

【日 時】 平成21年1月29日（木）13:30～15:35

【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】 豊田哲郎(作業部会委員)、神沼英里(理研)、国島直樹(理研)、浅田征彦(理研)、皿井明倫(九工大)、Shaji Kumar(九工大)、田畑哲之(作業部会委員)、中村保一(遺伝研・かずさDNA研)、藤澤貴智(かずさDNA研)、岡本忍(かずさDNA研)、中尾光輝(かずさDNA研)、菅原秀明(作業部会委員)、山中芳朗(文科省)、高木利久(作業部会主査)
堀田凱樹 (ROIS)、大久保公策(遺伝研・DBCLS)、永井啓一、西川哲夫、川本祥子、箕輪真理(以上、DBCLS)

(敬称略・順不同)

【議 事】

平成21年度業務計画・平成20年度プロジェクト進捗状況について

高木主査から作業部会分科会を開いた経緯について説明があった。「以前からPJ内には運営委員会と作業部会があり、作業部会は運営委員会にあげる提案たたき台を作ることを目的としてきた。昨年5月の中間評価において、関係機関の連携を深めるようにという提言があり、作業部会には実際に作業を進めている方に入っていただき、緊密な議論をしていただくようメンバー等を見直した。作業部会関係者を全員集めると人数が多くなりすぎて、各機関の発表時間も少なくなることが予想されたので、関係しているテーマを集めて分科会とすることにした。」本日は初めて顔を合わせるメンバーもいたので参加者が一言ずつ挨拶した。

引続き本プロジェクトの今後の方針等について、堀田ROIS機構長より文科省報告書に関する報告があった。統合DBは全省庁にわたるものであるべきだと言われているが、まず文科省での統合を進めそれから全体へ、ということで本PJが現在推進されている。DBは安定的な運用をしなければならないが、最初から安定的な新たな資金を調達することは難しいので、文科省委託費で立ち上げた。2年後(PJ後)の体制やPJの成果を今後も生かすための検討がすすめられたが、DBCLSが所属するROISの運営費交付金の範囲で手当てするのは難しいので、JSTに移しBIRDの予算の一部として安定的な運用ができるようにするというのが文科省からの提案。ROISとしては統合データベースセンターを(ROISの)第2期も継続したいと考えているが、仕事の内容とJSTとの関係については今後検討していく。一方でいろいろな施策を行うにも成果が必要であり、中間評価に関してはおおむね好評だが、指摘事項もある。今後2年間も引き続きPJの成果がしっかり出るようにご協力いただきたい。

➤ 予算配分額について

◇資料説明◇

21年度予算案について、資料1-2が中間評価会から評価の一部として提出された予算案であり、いくつかの機関については増減がある。昨年と今年の総額は同じになる予定。1-2を踏まえて、資料1-1の配分案ができている。総額11億のうち、2.5億分についてはJSTからの予算執行となる。これはプロジェクトとJSTとの将来的な統合への施策のためであり、そのうち0.3億については統合DBPJ内の遺伝研への課題分をBIRDからの課題と合わせJSTで扱うことになったためである。資料1-3が予算に関するスケジュールで、運営委員会での正式承認があとからになったが、すでに各機関から予算に関する資料を提出いただいているので、4つの作業部会分科会での議論を踏まえて、予算を確定していくことになる。また、資料1-4は中間評価に対する各機関の対応をまとめたもので、資料1-5は今年度の目標と進捗である。これらの資料を踏まえて、今年度の成果と来年度の計画について議論いただきたい。

◆質疑応答◆

とくになし

➤ 理化学研究所

◇資料説明◇追加資料を配布

1. 植物オミックス (情報注釈)

フォルダー型に情報をまとめて、ダウンロード (DL) できる形のアーカイブ化完了。2月には DL できるように公開したい。約束した DB 以外にも提供可能になる予定 (下記参照)。

2. 高等動植物由来蛋白質構造アノテーション

横山先生のデータが対象であるが、進捗が遅れているので加速する。

3. 微生物由来蛋白質構造アノテーション

かなり多くのデータを扱っている。データの搭載はほぼ完了している。画面表示 (詳細画面) と DL については一部のみ実施、引き続き完成を目指す。データの回収 (A)、整理 (B)、アノテーション (C)、公開 (D) の段階ごとに各データのステータスを説明。計測データだけでなく実験情報なども搭載。21 年度は各 DB について C の段階を進めたい。

4. アノテーションシステムの開発運用と変換データの中核機関への提供 (理研の DB 統合化のためのモデルケース構築)

理研から公開されている DB は 93 件 (資料では 92 件) あるが、その一部についてメタデータを整備した。メタデータの記述に関して、登録の仕組みを工夫した。当初から約束していた植物、蛋白以外に、FANTOM4 のデータを受取済みで、論文化が終了次第、公開可能。公開準備はほぼ完了している。

DB の公開を希望しても、実際にはアノテーション付与など公開に向けての作業に多大な時間がかかる。それを短縮するためのアノテーションシステムを構築中。これは中間評価での指摘 (他分野との連携) に対応するもの。また、もう一つの指摘 (理研内データの積極的公開) については、体制強化のための予算化検討中。アドバイザリーカウンシルからのトップダウンの働きかけもあり、理研全体の方針として進めていく。

◆質疑応答◆

○20 年度と 21 年度の目標について。シロイヌナズナの 20 年度の公開というのは DL までか？

→Yes。あまり複雑な視覚化などをしないで、公開することに注力。

○21 年度の数値目標は？ 次回の作業部会などでは公表可能か？

→同じペースを想定しているが、担当者変更との兼ね合いで数値が出せていない。消極的な意味ではない。各 Gr との交渉も必要になってくる。分野に限らず、全体として推進していく。公開はしたが、アノテーションが不足しているものもあるので、その整備も必要。国際的な合意のためのミーティングを日本で開いて、国際的なフォーマットを提唱するなどの活動も考えられる。

○他の項目についても数値がないのは、同じ考えに基づくものか？

→前回は数量的に読みやすかったので、達成目標として明確化した。このペースで進めていく予定だが、変動要因も多いので数値にしにくい。

○横山先生のところも達成可能か？

→理研の中からの声もあり、達成をお願いしている。

○21 年度の分については次回に向け数値的にも明確化してほしい。

○FANTOM4 で公開されるのは、ゲノムネットワークのデータとは別？ 同じものの見せ方が変わる？

→正確には不明。部分的な重複があるかもしれない。国プロ以外のものも入れてもらっている。理研全体としてはトップも含め協力的なので、どれに特化した形ではなく、公開を進めていく予定。

○植物なので、かずさとの連携は？

→今後連携が強くなっていくと思う。

- 植物については理研との統合を視野に準備。遺伝子 ID 等でつながるものはつなげようという計画。
- オントロジーやかずさのブックマークについて、連携を検討している。
- 微生物に関して、菅原先生から何かコメントないか？
- 統合 DBPJ 中の課題として実施。オンデマンドでの微生物由来の配列処理、辞書作りの 2 点。理研との整合性をとりながら進めたい。
- 辞書作りについてミーティングの企画ありとのことで、それが通れば進むだろう。
- ターゲットタンパクと、横山先生の仕事の関係は？
- ターゲットタンパク研究プログラムの中に情報プラットフォームという課題がある。タンパク 3000 のデータは PDB から取得したが、詳細な情報や実験プロトコールについては横山先生のところで整備予定でそれを待っている。まとまったデータはほとんど横山先生のところで、大学分は散逸してしまっているらしい。
- 統合 DBPJ の要望としては、横山先生のところで整備していただく必要があると思う。
- まず見えるものを作り、そこへの協力を促すのがいいと考えている。
- タンパク 3000 では自発的な登録があまり進んでいない。評価される対象となれば DB に登録するが、登録を進めるための仕組み作りが必要。
- 専門的な知識を持った人が必要。専門家を手当てするための予算が通ればクリアできるのでは？データの公開は契約時の条件となっていたのか？
- 条件ではあるが論文化との兼ね合いでなかなか出てこない。また、34 課題それぞれの特異性が高くアノテーターが個別に必要なので対応できるか疑問。
- しかし実際にはアノテーターを手当てするのか？
- 報告書からの情報抽出をシステム側担当で実施することで対応を予定している。
- データ公開についてももう少し強制力を持たせることは？
- DB に登録されない結果は成果と認めないといった通達により、成果が出てきたということもあった。
- (タンパク 3000 参加の経験から) あまり入力に積極的ではない人もいた。実験を行うと自動的に情報が入るといった仕組みを後半は作っていたところもある。
- LIMS の仕組みでやればいいが、個別課題にも適用するのは難しい。
- 実験情報はだれがユーザーなのか？
- よく似たタンパクがあれば、精製などの工程や反応条件の検討に実験条件が参考になる。
- 競争相手を利することになるのではないか？
- 巨額の (公共の) お金を使っているのだから、社会還元として当然だと思う。しかし、自ら DB を作った人には信念があって続けていくが、いろいろなところから集めただけになると続かない。プロジェクトからメンテナンスへのサポートはお願いしたい。
- 既存 DB には記述がばらばらなものが多い。だから、(PDBj の) 中村春木さんは記述のオントロジーを決めて始めた。中には実験途中で失敗したという情報もある。
- 遠くてぼんやりした Outcome ではなく、Output を明確に定義して、それを基準に評価する大事。
- ターゲットタンパクはタンパク 3000 の反省に基づき、難しいタンパクをやるという宣言がなされている。
- 他の人も使える基盤データを構築するのが目的なのだから、「難しいタンパク」という選択はおかしい。
- 化合物関係も PJ に入っていて、暗黙には Drug Discovery が目的になっているらしい。
- ということは個別研究をやっていることになる。
- タンパク 3000 の反省をもとにしているという意味は？
- 出口につながるものをやらないと、という反省。
- それでは Outcome をもとめていることではないか？
- シロイヌナズナとしてはどのくらいのデータが公開されるのか？

→発表時に DB としてまとめたものはそのまま公開。数も増えていくと思うが、次世代シーケンサー（実験条件が異なる）のデータも出てくるので、種類も追加されていく。データの種類やジャーナルごとのデータセットといった分類をすることが誰でも可能になる。

○データの種類（計測器の種類に依存）はどのくらいあるのか？そんなにないのでは？

→メタデータはかなりヘテロになる可能性がある。

○DB の形にすると蓄積することが難しい。なら、ばらしてしまえば簡単に再利用可能になるのでは？わからない前処理をいろいろするより、処理していないデータを置いておくほうがいい。

→多面的に見せることが必要になるのではないかな。配列についてはシグナルを出せばいいのか、ゲノムにマップしないとわからないのか？多目的になっているので出し方が複雑になる。メタデータをつける時に、ゲノムにマップすることも親切では。

○いろいろなやり方で重複作業が発生していることが問題。ばらせないので、セマンテックでつないでいる。

→ データを公開する側では検索が付いていないから **Reject** されることもある。利用可能な形態にするためには、あるレベルまで何らかの処理をしてあげることが必要。

○使いやすい形で提供された DB の賞味期限は短い。最終的には統合に使えるデータがあればいい。

→統合に求めるものは人によってそれぞれなのでは？

○だからその中間をとって、整理したデータと生データと両方得られるようにしているということか？

→様々な段階のデータが取得できるようになっているつもり。

○生データと実験条件を付けておけば誰でも使えるはず。

○使えないひともいる、ということだ。

→データそのものを使いたければ集められるようになっている。

○RDF(Resource description framework)の書き方などは決まったものがあるのか？

○RDF のオントロジーをどうするのか、ということか？

→フレキシブルなツールがある。オントロジーについては、実際に使われているもの（よそで使われているもの、国際的なもの、あるいは大家の提唱しているもの）を取り込んでいる。

➤ 九州工業大学

◇資料説明◇

すでに構築されている熱力学 DB について、本 PJ では構造情報との統合を目指す。PDBj との統合化のためにクロスレファレンス付与、オントロジー整備、データ交換のフォーマット整備を一部実施。文献からの情報抽出としては、情報の載っている文献を選び、具体的な情報を抽出する作業がネック。これを自動化するためにセンターと連携して開発中。ウェブサイトを作成し、成果の進捗を載せるとともに、データの DL サイトにリンクを張った。21 年度は、今年度開発内容をさらに進めるとともに、蛋白質相互作用 DB のプロトを作成、XML フォーマット公開、オントロジー整備、データ抽出の自動化を予定。

◆質疑応答◆

○データの公開状況は？

→すべて検索可能。アグリーメントを結んでから全体のデータを DL 可能。アカデミックは完全フリー。

○コマーシャルやアカデミックへの利用許諾については要調整で、DBCLS でも検討中。

→こちらも悩んでいる。

○やられた内容からみると著作物に近い DB ではないかと思う。

○キュレーションそのものは別の予算で実施。統合 DBPJ ではクロスレファレンスの付与、オントロジーの整備の部分。権利関係が複雑になっている。

→キュレーションは科研費中心。著作権はどこにあるのかが不明確。

- 文献から取ってくるときにどれだけ独自性があるかということも著作権としては問題になる。
- 研究者が目的に沿って選んでいて、追加情報を加えているので、著作権はあると思うが、要検討。
- オントロジーは生物などに比べて熱力学のほうがすっきりするのではないか？
- 用語の使い方や定義が人によって異なる。まとまる動きがない。
- 国際的な連携の動きは？
- 熱力学としてはないが、作ろうとしている米グループがあるので、協力しようと思う。
- 入れるべきデータの総数は？データの増加率は？
- 増加数は年間 1500 件くらい。総数としては、蛋白質変異体のデータは 23000 件、蛋白質の相互作用データは 8000 件、化合物との相互作用は 20000 件くらい。
- PJ を続けていけば、入れるべきデータは網羅できる予定か？漏れがないかどうかについての検討は？
- かなり専門的な情報なので、ジャーナルも限定的（10–15 位）であり、網羅できていると思う。
- この DB が国際標準になって、みんなが入れるようになればいいと思う。
- ジャーナルが限定的なら、すでにデータをカバーした段階で、論文投稿の条件にしてもらったら。
- そうなるといい。

▶ まとめ（総合討論） 他

<理研>

- 微生物由来蛋白質構造 DB を検索するためにはどういうルートが使えるのか？
- 個別のキーワードを Query として検索できる。一括 DL もできるが、Web も使いやすいように作った。
- 特殊な検索には播磨の Web サイトも使える。
- 統合 DB とメタデータを介して検索できるか？
- メタデータはどのようなデータがあるかということが分かるもの。Web サービス化する予定はあるので DB 指定の上、検索することができるようになる可能性あり。DB 開発の段階によって、検索できるものとできないものがあり、人によって見える結果が違う。認証やログインでアクセス権を区別する必要がある一部ある。
- そこをシンプルにしてもらって。
- 作るほうが面倒なだけで、使う人にはかわらない
- 全部自由にするしかない。人のデータ見て論文化はできないのだから、全部早く見せてしまえばいい。アクセス権コントロールとか、利用契約を無限に作るほうがよっぽど費用がかかる。
- 今すぐにも ROIS にイニシアティブを取ってもらってポリシーを整理してほしい。
- シロイヌナズナのデータ統合はどんなイメージか？一つのゲノムブラウザーに全部つながるのか？
- マップするものと結び付けるデータとはやや異なる。ゲノム上にマップするものは可視化が必要。データの関連付け情報があれば、ゲノム上に統合できる。データの型から判別して自動化できるものを考えている。
- シロイヌナズナは、理研でも多種のオーム解析が実施されており、データサイズがあまり大きくなくて、きちんとアノテーションが付いているので、全体の情報をまとめて提示するためのいいモデルになる。
- 理研だけでは範囲が狭いので、比較ゲノムのデータをつけて出すこともできる。
- 発現情報については一般に出ている物以外のものが理研のデータには入っているのでは？
- 論文を出したものについて発現データはすべて公的サイトに寄託済。ただし、表現型の寄託 DB はない。
- マイクロアレイは GEO にある。ものはバイオリソースにあるが、その表現型についてもあるのでは？
- 一応持っているが、いくつか問題あり。1. 書き方が古い、2. 情報についても再配布可能であるという了解を取っていない、さらに 3. 新しい表現型の記述を追加したときに元の寄託者が責任を持ってない。クリエイティブ・コモンズのマーク(著者名記載)をつけるとむしろ問題。
- 理研のデータはどの程度 TAIR(The Arabidopsis Information Resource)とリンクされているか？

→TAIRのIDから対応ページへのリンクが張ってある。TAIRについては利用条件が不明。TAIRのデータを再現できる情報すべてが公開されていない。こちらとしてはアノテーションを充実するためにデータを出しているのだが、こちらのIDを用いデータを参照してもらうようにするのがいい。TAIRに限らず、組織的にDB化をちゃんとやっていることを示して、国際会議でも発言できるようにしたほうがいい。

<九工大>

○皿井先生のDBの使い道として最もわかりやすい形は？

→蛋白質の安定性のDBについては蛋白質工学。リガンドとの相互作用については創薬もターゲット。

○企業からのオファーは実際にあるのか？

→安定性のデータを全部欲しいという申し出が1件あり。著作権等の権利を整理する必要がある。

○文科省から国家PJの成果の扱いについてのルール・方針については出してもらえないか？ 国際連携の際にはそこが肝心。ローカルにカスタマイズしたルールでは個人的に国益を売るようなことがまかり通っている。PJの中で統一したものがほしい。

○統合DBの中で作ったものについての検討か？

→今研究運営委員会でも考えようとしているのだが、いろいろなケースが出てきて悩ましい。

○原則は企業にもどんどん出す。でも特定の狭い企業を利するやり方はいけませんが、お金を取るとはいいことだという考え方もあり、整理されていない。

○運営委員会でも議論を開始しているところ。皆さんのお知恵を拝借したい。

○JSPSの公開への助成金（研究成果公開促進費）の成果はどこに？DB作成についての補助金だが、公開についてのルールがないので。

→以前の大久保調査の際にリストするときソースとして使った。

○10年くらい前に成果の取扱いについて話をしてしたが、「難しい」という結論で終わっている。

・2009年6月12日(金)に成果発表を兼ねたプロジェクトのシンポジウムを予定しているので、ご協力願いたい。ポスター発表など、PJ関係者に閉じずにいろいろなDBの活動を紹介したい。

(15:35 終了)