

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分科会【立体構造・植物・化合物関係機関】議事要旨

【日 時】 平成22年11月2日(火) 10:00~12:25

【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】 田畑哲之(かずさ DNA 研)、中村保一(遺伝研)、藤澤貴智(かずさ DNA 研)、岡本忍(DBCLS)、中尾光輝(DBCLS)、五斗進(京大)、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)、田中康博(文科省)、高木利久(作業部会主査)
米澤明憲、大久保公策、永井啓一、坊農秀雅、河野信、高祖歩美、山本泰智、吉羽洋周、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(資料確認含む)

議事要旨(作業部会分科会参加機関の組み合わせが前回と今回で異なるため2種類あり。事前にメールで確認済み)の修正があれば本会議終了までにお申し出いただきたい。また、委員名簿の変更(JSTからの委員が変更、松原委員の職名変更、高木主査の所属変更)についてご確認いただきたい。

(高木主査より)プロジェクト(以下、PJ)周辺の近況報告を行う。半年後のPJ終了後、事業はJSTに設立される新センターへ移行する。DBCLSは技術開発を目指す。10月1日にJST新センター準備室が設立されるに伴って、高木主査はDBCLSセンター長の任を解かれ、JSTの準備室へ参加した。ただし、引き続きDBCLS特任教授として本PJを担当する。DBCLSの新センター長には米澤東大教授が就任した。準備室の情報はJSTのホームページで発信される。

2. プロジェクト平成22年度上期進捗の評価依頼について 資料2-1、2-2

文科省ライフ課田中調査員より、すでに10月15日付で作業部会の委員に依頼済みの22年度上期進捗評価について説明があった。

各機関(評価単位は機関ごと)について、9月末目標の達成状況、下期計画の妥当性、22年度当初の評価に対する意見・コメントへの対応をポイントとして、作業部会委員に相互に評価いただきたい。年度当初の評価ではすべての事業が継承するという結果だったので、確実に成果を出せるよう、建設的なアドバイスや提案をお願いしたい。

フォーマットは年度当初のものと類似したもの(会議資料参照)。メ切りは11月10日。評価内容は無記名で集計し、まとめた情報を各機関へのフィードバックに用い、文科省もそれに基づきアドバイスする。また、事後評価委員会の参考資料とする予定。(本来PJ終了後に行う事後評価であるが、本PJについてはその後もJSTへ移管して継続されることから、その実施内容の参考とするためにPJの終了を待たずに実施するものである。)

事後評価について補足すると、それへの対応作業として、実施機関向けに成果報告書(毎年提出しているものの9月末までのアップデート版でよい)の提出を昨日依頼した(提出は2-3週間後)。また、今回まとめる相互評価の内容についても参考資料として提出する予定である。

◆質疑応答◆

○相互評価と事後評価の直接的な関係は?

→特にない。相互評価の趣旨は、PJの成果を最大限に出せるように建設的な意見を出すというものである。

3. DBCLS の成果概要 (永井)

◇資料説明◇ 資料 3-1、3-2

PJ 開始当時の複数年にわたる年次計画及びその後の調査に基づいて設定された平成 22 年度の 9 月末の定量的目標については達成した。これらを踏まえ、PJ 終了時の目標として、経産省 MEDALS との共通化、統合検索のプロトタイプ構築、DB 受け入れ 45 件を設定した。ポータルサイトで紹介している各種サービスの利用状況を資料 3-2 にリストアップした。ポータルサイト全体では、ユニーク訪問者 4.8 万、ページビュー 100 万(いずれも H22.1-9 の月間ピーク値)である。サービス毎では、横断検索、日本語コンテンツ公開、新着論文レビュー、DNA DB 総覧、BodyParts、統合 TV へのアクセスが多く、その他開発系サービスにも一定のアクセスがあるので、それらが浸透しつつあることが分かる。

◆質疑応答◆

(特になし)

4. プロジェクトの平成 22 年度 9 月末時点進捗状況について

以下、資料 4 を参照しながら議事を進める。現状の公開サービスと、今後の予定を中心にお話しいただきたい。

▶ 京都大学 (五斗)

◇資料説明◇ 資料 5

本課題には大きく 2 つのテーマ(共通基盤技術開発と化合物・医薬品 DB の構築)がある。9 月末までに、類似構造検索インタフェース(IF)の開発とパッケージ公開、類似反応経路検索システム(化学反応ネットワーク予測システムの改良版)の開発(現在テストバージョン公開)、構造検索結果でのアライメント表示、お薬手帖での併用注意・禁忌の表示を達成した。アクセス状況(9 月)を見ると製薬会社からのアクセスも多い。プロバイダー系のアクセスも多いことから、一般ユーザーのアクセスもあるのではと思う。年度末までは、全体的なデータ更新を継続するとともに、化合物情報は充実してきたので、反応に関する情報の入力に注力する。未公開分の IF の整備公開を行い、HP 全体の日本語の説明加えて国内ユーザーが使いやすいようにする。

◆質疑応答◆

○本 DB が対象としている化学反応は？

→生化学反応がメイン。KEGG が代謝関係の DB なのでその延長としてデータが収集されている。有機化学はあまり入っていない。

○情報源を掲載すればスコープがはっきりすると思うので、掲載してほしい。

→KEGG では今は文献リストなどを整備している。Pathway の知識は教科書的な情報なので、レビューや教科書を参考文献としてリストしている。

○人が読んでこの DB に入力しているのか？

→基本的に反応の経路の情報を手作業で入力し、反応の始点終点の情報を重ね合わせるなどの計算に使えるようにしている。計算のチェックも人の目で行う。

○文献が集積すればするほど矛盾が出てくることもあると思うが、その解決方法は？

→反応自体がよくわからないものは載せていない。矛盾がある場合はどちらの文献が正しいと判断できればそちらを載せるが、判断がつかなければ Pathway 情報としては載せずに、個別の反応として登録しておく。

○そのような情報にもちゃんと Reference がついていれば後から振り返ることもできるので有益である。

○JAPIC のデータについては利用制限があったと思うが、何かそれについての進展はあるか？KEGG からの再配布は不可で閲覧のみ、という条件のままか？

→システムとして全体のダウンロード(DL)は不可であるが、個別データの閲覧と PDF の DL は可能である。

○反応の分類は無いのか？オントロジーとか、反応どうしの関係情報は？

→例えば、酵素については EC 番号による構造化表現はあるので、参照できる。

○DB の網羅性の観点から、JAPIC のデータが入っていれば医薬品は全部カバーできていると言えるか？

→日本における処方薬と市販薬はすべて入っている。ただし外国独自のものは無い。

○検索について、糖鎖の課題（産総研）の中でも類似検索のしくみを開発している。全体としてアルゴリズムがまだ確立していないのか？あるいは対象が違うから実装レベルで別に開発しなくてはいけぬのか？

→糖鎖は単糖を基本としており、こちらは原子をベースにしているので、アルゴリズム的にも違ってくるし、こちらのほうが計算上の負荷が大きいものである。

○化合物については原子ベースでやらないといけぬのか？

→そうである。糖鎖のようなアプローチと組み合わせて、コアの部分を原子ベースで行い、部分構造をモジュール的に計算することもできるかもしれないが、そういったことをやっている人はあまりいない。

○糖鎖グループでやっているのは補完的であって、重なりのあるものではないということか？

→そう考えている。KEGG でも糖鎖を扱う部分があったが、本 PJ 内に糖鎖のグループがあるので、現在、新規開発はしていない。

○有機化学の分野では糖鎖は扱われていないのか？構造類似検索の手法は無いのか？

→有機化学では扱わないと思う。有機化学では化合物の反応経路などは扱っているが、ほとんど有料のシステムである。類似検索はいろいろ手法があり、PubChem など公共のものも一部ある。

○これらの DB の想定ユーザーは研究者か？

→化合物 DB については研究者向け。医薬品 DB も当初は研究者向けであったが、添付文書は一般ユーザーも使える。また、処方の際の禁忌を調べる簡単な仕組みとして薬剤師が使っていて、あとで実際の添付文書で確認する必要があるものの、かなり役立っているようである。

○その意味でお薬手帖の機能が出てきたのか？

→もともと一般向けに作ったが、薬剤師にも使えるようだ。

○この先の展開としては何をイメージしているのか？

→副作用（相互作用）の情報が出てくるので、それらの整理をしたい。化合物については、反応を中心に整備。漢方や生薬については構造レベルの情報がないが、それらとリンクできるように DB を拡張したい。

➤ かずさ DNA 研究所（中村、岡本[デモ]

◇資料説明◇ 資料 6

DB 構築時に付与した情報の陳腐化を防ぐためにウェブ上で多くのユーザーが共同でアノテーションをできるしくみの提供が目的。モデル DB 上での集積情報数(18 万件)、登録利用者数(100 名)とも目標をクリアし、今年度からかずさの課題となった植物ゲノム情報の統合についても対象植物・微生物数(各 8 種・16 種)をクリアした。その他、講習会や学会展示等で広報活動を行った。PJ 終了までにはさらにそれぞれ数を上積みできる予定。デモとして、かずさアノテーションでの実際のアノテーション作業、それらによって集積された情報の既存 DB(RizoBase, CyanoBase)への統合の様子、かずさ Mart によるデータの利用法について示す。

◆質疑応答◆

○Gene Index の更新頻度は？最近は全文の公開も早いので、結局文献検索しないといけぬことになるのでは？ただし、検索では個人レベルで終わってしまうので、共有の仕組みとして依然として有用性はあるが。

→頻りに論文が出てきて高頻度の更新が必要な分野もあると思うが、シアノバクテリアについては今のペースでやられていれば、常に最新レベルの情報が入手できると思う。情報共有のためにもこの仕組みが必要。

○ビューワーの genoDive について、このような表現方法はこの分野の研究者の思考にあったものか？

→シンテニーに注目するケースでは有用な表現で、研究者の頭の中にあるものに近いと思う。

○登録利用者について、登録すればアノテーションする権利を持つということか？

→そうである。

○9月末や年度末の登録利用者数の目標からすると、ほぼこの分野の研究者の大半をカバーしているか？

→参加者を募るには働きかけが必要で、まだ増える可能性もあると思う。当初は、サービス公開すれば登録があるかと思っていたがほとんどなかった。今回は、アノテーションジャンボリーに利用したいという申し出があって、登録してもらうことで人数が増えた。論文作成のプラットフォーム(PF)としても使われているので、利用方法により、利用者も増える可能性がある。

➤ 理化学研究所（豊田）

◇資料説明◇ 資料 7-1、7-2

植物オミックス統合 DB とタンパク質実験・構造統合 DB が課題である。植物 DB については、理研の DB に加え世界的な標準 DB の情報も統合されており、論文公開に合わせて、公開される理研データが増えていく予定。各 DB のダウンロード以外にユーザー向けの機能として検索機能も付け加えた。また、表現形情報については、Plant Ontology や PATO を用いて統一しており、文献キュレーションによる Plant Phenome DB も整備した。タンパク・構造 DB についても理研+世界標準の DB が統合され、理研で実施されたタンパク 3000PJ 関連データや、微生物タンパク質のデータに加えて、それぞれの結晶化条件等実験関連情報も含まれている。実サイトを用いたデモを実施。

◆質疑応答◆

○資料 7-1 の PJ 終了時の目標について、横山データ(タンパク 3000)の公開については「可能なものについて」となっているが、可能でないものもあるのか？

→解析施設において、外部からの依頼に基づいて受益者負担で実施した内容もあるので、それらは公開不可であるが、タンパク 3000 PJ で実施された実験結果は出せるはずだし、担当者には「データの紛失が怖いので出したい」という気持ちがある。データ量が膨大で、順次送られてきているので、全体がまだ見えない。

○タンパク 3000 の全部の情報が見える統合 PF があつたはずでは？

○ターゲットタンパク PJ で中村(春木)先生、菅原先生が中心でまとめているはず。中村先生は DBCLS の客員教授なので、この課題とも連携して進めることになっている。

→本課題でやっているのは理研内の個別の研究者で持っているものを無くならないようにするもの。ターゲットタンパク PJ では理研以外の部分も扱わないといけないので、分業してやるという体制。

○実験条件のオントロジーはターゲットタンパク PJ のほうで設計することになっていると思ったが？

→同一の研究者が関与しているので、一致しているはず。理研では、言葉の一致、標準化を考慮している。

○植物のオントロジーについては理研で独自に作っているのか？

→PO と PATO はすでに確立されつつあるものを利用している。植物とマウスについては国際的関係の中で標準的なものを採用する方針。上位オントロジーについて、溝口先生が押されているのが YAMATO だが、国際的には別のものが推奨されており、将来的に移すこともあり得る。

○PATO は病気を表すものか？

→Quality(形質・特徴)を表している。PO(解剖学的用語)との組み合わせで表現形を示す。

○解剖学的用語は各生物に固有だが、形質については生物ごとに固有ではなく一般的に使えるものか？

→そうである。

➤ 九州工業大学（皿井）

◇資料説明◇ 資料 8-1、8-2

熱力学データの構造データの統合に必要な技術開発が課題。熱力学 DB には 25000 件、蛋白質・核酸相互作用 DB には約 10000 件、蛋白質・リガンド相互作用 DB には 24000 件の定量的データが格納され、すでに 200 件を超える引用がある。データの収集については関連論文の収集とデータの抽出に一番手間がかかるので、TogoDoc のカスタマイズにより、関連論文の収集とスコア付けを試行中。今後も、データの収集を継続し、TogoDoc の応用についてさらに進める。

◆質疑応答◆

○TogoDoc は要旨だけで熱力学的データの有無を検出するのか？ 式などが出てこなくてもわかるか？

→これまでの経験から要旨にも熱力学的な専門用語が出てくるので、ある程度わかる。取りこぼさないようにリコールを上げている。

○TogoDoc を利用する以前はどのように論文を収集したのか？

→PubMed の検索や主たるジャーナルを閲覧して収集していた。TogoDoc では論文ごとにスコア付けするが、載っていそうな論文はやはり上位に来るので、上から見ると主なものが取れてくる。

○テキストマイニングの仕組みは？

→サポート・ベクター・マシーンである。学習するタイプのもの。

○このような DB は商用でも似たものは無いのか？

→ない。

○引用数 200 というのは立派だが、どのように調べたのか？

→ときどき検索している。Ref のところに引用されているものを調べている。検索漏れもあるかもしれないし、URL で引用されているものも他にある可能性がある。

○データも共有できるように公開されており、有用である。

→全体を DL してマッシュアップすることも可能である。

5. その他

事務局より、シンポジウムご協力へのお礼、シンポジウム発表ポスターデータ送付(CC ライセンス付与)のお願い、ユーザー評価開始の案内と対応へのお礼があった。

○評価のサイトや結果については英語化も検討してほしい。機関の海外アドバイザーなどに見せることもできるようにしてほしい。

→現在は評価者が日本人のみなので今回は日本語で実施する。サービス全体の英語化(英語化する意味のあるものについて)とも併せて、今後海外の方にも評価していただくかどうかも含め、今回の評価後に検討したい。PJ の当初の目的が日本の研究者の DB 環境向上であったのと、国プロで実施していることから、納税者に向けた国内サービス中心で実施しているが、次のステップとしては海外のことも当然視野に入れて考えるべきだとは思う。

(12 : 25 終了)