

**「統合データベースプロジェクト」
第4回研究運営委員会 議事要旨**

【日 時】 平成20年4月3日（木）15：30～18：30

【場 所】 情報・システム研究機構事務局会議室

【出席者】

（運営委員）秋山委員、浅井委員、勝木委員、金岡委員、久原委員、田中委員、田畑委員、中村桂子委員、中村春木委員、長村委員、大久保委員、小原委員、五條堀委員、菅原委員、高木委員、堀田副委員長

（分担機関、補完課題出席者）

京都大学	：五斗准教授
東京医科歯科大学	：中谷准教授、高井准教授
（株）日立製作所	：小池主任研究員
（独）理化学研究所	：豊田部門長
（独）産業技術総合研究所	：成松センター長、新聞研究員、鹿内研究員

【陪 席】

内閣府	：鬼頭調査員
厚生労働省	：石原調整官
農林水産省	：宇井課長補佐、枝川専門官
経済産業省	：諸橋係長
文部科学省	：生田課長補佐、澄川係長、田中調査員
（独）科学技術振興機構	：菊池次長、河村課長代理、藤田係員、酒井主任調査員
ライフサイエンス統合データベースセンター	：永井特任教授、西川特任教授、川本特任准教授

【事務局】 高野事務局長、加藤企画課長、石田総務課長、植田財務課長、植田事務室長

【挨 拶】

堀田研究運営委員会副委員長から、松原委員長欠席のため議事進行は堀田副委員長が行う旨の発言と挨拶があり、開会が宣言された。

【議 事】

1. 研究運営委員会（第3回）議事要旨（案）について

堀田副委員長から、1月8日に開催された第3回研究運営委員会の議事要旨（資料1）に関して、意見があれば会議の終了までに事務局まで連絡して欲しい旨の発言があった。特に意見はなく、議事要旨は承認された。

2. 平成19年度成果のとりまとめ（中核機関、分担3機関）

堀田副委員長から、「統合データベースプロジェクトの平成19年度の活動と成果について、中核機関、各分担機関から説明をお願いしたい」旨の発言があり、中核機関、京都大学、東京医科歯科大学、東京大学の順に説明を求めた。

（1）中核機関

高木委員より、資料2-1を用いて中核機関の平成19年度の活動と成果について説明があった。質疑応答は以下のとおりである。

●DBCLSのホームページとLSDBとの違いは何か？

→センターのホームページ（DBCLSのホームページ）とプロジェクトのホームページ（LSDB）を分けて運用している。

（2）分担機関

1) 京都大学

京都大学五斗准教授より、資料2-2を用いて平成19年度の活動と成果について説明があった。質疑応答は以下のとおりである。

●化合物の立体構造の座標データもデータベースに入っているのか？

→入っていない。リンクのみである。

●バイオインフォマティクスの研究では、構造情報を用いる機会が増えている。PubChemには構造異性体の情報が多く含まれており、どの構造を用いればいいのか専門家でない研究者にはよくわからない状況がある。是非構造自体も扱っていただけたらと思う。

→検討したい。

●サービスに用いている材料は、どのような条件で使えるのか？

→京大で作成したものは、再配布可能である（KEGGは別である）。

2) 東京医科歯科大学

東京医科歯科大学田中教授より、資料2-3を用いて平成19年度の活動と成果について説明があった。質疑応答は以下のとおりである。

●具体的には、どのような人がどのような時に使うデータベースなのか？

→利用者は研究者と臨床家を考えている。研究者の場合は、分子データと臨床データの関係の検索を想定している。例えば、疾患の発現パターン類似性をキーにして、患者の予後情報を抽出して比較するというような検索に用いる。また、病情報に分子情報を加えることで、予測情報を増加させることができる。

●そのような検索の仕組みはすでに構築してあるのか？

→まだ構築していない。現在は、臨床関係の検索のみである。分子情報はこれから入れていく予定である。

3) 東京大学

(株)日立製作所中央研究所小池主任研究員より、資料2-4を用いて平成19年度の活動と成果について説明があった。質疑応答は以下のとおりである。

●これは、データ統合というよりも、ある研究グループのデータを公開するということか？もっと広くデータを集めるべきではないか？

→もっと広くデータを集める予定である。

3. 補完課題（理研、産総研糖鎖）のDB整備、統合化の取り組みについて

堀田副委員長から、統合データベースプロジェクト補完課題のデータベース整備と統合化の取り組みについて、理研、産総研糖鎖医工学研究センターの順に説明を求めた。

1) 理研

理研豊田部門長より、資料3-1を用いてデータベース整備と統合化の取り組みについて説明があった。質疑応答は以下のとおりである。

●新しい部門で公開するデータを3倍に増加させるという話があったが、予算はどれくらいか？

→今回の予算は、部門全体の予算の約1/4である。

●補完課題としての意義について、技術が中心なのか、理研のリソースの公開が中心なのか？

→中核の指示に従い、仕組み自体は理研の予算で実施し、補完課題ではアノテーションに特化している。

●理研にある大小様々なデータは全て公開するのか。

→一番大きいのはタンパク3000のデータが30テラバイトほどある。それらを整理して公開すると共に他のデータも公開していく。公開を了解してもらったデータベースからこの仕組みに入れていく。

●SNPのデータに関しては理研としてはどうするのか？

→一般的にデータを持っていかれることに対する拒絶感が大きいので、徐々にやっていく。植物や蛋白構造のデータでは共有してもよいということだったので、今回補完課題として申請した。

●理研では、リーダーが理研以外に異動した場合、データベースの権利はどうなるのか。

→理研の就業規定では被雇用者の創作物に係る著作権は原則として理研に帰属する。異動に伴い著作

権を譲渡する規定はない。

●公開の交渉の相手は、理研でいいのか？

→著作権は理研にあるが、ライセンス条件については研究代表者の意見で決めている。研究者に公開化を強いる規定は理研にない。

●データの商用利用は可能か？

→上記ライセンス条件で認めれば可能。ロイヤリティは理研の知財センターを介して受領する。

●DBCLS では、ライセンスはどうなっているのか？

→DBCLS では、他者の権利が入っている場合以外は、DBCLS が作成したものについてはデータベースもツールも原則的にクレジット表示のみでよく、再配布も商用利用も可能である。

●DBCLS では、著作権は放棄されているのか？

→著作権は日本の法律では放棄されない。

●サービスに用いている材料は、どのような条件で使えるのか？

→ツールはオープンソースで出している。

●DB数を3倍に増やすと言われていたが、最終的な目標として具体的な数字で言って欲しい。

→1～2/月のペースでやっていきたい。

2) 産総研糖鎖医工学研究センター

糖鎖医工学研究センター成松センター長より、資料3-2を用いてデータベース整備と統合化の取り組みについて説明があった。質疑応答は以下のとおりである。

●中核機関との連携が大事。中核との間で緊密に話をされているか？

→頻繁に打合わせを行っている。糖鎖センターでデータがまとまってくれば、具体的な連携を図る予定。

●コンソに入る条件はどうなっているのか。また、コンソ内とセンター内ではデータの利用に関して違いがあるのか？

→この問題は、省庁間でも同意がとれていない事項だと思う。コンソーシアムをどう考えるか？よりフリーに近くするのか、会費制にするのか、等について、文科省と経産省で考え方が異なると感じている。

●会費制か。

→会費制ではない。論文も書いていないデータについては、コンソはクローズドにせざるを得ない。すぐに公開すると、実施者のメリットがなくなってしまう。

●知財をとって論文を書いた後のデータについてはどうか。

→論文発表後であれば公開可能である。

●公開までにはどれぐらいかかるか。

→データに依存する。このコンソに関わる問題はきわめて本質的な問題である。考え方に省庁間の違いがある。

4. 契約・著作権への対応について

堀田副委員長から、「中核機関よりデータベースを受け入れるにあたっての契約・著作権に関する報告があるので、説明願いたい」旨の発言があり、中核機関の高木委員に説明を求めた。

1) 契約・著作権への対応

高木委員より、資料4-1を用いてデータベースを受け入れるにあたっての契約・著作権に関する報告があった。質疑応答は以下のとおりである。

●公募要領案にデータの提供時期が明記されていない。論文発表後に出すということか。論文発表に至らないものもあるが。

→プロジェクトの採択時から緊密な連携をとりたいと考えている。

2) データベースを受け入れに関するアンケート調査について

高木委員より、資料4-2を用いてデータベース受け入れに関するアンケート調査についての報告があった。質疑応答は特になく、報告内容が了承された。

5. 平成20年度予算配分について

堀田副委員長から、「来年度の統合DBプロジェクトの予算配分について説明願いたい」旨の発言があり、高木委員に説明を求めた。高木委員より、資料5を用いて平成20年度の予算配分結果について説明があった。質疑応答はとくになく、平成20年度の予算配分について了承された。

6. 中間評価について

堀田副委員長から、「統合DBプロジェクトの中間評価のスケジュールについて説明願いたい」旨の発言があり、文部科学省に説明を求めた。文部科学省生田課長補佐より、以下のように中間評価のスケジュールについて説明があった。「中間評価の調査票をまもなく送付し、審査委員によるヒアリングを実施する。6～7月までに終了する」。質疑応答はとくになく、中間評価のスケジュールについて了承された。

7. 3年後のプロジェクトのあり方について

堀田副委員長から、「統合DBプロジェクトは、3年後の平成22年度末で終了するが、その後のあり方について意見を伺いたい。資料については高木委員に説明願いたい。」旨の発言があった。これを受け、高木委員より、資料6を用いて統合データベースプロジェクト終了後の体制等について説明があった。意見、質疑応答は以下のとおりである。

- アーカイブ化と統合化のどちらを優先するかに関しては、両方やるべきである。集めるだけだと他でもやれる。統合化技術だけではユーザーから感謝されない。そう多くない数のデータベースを統合化して見せるというパイロットスタディが重要。全てのデータがつながる必要はなく、インターフェースの統一化程度で良いのではないか。それ以外のデータベースについては、ポータルで良いのではないか。
 - 一番重要なことは、競争的資金がつくことはないということに対して、何を提案すればいいのかということ。そのための体制はどうあるべきかを考えながらやるべき。他に組織を作るのは困難である。DBCLSをベースに競争的資金でないやり方はないか？DBCLSをいかに運営することが大事である。
 - いかに戦略的に恒常的な組織に持っていくかという政策的な問題ではないか。研究者同士で頭をひねっていても仕方がない。もっと研究者が政治に参加すべきである。
 - 2月にノルウェーの種子バンクができた。その理由は、作物の種子の有用性に対する理解がヨーロッパ文化にあること、及び種子銀行ネットワークがすでに存在し、そのコミュニティーの意思が存在したことが考えられる。分担機関や補完課題機関が統合DBセンターを支える発言をすることや、DBの重要性を理解させる努力が必要。ヨーロッパには1000年～2000年間種子を保存するという考えがあり、それと国内の意識とのギャップを埋めるには、アカデミアの努力の余地がある。
 - 文科省が率先して作ったこのプロジェクトの位置づけを再確認する必要がある。もともと、総合科学技術会議の中で始まり、大久保先生の調査に引き継がれ、それを受けて文科省がいち早く開始した経緯がある。このプロジェクトは文科省が作った単なる一つのプロジェクトではない。今後バイオ戦略会議の議論にのせていく必要があり、専門家が行政官をサポートしていく必要がある。
 - 維持管理に必要な費用と新規構築分は分けて説明した方が良いのでは。維持管理に競争的資金は合わないのでは。
 - 統合開発を競争的資金でやるのは無理がある。
 - 日本に世界に発信できるDBがあり、それを維持していく必要があるという認識が4省庁で一致しているということであれば、総合科学技術会議でも理解してもらえるのでは。そのためのルートは、内閣府が作っていただけののでは。
 - 理研の場合、アドバイザリーカウンセルのメンバーがほとんど外国人であり、そのため、欧米のセンターとの連携が常に問われる。国際的な視点の方針を立てていただけるとありがたい。
 - 最初に4省庁のデータを統合してみせましたというのをやれば、DBCLSのオリジナルな成果になるのでは。
- すでに、文科省と経産省のcDNAの受け入れを実施している。これがその例になるのでは。
- アーカイブするだけだと他でもできる。ライフサイエンスや情報技術に関する知識がないとできない例をまず明確に見せることが重要。
- 現状はそのアーカイブすらできていない。そこで、現在はデータを集めることにエネルギーを注いで

いる。

- データをオープンにしてもらうためには、統合がうまくいくことを示す必要がある。
 - 連携施策群が終了するということだが、今回フィージビリティースタディーだったので、次があるはずではないか。
- 打合わせの会合は継続する予定である。
- 統合するメリットが見えていない。病気から始まって物質までという一般受けしやすいもので、統合のメリットをアピールしたらどうか。
 - 成功事例の重要性には同感である。しかし一般受けは危険な面もある。技術者向けのものを作らないと、国際的には評価されない。
 - 利用の制約の少ないデータを増やすことが重要である。
 - それは大前提。さらにデータが集まったことによる成功事例があった方が良い。
 - 以下の3つが重要、
 - 1) 制約のない共有
 - 2) 操作感の統一
 - 3) 新しい生物学へのブレークスルー
 - 統合DBの概念が人によって異なる。生物学への貢献で十分なのか、もっと社会への貢献まで必要なのか。そうだとすると、とても大変な問題である。このプロジェクトで研究として大きな成果を生み出すようなところまで示すのは、現在の100倍以上の規模が必要。
 - すぐに何かを作ろうとすると、公開度の高いアメリカのデータを使うことになり、英米データの加工合戦に陥ることになる。これまでかなり多くの予算が測定型プロジェクトに投下されており、それは基盤データの取得ということで始まっている。そのようなデータが何千億円分公開されていない状況に目をつぶり、英米データの加工合戦をやるのはいかになものか。
 - 企業の立場からまっさきに期待したいのは、大量の資金をつぎこんだデータである。企業から見てアクセスできないところにある物が多い。NEDOのデータや厚労省のデータを組み合わせると何か期待できそうだが、両方をみることができない。まず見えるようにしてほしい。経産省と連携しているという話があったが、有用なデータは入ってきていない。壁をとる働きかけが重要であり、恒常的な機関がアーカイブ化をやってくれることを期待する。
 - 先ほどの三つの項目は我々（DBCLS）もよく理解している。現在の状況で新しい生物学への貢献を実施しようとする、限られた予算の中で共有化の作業がおろそかになる可能性がある、現在は共有化に力を注いでいる。その次に操作感の統一をやる。さらに、それを使った新しい生物学への貢献は、外部の機関や研究者にやってもらう。そういう姿勢でやっていく。

8. その他

堀田副委員長から、「中核機関より提案がある」旨の発言があり、高木委員より、1)「研究運営委員会の資料及び議事要旨のウェブサイトへの公開」、2)「研究運営委員会委員の見直し」について提案があった。1)については、議事録を公開すると発言がし難くなるのではという意見も出たが、議事録内容の確認のステップを設けることで、了承された。2)については、現在オブザーバーとして参加していただいている各省から委員として参加してもらう案に関して、「委員になると省としての正式な意見になるので、難しいのでは」という意見も出たが、各省に持って帰って検討してもらうこととなった。また、それ以外の委員の追加案があれば出していただくこととした。

堀田副委員長より、「次回の運営委員会の開催について7月頃を予定している。」旨の発言があり了承された。

以上