

「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(第3回) 議事要旨(案)

【日 時】平成19年12月6日(木) 15:30～18:00

【場 所】情報・システム研究機構 事務局会議室

【出席者(委員)】浅井委員、金岡委員、黒田委員、田端委員、松原委員、大久保委員、高木委員、永井委員

【出席者(補完課題研究代表者)】

(独)理化学研究所 : 豊田ゲノム科学総合研究センターオミックス情報統合化研究チームチームリーダー

(独)産業技術総合研究所 : 成松糖鎖医工学研究センター長

情報・システム研究機構 : 池尾国立遺伝学研究所准教授

【陪 席】

文部科学省 : 澄川ライフ総括係長、田中調査員

(独)科学技術振興機構 : 藤田係員

情報・システム研究機構 : 堀田機構長、川本特任准教授、西川特任教授

【事務局】高野事務局長、石田総務課長、笹島総務課課長補佐、植田事務室長

【議事】

1. 研究運営委員会作業部会(第2回)議事要旨(案)について

高木主査から、10月1日に開催された第2回研究運営委員会作業部会の議事要旨(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. 補完課題の進め方

(1) 補完課題について

文科省田中調査員より、補完課題に関するこれまでの経緯について説明があった。その後、この作業部会の中での補完課題の位置づけについて、永井委員より説明があった。

(2) 中核機関と各機関の連携方針

高木主査より、中核機関と各機関の連携方針について説明があった。中核から各機関への依頼事項として次の3点をお願いしたい旨の発言があった。1)それぞれの分野での日本の取りまとめ(ハブとしての機能)、2)データ公開の促進(できれば全データのダウンロード)、3)統合検索に関する中核との連携(インデックスやAPI)。

(3) 採択各機関の計画

・(独)理化学研究所

豊田理化学研究所チームリーダーより、理研のテーマについて説明があった。質疑応答、討論の内容は以下のとおりである。

統合化と公開のプロセスはどのようになっているのか。

公開しないというポリシーはない。論文のこともあるので、4年間のうちに順次出していく予定。

広く植物研究者の役に立ちそうなので、期待している。

蛋白3000のデータが公開されていない状況があるのか。

原子座標データは全てPDBjに公開している。それ以外に、実験条件や実験の生画像データ等が大量にある。これらを再解析に活用可能なようにDB化して公開する。NMRのデータは、蛋白3000でやれなかった構造に関するアノテーションを実施して公開する。

PDBjとの連携はどうするのか。

PDBj代表の中村春木先生と、オントロジーや辞書等に関して連携する方針である。

・(独)産業技術総合研究所

成松糖鎖医工学研究センター長より、実施するテーマについて説明があった。質疑応答、討論の内容は以下のとおりである。

経産省のPJの成果を文科省から公開するという点に対して、経産省として問題はないのか。

NEDO は日本企業の育成が大きなミッションであり、企業サイトの感覚ではデータはできるだけ出したい状況がある。しかし、国民の税金で実施されたプロジェクトの成果は公開すべきであるという考え方もあり、参画企業にはある程度優先権を与えるというところで折り合いがみついている。むしろ、経産省は文科省から公開してもらえることに対して歓迎している。

・情報・システム研究機構国立遺伝学研究所

池尾国立遺伝学研究所准教授より、実施するテーマについて説明があった。質疑応答、討論の内容は以下のとおりである。

ユーザーは誰か？登録者か？

色々なケースがある。バイオインフォの人がローデータから解析しなおすケースや、登録者がユーザーのケースとしては、ローデータの保管場所として、あるいは複数の組織にまたがる情報のハブとしての機能等がある。

一般ユーザーが用いるとき、DDBJのエントリーとの対応はわかるようになっているのか？

そのようにする予定であるが、具体的な仕様検討はユーザーにヒアリングしながら進めていきたい。

登録するデータは過去に遡るのか？

バンクとしては当然過去分もアーカイブすべきであるが、ストレージのサイズの問題があり、まずデータがどれくらいあるのかを個別ヒアリングによって調査する必要がある。

データ量の伸びが、ディスクが安くなっていく速さよりも早いので、戦略が必要である。

現体制ではサンガー法はなんとかできるが、次世代型には足りない。

トレースデータからデータの質がわかるので、歓迎しないユーザーがいる可能性がある。バイオインフォを活用すれば死んでいたデータを活用できる可能性がある。

NCBI や EBI をみると全部は入っていない。質の悪いデータははいっていない可能性がある。

データ登録に対する義務化が必要。今は紳士協定。

トレースデータには、タンパク 3000 の実験データと同様な意義がある。

バイオインフォの立場からは元のデータがわかるのは興味深い。しかし、個々のPJの質のコントロールになるのはおかしいのでは。シーケンシングPJの議論と分けて議論すべきである。逆にきたくないデータがほしい場合もある。目的とコストを意識する必要がある。

アーカイブするデータについてはどこかで取舍選択が必要。グレーゾーンが役に立つ場合がある。

将来は、プロジェクトを立てる際に、どういう質のどういうデータを出すかということカップルさせる必要があると考えられる。

・九州工業大学

西川ライフサイエンス統合データベースセンター特任教授より、皿井九州工業大学教授の代理で、九州工業大学が実施するテーマについて説明があった。質疑応答は特になく内容について了承された。

補完課題全体に対するコメントとして、以下のコメントがあった。

補完の採択金額は公表しているのか。コストの議論をしないと、実効的な議論ができない。公表する方向で検討願いたい。

文科省から 補完以外の予算も含めて、検討する。

3. 契約・著作権、個人情報問題への対応

永井委員より、契約・著作権、個人情報問題への対応について説明があった。質疑応答、討論の内容は以下のとおりである。

「研究委託業務推進室の感触」として、バйдール法をベースにした案件については、文科省として特に声明をださなくても、データ産生者と統合DBセンターの当事者間でデータ提供について合意すれば問題ないというスタンスである。個別に話しをしてほしい。データ提供が断られるケースが多ければ、

文科省も何らかの手をうつ（公共の利益のために協力して下さいというのはあるかもしれない）。現実的に本当に断られるのか。

依頼のフィービリティスタディーは必要だが、いちいち個別に最初から交渉していたのでは、らちがあかない。データ産生者に対して依頼の文書を送ると状況が一気にわかる。その際には、背景や論拠の説明が必要である。

今の独法化の時代に、「独法」から「独法」へそういう依頼がきても、まともに対応してくれないはずである（決済をとるのが大変である）。

文科省から委託されているわけだから、本当は、文科省がやるべきことである。

それも含めて委託していること。委託先が力不足で壁に当たっているということであれば、文科省として動く。

力不足ではなくて、筋違いではないか。

> 経緯的な説明をつけ納得を求めるのであれば、もともと内閣府調査もあり、第三期基本計画の中で世界最高水準のDBを目指すということがあるので、理解が得られるのではないかと思う。

断られることを想定してやってみるという議論は以前からある。一方で、データ産出者をリスペクトした形でやらないと、このプロジェクト全体がソッポを向かれてしまうという助言、忠告を非公式に受けている。

もし産総研にそういう依頼がきても、組織上やりにくいと思う。出したいと思っている人に、もう少し動きやすい誘い方をしてほしい。現状では手間とストレスがかなりかかる。

データ産生者側で、データをダウンロード可能な形で公開してもらえば、解決する。

どこの組織にも公開のプロセスはある。統合DBプロジェクトに参加したという形であれば、すぐにデータを提供できるが、別プロジェクトから頼まれたから提供するというのは難しい。

手紙のあて先の問題もある。受託の長であるはず。実態は担当者だが、オフィシャルな契約者に出せばいい。

永井委員の話は、文科省統一契約書の範囲内の話である。それ以外のプロジェクトの場合は、別途検討する必要がある。

科研費には著作権に関する条項がなく、報告義務があるくらいである。

だんだんと進めるしかない。まず、文科省分から手紙を出していくべきである。

最初に新規プロジェクトで公開を約束させる方からやるべきである。「新規プロジェクトについては、データ公開のルールが決まって適用され始めている。ついては、過去分についてもさかのぼってこの精神を適用したい」という形でないとうまくいかない。

過去の分は個々の権利者との交渉が必要だが、これからの分は対応が可能である。

まず、これからの方針を固め、その精神にのって過去分もやるほうがいい。

データ公開のためには、強制力が必要である。そのためには、総合科学技術会議に働きかけるべきである。

公開の範囲の検討が必要である。文科省以外のプロジェクトのデータベースに関しては、文科省だけでは決められない。別途検討する必要がある。

大上段にデータ提供を要請するのは、やりにくい。こちらからデータ提供をお願いするよりも、データ産生者側でダウンロード可能な形で公開してもらおうほうがやりやすい（みんなに提供するというスタンス）。データ提供に関する成功PJの例をだして、これを根拠に他省庁のPJや過去のPJへと広げていけばよい。

プロジェクトに民間の資金が入っている場合も含めて議論すべきである。

どこまでが、研究でどこまでが基盤事業なのかの議論が必要である。

トレースデータのような生データでも知的所有権があるのか。

基本的に権利は受託者側にあるというスタンスであるのか。

権利に関しては、研究受託にする分と事業委託にする分を分けて議論するべきである。

国の事業と研究を分けてきちんと評価するとか、事業に関しては国に出すとか、そういう大きなルール作りやコンセンサスが合った上で、覚書や公募要領に反映させる必要がある。それができた後に過去分や他省庁分についてもやったほうがよい。

以上のことを文科省ライフ課だけでやろうとするととてつもなく大変なことである。文科省だけでモ

デルをつくらうと思わないで、より上のレベルから作らせたほうがいい。

総合科学技術会議でワーキンググループを作ったほうがいい。

総合科学技術会議にもっていくストーリーとすると、連携施策群から、五條堀先生、本庶先生という形で問題を落としてもらうというストーリーになる。

時間がかかる話になったが、そういう方向でいくということにしたい。

公募要領は、大きな話とは別途すすめるということでのいいのか。

> よい。

(まとめ)

委託プロジェクトを事業分と研究分に分けて、事業分に関してはダウンロード可能という形で国に出すということ、プロジェクト契約時の覚書や公募要領に反映させる。

その後、この方針にのっとり、過去分や他省庁のプロジェクト成果に適用すべく交渉する。

総合科学技術会議へ働きかける(事業分の委託成果のデータ公開に関してワーキンググループをつくり議論する)

(今後の進め方について)

公募要領への反映や総合科学技術会議へ働きかけについては、作業部会の中でワーキンググループをつくり検討を進める。

4. その他 今後の予定等

特にアナウンスはなく、高木主査より、閉会が宣言された。

以上