

「ライフサイエンス分野の統合データベース整備事業」 第2回研究運営委員会 議事要旨

【日 時】 平成19年10月5日(金) 15:30～17:30

【場 所】 東京大学山上会館201・202会議室

【出席者】 秋山委員、浅井委員、大倉委員、五斗委員代理、久原委員、田中委員、田畑委員、徳永委員、中村桂子委員、中村春木委員、松原委員長、大久保委員、小原委員、五條堀委員、菅原委員、高木委員、堀田委員

【陪 席】

内閣府	: 重藤参事官、鬼頭調査員、堀部主査、柴田主監補佐
厚生労働省	: 坂西主査
経済産業省	: 諸橋係長
文部科学省	: 菱山課長、生田課長補佐、田中調査員、石塚調査員
(独)科学技術振興機構	: 黒田課長、酒井主任調査員、橋本主任調査員
東京医科歯科大学	: 中谷准教授
(株)日立製作所	: 小池主任研究員

【事務局】 植田財務課長、笹島総務課課長補佐、加藤財務課課長補佐、丸山遺伝研管理部長、川本特任准教授、西川特任教授、植田事務室長

【挨拶】

松原研究運営委員会委員長から簡単な挨拶があり、開会が宣言された。

【議 事】

(1) 研究運営委員会(第1回)議事要旨(案)について

松原研究運営委員会委員長から、4月27日に開催された第1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) 人事報告

松原委員長から、中核機関の人事についての報告を高木委員にお願いしたい旨発言があり、高木委員から資料2を用いて説明が行われ、承認された。

(3) 統合データベースプロジェクトの進捗状況について

松原委員長から、統合データベースプロジェクトのこれまでの経緯と目指すところについて、堀田情報システム研究機構長に説明を求め、堀田情報システム研究機構長から説明があった。松原委員長から、統合データベースプロジェクトのこれまでの進捗状況について、中核機関、各分担機関から説明をお願いしたい旨発言があり、初めに中核機関の高木委員に説明を求めた。

1. ライフサイエンス統合データベースセンター(中核機関)

高木委員の簡単な紹介の後、川本准教授より、中核機関の目指しているサービスについて、特に横断検索と統合TVについてデモを用いた紹介があった。川本准教授の説明とデモに対する討議内容は下記のとおりである。

(横断検索に関して)

内部にデータをもってやっているのか。

Google的にクローリングを行い、データを集めている。

横断検索の規模は？

現在、国内の主要なデータベースを網羅している。

Google を用いる以上の意味は何か？

ドメインを限ることによって不必要な情報をカットできるので、大事な情報が埋もれてしまわないようにできる。また、ドメインを限ることで、オントロジーを適用することが可能になる。横断検索といっているものとは別に、統合検索というサービスを始める予定である。そうすると、もう少し意味に踏み込んだ検索ができるようになる。現在は、横断検索を開始したところである。

横断検索には意味がある。アルツハイマーとかマウスとか検索したときに、国内にどういうデータがあるのかを目の当たりにみることができるとは意味がある。

横断検索に関して、外国 DB のオプションがほしい(デフォルトは国内で、外国のブックマークをつくる) クローリングの時間とかインデックスのサイズの問題等の若干の検討が必要。できることなので、対応したい。

(統合 TV に関して)

DB のアップデートへの対処は？

マニュアルで対応する。

項目数が相当大きい数になりそうだが、大丈夫か。

現在の生物学者の解析のパターンは、せいぜい100～200とおりではないかと考えている。

項目の粒度がばらばらにみえる。プログラムの制約情報を載せたほうがいい。統合 TV の作成をオープンにしたほうがいい。

検討して対応する。

2. 京都大学(分担機関)

分担機関である京都大学の五斗委員代理より、これまでの進捗状況について説明があった。この報告についての討議内容は、下記のとおりである。

データベースセンターとの連携は？全体としてのロードマップの中で何を優先してやっていくのか。

PJが始まってからは、1～2ヶ月に1度のペースでセンターとの間でミーティングを行っている。検索に関して似通った機能があるため、その対応を議論しているところである。もししばらくしたら、方針が出せる。

3. 東京医科歯科大学(分担機関)

分担機関である東京医科歯科大学の田中委員より、これまでの進捗状況について説明があった。この報告についての討議内容は、下記のとおりである。

データベースの範囲を、当初の2疾患から、全疾患のデータベースへ拡大するのはいいことである。データの収集はどうやってやるのか。

まず学会を活用して調査を行う。

疾患情報の公開に関して、100人以下に特定化できないような方法はすでに確立しているのか。

現在、国際的にも議論がなされているところ。

SNOMED(米国臨床病理医協会が作成した医療用語集)は、利用料金が非常に高価という問題がある。連携する厚生省のプロジェクトではSNOMEDと同じようなライセンスを行うのか。

SNOMEDを使わないでフリーで同じことができるようにすることが目標である。

4. 東京大学(分担機関)

分担機関である東京大学の徳永委員より、これまでの進捗状況について説明があり、了承された。

(4) 作業部会(第2回)の報告について

松原委員長から、10月1日に行われた第2回の作業部会について、報告をお願いしたい旨発言があり、作業部会の主査である高木委員に説明を求めた。高木委員から、研究運営委員会作業部会(第2回)で議論された「契約・著作権、個人情報問題への対応」と「データベース受入の方針案」について、資料4-1と資料4-2を用いて説明があった。

「契約・著作権、個人情報問題への対応」では、以下の方針が示された。

現在の委託契約では、権利は受託者にある。今後の対応として、過去分については個別に協力要請を行い、新規分については契約時にセンターへの協力を盛り込む方向で検討する。

個人情報保護に関しては、過去分は個別対応し、今後については三省指針の改定時等にセンターへの協力を盛り込む方向で検討する。

上記の報告について、問題点が明快にされているので大いに期待したい旨の意見が出され、報告内容が了承された。

「データベース受入の方針案」については、以下の方針が示された。

検索方法について4段階のスキーム(1)DBポータル、2)横断検索、3)統合検索(分散型)、4)統合検索(集中型)を考えており、分担機関、補完課題、個別プロジェクトDB(独自実験データ)について、状況に応じて4段階の各方法を併用して統合していく。また、データベース受け入れに関しては、独自1次データを有するデータベース(遺伝子発現等)を優先して統合化し、それ以外に管理ができないものも順次受け入れる方針が示された。

この報告についての討議内容は、下記のとおりである。

しっかりとした調査に基づいていて、ロードマップがみえてきている。やるべきタスクがしっかりしている。今の与えられた条件の中では、やるべきことがしっかりと見えてきていると感じられる。しかし、それを超える問題に関しては、具体的な問題を解決していく必要がある。

統合DBの意義について、今日の議論は誰が聞いても説得力がある。もっと色々なところで紹介すべきである。

文科省以外からは統合DBへの参加はやりにくい面があったが、本日紹介された検索タイプの分類に基づいた統合の方法では、他省庁のDBを統合していく際にやりやすくなると考えられる。

目指すべきはAll Japanのセンターをつくることであるが、そのためには、総合科学技術会議を実績で説得することが必要である。

本日までにかかなりの進展があった。他省庁を統合DBに誘う準備体制ができてきた。そのために重要なことは、分担機関、協力機関との連携性をよくすることである。

All Japanに向けて、今日の話はすでに存在しているDBをどう統合するかという話である。これから技術が発展して新しいデータ、新しいDBをつくる人たちにに対するガイドライン、あるいはお手伝いする機能を持つといい。

そのための準備として、受入のテンプレートやフォーマットや辞書の準備を進めている。

DB側、クエリーを含めて、ある程度秘匿した状態で実行するツールの開発が重要である。

海外への発信が重要である。ポータルの英語化等の宣伝できる仕組みが必要である。

検討する。

(5) 科学技術振興機構(JST)バイオインフォマティクス推進センター(BIRD)

との連携：公募結果について

大倉委員から、前回の研究運営委員会でJSTから提案があった統合データベースプロジェクトとバイオインフォマティクス推進センターとの連携、及び公募を行った結果について、報告があり了承された。

(6) その他

五條堀委員より、11月29日に内閣府主催の「統合データベースのシンポジウム」を開催する旨のアナウンスがあった。

高木委員より、今年度の運営委員会をあと2回程度開催したい、次回は年末から年始にかけて開催したい旨の発言があり了承された。

以上