

【日 時】 平成22年6月1日(水) 15:30~18:10

【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】 中村保一(遺伝研兼かずさ)、藤澤貴智(かずさ)、岡本忍(DBCLS)、中尾光輝(DBCLS)、豊田哲郎(理研)、土井孝爾(理研)、国島直樹(理研)、皿井明倫(九工大)、田畑哲之(かずさ・作業部会委員)、田中康博(文科省)、中村春木(阪大)、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、金進東、吉羽洋周、河野信、高祖歩美、坂東明日佳、八塚茂、平井信一、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。

◆質疑応答◆

○展示活動について、中核以外のメンバーの参加はあるのか。

→リーフレットの作成など協力いただいている。一時は展示会対応人員として協力いただいたこともあるが最近では当日対応については中核のメンバーだけで行うことが多い。

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

➤ かずさDNA研究所(中村)

◇資料説明◇ 資料4

植物および植物関連微生物のゲノム情報DB統合と高度化として、アノテーション蓄積の仕組みを構築し実際に運用しながら、改善を実施。この1年で7(14-7)生物種を追加し、3,029(6,468-3,439)報の論文から54,399(161,373-106,974)件のエントリを追加。H22年度の実施計画としては1. 高度情報集積DBの運用と改良、2. ゲノムアノテーション情報の蓄積と高度化、という21年度の継続内容で、9月末までには1については、システム改善・登録ユーザを100名規模に増加・BioMartによる情報追加、2については、1000報分の情報追加・複合体情報を500件追加・可視化改良を予定している。

コメントへの回答は以下の通り。

<システム開発の必要性>

大量のゲノム情報へのアノテーションづけに対する効率的な支援システムが必要。

<アノテーターの確保に依存。>

人材確保4割、作業デザイン2割、システムデザイン2割、コミュニケーションデザイン2割と考えており、人材確保は確かに重要。そのためにも人材の教育、地位向上も必須。

<他機関との連携>

ゲノムに関して言えば、名寄せを実施するのは有用。連携についてはこれからも議論する必要あり。

◆質疑応答◆

○複合体と要素タンパクの両方にアノテーションを付けるのか。

→両方のやり方が可能であるような情報のつけ方をしてある。

○コンテンツについての計画は予想通りで良いと思う。一方、プログラムについてもPJの成果という位置づけであれば、それらも配布可能とするのか。

→他のところでも使えるようにしてほしいというご意見もあった。最終年度にはかずさ以外のサイトで利用する案件が出てきている。

→配布するようにまとめることは可能だが、作業が必要。プライオリティは低い。権利関係としては整理すれば隘路事項はない。

→かずさのサイトを利用してもらえれば、他にシステムを立ち上げる必要はないので、システム配布自体を積極的に進めるメリットはあまりないため、検討していない。

○アノテーター、キュレーターについて指し示す対象者を確認したい。

→言葉を混同して使っていて申し訳ない。実際にデータを入れているのはキュレーターとして紹介した人たち。登録ユーザというのは、書き込みもできるので、単に見るだけではなく、これを利用してコミュニティを立ち上げたりしている人。アノテーターとして機能しうる立場ではあるが、それほど多く書き込みがなされているわけではない。現在あるエントリは雇用されている人がやったもの。

○1エントリの単位は。

→ある遺伝子についてどの論文のどこに書かれていたよ、という情報。論文数にすると減る。

○最初に論文を選ぶのは、読んだら出てこなかったということもあるのか。

→ある。遺伝子の名称が基準化されていないので、人手でやる必要がある。例えば、別名があったり、表示がⅢだったりⅣだったりする。

○特定の遺伝子についての一連の参考文献が出てくるという使い方でいいのか。

→それも一つの見方。論文単位で見える見方もある。ただし、現在は資料4の2ページの下にあるような見方（論文そのものにマーカーが引いてあるような図）はできない。論文の権利関係もあるので。

○資料2-3で掲げた9月末の目標（研究者コミュニティの認知、中心的なツールとして機能すること）については、例えば、登録ユーザが100人規模になれば達成したと置いていいか。定量的な指標は無いのか。合格点は設定できないか。

→判断が難しい。登録者数で見るとは、すでに世界の中心的なDBとなっているかずさのシアノベースからのリンクの数で有用性の認識とすることはできるとは思う。

○どこかの時点で、9月末の評価に向けた具体的な数値目標を設定してほしい。別途検討してほしい。結果だけではなく、努力の過程で評価するので、達成できなかつたらという心配は必要ない。

→確認したいのだが、目標として掲げられている「統合DB基盤システムのツールとして…」というのは。

○一連のサービスがかずさのサービスとして見られているような気がするので、統合のシステムの中に組み込まれていることを明示的に示してほしい。

○それはプレゼンの問題もあるかも知れないので、検討していただきたい。

○かずさアノテーションにはすでに蛋白質間相互作用（PPI）の情報も入っているが、それらの情報と複合体情報との関係は。

→PPIのほうはずさの持っていたツーハイブリッドデータを一括登録したもの。複合体情報は個別の蛋白質に限定されない情報として登録の必要性が出てきたもので、確認されている情報を登録するもの。別の概念である。

○データの増加への対応がひとつの大きな問題であると思うので、別途全体的に検討する必要はあるだろう。

▲9月末時点の到達目標、要望について▲

- ・9月末の目標（研究者コミュニティの認知、中心的なツールとして機能すること）について具体的な評価目標（定量的な指標）を設定する。
- ・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。

➤ 理化学研究所（豊田）

◇資料説明◇ 資料5（追加配布資料あり）

①シロイヌナズナ

今年度の予定としては、9月末に向けていくつかのデータの整理・公開を予定（いくつかの成果については、Plant and Cell Physiology[日本のジャーナルだが、認知度は高い]に特集号として発表予定）しているとともに、DB統合化の範囲を拡大しつつある。理研外のDBへの対応、マウスへの適用も検討中、国際対応、FANTOM4データの統合も進めている。

②横山データ

ダウンロード（DL）可能な形で一部公開済み、横山研で付与したアノテーション情報付きでDBCLSへ直接データを渡すことになっている。

③蛋白質構造データ

これも公開済みなので、データの追加に従って順次公開。すでにDBCLS側ではデータをサイトから回収して使用していると思うので、追加データについても同様に対応していただきたい。

理研のミッションはデータを作ってDBCLSに渡すというものであったが、ユーザ向けの対応も必要となってきたことから、簡単なAPIとその仕様も提供する。利用プログラムも公開利用できるようにする。利用者がサイネスの中にワークスペースを持てるようにし、オープンな参加型のDBシステムとして、ゲノム設計コンテストを開催(募集中)。理研データを一括して検索できるシステムを構築中で、DBCLSの横断検索とも連携する予定。

コメントへの対応としては、

<理研のものしか入っていない>

いくつかは外部との連携もあるが、あまり積極的には宣伝していない。文科省PJとの区別も必要かと思う。より大きな統合という意味でも、国際的な広がりを検討している。

◆質疑応答◆

○タンパク質関係で、資料5の12ページにあるような件数を9月末までに全部DBCLSにいただけるということだろうか。

→この件数は横山先生のところの21年度の実績なので、サイネスで公開できているものとは件数が違うかもしれないが、この件数をDLできるというつもりである。ところで、データを渡した後は、オリジナルサイトでは削除してもかまわないのか。（将来的にサーバなどがひっ迫した場合など）

○まず、現在は回折データがDLできることを確認しているが、そのほかのデータについては不明なのでどういう状況か。

→資料 5 の 12 ページのデータについては計画上横山さんから DBCLS へ渡してもらうことも可能なので直接話していただいてもかまわない。

○回折データに関して言えば、こちらに置くのはメタデータで DL の際に理研サイトに行くということにしてはいるはず。データの全部をこちらで引き取るという話になっていないはずだが。

○理研では維持できないということか。

→長期的に考えたときに、どうするかを考えたい。将来的には理研で保持できなくなる可能性もあるということだ。

○理研が消すか消さないかについてはそちらの判断である。DBCLS にあずかったものは組織がある限り消すことは無いと考えている。

→計画上もともと渡す、ということで準備している。理研ではこのデータを用いたユーザーサービスを想定していないので。

○メタデータは整備されているのか。だとすると、回折データについては URI で指定すればいいので、どこに置くかは問題ないのでは。

→データを”渡す”といった意味は何かということ。理研でメタデータを付ける必要もないので、ちよくせつ DBCLS が受け取るのであれば、そちらでメタデータを付けてもらって構わない。

○横山先生も交えて、後で別途相談する。

○（蛋白質構造 DB について）PDB の URL が正しくない件についての対応をお願いしたい。

→PDB の URL の変更については、HATODAS II への更新（DL できるようにする変更）の際に、一緒に行うつもり。

○DB Registry の CC ライセンスは、CC-BY-NC-SA なので変更して NC をはずしていただきたい。

→DBCLS が NC なので、問題ないのでは。

○DBCLS 自体が NC だから使用に際しては問題ないという意味か。そうではなく、再配布の際に支障になるのではずしていただきたいという趣旨である。DBCLS では公開に際して NC はつけていない。

→外すのは問題ないと思う。他の例で PDF などでは NC となっているものがあつたので、混同しているのかもしれない。確認して整理する。

○9 月末までにフェノタイプ情報を統合する点について、オントロジーの変更に伴って段階的に変化させるのか、9 月にまとめて変わるのか。

→作業が進むにつれ、新たなオントロジーがリンク付きで追加される。もともとあるものはリンクを含めて不変なので、追加されるイメージ。

○TAIR のバージョンについては、7→8→9 と変えていくのか、一気に 9 に置き換わるのか。

→これもリンクが増えることになると思う。前の情報は残る。

○資料 2-3 における到達点（統合 DB のミニマムモデルの成功事例として認知される）について、具体的にどうなっていれば認識されると思うか。

→認知されるための統合検索サイトを公開する、ということを目指にしたい。もともとの計画ではデータを渡すというのがミッションで、一般ユーザからの評価の対象となることを想定していなかったもので、評価可能なサイトを公開することが目標。認知のための努力（広報など）は別途実施する。

○検索サイトというのはサイネスのこと。

→サイネスそのものではなく資料 5 の 19 ページにあるような検索サイトのことを想定している。9 月末までには雑誌の特集号に投稿予定なので、その時点で公開できるもので評価してもらうことにしたい。

○DBCLS の横断検索との連携について、現在は理研の DB は全く入っていないのか。

→（DBCLS）FANTOM など公開されているものはすでにはいつているものもある。仕組みとしては、キーワード検索になじむものなら、こちらで検索するというものもあるが、こちらから理研全体を網羅的に検索し

ようとするとは出来ないこともあるので、理研で統合検索の仕組みを作るのであればそれと連携することもできるのではないかと思う。

○本統合 PJ の委託の範囲は、対象 DB が決まっていると思うので、まずはそれについて実施してもらい、それを RIKEN 全体に広げるのはその後やってもらえれば。

→委託の範囲の検索が 9 月末に実現できると思っている。検索クエリが理研側に投げられ、対象案件のカウントが出て、詳しく見ることもできる、という仕組みにしたい。理研全体のサイネスに入っているもの全部について検索時のカウントが見られるようになる。DBCLS が経産省と連携しているやり方と一緒にするのはないか。

→(DBCLS) 提案内容は経産省とは少し違う仕組みになる。植物やタンパクについて網羅的検索はもうできると思う。統合 PJ への期待としては、横断検索の際に機関(たとえば理研全体)ごとに網羅するという要求もあると思うのだが、それが実現できるかどうかについては判らない。

→理研の中で検索できるようにして、網羅的にできるようにするつもり。

○資料 5 の 16-18 ページの取り組み(API 公開、参加型システム)は、本 PJ の外の部分か。

→計画に出している範囲である。コンテストの費用については理研が出すが仕組みづくりは本 PJ で対応。

○追加の資料を見る限り、PJ での対応とは思えないが、何処にも文科省 PJ の名称が記載されていないので、もしそのような位置づけであれば記載していただくほうがいい。

→コンテストについては PJ 対応ではない部分もあるので、内部での相談のうえ、理研の事業として公表されている。成果利用として報告するという判断だった。システムはコンテストに限ったものではなく、汎用性のあるものであって、別途提供している。その認知度を高めるためのコンテストである。PJ の名称を入れるとかえって、コンテストそのものに委託費が使われているという印象を持たれるのが心配で省かれている。コンテストで集めたプログラムについては DB 公開していくつもり。そういった意味での成果利用である。文科省にはコンテストについて相談した際に、文科省統合 PJ の成果利用であることを出すか出さないかも相談しており、出さないという判断を貫いている。

○出すことに際して、支障はないはず。かえって宣伝になるので。文科省サイドで一応確認する。

▲9 月末時点の到達目標、要望について▲

- ・横山データについて、DBCLS への引き渡し内容を整理したうえで引き渡しを完了する。
- ・研究者コミュニティから認知される統合検索サイト(委託事業で指定されている範囲の DB をカバー)を公開する。
- ・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。

➤九州工業大学(皿井)

◇資料説明◇ 資料 6

物理化学データの統合化、文献からの情報抽出がテーマ。専用 HP から成果を公開。各種データベースでは確実に件数を増やし、引用も多く行われている。データの一括 DL も可能。22 年度の計画としては、クロスリフェレンスの充実、蛋白質間相互作用 DB の作成などのほか、TogoDoc を用いたデータの自動収集ができるように改良したい。本事業終了後も継続を検討。

コメントへの回答としては、以下の通り。

<遺伝研 PMD との連携>

PMD は更新ストップ、他の変異 DB との統合を進めている。

<長期運用への対応>

作業軽減のための自動化を検討中。

<個別 DB の統合のための検討>

個別 DB には専門家が必要。維持のための資金や体制をオールジャパンで検討してもらいたい。

◆質疑応答◆

○データ収集の省力化について、数値的にどのくらい省力化できそうか。

→文献収集に関してはかなりの論文が取れる。ただし、ノイズ(関係無い論文)が多い。半分くらい楽になった。抽出に関してはあまり省力化できない。WiredMarker も使っているが、論文中の表からの抽出は無理。

○ユーザ数は

→月に 2000-4000 件くらいのページビュー。但し専門性が高いので、ユーザ数は数百程度と思われる。

○22 年度計画について、クロスレファレンスというのは、新規の部分を対象に追加する、ということか。

→そうである。内容を更新するというので、新規な項目追加ではない。

○PPI のデータについてはその特徴は。

→九工大で実験している研究者がおり、そのハイスループットデータを入力することを想定して仕組み作りをした。実験が遅れているためデータはまだ格納されていない。生物種は限られていない。

○データ公開について「原則的に」というのはどういった制限か。

→誰が DL したかは判るやり方(サインをしてもらって、パスワードを提供)でやっている。ただし、LSDB の横断検索では、すべて検索できる。

○サインをしてもらうのも、PJ の方針とは違うが、それについては。

→PJ に参加する前から 10 年以上このやり方をしているので、そのままになっている。止めてもいいが、誰が DL したかを把握したいので。

○商業利用不可については。

→これも同様。ただし、DBCLS から DL できるデータについては NC を外している。

○DB の更新やアノテーション作業はラボ内で閉じて実施されているが、オープンに実施できないか。

→可能性としては考えている。誰でもできるというわけではないが遠隔でできるという意味では、CBRC から協力してもらっている方もいる。

○このままでは、九工大がそれまでに手掛けていた DB 構築作業に対して補助金を出しているようにも見える。本来は統合 DB に必要な DB の構築という位置づけのはず。その意味では、PJ 終了時に九工大の DB として別途進んでいくというのは趣旨ではない。DB を作ったのではなく、アノテーションの仕組みなどインフラを作ったと言えるようにしたい。そういった形の事業貢献をお願いしたい。独り歩きできるように。

○CBRC 以外にもリクルートして、全体をオーガナイズするという立場でやっていただくのは可能か。

→システムを使って、あまり人手をかけずにできるようにする、システムは他のところに移すということも可能だと思う。その意味での資産は残ると思う。

○(ビュー数からみて)かなり小さいサイトだし、ユーザも限定されて、コミュニティは小さいのでは。となると、他のドメインでも使える、というメリットを出してもあまり意味がないのではないか。

→似たような小さなコミュニティがあった時に、使えるシステムであるということ是可以する。

○小さい DB で成功例を見て、他のところにも使えるようにすることには意味がある。

○ProTherm 自体はユニークなので、うまくリンクするなどすれば、アクセス数の増加もあると思う。

○(DBCLS から) TogoProt ではホモログとの関連付けなどで少し工夫しているので、今後は多少アクセスが増加することが期待されると思う。

→データの見せ方の工夫などでも改善できるのではないかとと思う。

○13 年もたつて 2000 ページビューというのはあまり多くないだろう。人手をかければこの手の仕事はどのドメインでもできる。やはり利用度など、客観性のある評価基準が見当たらず、そのようなものが必要。国の予算の使い道としてはそのような判断基準が必要だろう。

○やはり DL の際にサインするというのがネックなのではないか。はずしてみてもいいかもしれない。今後検討する。

→やってみてもいいかもしれない。今後検討する。

▲9 月末時点の到達目標、要望について▲

- ・アノテーションの仕組みなどインフラを作ったと言える、そういった形の事業貢献をお願いしたい。
- ・ダウンロード時のアグリーメントへのサインの撤廃など、利用度を促進する試みを検討する。
- ・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。

最後に、LSDB 全体のパンフレットが更新されたので、機会があればご利用いただきたいという連絡があり、閉会した。

(18 : 10 終了)