証的統合についてはデータの拡充、公開準備などを行った。22年度はシステムの部分改良を進めながら、評価の意見にもあったロールモデルとしての知見のまとめなどを実施し、他の機関が今後このシステムに参加するための情報提供の準備などを行う。

◆質疑応答◆

○ICODのデータソースがどこにあるかはどうすればわかるか。計測データを取得したところはどこか、など。例えば論文を参照できるなどすればよいと思う。

→すべて東京医科歯科で取得したデータである。現在、論文は参照できないので、Referenceを付けるようにする。9月までの予定に入れてないが、共同拠点での取得データも盛り込んでいく予定である。

○(阪大へ)統合医科学DBの「統合」という意味は。

→(阪大)2つの概念が混ざっている。医科歯科と阪大をマージしていること、その前段階として阪大で複数病院から集めたデータをまとめることの2つの概念を指している。

○後者は判るが、前者についての「統合」とは具体的には何か。

→(阪大)同じ「臨床」データであるが、疾患が異なる場合にどこまで統合できるかを模索した結果である。

○そういうニーズがあるのか。

→(阪大)以前から議論はあった。今後より多くの病気が集まるケースでは生かされる可能性がある。

○パフォーマンスが出ない理由は何か。

→(阪大)いくつかの技術的な問題があると開発者から聞いている。

○メディカル特有の問題ではないということか。医科歯科と阪大で分散してデータをもつことなど、ハード的な問題も含めた問題だと思っていかが。

→(阪大)希望するパフォーマンスで出ていないので、改良したい。システムやネットワークの問題である。

→(医科歯科)「統合」については、当初の計画ではいくつもの疾患DBを取り込むことにしていたのを、計画を縮小したため、現在の対象疾患はかけ離れているように見えるが、今後このシステムを利用して、現在の対象疾患以外のものも受け入れることができるように考えている。

○ガイドラインについては、もっと具体的なものが欲しい。特に“個人を識別可能な情報”について、具体的な例をあげていただいて、このPJとしてはどう考えるかに結びつけたいという趣旨であった。

→(医科歯科)大量の分子情報では個人が識別できるかもしれないが、このDBの中で扱っているのは発現情報であるのでそれだけでは個人の識別はできないと思う。投薬情報などは資料4の6ページにあるように100人程度以下にならないようにするのが基準。病歴情報などで過度の絞り込みができない情報だけに限定しており、その程度であれば全部公開することになっている。

○ICODはカルテの内容はよく整理されているが、公開する値打ちとしては、分子情報とカルテ情報が連結されているところ。分子データとしては症例を絞り込んだ結果からヒートマップの絵は見えるがアレイやCGHの個別のデータが取れないのは不便。

→(医科歯科)アレイのデータに関してはICODのトップページから全部ダウンロードしてIDで抽出する方式になっている。CGH-DBは医科歯科内の別の研究室で管理されているため、同じように公開できない。

○医科歯科の中での統合も進めてほしい。それぞれが抱えたまま統合しようとする、連携のたびに仕組みが増える。論文化されているデータであればアレイであれCGHであれ、少なくともNCBIには入っているはずなので、そのリンクを書くだけでも、NCBIには無い臨床情報がこちらのDBにはあるので、非常に価値があると思う。

○倫理規程案の策定について、何を目的とした倫理規程案か。Clinicalデータの公開や共有のロールモデル

になるようなものか。

→ (医科歯科) この DB はあくまで症例の DB なので、症例データを公開するためのものである。

○まとまった後は、統合 DBPJ で議論していくための提案を作成するということか。

→ (医科歯科) から提案して、後は中核の倫理委員会で検討いただくものである。

○アクセスはどのくらいあるか。

→ (医科歯科) 正確なデータは現在把握していないので、最終的に報告書で報告する。

○本 DB が成功と言えるのは客観的にはどうなった時か。当初の壮大な計画からミニマムにして成功事例を作るように、と中間評価では言われたと思うが。その成功事例とは何であると考えているか。

→ (医科歯科) 今後ほかの臨床 DB が入る可能性を持つ仕組みができた状態にし、少なくとも2つの DB について統合検索ができ、他の DB が入ってきても同様なことが可能になるようにする、ということだと思う。

○共通の入れ物としては、形と意味(オントロジー)の両方の観点があると思うが。

→ (医科歯科) 形もそうだが、オントロジーの整備も国際的にも進みつつある。別のものが入ってくるとまた別のオントロジーが必要になる。

→ (阪大) これだけ違う疾患でも共有項を見出して、一緒に検索できる仕組みを提供するというのが一つの達成点。それに加えて、技術的ではないハードルとして倫理の問題を超えられるようにすることも必要。

○9月でのゴールを見据えて最重要課題を考えるにあたり、疾患を絞るとか、課題を絞るとか、いろいろな方策が考えられると思う。例えば、大久保先生がおっしゃったように、ユーザが使いたいと思われるようなものを作る、というターゲットを定めるというやり方もあると思う。

→ (医科歯科) 臨床情報がついたマイクロアレイ DB の要望が高いのであればそれを作ることも検討する。

→ (阪大) 本事業で「統合」が意味するものは何なのか。形を整えて一緒に置いておいて、使いやすい DB を作るだけなのか、異質のものをいっしょにすることが統合なのか。このグループがやっているのは後者であると思う。

○一般ユーザから利用度が高いという評価を受けられるかという観点にはその両方が含まれていると思う。

→ (阪大) であるから、頂いた意見を基に、我々のグループとしてはどの部分に注力するかを考える。

→ (医科歯科) ヒートマップについては引き続き検討する。

○医科歯科大グループには厳しいことを申し上げているが、このままではちょっと安心できない部分もあるので、個別に議論して着地点について話をしたい。次回の会合を待たず打ち合わせを持ちたい。

→ (医科歯科) 要素 DB を見せていく方向と、他の DB の受け入れも見据えた統合を優先するといった方向、どちらを重視すべきか、について整理したい。中途半端になっているところはある。

○中核の倫理検討については進んでいるのか。

→ (DBCLS) 医科歯科 G の提案を待っている段階。

▲9月末時点の到達目標、要望について▲

- ・アクセス制御機能を高度化したうえで利用者に向けて9月にDBを完全公開する。
- ・一般ユーザから利用度が高いという評価を受けられるための着地点について、個別に(文科省および中核と)議論する場を設ける。
- ・本委託事業で実施したツールなどは再利用可能なようにする(DBCLSが調査する)。
- ・新サービスの公開時には、中核からの広報活動による紹介や使い方説明として統合TVを活用する。

➤ 東京大学グループ(徳永)

◇資料説明◇ 資料5

21年度末にはかなりの機能が公開され、高機能化を引き続き進めている。公開対象疾患・データも継続的に追加中。結果表示など新規の機能も公開。また、限定公開データのための受け入れ・公開についての審査

のための仕組み（ルール・組織）を作った。22年度は、9月までにはインターフェースの改良、計算手法の改良、さらなる宣伝活動、海外機関との連携開始などを予定している。

◆質疑応答◆

○“リシークエンシング”という名称は正確には MutationDB なのではないか。

→計画書時点の名称がそうだったので日本語ではこうなっているが、英語では MutationDB としている。

○国外からのコラボについては、国内の研究者コミュニティ全体の利益になるような形になるように気をつけていただきたい。国外への提供についてよく批判されることに、「国益を失う」というのがあるのが、海外の機関等へ提供される際には少なまったデータになるからそんなことは無いといつも説明しているので、海外へ詳細データが行かないよう留意していただきたい。

→提供元の了解を必ず取りながら提供する。通常は RawData のレベルまでは提供しない。

○今の議論は省庁によって違うのではないか。そのために内閣府がまとまらない。

○国益というけれど守っていても国益にならない。外部公開してこそ、国のステータスを上げることになる。

○連携の中身はどのようなものか。DDBJ のように 3 極連携まで行くものか。

→そこまで本格的なものではない。最初はリンクを張る、あちらで公開してもいいレベルのデータはこちらから提供する、反対に向こうからもいただくといったことの検討を開始したところ。互いに DB の中身を見て検討できるように、研究者が交流を開始するための相談が始まったところ。

○話が固まりそうなら、報告してほしい。省内では手続きが必要なこともある。文部科学省が実施している事業なので、あくまで形式的なものであるが文部科学省が相手機関との交渉を決定するという形にしてほしい。情報は事前にほしい。

→先方は日本学術振興会の相互交流のためのグラントに申請しているはず。情報はお伝えする。

○徳永先生のグループで、井上先生が計画書の中で“疾患 DB を人類遺伝学会での登録先とする”ことを約束されていたが実現可能か。

→編集委員の中でその方向で議論を始めているところ。断言は難しいが、その方向を希望している。各雑誌でも基準はばらばらだが、どこかの公的な DB にデポジットされることを義務付ける方向。顕著なデータは良い雑誌に投稿されることが多いが、顕著でないものでも、公開の意思があれば受け入れる、というプロモーションをしている。

▲9月末時点の到達目標、要望について▲

- ・疾患 DB を人類遺伝学会での登録先とする方向に努力する。
- ・本委託事業で実施したツールなどは再利用可能なようにする（DBCLS が調査する）。
- ・新サービスの公開時には、中核からの広報活動による紹介や使い方説明として統合 TV を活用する。

➤ 九州大学（林）

◇資料説明◇ 資料 6

作業部会委員による評価への対応について、“PJ への貢献”についてはまだ実データでの運用が無いので未確認、“DB ごとの QC が必要”という点についてはむしろ一定の基準でやるのが重要だと思う。

○9月末までにはこれらの点については実行されるか。

→全部は無理だと思うが、一部は実行していきたい。東大と相談していく。

現在はハプロタイプデータの構築に注力しているため、SNP 関係については進捗があまりない。CNV についても貢献できる標準データを提供していきたい。

◆質疑応答◆

○多くの機関から提供される SNP データを今後 QC していくということか。

→そうであるが、まだ実施できていない。多くの機関というが、実際には大量解析を実施していてかつ信頼

できる GenotypeCenter としては東大以外にあまりないと感じている。

○ゲノム特定の範囲で、他の施設で解析が実施されたデータもすでに DB には入っているはず。イルミナからのデータや、九大からのデータもある。

○資料 2-3 に示された 9 月末の達成状況（有用なツールとして機能する）は可能か。複数機関から出たデータを QC して、比較できる状態にすることはできるか。

→2 か所以上の QC データが出てきて、比較できる状態になる予定と考えている。

▲9 月末時点の到達目標、要望について▲

・有用なツールとして機能することを示すために、2 か所以上の QC データが掲載され、比較できる状態にする。

・本委託事業で実施したツールなどは再利用可能なようにする（DBCLS が調査する）。

・新サービスの公開時には、中核からの広報活動による紹介や使い方説明として統合 TV を活用する。

➤ 京都大学（時松）

◇資料説明◇ 資料 7

21 年度は技術開発に加え、リンクされる DB の追加を行った。22 年度は情報の統合的な利用、医薬品・化合物の解析への応用に向けた枠組みの構築を目指している。具体的には、SIMCOMP/SUBCOMP のパッケージ化、複数化合物入力の実装 DB 間の関連検索強化、薬物相互作用・パスウェイを検索可能なインターフェース、ユーザ定義のリンクを使った検索の開発などを予定している。

◆質疑応答◆

○今回の開発に関連して想定される DB ユーザ(たとえば製薬など)があると思うが、その利用状況は。

→昨年度、そういう情報が必要という指摘があったと思うので、五斗がまとめている。

○想定ユーザに本当に利用効果が出ているのかを知りたい。そういった解析をお願いしたい。ターゲットユーザにうまく使われているかがわかる統計情報が欲しい。

○エンドユーザーとして使うよりもダウンロードして使いたい。自前データと統合したい。その場合 KEGG はライセンス料が高いので、二の足を踏む。国プロのあり方としては、料金を取るのはいかがなものか。

→ゲノム分野ではそうかもしれないが、化合物などの DB では有償のものが多く、ここでもライセンスを払って統合 DB に入れているものもある。

○民間が作ったものと、国の予算で作ったものとは違うはず。

○基盤技術開発で作成した汎用ソフトウェアは公開できるのか。

→SIMCOMP/SUBCOMP はパッケージ化して DL できるようにすると聞いている。ソース公開は不明。

PathPred については Web サービスなので、パッケージ化できないと思う。

○KEGG のデータ構造に依存しているものは他のデータフォーマットに応用する場合はソースを公開する必要があると思うが。

→検討する。

○（全体に）委託事業で実施したものについては、ツールなどは再利用可能なようにしていただきたい。

○（DBCLS）9 月末に向けて何が公開可能になるかを、プロジェクト全体で調査したい。

▲9 月末時点の到達目標、要望について▲

・想定ユーザに本当に利用効果が出ているのかを検討できる統計情報を示す。

・本委託事業で実施したツールなどは再利用可能なようにする（DBCLS が調査する）。

・新サービスの公開時には、中核からの広報活動による紹介や使い方説明として統合 TV を活用する。

➤ 産業技術総合研究所糖鎖医工学研究センター（成松）

糖鎖は膜蛋白や分泌蛋白以外にも核蛋白、細胞質蛋白までついている可能性が高いことが判ってきた。蛋白質のどの部分にどんな構造の糖鎖がつくのかの構造解析は今のところ質量分析装置でしか行えないためある程度のサンプル量が必要。糖鎖遺伝子の発現パターンだけでは糖鎖構造まではわからない。そのためにこういった DB は非常に有用。また、経産省の組織なので、成果は産業に結びつけることが肝要で、特許なども含めた DB にしている。アクセスについては、国内のユニーク ID は 2000 程度。糖鎖の研究はもともと日本がリードしている面もあるが、この DB をきっかけにさらに日本のステータスを上げたい。統合 DB の方向とは何かを 4 省で考えるべき。温度差なんて言っている場合ではない。米国、欧州の糖鎖 DB については、資金が滞り気味で、逆に今がトップに立つチャンス。

◇資料説明◇ (鹿内) 資料 8

22 年度は、研究プロトコル (現在は英語のみ)、オンラインレポート、新規 DB の整備、統合検索の完成を計画している。プロトコルについては研究者が以前書いた文章の流用など著作権に留意する必要があるかもしれない。9 月末、年度末、それぞれ目標を決めて、実施予定。

◆質疑応答◆

○オンラインプロトコルやレビューの充実は賛成だが、実験データをやっていればいいという声もある。お考えは。

→できるだけ多くの研究者・企業人・一般の方へ情報を届けるために DB はあるので、多様なデータがあることが重要。

○書かれたものを収集する場合は、この PJ で集めたものはマッシュアップできるように、権利を放棄してほしい。

○米欧で予算が厳しいという話だが、他でもそのような話を聞いた。米欧で落ちている理由は。

→理由はよくわからないが、米は何とかなりそうだが、欧はまとまっていないので。

○Elixir も Feasibility Study 以降の予算が出ないという話がある。

○利用状況は。糖鎖コミュニティで使われている部分と、蛋白質関係のコミュニティから使われている部分はどのくらいか。

→蛋白質と糖鎖のコミュニティはかなり混ざっているので、判別不能。プロテオミクスだけでは解決できないことが分かってきて、グライコミクスが注目されている。

○ユーザープロファイルに関する何らかの数値的な情報を次回ご提出いただきたい。

▲9 月末時点の到達目標、要望について▲

- ・ユーザープロファイルに関する何らかの数値的な情報を提出する。
- ・本委託事業で実施したツールなどは再利用可能なようにする (DBCLS が調査する)。
- ・新サービスの公開時には、中核からの広報活動による紹介や使い方説明として統合 TV を活用する。

➤ 全体へのコメント (一部の議論は、各機関の質疑応答の最後に追加した)

○サービスを公開するにあたって、中核などでのイベントで紹介に加えて、その使い方の説明としてマニュアル以外にも統合 TV などでもっと紹介してほしい。

→ (DBCLS) サービス提供者からのリクエストがあれば対応する。

○この分科会は参加機関が多くて時間がかかるので、次回は発表順番を逆にする。

(18 : 40 終了)