

【日時】 平成22年1月22日(水) 10:00~13:30

【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)、五斗進(京大)、中谷純(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)、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)、新間陽一(産総研)、鹿内俊英(産総研)、本間善之(文科省)、田中康博(文科省)、山中芳朗(文科省)、堀田凱樹(ROIS)、大久保公策、高木利久(作業部会主査)
永井啓一、西川哲夫、川本祥子、箕輪真理、吉羽洋周、畠中秀樹、河野信、八塚茂(以上、DBCLS)
(敬称略・順不同)

【議事】

1. 平成22年度予算配分について(文科省、田中氏より説明) 資料1-1~4、資料2

この内容(資料1-1~2)は22年度の計画策定にあたって文部科学省が策定したもので、研究運営委員会に年末に周知した内容である。22年度委託費の配分について、21年度の達成水準と22年度の達成見込みを基に実施したいが、その際の評価を、成果報告/22年度の計画(案)/ユーザ評価の内容を参考に、作業部会メンバーの相互評価によって行いたい。22年度の計画については、開発は上期(9月末まで)に終了し、後はブラッシュアップに注力してほしい。内容についてもJST内に設立予定の新センターへ引き継ぐべき内容かどうかと観点で判断してほしい。文科省が決定する範囲は、文科省からの委託費の部分。JSTからの予算の配分については、参考意見を届けるにとどめる。評価の単位としては、機関ごととする。ただし、医科歯科Gと東大Gについてはグループ単位での評価とする。一人の評価者が最大14件の評価を行うことになる。

資料1-3はスケジュール、1-4は各機関の委託費の内訳(概要)であり、資料2は21年度の計画と12月末の時点での達成状況をまとめたものである。

2. 中核機関より

プロジェクトの最終年度の取りまとめを前にして、全体として同じ目標を共有するためにも、中核機関の活動内容をご理解いただくということで、川本特任准教授から現在の取り組みと来年度の計画が紹介された。

◇資料説明◇ 資料3

取り組みとしては、統合ホームページの概要、アクセス統計、カタログ系サービス、横断検索、アーカイブサービス、統合TV、共通基盤開発(その情報流通における位置づけ)、これらの年次計画上での進捗を紹介した。次年度の計画としては、統合化の全体像とステップを提示した上で、横断検索の機能向上から統合検索への展開、およびそのプロトとしてのTogoProt、ツールとしての辞書シソーラス、を紹介した。また、ユーザ評価について、各機関の協力を感謝するとともに、結果の公開について紹介した。

◆質疑応答◆

○DBCLS以外の中核の他の機関との連携で生まれた成果についても説明いただきたい。

→辞書については奈良先端大との情報交換を開始した。自然言語処理においてははずさで利用されている等、いくつか連携の結果は出ている。

○本年度の成果報告書にそのような内容を盛り込んでいただきたい。

→了解した。プロジェクト外ではあるが、経産省と連携して項目の統一化などを実施しているので、そのようなことも盛り込んでおく。

3. プロジェクトの平成 21 年度進捗状況および平成 22 年度業務計画について

▶ 東京医科歯科大学グループ

◇資料説明◇ 資料 4

【医科歯科大】21 年度の成果として、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、②疾患 DB の高度化と実証的統合に関する研究、③プロジェクトの統合的推進、④その他を行った。以下詳細。

①インターフェース (IF) の高度化として、阪大 DB を外部モデルとして HTTP プロトコールでの情報検索フローを作成。ただし、詳細データの取得における認証についてはさらに検討が必要で、各機関での倫理基準のすり合わせが今後必要になる。要素技術の国際標準化の推進として、GSVML が国際標準に採用、NCK が WHO で検討された。②アクセスレベルの設定として、DB のサブセットを設定し、フリーアクセスレベルの試験運用を開始した。③倫理課題検討会を実施し、現在第 3 案となる原案が作成されている。④HP デザインを変更し、成果報告を掲載した。

【阪大】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として、医科歯科大開発の検索エンジンとの統合化と、使い勝手向上のための改良 (結果表示、クロス検索の IF 改良) を行った。②疾患 DB の高度化と実証的統合に関する研究として、患者からの同意取得を含め、データの拡充・公開に向け作業を進めている。22 年度は、統合技術開発に関しては、技術上の課題、他の疾患への適用時の課題を検討し、知見をまとめる。実証的統合については、連携システムの効率化を図るとともに、公開症例数を追加する。

【全体としての 22 年度の計画】

使い勝手の向上 (病情報など用語連携の強化)、ICD 新版へのマッチング、知名度向上、倫理案に従ったアクセスレベルの設定、フリーアクセスレベルの利用評価の実施および倫理案へ反映を予定している。

◆質疑応答◆

○多型情報の統合については PML(Polymorphism Markup Language) が世界標準となりつつあるようだが、GSVML とは異なるカテゴリのものか？

→関係者間で情報交換中である。PML は SNP に特化。GSV(Genomic Sequence Variation)ML は、範囲は広いが少しラフなもので、臨床情報とリンクしているところが異なる。

○GSVML は LSDB の成果として発表されたのか？

中谷 GSVML はすでに確立されたもので、その一部を LSDB のほうへ採用したということである。

○東大 G ではこの ML については検討したのか。

→[東大 G]GSVML についてはあまり知らなかった。

→学際的というよりも、Clinical よりの Genomics 情報の標準化のためのものである。

○同じ情報でもドメインが異なる場所では異なる標準化がされるものなのか？

→そういうことが無いほうがいいと思うが、GSVML は Clinical Genomics の中でもかなり Clinical に寄ったところから出てきた標準化の動きで、すでに 180 カ国くらいで採用されている。

○PML については dbSNP での採用を働きかけられたが不採用だったという経緯がある。ルールとしてはあまり具体的なものではなく、多くの人が採用しなくては実効性があまりないので、PML はまだ標準とは言えないのではないか。

○多型 DB としては最大の dbSNP 形式で書かれたものを GSVML に書き換えることはできるか？

→できる。多型情報と組み合わせる臨床情報のほうにもある程度共通のフォーマット (HL7 など) があり、GSVML はそれにも対応ができている。

○技術的な課題以外に、前回の議事録にあるように、どういうデータをどのように公開することを考えるのが大事という指摘があり、そちらの検討を進めていただきたいという意見があったと思うが。

→米国 HIPPA ルールの項目と症例に関する情報をフリーアクセスレベルでは削除している。それについて

の評価を元に公開の基準を考え、来年度はひとつの結論（案）を出したい。一方で、HIPPA で特定できるかできないかについては2年ほど前から世界的に議論があるので、それについても検討を継続する。

○マイクロレイの画像もフリーアクセスで閲覧できるが、これをデータで提供することができないか？

→ICOD の試験運用では一部データを取れるようになっている。公開してみても、問題無いかを検討する。

○他のデータも解析可能な形で提供してほしいと思うはずなので、公開に向け検討をお願いしたい。

○東大 G でもきちんとルールを作ったが、それは委員会等を開いて検討を重ねた結果なので、同じ手順でルールを作ればよいと思う。プロジェクトの成果として一番求められていることではないかと思うので、わかりやすいルール作りをお願いしたい。それをもって、公的な場に提案することはできると思う。

○22 年度の最終ゴールとしては何を設定しているのか？

→最も大きなゴールとしては倫理原案ではないかとおもう。

○スケジュール的にはいつごろ原案が出され、最終案がまとまるのか？

→上半期のうちに中核機関に審査してもらえるように提出したい。

○倫理原案がゴールなら、極端な場合、技術開発は中止して、原案作成に作業を集中させることも可能か？

→フリーアクセスの利用評価は倫理原案作成にも関連する。心因的な要因もあるので、正しい利用評価を得るためには開発も一部継続する必要ありと思う。

○倫理原案は重要である。重要性を説得するためにも検討時にある程度完成度のある DB が必要だと思う。

○フリーアクセスの部分はダウンロードして使ってもいいのか？

→評価期間中は OK。

○転用できるもののほうが、値打ちがある。ブラウザだけではだめ。

○但し、稀な疾患であればある程、写真などの情報が別ルートの情報と合わせて個人特定のリスクがある。そのために倫理関係での懸念がある。ある程度稀な疾患についても説得できるような議論のための成果を出せるようお願いしたい。

○ダウンロードして再解析しようと思うのはそれほどまれな症例についてのデータではない。逆に稀なデータについては、写真を含め、これまでもかなり詳細な情報が出されていると思う。プライバシーの問題はこれまでそれほど起こってはいないのではないか？

○個人情報保護の意識の高まりなどで、だんだん事情も変わってきている。

→これまでの議論の中でも利用の形態について検討しておかなくてはならないという懸念が出てきている。

1 次利用（DB 作成）、2 次利用（統合など他の研究で利用）、3 次利用（商用利用）については、HIPPA では 3 次を禁止している。フリーで公開した場合、3 次利用として商用に転用される可能性があるかどうかを考えなくてはならない。

➤ 東京大学グループ

◇資料説明◇ 資料 5

先ほど議論のあった倫理問題については、東大 G では今年度共有方針他一連の関連文書を作成したので、メインの共有方針文書を追加資料として配布する。

開発項目についての新規報告として、標準 SNPDB（追加検体データ登録待ち）、Case-Control DB については量的解析に関する検討、CNV については表示改訂（3 種類のアルゴリズムの計算結果など）、等を実施した。共有ルールについては外部の委員を多く招き、検討して作成・公開（日英）、具体的な案件が出てきているので近日中に委員会を開催し、具体的手順について確認して進める。22 年度は、DB の改良、追加登録と公開、機能拡張を行い、データの受入と提供サービスの運用、海外機関との連携を進めたい。

◆質疑応答◆

○データの提供ルールについて、レベル 2 は申請のみ、レベル 3 は許可に審査が必要ということだが、CNV

はなぜレベル 2 なのか？

→CNV については、個体レベルの情報ではあるが、個人情報には該当しないので、レベル 2 となっている。

○Resequencing という意味は？この言葉には一般的にいくつかの意味があるようだが。

→標準ゲノムが決まっているので、個人について再度特定の部位を再度解析するという意味での Resequencing である。検体を増やすという意味合いがある。

○Resequencing でも WholeGenome でやると GWAS 目的にも使えるので、WholeGenome であるかないかをはっきりさせたほうがいいのではないかと？

→その意味ではここで扱っている Resequencing は、Locus 特異的なもので、WholeGenome ではない。

○いくつかの疾患が取り上げられているようだが、その選択については何か優先順位があるのか？

→提供してくださる研究者に依存している。

○罹患数の多い疾患と稀なものでも、提供してくれればすべて受け入れるということか？

→GWAS についてはある程度の頻度の疾患でないという意味が無いが、原則すべて受け入れる。

○データ共有委員会の位置づけを教えてください。

→[DBCLS]これは統合 DB プロジェクトの委員会である。各大学で倫理審査委員会を作って審査するのと同じで、3 省の指針の範囲内で具体的な審査するものである。

○データ登録は増えているのか？人類遺伝学会で論文化の際にデータの登録をこちらにという働きかけをしていたと思うがその効果は？

→いくつかのアカデミックな会合などでは、宣伝はしており、認知は進んでいると思う。

○厚労省関係のデータの登録については？

→1 つ登録済み。研究者間では、登録協力の話は出ているが、具体的にはこれから。論文化されてから、という条件がつくので時間的にラグが生じる。

○医学以外の分野では研究素材の再現性が重視されるが、医学分野では、データの再検証のためにデータを登録しなくてはならないという学会標準のような動きの見込みはないのか？

→まさにそのための DB だと思っている。理研の SNP-DB 以外は網羅するように、研究者コミュニティで働きかけているが、上記の理由で多少時間はかかる。

○この DB に登録したら、一定の QC を経て Q 値がついて公開されるのか？それだけでも値打ちがある。

→そうである。一度利用していただいた先生は、また次もと考えていただいているようである。

○この議論はつぎの九大の発表とも関連するので、続けてご発表いただきたい。

▶ 九州大学

◇資料説明◇ 資料 6

DB とはブラウザで見るというよりも再解析が可能なものであるべきである。そのためにデータの信頼性、後の使われ方が大事である。信頼性を高めるためには、1 か所でデータの受け取りから公開を行うのではなく、QC のステップを外部に出すことが必要であると思う。今後は、手続きを踏んで東大 G のデータを入手して、実データを用いて構築した QC パイプラインを検証したい。また、SNP の標準データとして、統合 DB に登録されている情報に、我々が有する日本人確定ハプロタイプ情報を提供して、高精度のプロタイプ情報を含む日本人標準 SNP データベースの確立に役立てることを予定している。

◆質疑応答◆

○共有方針を決めたので、手続きを取っていただければ必要な情報は出すことが可能になっている。

○再度提案するが、Resequencing という言葉は意味が混同している（同じサンプルを再度解析する、あるいは標準ゲノムが既にあるそれを参照しながら別の個体を解析する）ので、やめたほうがいいのでは。別の適切な言葉があるように思う。

→[東大 G]DB 名称としては MutationDB となっている。提案時の名称が Resequance であったため、今でもその名前でプロジェクト内では話をしているが、外向けの DB 名称としては使っていない。

➤ 京都大学

◇資料説明◇ 資料 7

共通基盤技術開発として、プログラムの改良、表示機能の追加、反応経路予測ツールの公開を実施、統合 DB 開発・運用として、LinkDB 対象 DB の追加、LinkDB 検索機能の追加を行い、医薬品 DB と LinkDB については定期的な更新を行った。22 年度の予定として、医薬品 DB の高機能化(他の医薬品 DB, や副作用 DB との関連付け)、化合物 DB の KEGG DB との関連づけ他、LinkDB の自動化他、共通基盤開発 (複数化合物入力の実装、予測結果パスウェイの関連付け、DB 間の関連検索の強化) を検討している。

◆質疑応答◆

○AERS (Adverse Event Reporting System : 米国 FDA が公開している医薬品の有害事象データベース) に相当する物は日本にはないのか?

→探してみたところ、無さそうであった。

○厚労省関係では、医薬品基盤機構から一部公開しているはず。ただし、公開の基準があまり明確でない。

○開発は上期までという区切りをつけることが可能か?

→きりが無い工程 (開発→評価の繰り返し) になってはいるが、可能だと考えている。

○アクセスの内訳はわかるか? 製薬会社からのアクセスはあるか? また、医薬品 DB の何のサービスを使っているかはわかるか?

→ドメインとしては製薬もあり、詳細解析はしていないが、10~15 くらいのドメインがある。利用サービスの内訳は不明。化合物 DB のツールのアクセスは医薬品 DB の半分くらいである。

○開発項目 3 つについての優先度は? 死守したいものはどれか?

→PJ の目的としては医薬品 DB の開発は必須。LinkDB が無いと関連情報が取れないので、こちらも重要。

○経費内訳として、オリジナルコンテンツの作成費ではなく、ソフトウェア開発の外注がメインか?

→そうである。人件費もソフト開発が多いが、RPAIR などの DB 作成にもかかっているため、一部コンテンツもできている。

➤ 産業総合研究所糖鎖医工学研究センター

◇資料説明◇ 資料 8

プロジェクト雇用の 5 名体制で本 DB システムが出来上がっている。米国、欧州ではそれぞれコンソーシアム体制で糖鎖関係の DB ができ始めており、JCGGDB は日本の DB として、アジアへの展開を考えたい。

それぞれの DB は、特徴があり、米国は Chemistry 中心、欧州は NMR データが多い。日本では Biology と Chemistry を中心にしている。対等な 3 極構造として合体させ、世界的な DB へと発展させたい。

20 年度は書式等の統一を行い、各機関からのデータを収集し、キーワード、構造検索ができるようにした。

21 年度は、専門家によるプロトコルの収集を行い、その機能の一部として、糖鎖合成支援のツールを作成した。また、糖タンパク質統合 DB として、タンパク質上に実際にある糖鎖修飾と糖鎖構造の情報を収集し、生物種間で閲覧できるようにした。糖鎖結合性の病原体の情報、糖鎖関連疾患と原因遺伝子の情報、K/O

マウスの情報、ミニレビュー閲覧サイトを公開。統合検索のプロトタイプを開発した。22 年度は、オンラインプロトコル、ミニレビューを充実させ、新規 DB を整備するとともに、統合検索を完成させる予定。

国内の糖鎖の研究者が協力的なのが、非常に強力である。担当者の努力により、非常にユニークかつ、有益な DB ができつつある。

◆質疑応答◆

○ユーザーアクセス数はどのくらいか？

→GGDB だけで、外国も含めて年間 30 万アクセス。(国別集計は未解析)

○今後、経産省からのサポートは無いのか？

→PJ の終了時に DB としてまとめるということは経産省では行わない。経産省と文科省の文化は違っており、成果の公開の範囲についても世界か、日本限定かという考え方も違う。

○農水省もまた違った考え方を持っている。

→実際に、知財化、論文化されてないものは DB には載せていない。

○知財を取ってからとなると経産省の考え方でも折り合うのではないか？

→知財を活用するステップも独占したいので、知財を取ったあともすぐに広く公開することはしないようだ。

○(糖鎖のグループには限らないが)各機関で公開したものについて今回の報告で初めて知った例も多い(例、PACDB)。LSDB トップページや JST のサイトを活用した情報発信を積極的にしてほしい。LSDB の成果について外から見えにくいためにアピールしにくいので、ご協力いただきたい。

○統合検索については LSDB のものに先駆けているのか？

→[DBCLS]統合検索は分野を限ったところと全体的なところではやり方が異なる。LSDB でもタンパクに特化して先行的に実施している。

○ユーザ評価の中に糖鎖への厳しいコメントがいくつもあったが、それは期待や評価の表れと考えていただきたい。使い勝手についての厳しいコメントがあったが、それは分野の専門家でないために使い方がいまいちわからないということなのか？

→マニュアルなどの整備がまだなので、その点是对応したい。

○例えば、この使い方を統合 TV などに投稿するとか、そのようなことも検討してほしい。

→時間があれば対応したいと思う。

○開発の区切りはどのあたりと考えられるか？

→ゴールの設定は難しい。研究の発展に伴って必要なものはどんどん出てくるのできりが無い。

○この PJ は受け皿の骨格を作るためのものなので、ベースができたことで成功と言えるのではないかとおもう。それが海外とのハブにもなればさらに良い。

○日本のものは全部入るのか？

→今のところ、1つを除きほぼ全部入る予定。

○多型データ (GWAS) については、なかなかそうはいかないですね。

→[東大 G]全部とは言えないが、多型解析をしているグループも限られているのでかなり網羅できると思う。

○人類遺伝学会での論文投稿時の義務付けがベースになるのではないか？

→[東大 G]雑誌のほうはそのようなルールで運用。ただ、先に国際的な DB にデポジットされたケースもあるので、その場合はお願いベースで進めていく。ところで、NIH の場合はグラントで出したものは提出義務があることになっている。日本の場合もデポジットを促すように各省で方針を出してほしい。

→[文科省]文科省ライフ課としてはその方針でお願いしているが、省によっては文化が違うので、その後の開発などのことを考えるとなかなか足並みがそろわないのが実情。

(13 : 30 終了)