

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分科会【立体構造・植物関係機関】議事要旨

【日 時】 平成21年5月29日（金）10:00～11:30

【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（理研）、土井考爾（理研）、国島直樹（理研）、浅田征彦（理研）、吉良聡（理研）、皿井明倫（九工大）、岡本忍（DBCLS）、藤澤貴智（かずさ）、中尾光輝（DBCLS）、田中康博（文科省）、大久保公策（遺伝研）、高木利久（作業部会主査）  
永井啓一、西川哲夫、川本祥子、畠中秀樹、箕輪真理（以上、DBCLS）

（敬称略・順不同）

➤ 理化学研究所

◇資料説明◇

・全体および植物 DB 統合

理研のデータの一部公開とともに、理研全体で公開に向けた取り組みができるようにとのことで推進中。また、中核から求められている理研 DB のメタデータについて作成中。

DB 統合についてはシロイヌナズナをモデルとして進めている。民間との協力でこれまで運営してきた DB については企業ロゴを出すなどして個別対応。実験データファイルも公開。21年度は既公開の DB についてのアノテーションの充実、その他のデータの公開、アノテーション情報の国際基準適用などを実施する。蛋白質構造情報 DB の公開を推進中。

・微生物（結晶解析、重電子実験）

H20年度はデータのクリーニングを進め、公開は完了した。数量としては3種類の情報について目標を達成。7月にダウンロード（DL）可能にする。H21年度は公開データについてデータやアノテーションを追加する予定。

・構造解析実験

横山ラボのデータについては、20年度は、発現検討実験（大量合成、大量発現、精製）、結晶化実験に関するデータを整理し、2万件の回折実験データを公開した。21年度はアノテーション付与を実施、DLを可能にする。

◆質疑応答◆

○（かずさ）植物関係の連携としては、シロイヌナズナのデータをいただいて BioMart での連携を行った。今後の課題として国際標準データ TAIR のバージョンとの関係づけがある。

→データは随時変わるので、また改めてお渡しするが、とりあえず連携の実験のためにお渡しした。TAIR のバージョンがまたすぐ上がる、という話もあるので、合わせていく必要はあると思うが、今後どうしていくかご相談したい。TAIR の中で日本の貢献分についてクレジットが残るような形にしたい。

○ダウンロード（DL）のほうが開示に対して遅れるのは、制度的問題か、テクニカルな問題か？

→両方ある。データファイル数が大量なためデータの作成自体に1カ月くらいかかるのがテクニカルな問題。ライセンスの設定については弁護士と最終相談中。

○回折データは DL できるようにする予定か？

→横山ラボのデータは数も多いので、ある程度まとめた単位で DL できるようにしたい。

○（結晶解析実験の）①と②のデータ間で同じタンパクについて変異の有無で比較したい場合はどのようにすればいいのか？

→タンパク質名で検索すればよい。

○アノテーションとは具体的にはどのような内容か？

→今は実験データしかないので、機能や変異といった情報を付けていく。理研の中でまとめていくためには個々の DB 間の対応付けが必要なので、用語をそろえるなどの作業も必要である。横山ラボのデータは、ノートレベルから起こさなくてはならない。その作業を実施中。

○許諾条件は表示継承とのことだが、DL データについては改変禁止の要望もあったと思うが？

→個別の DB については作成した先生の意向を重視しているため対応がいろいろ。著作権について怪しいものについてはどうするのかという議論もある。メタデータについては改変禁止を削除する予定。

○メタデータの定義はどのようなものか？

→いろいろな様式がある。セマンティックウェブを用いれば項目の意味関係を表すことも可能。コンテンツの中身まで表現するようなものもある。

## ➤ 九州工業大学

### ◇資料説明◇

熱力学データと構造データの統合のために、統合化技術の開発と文献から情報収集するためのテキストマイニング技術を利用した自動情報収集の開発を行っている。20 年度は、一部数量目標を達成できなかった部分があるが、おおむね達成した。数量目標未達のクロスレファレンスについてはこれまで科研費で行っていた、元となる情報の収集のための予算が取れなかったため、元データの数が足りなかった。今年度も当該予算申請が通っていないので、元データの作成に支障が出るのが予想される。試算してみたが、この費用についてプロジェクトから追加で補助を頂けるとありがたい。21 年度についてはとりあえず自助努力でできる範囲の数量目標を立てている。

### ◆質疑応答◆

○ProTherm (蛋白質の安定性の熱力学データベース) データ 700 件と ProNIT (蛋白質・核酸相互作用の熱力学データベース) データ 700 件となっている登録件数は予算増の場合どうなるか？

→本来の目標はそれぞれ 1000 件と 1300 件である。

○増分 (300 件と 600 件) だけの予算補助でも、その目標が達成可能か？

→各 700 件というのは自助努力で何とかなるかもしれないが、もともと科研費では 1000 件と 1300 件分の予算を申請していたので、できれば全件分の対応をお願いしたい。

○1000 件と 1300 件が実施できれば、過去の分も含めて網羅した DB が作れるということか？

→ProTherm のほうは毎年の増加分を含めてカバーできると思う。ProNIT のほうは後から開始したので、件数が多くなっているが、2-3 年あれば過去分を網羅したものができると思う。

○ユーザー数はどれくらいいるか？ユニークユーザーとしては？

→アクセス数として年間 3 万件くらい。ユニークユーザー数は実際には数えていないが、感覚的には 100 人のオーダーよりは多い。

○現在のクロスレファレンス数は？中に含まれている情報はどのようなものか？

→ProTherm は 2 万 4 千件と ProNIT は 8 千件。50 種類くらいの項目のデータがある。

○データは一括で DL 可能か？DB としてのウェブ上のサービスは、DL したデータを扱うよりは高度な

検索等ができるのか？

→DL 可能、ただし、アグリーメントにサインを要する。ウェブ上でできることはそれほど複雑なものではない。文献からデータそのものを作るところに労力がかかるので、データに価値がある。

○熱力学データはハイスループットにはならないか？サンプルを作るところが手間か？

→個々の研究者が個別の対象について実施しているレベル。個別の変異を入れて、安定性がどう変わったか、といったデータなので。

○論文を読むのと、この DB を使うのと、どちらが効率的か？

→個別タンパクが対象か、網羅的解析をしたいかといった研究の目的による。網羅的解析をするのであれば、個別の論文を読むのは大変。1 論文あたり 30 (蛋白分) くらいのデータが入っており、論文を読むところに時間がかかる。この DB を使ったと言及している論文がすでに 100 報出ているので、それらの研究者は価値を感じてくれたと思う。

○DL のアグリーメントの内容は？

→引用など、一般的なこと。再配布は禁止している。アカデミックはフリーだが、企業に対しては有料 (交渉次第) にしている。ただ、今後はどうしていくかと検討が必要。DB の著作権者は誰か、という整理から行っている。データを選別しているところに意匠性があり、著作権があるとされている。

→ (DBCLS) 大学の知財の方と検討していただいているという状況。

○個々の論文は著作物だと思うが、著者には断っているのか？DB として集めていること自体は著作権が発生しないのではないか？そのため、センターでは、アカデミックも民間もフリーとしているのだが。

→実際には手間ばかりかかるのでフリーにしてもよいのだが。大学の方針や要請もあるので。

○引き続き、中核機関と協力しながら検討していただきたい。

○予算的な補助について文科省からご意見をいただきたい。

→ (文科省) 統合 DBPJ でやる業務範囲を考えると、これまで科研費でやられてきた内容は PJ 外の部分になる。調達できなかった外部資金の補てんという説明では厳しい。ただし、これをやらないと PJ の成果として立ちいかない、という運営委員会等の合意があれば、業務範囲を拡大することも可能。客観的にそれを説明できるようにお願いしたい。助成金の申請が今回通らなかった理由があるはずなので、それは何かを踏まえて、この PJ 内で必要かどうかを委員会で審議いただきたい。

○PJ の趣旨はすでにどこかにデータがあることが前提でそれを統合化するところがタスクということなので、(文科省の) 言われることはわかるが、このケースはすでにある文献データからまた編集するので統合化作業の一環と考えてもいいと思う。このあともう少し議論して、委員会に出す出さないを決める。

(科研費の) 目的のひとつは情報をデジタル化することと思われるが、文献情報はすでにデジタル化されているので、と理解されたのではないか？

## ➤ 総合討論

○ (文科省) PJ の成果が出てくる際に強くアピールしたいので、理研等個別機関での公開告知以外にも、PJ のサイトからも発信したい。DBCLS の広報担当にそういったニュースを集める仕組みを作ってほしい。成果報告書にはいろいろ書かれているのだが、タイムリーに発表してほしいし、発表する際にはプレスにするなど、アピールできるようにしたい。毎日何らかのニュースが出てくるくらいにしてほしい。JST のプロジェクトのページもなかなか更新されていないので、更新されるようにしてほしい。オーバ

一に宣伝する必要はなく、発表の事実を公表する形でいい。

○皿井先生や国島先生がやられているような文献からの情報収集のノウハウなども共有すれば役立つと思うので、あとでヒアリングさせていただきたい。

→（ヒアリングについて）協力する。

最後に高木主査より、1)前回の議事録確認依頼、2)シンポジウム(6月12日)への協力御礼と依頼があり、会を終了した。

(11:30 終了)