

## 「ライフサイエンス分野の統合データベース整備事業」 研究運営委員会作業部会(第1回) 議事要旨

【日 時】 平成19年6月15日(金) 15:30~17:00

【場 所】 情報・システム研究機構 事務局会議室(秀和神谷町ビル2階)

【出席者】 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

【陪 席】

文部科学省 : 坂下課長補佐、田中調査員

情報・システム研究機構 : 堀田機構長、

【事務局】 高野事務局長、石田総務課長、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、本作業部会においては、4月27日に開催された研究運営委員会における宿題事項である補完課題の公募等についての検討を行うこと、さらに作業部会のメンバーは、堀田機構長、高木センター長、文科省と調整を行った上で選出した旨の発言があり、開会が宣言された。

【議事】

(1) 統合データベース整備事業研究運営委員会作業部会について

松原研究運営委員会委員長からの要請で、事務局から資料1-1を用いて作業部会についての説明があった。その後、資料1-2の研究運営委員会規程に基づいて、松原研究運営委員会委員長の指名により、本作業部会の主査として高木センター長が指名され了承された。

(2) 補完課題について

高木主査からの挨拶の後、坂下課長補佐から、本事業の審査会からの提言に基づいて補完課題公募の検討を本作業部会において行うことになった経緯の説明があった。その後高木主査より、資料2-1~資料2-5を用いて、補完課題について検討すべきデータベースについての説明があり、討議が行われた。この報告についての討議内容は、下記のとおりである。

(受入戦略に関して)

●受け入れを具体的にどうするのか？

→一つはサーバー毎にあずかる。もう一つは、データのみ受け入れて管理する。内閣府の調査によって、データベースの種類が分類されていて、パターンによってどのような受け入れ方法が良いのかを検討することが、これからの課題である。

●やりやすいところから手をつける方針はいいが、その年次計画をしりたい。何を公募すべきかの基準を具体的に提示してほしい。受け入れ戦略が必要である。

→よく設計された整理棚で受け入れる部分と、整理しながら順番にやっていく二つの部分がある。受け入れには色々なレベルがある。大規模DBはすぐには取り入れず、相互利用可能にするのが考え方かと思う。大規模なものを公募したらどうか。

●中核に統合化する時に、どうやって統合化していくのか。作業の規模感がわからない。統合作業の費用は？

→センターのほうで多少の仕事は発生する。たくさんDBはあるが枯れたところから手をつけようという絵をかいている。権利関係の問題もあるので、受け入れチームで、順番にやっていく。

●標準化は、DBの保有者側には、技術者がいない場合が多いのではないかと？費用はセンター側に発生するのではないかと？

→インデックスを用意すれば、DBの保有者側に技術者がいない場合でも、標準化が可能ではないか。また、大きな組織にはそれなりの技術力がある。

●気になるのが、機構側に作業が発生し、負担が大きくなるのではないかと。募集しそうなところのイメージは？

→国家プロジェクトで動いているものがよい。小さなものはセンターでできる。

●早い段階でまとまった目に見えた統合 DB の姿がないといけない。250 の DB がつながって欲しいと見られている。本年度、どういう形で成果が出てくるのか。5 分→1 分 のような目標値が欲しい。  
→統合 DB の成果は DB の数では測れない。ユーティリティーで測るべきである。5 分→1 分といった評価は望ましくない。

→以下の対応を行う。

●250 の所在がわかるようにする。

●メジャーな DB は、横断検索によって文献も統合し、ユーティリティーを上げる。

●小規模 DB は、フリーズしてもってきて読み解いて、検索可能にする。

●終わりにかけているもの以外でもいいのか？

→Yes。将来的には、棚を提供して、それに合わせて、プロジェクトをデザインしてもらおう。将来の DB に関しては統合の棚を踏まえて提案してもらおう。

●セキュリティの面でも、技術の開発が大切である。

(小さな DB に対する対応に関して)

●小さな DB にかまっていられなくなるのではないかな。

→規模感は、大きなところを抑えると、データ量として 70~80% を抑えられる。小規模な DB は中核で容れものを用意して散逸しないようにしたい。

●小さな DB は負担大きい。

→データをフリーズさせてしまえば、受け入れの負担は小さくなる。最初は横断検索のみで統合検索も提供しない。あることが大切。著作物としてなくなることが重要。

●見えるようにしてもらおうのは重要。小型のものをたくさん入れた時に 1 回で検索する横断検索技術はある。

(構造アノテーションについて)

●構造アノテーションとは何か？

→構造アノテーションは、遺伝子領域等に関するアノテーションである。

●構造アノテーションはどうするのか？

→構造アノテーションも一緒に実施してもらおう。構造アノテーションを独立に実施するのは現実的ではないので、DB 作成と連携するのがいい。

(補完課題公募に関して)

●補完課題では、何を課題とするのか？

→大型プロジェクトや大量データ産生機関を対象として、センターの標準化仕様に従い統合化するための諸作業を、課題化する。

●どういう形で選ぶのか？

→公募で選択する。

●補完課題の期間は？

→事業の終了年度までである。

### 【まとめと宿題事項】

●まとめ； 補完課題では、大型プロジェクトや大量データ産生機関を対象として、センターの標準化仕様に従い統合化するための諸作業を課題化し、公募によって採択する。

●宿題事項； データベース受入戦略の具体化

### (3) 海外連携について

高木主査から、資料3を用いてライフサイエンス分野の海外連携に関して説明があり、センターでどういう海外連携をしていくのか意見を頂きたい旨の発言があり、討議が行われた。討議内容は、下記のとおりである。

●イネの場合は、現在、連携はない。連携をやる必要性あまりない。

●前回の委員会で、PubMed Central のミラーの議論があった。日本には、このような問題に関して対応するための機関がない。委員会で個別の議論をすることになる。色々な分野で海外連携の話が出てきたときに、意見を求められる可能性があるので、ある種の方針があることが望ましい。

- DDBJ は、毎年実務者会議、隔年で戦略会議をやっている。統合 DB が何か寄与できないか。
- 要請があった時に何かができるフレックスな仕組みをつくったらどうか。
- ある DB を外国からみたときに、その DB のプレゼンスがあれば連携のアクセスが出てくるのは自然である。トップダウンの仕組みを作る必要性はなく、個々の対応にまかせた方がよい。
- 窓口は開けておく必要がある。センター長の判断、作業部会の判断で、そのときどきの判断をすればよい。
- まずは国内の DB の統合が先決である。しかし、現在の動向の把握は必要。

**【まとめと宿題事項】**

まとめ； 現在、特に海外連携のための施策は行わないが、現状動向の把握は必要であり、センターで調査を継続的に実施する。

宿題事項； 海外連携に関する現状動向調査の継続。

**(4)その他 今後の予定等**

文科省から、次回の作業部会について以下の要請があり、了承された。

来年度の概算要求に向けて、早めに議論して運営委員会にかけ、その後のライフサイエンス委員会で承認を受けたい。8月22日か9月のライフサイエンス委員会に出したいので、来月には内容を固め、作業部会を7月下旬から8月上旬にかけて開催する。

また、補完課題の公募のスケジュールに関して、最速で、7月中旬半ば公募開始、今月中には内容を固めて事務作業を開始したい旨の発言があり、了承された。

以上