

**「統合データベースプロジェクト」
研究運営委員会（第6回）議事要旨**

【日 時】 平成21年1月19日（月）14：30～17：10
【場 所】 情報・システム研究機構事務局会議室
【出席者】 大倉委員、勝木委員、豊田委員、長洲委員、中村桂子委員、中村春木委員、
松尾委員、松原委員長、湯元委員、吉田輝彦委員、吉田光昭委員、大久保委員、
五條堀委員、高木委員、堀田副委員長

【陪 席】

厚生労働省 : 坂西主査
農林水産省 : 浅野係長
経済産業省 : 諸橋係長
文部科学省 : 川上調整官、山中調査員
（独）科学技術振興機構 : 黒田課長、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
箕輪特任研究員

【事務局】 高野事務局長、植田財務課長、笹島課長補佐、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. プロジェクトの経緯と課題

最初に松原委員長から平成20年8月11日開催の第5回研究運営委員会において了承された見直し案に基づく新しい研究運営委員会委員名簿が読み上げられ、各出席委員から簡単な挨拶が行われた。今回から新しく参加する委員もいるため、高木委員から資料1に沿って、本プロジェクトの経緯が説明された。

本プロジェクトの開始に先立ち、内閣府の連携施策群で「ライフサイエンス分野のデータベースの統合化に関する調査研究」（担当：五條堀先生、大久保先生）が行われた。平成18年5月には本プロジェクトのベースとなる報告書が文科省のライフサイエンス委員会のデータベース整備戦略作業部会から提出され、それを受けて同9月に、いわゆる統合データベースプロジェクト（半年分の予算約3億円）が公募でスタートし、情報・システム研究機構を中心としたチームが受託した。その後、平成19年4月にライフサイエンス統合データベースセンターが設置された。同月に4年間の本プロジェクトがスタートし、中核機関1件と分担機関3件が、同10月には補完課題4件が採択された。（平成19年度プロジェクト予算総額は16億円）今年度20年4月には予算規模が11億円になった。同5月には中間評価が行われ、7月にはその報告書が出された。報告書は、「順調に進行しているが、中核機関が必ずしもイニシアチブを発揮していないので、より緊密な連携をするように」という内容で、これを受けて研究運営委員会と作業部会を刷新し、作業部会は開発担当者がより緊密な議論をする会に、研究運営委員会は大所高所から日本の統合データベースのあるべき姿を議論する会に変更した。

報告書ではさらに、「本プロジェクト終了後のデータベースを如何に維持していくか」について言及され、それを受けて平成20年12月から五條堀先生を中心とする内閣府ライフサイエンスPTの中で統合DBタスクフォースが設けられた。また、文科省のライフサイエンス委員会の中にライフサイエンス情報基盤整備作業部会（委員長：勝木先生）が設置され、平成21年1月には「ライフサイエンスデータベースの統合・維持・運用の在り方」という作業部会報告書が出された。いずれにせよ平成23年3月には本プロジェクトそのものは終了予定である。

現在プロジェクトが抱えている課題としては、①中核機関がイニシアチブをとってプロジェクトを如何に統合的に進めていくのか、②データベースの受け入れに関する共有化の進め方、公開に関するルールづくり③次世代型のシーケンスなど、生命科学の新たな展開におけるミッションの明確化、さらにもっと大きな課題として、④本プロジェクト終了後のデータベースの恒久化のための体制（4省

連携の体制含む)、⑤データベースの運用に際して受益者負担的も検討すべきか、などが挙げられる。

本日は、このうち④プロジェクトの終了後の体制について、予算について、②データベースの受け入れ・公開のためのルールづくりについてご議論いただきたい。

上記の説明について委員長から質問を求めたが、特に無かった。

2. プロジェクトの体制・進め方について～文部科学省「ライフサイエンス情報基盤整備作業部会」の結果を受けて

松原委員長から、ライフサイエンス情報基盤整備作業部会の報告書について、文部科学省ライフサイエンス課に説明を求めた。

ライフ課の川上調整官から資料2を用いて以下の説明がなされた。

資料2は報告書本体と参考資料(提案資料や作業部会での議論の資料および総合科学技術会議[以下、CSTP]の関連資料)から成る。報告書の前半は、データベース整備の意義・必要性、あるいは現状・問題点、諸外国の状況等についての提言内容なので簡単に触れ、その後のデータベースの統合・維持・運用の在り方(機能及び体制)を中心に説明する。

データベースの整備の意義及び必要性について、今後も大量のデータが出るため、それらを管理し、発見に結びつけるために必要で、特に医療界や民間企業などにおいて必要性が高いことが指摘された。

現状及び問題点について、文科省では統合データベースプロジェクトとJSTのBIRDのファンディング事業を実施しているが、全省的には今後どのような取り組みをするかが問題であると指摘され、これについては一部CSTPの下でも各省連携の施策を推進している。プロジェクトに関連する具体的な問題点として、基盤整備の弱さ、データの共有認識の弱さ、あるいは中核機関のイニシアチブが発揮しにくい体制、予算規模、事業運営の恒久化の問題が特に指摘された。

諸外国の状況についても参照しながら、データベースの統合・維持・運用の在り方について、まず機能面においては、

- ・文科省が率先して実施し、その後の全省的連携には総合技術会議の議論を踏まえる
 - ・担う機関の基本的な役割として、将来的な国内のデータベースの集約および継続的かつ主体的な運営機能、そのために必要な技術開発、付随情報も併せた統合化を進めていく必要がある
 - ・当面、研究開発機能は持たずに、データベース統合の事業に徹する
 - ・制度に関する検討も必要となる
 - ・開発者と利用者への支援を両立させるためのコミュニティからの協力体制を構築する
 - ・プロジェクト成果実績を継承していくにあたり、開かれた運営や継続的な人材育成に留意する
- といったものが要求されると指摘された。

これらの機能を担うための体制としては、海外の事例のような独立機関の設置も考えられるが、本プロジェクトの終了までには実現は困難なので、将来的な理想像についてはCSTPの今後の検討に期待しつつ、CSTPから出された平成21年度概算要求の優先度判定等における「本事業の成果を重視。恒久的な体制整備に向けて統合データベースプロジェクトとJST-BIRDとの一本化を目標として具体的検討を進めるべき」という指摘を踏まえ、ファンディング機能のノウハウを持ち、文献情報等のデータベース利用者支援も行ってきたJSTにおいて新たな組織を設置し、そこでROIS-DBCLSをはじめとする関係機関各々が持つポテンシャルを最大限生かしつつ柔軟な運用を可能とする仕組みを検討することが適切であると指摘された。

新たな組織の運営については、

- ・広く研究者コミュニティの意見を踏まえる
- ・利用者からのニーズに的確にこたえられるように配慮する
- ・現在の統合データベースプロジェクトで培われた成果を効果的に引き継ぐ
- ・透明性、公平性、客観性に十分配慮したオープンな運営体制を確保する
- ・データベースを有する関係機関とネットワークを密に図る
- ・総合科学技術会議の検討結果も踏まえる

といった在り方が提言された。また、本プロジェクト終了時の移行のために21年度予算について経過措置を講じることが望ましいという指摘もあった。

●作業部会の主査であった勝木委員から補足があった。

作業部会の主題は本プロジェクトが5年間の期限付きプロジェクトである一方、統合データベース

は継続しないと全く意味がない、それではどうやって移行するかというもの。

その結果、これは国家的な事業であるので、各省庁で議論するよりも1つの統一した統合データベースを扱う、継続的な組織をつくるのが最善であるという結論にまず落ち着いた。しかし、まず5年のプロジェクトを繰り返すことは継続性の保証がないという問題点があった。さらにそれをROISで継続するという点で、毎年削減が義務付けられる運営交付金の形では、ROIS自身の組織に影響することから現実的ではない、という指摘があった。一方、「JSTのBIRDを一緒にして、この事業を統合データベースの中でやりなさい」という総合科学技術会議の意向があったので、JSTへ移行する場合を検討すると、ROISに比べて7倍ぐらいの運営費交付金を持つJSTなら、20億円程度の予算もバッファの中で解消しようと(文科省から)説明があった。

このように、行政の委員会として現実に可能な施策についての検討を行ったのみで、統合データベースの枠組みを政策的にどう変えるかについては、(文科省という)組織の中では我々は議論できなかった。であるから内閣府のタスクフォースへ、「統合データベースは国家的な事業であるべきで、暫定的なプロジェクトとしての対応を続けていくべきではない、今後はぜひタスクフォースでプロジェクト終了までの2年間に徹底的にそのことを議論して欲しい」という報告をした。今後の社会の情勢の変化、データの爆発的な増大も視野に入れた検討をして欲しい。

●当事者であるROISとしてはこの報告書により、2年後の姿を検討しながらJSTとの協力体制を確立していくという形が示されたと思う。報告書が出たばかりなので具体的な相談は始まっているが、協力をしたい。まずこの2年間の問題と、それからそれ以降の問題2つに分けて考えたい。この2年間はあくまでも中核機関として、JSTと協力しながら引き続き現在の事業のレベルを最大限にまで上げて事業としては終了したい。終了後については、センターそのものはROISの中で継続し、データベースの統合化についてもっと広い意味でいろいろな開発研究ができる組織としたい。コミュニティがみんなで賛成するために中立的立場のROISがプロジェクトを引き受け、広い範囲から大勢の方に集まっていただけだったので、そのことはぜひ新しい組織にも引き続き伝わって欲しい。また、統合データベースに関わる仕事には科学の知識と能力が必要であるが、サイエンティストとしてキャリアにはなりにくいので、有力な人が集めるためにもこういう仕事がちゃんとキャリアパスとなるように考えていただきたい。

●(委員長から)引き続きフリートークで現状問題等についてご意見をいただきたい。

●内閣府連携施策群について若干補足する。平成17～19年度に実施された内閣府の総合科学技術会議の連携施策群は、科学技術関連施策における各省間のオーバーラップや抜けを8分野において調査し、オーバーラップは是正し、抜けについては補完課題として推進するというもの。そのうち1つがライフサイエンスの基礎。ライフサイエンスにおけるデータベースについては重複の問題はあまりなかったが、各省連携が不足していると考えられたので補完的課題として調査を行った。それが「データベースの統合化に関する調査研究(代表:大久保委員)」であり、その中で我が国におけるライフサイエンスのデータベースは大体250ぐらいあること、米国と我が国の違い(ルールづくり、体制、データベースの内容)などが明らかになってきた。調査で明らかになった現状について、いち早く呼応したのが文部科学省で、平成18年には本プロジェクトが始まった。その他の省も省内での統合化や連携可能な形のまとめなどを始めたが、文科省の本プロジェクトが先導しているので、ここを核として進めていくことになった。何もなかった状況からすれば、大きな進歩。本プロジェクトの運営会議には総合科学技術会議並びに各省とも参加して関心を持っているが、時限付のプロジェクトなので2年先の終了時のことを全体で考えなくてはならない。

一方、調査を実施していた連携施策群も昨年3月に終了したが、この重要な問題については放置しておけないので、内閣の総合科学技術会議のもとにあるライフサイエンスPTの中に統合データベースタスクフォースを設けた。昨年の12月から4月までに4回開催予定(すでに1回終了)で、何らかのビジョンを出すことになっており、タスクフォース内で2年かけて議論できるわけではない。現在はまだ、将来のあり方(文科省プロジェクトとしてのあり方と各省との連携)について、ビジョンやロードマップがなく、研究者にも不都合であるが、行政担当者も先に進めない状態。大きな課題としては3つあり、1つめは公開性の問題(ルールづくり、研究者レベルのルールと、行政的なルール)、2つめは継続性、体制の問題(資金含む)、3つめはどういう統合を求めるか(コンテンツ)の問題。これらについてビジョンが必要と考えている。

●統合データベースプロジェクトを立ち上げ時の文科省担当者として一言。立ち上げ当初、どういう組織体制で実施するかについて、堀田先生とも意見交換をしたが、継続性が最大の課題で、行政としても最大の悩みであった。文部科学省の中にも基盤的なプロジェクトが幾つかあり、本来こういった

ものはしっかりとした継続できる組織体制でやるべきだが、運営費交付金には暫減という大きな縛りがあるため、プロジェクトを立てた。今回の統合データベースも当初からJST/BIRDとの連携なども考えていたが、一番アカデミックに認知され、協力が得られるということで、まずは第1期の5年をROISをお願いをした。数年やってみたところで、期限付きのプロジェクトで2期、3期とつなぐのはいかがなものかという議論があり、今回組織体制について検討されているのだと思う。文科省でも統合DBをその他の基盤プロジェクトの継続性の担保に関するいいモデルとしたいということだと思ふ。そして世界的にも納得される形にしないと、今まで費やしたこのプロジェクトの経費が無駄になる。社会的に大きなインパクトがあるので、次にどうつなげるか、組織体制をどうするのかを真剣に考えねばならない。次のプロジェクト、次の体制とこの運営委員会が密に連携して、どうつなげていくのかを議論していただきたい。

●昔は日本のデータベースがなぜ必要かという議論をする人は結構多かった。何でも海外のものをもってきて使えばいいというのは、科学者ではないコミュニティに強く、その議論を突破するところから始まったと思う。今でもその考えの人は結構多いと思う。

それから、国としてデータベースが必要だとすれば、恩恵を受けるはだれかがはっきり言われていない。生命科学の研究者だけだと思う人も結構多い。その点については、報告書の表に書かれている科学技術・学術審議会、研究計画・評価分科会、ライフサイエンス委員会、それから、基盤整備作業部会の委員の間で意思が共有されているのか。そこがはっきりせず費用がかかると騒いでいてもだめなのだが、この点について意見を述べる、説明していただける方はいらっしゃるか。

●いわゆるサイエンスコミュニティよりは産業側のユーザーとして一言。我々はいいいデータベースさえあれば、日本であろうとどこであろうと構わず、使いやすいうところに飛んで行く。そのため今まではNCBIなり、EMBLなりのデータを主に使ってきたのは事実。では、それでいいのかというと、今後は2つほど問題があると思う。1つめは非常に多種多様なデータで、NCBIやEMBLでも賄い切れずに散らばって、日本がリーダーシップをとれる分野の日本にしかないデータベースが出てくる。2つめは横串で見なくては解析できないデータ。また、今後、個人個人のゲノムのシーケンスが全部明らかになる時代を考えると、ゲノムシーケンスは国のものか、個人のものか、を考えていかななくてはならず、少なくとも日本人のゲノムはやはり日本で1つにまとめていかなければいけないだろうと思う。そのためには日本としてやっぱりデータベースを持たなくてはいけない。それは文科省あるいは経産省という省庁別のものでなく、日本のものとして早く考えていただきたいというところは、いろいろな会議で言ってきた。とにかく産業側は省庁関係なく、データがまとまっていてどこかにあってほしいという要望が非常に強い。

●今のような議論は、それぞれの委員会の委員は体感されているということだろうか。
→そう信じている。

●日本だから日本人のデータベースをつくらなければいけないというと、それではそのほかのところは外国のものをもってきてくれれば良いという議論にもなり難しい。日本人のというよりむしろ、日本固有のものを目指すべきだ。これからのライフサイエンスも、バイオテクノロジーも、今までとは異なったやり方のデータ駆動型のサイエンスの面が非常に強くなる。そのときにどんなデータを持っていて、どのくらい自由にデータにアクセスできるか、解析のパワーがあるかが、重要になってくる。

これからのサイエンスに非常に大きくかかわるところなので、日本人固有の遺伝子データベース、国産の植物のデータを持っていれば良いとはせずに、それぞれの委員会で議論を共有してほしい。そのためにはどういうセンターが必要かという考え方について各委員会が共通のコンセプトで議論していただくように、今後ともご協力願いたい。

日本のライフサイエンスを進めるために日本固有のデータベースは必要なのだというコンセプトに基づいて、データをどう共有化するか。体制を、ルールを、継続性をどうするかというような考えが出てくると思う。これから議論の方向性を整理せずに自由に議論いただきたい。

●基盤整備作業部会の報告書に真意が書き切れていないとおっしゃることが、まさに委員会間でそれぞれの真意が共有できないということではないか。

→作業部会という制約があり、継続問題について議論できるのは運営費交付金のレベル。コンセプトの話はできない。

●現実的な答えを出さなければならないが、一方で原理原則も求められるということか。

→原理原則については内閣府のタスクフォースをお願いしている。

●連携施策群での調査もサマリーになると真意が見えなくなる。問題は科学の問題というよりも、政府施策としての大型プロジェクトのアウトプットが無計画に積み重ねられていて、本プロジェクトで

費やした経費がこの先が計画されないと無駄になるのと同じように、その何百倍もが無駄になる可能性があるという日本の特殊な状況。

→だからこそ統合データベースプロジェクトがあるのではないか。大きなデータを放置しないで、一番求心力のある非常に単純な目標を出して、その上でそういうものをどう処理するかということをやらない限り無理。ボトムアップで幾らやってもだめ。大きな哲学を出して、それを批判する形で削り込んでいかないと。

●科学者である我々だけがわかればいい問題ではない。

→科学の話から出発して、行政をどうするかという話に持っていかない限り、実現できない。制度をどうしようという話ではない。

●これまでライフサイエンスの研究につき込んだ何千億円分のデータを日本が持っているのに、国民にその一覧も提示できないのはおかしいということ。国民の財産だから日本国民が見られるところに置いて、その扱いに関しては国民に考えてもらうべき。

→統合データベースとはそういうレベルの話なのか。もう少し学問的に加工する、あるいは実用的に加工することが非常にトップダウン的に明晰にされていない限り、集まろうという気にならない。

●それは集めたアウトプットのアプリケーションで、アウトプットをアウトカムにつなげるための技術で、そこはアイデアや知恵が盛り込まれる競合可能な部分である。まずはアウトプットがだれでも使える形で集められていることが必要でそれができていない。

→統合データベースとは、そこから新しいサイエンスが出てくるもので、そのためにはやっぱり加工が必要なのではないか。

●実際に始めてみると、このプロジェクトが加工に使えるデータがないということに気がついた。

→そこは理解できない。もしそうならば本プロジェクトは無駄ではないか。やはり問題の核心は全然つかまれている。省庁連携や継続といった小さな話から出発したら、結果的にプロジェクトが成功しなかったときに「省庁間の壁が厚かった」で終わりになる。だから、最初からビジョンをきちんと出して、後で具体策を肉付けすればいい。ただ、それは作業部会でやれる範囲を超えている。その政策を行うのはCSTPで、そこからおりてくれば我々は具体的にやる。今、リーダーとなっている人たちがちゃんとシンプルな原理原則を出して、それから組織をどうするかという議論に入れば、求心力が出るはず。そして、そのデータベース関連費用は政策として1カ所に集めて、それを役割分担して分けるぐらいの力は持たなければこれは無理だと思う。

●作業部会はやはり文科省が主催なのでその範囲内でしかできない。今の問題は今後どうするかというよりも、そもそも統合データベースはどうあるべきかという一番大事なところを少し議論しようと委員長は言われていると思う。2年前にも委員会で言ったのだが、統合データベースそのものの概念がどういうものかがよくわからない。ただ、山ほどある正確でないデータベースを1カ所にまとめてつなぎ合わせるだけではだめ。だから、統合データベースとはどういうものか、その意味、やらない場合のデメリットをもう1回きちんと議論する必要が出てきたのではないかと思う。統合データベースのイメージは各人で大きく違うのではないか。その異なるイメージを持ちながら1つの結論を出そうという議論を各省庁の行政の枠の中でしている。大事なものは研究者が一生懸命研究して自分たちの成果をどうするかという研究者の意思なのだと思う。しかし、日本の生物学者には、自分たちの研究の推進や効率化についての戦略を自分の立場を超えた立場で考えるという訓練がほとんどなされていないと思う。だから統合データベースの議論になったときに、そこにかかわる人が自分の研究を優先してはならないとか、言わずもがなのことを改めて言わなければならないということが非常に問題。統合データベースって、一体どういうものをつくらうとしているのか、どういう価値があって、やらないと日本としてどう損するのかというのを改めてまとめてほしい。イメージが個人ごとに異なることが問題。

●何かまとめがと出されていたと思うが、お考えのある方は。

●生命現象における分子から環境間での様々なレベルの情報の階層間をいかにつないでいくか、その基盤づくりが統合データベースのあり方を決めると思う。ライフサイエンスの発展には、文科省を中心の分子情報、厚生労働省が持つ病気の部分など、省庁別になっている情報をつないでいく必要がある。応用については農業や進化などいろいろあると思うが。

●それだけの話なら別に日本でなければならぬ必要はない。世界に1個あればいいという話になりかねない。日本らしさはない。

●日本固有については別の問題だと思う。統合データベースとしてのあり方としては、論理的なつながりかた、因果関係あるいは整合性でもいいからそれらをつなげていくことが大事なことだと思う。

●この点は抽象的な議論なので、打ち切ることにする。しかし、基盤整備作業部会の報告書には統合データベースはどういうものかということは書いていないので、集約して何らかのたたき台を出すことが必要だと思う。3つか4つたたき台をたたいていくことでミニマムの要件が出てくるといい。ここにいる人たちが納得して、外の委員会にも「これが本来あるべき姿です」と提示できるものができる可能性がある。これをいずれ何かの方法でやることにしたい。

●平成18年のライフサイエンス委員会データベース整備戦略作業部会の報告書にはなぜ統合データベースが必要かということは書いてある。もちろん大久保先生の連携施策群調査の結論というものもそれに加えればよろしいかと思う。

●その報告から数年たっているの、それを1つの参考のデータとして、最後のときにこれをもう1回まとめさせてもらいたい。いろいろな委員会で共有意識のもと議論する方法が大事で、この委員会が一番広い範囲の関連のある方々が集まっているところなので、考えていきたい。

●データ駆動型サイエンスがこれから中心になっていくとやはりトレーサビリティが重要。論文の結果を導くのに使われたデータはやはり公開されるべき。科学の再現性はトレーサビリティがあってこそだと思うので、トレーサビリティのあるデータを入れられる統合データベースは、サイエンスからうそが出ないようにする意味でも、データの再利用という意味でも必要だと思う。データの提出には、各ラボでデータのガバナンスできていないといけないが、現実には特定の研究代表者（PI）にしかデータの在処がわからない、その人が移動したら無くなってしまい、データの提出ができないということになる。人のデータまで理解して出すのは難しいが改善が必要。ガバナンスがきちんとしていけば、公開もそんなに苦もなく、人が入れかわっても新しいPIがちゃんと結果をガバナンスできて、間違った論文が出ないように管理できるし、これはデータ駆動型サイエンス非常に重要な点なので、そこを強化すべきだと思う。その延長で統合データベースはうまくいくのではないかと思う。

●皆さんの議論から、統合データベースのあるべき姿について幾つかのフレーズにまとめられるような、必要なものが何項目か出てくるのではないかと思う。機能に関する議論もあるし、何のためにつくるのかという議論もあると思う。しかし、最も重要なのは、日本の将来のライフサイエンスに不可欠な省庁を超えた何ものかが絶対に必要であるという意識を、ここにいらっしゃる皆さんに共有していただくことだ。そのような共有意識を作り出せるのは多分ここしかない。予算的なこととか、2年後の体制のこと、具体的な統合のやり方やいかにして象徴を超えるかの行政的な手段なども考えなくては行けないが、具体的な形は誰がリーダーになるかで決まってくる。

●質問だが、CSTPから「省庁を超えたものにしなさい」というコメントがあり、文科省の報告書ではBIRDとROISとの統合化やJSTに新たな組織を設置という提言がされているが、これらの提案は他省庁からのポジティブな意見に基づいて出ているのか。

→報告書では、「次善の策」であると言っている。つまり、これはやむを得ず、継続ということが絶対に必要であるから、実現可能な一応の対策としてBIRDと統合データベースが一体化するにあたり、統合データベースについては引き続きROISが責任を持ち、予算の組織としてはJSTが担当するということである。他省庁の意見という形では聞いていない。

●他省庁からも参加いただいたということだったので、意見が反映されているかと思った。

→他省庁の研究所の方が委員で意見を言われたが、他省庁がこれを了解しているということではない。

●他省庁の方は本日もオブザーバーとして出席なさっているが、やはり他省庁の考え方もあるので議論を勝手に進めてここで決めるのは難しいのではないか。

→おっしゃるとおりだ。

→作業部会では、本プロジェクトの継続性の問題が中心だったので、議論の中で他省庁からの視点は念頭になかったが、他省庁からJSTやROISではどうかという意見も確かに幾つか出た。

●JSTやROISをどうするかという中途半端な問題ではない、という指摘もある。小手先の対応ではなく、苦しんでもやるだけの決意の価値が本プロジェクトにあるのか、その論理的根拠があると向かっていける。その論理的根拠がつけられるかが非常に大きな議論ではないか。

●先ほど指摘のあったようにこの委員会の権限は限られている。私が時間を限ってやろうと思ったのは、せっかくこれだけのメンバーが集まっているので、その権限外の議論を自由にしてほしいということだ。「かくあるべき」ものに向かって、コアになって考えていくメンバーの一部であるという気持ちを持っていただきたい。

●シームレスに使える新しい発見をできるのが統合データベースだと理解していいか。今はまだ本格的ではないが、あと2年たったらできると私は思っている。日本の中で統合データベースの内容をイメージできるのはプロジェクトに携わっている人ではないか。みんなが納得できる統合データベ

ースのモデルを中核機関がこの2年間にきちんとつくるのが大事。日本の統合データベースの見本と作り方を示すことが肝要。

●アウトプットの確保がまず非常に大事ということはわかるが、一方でアウトプットの提供方法にアプリケーションが影響することもあると思うので、本プロジェクトとして、「DB開発者と競合する研究開発機能は持たない」という宣言は足かせにならないかと心配。統合した新しい知識の発見こそ研究開発そのものなのではないか。そういった研究委託事業をすることが若手の人材開発にもなるのではないか。

●第1期の作業としてはデータを出すことが良いことになるモデルをまずつくるのが重要。ただ、モデルを示すのにも素材が必要で、そのための協力をいただいている。だから、第1期からは限られたものしか出てこないことは仕方がない。しかし、デファクトスタンダードのOSのように、仕組みが整ってくればそこに載せることが重要だという認識が広まっていくのではないかと思う。第2期には良いプロダクトが出始めて、結果としてもっと載せることになればいいと思う。物理の世界とは違い、ライフサイエンスは個人が研究するスタイルだったのでコミュニティのために何かをすることが非常に苦手だが、チャレンジをぜひ成功させていただきたい。

●私が思う「統合データベース」は省庁を超えて日本で生産されるデータをまとめ上げて有効に使えるような仕組み。BIRDも含めて他機関のいろいろなデータベースをうまく連携させて使っていく次の組織をつくるための方向性を2年後には出すことをやるべき。具体的なミッションを達成すれば良いというものではないのでは。

●本プロジェクトが動き始めてやっと「まとめ上げて使えるようにする」ということがどういうことかが見えてきた。具体的な事例を出すことが今とても大事ではないか。

→資料4にはアウトプットとアプリケーションとアウトカムの説明をするために例をわかりやすく書いているまでで、これに限定されるものではない。

→まとめ上げる前の元となるプライマリーデータベースは非常に大事で、ここに挙げられていないからといって、それをないがしろにするということではない。

●データベースはどれでも同じだと思っている人たちが「統合データベースをやるなら（プライマリー）データベースを削ればいい」と考えれば、統合データベースプロジェクトゆえにプライマリーデータベースがつぶれることが起こりかねない。プロジェクトで作る大きなプライマリーデータベースと統合データベースは全く異質で、その両方が必要。ほんとうに理解しているのはライフサイエンスの研究者だけなのではないか。文科省や総合科学技術会議からの啓蒙も必要。

●両者の違いは特定のコミュニティでしかわかっていない。それぞれのデータベースの概念の違いを明確にしてアピールすることが大切。報告書に書いてあるだけでなく、みんなに説明できるような絵を描いて、国家的な運動を国もアカデミアもやらなければいけないし、関係者皆がそれぞれの立場でいろいろなところでちゃんと言っていくことが重要。統合データベースのイメージのアウトラインはみんな同じにして説明をしていくことが必須。

●委員会の報告ではないものでもいいか。報告書にまとめると、意図するところが伝わらなくなる。

●委員会の報告ではないものでもいいが、読んで皆が実行できるものでないといけない。

●有志でそういうのを出してはどうか。

●その形が結論に近い提案ではないかと思う。ここに集まった方は、統合データベースとはいかなるもので、国際的にもどうあるべきか、これからのサイエンスにおける役割といった期待感とかイメージを漠然と持っているのではないかと思う。まとめて書くのは大変だが、その中でぜひ強調したい1、2点を持ち寄り、何かの方法でそれについての議論のやりとりをしたり、半日ぐらいかけて幾つかのテーマに絞って議論をする機会をつくれれば、それは非常にいいコアになると思う。制度的、委員会的な枠を超えて自由に発言できるようにしたい。ライフサイエンスの統合データベースに関して、ワイズメンの知恵を集めて最低限こういうものが必要だ、こういう精神が要るのだというものを、1行でいいから各人が精魂込めて書いた意見を持ち寄って、その中でも特に重要なものについて意見が分かれる場合には議論することを、今日約束できればいいのではないかと思う。それぞれ違ったスタンドポイントがあるだろうが、2月の中旬ぐらいまでに高木委員のところに意見を送っていただき、集約して箇条書きに取りまとめてはどうか。各人が重要視した問題のリストがあれば、次に何か言ったり、考えたり、ほかの場所で議論したりするときの核に必ずなるので。

●だれかが骨子を書いて、それに対してみんなが意見を言うほうが易しいと思う。この場合は批判することも非常に建設的なことだと思う。

●皆さんの意見のリストができた上で、どなたかに骨子の作成をお願いするという可能性もあるが、

まずたたき台ありきでやらないで骨子の種を出していただきたい。内容は統合データベースというものがどうあるべきか、最低限何を求めるか、などという項目でもいい。リストをどのように扱うかについては、委員長、副委員長、高木委員に任せていただきたい。リストの公表の仕方についても意見がまとまったら、議論したい。

3. 平成21年度プロジェクト予算の方針について

松原委員長から、「平成21年度プロジェクト予算の方針について説明をお願いしたい」旨の発言があり、資料3に基づき、高木委員から説明がなされた。

資料3-2は昨年の5月の中間評価による本プロジェクトの評価の際に、平成21年度の予算金額について具体的に出された案をまとめたものである。平成20年度分11億円の配分が記載されているが、これに対する評価の結果として、中核機関分を5,000万円増額、京都大学分を3,700万円減額、医科歯科大学と大阪大学のグループ分を2,000万円減額、また、東京大学グループ分を700万円増額、それ以外の補完課題は始まって間もないこともあり、特に変更無しという予算案をいただいた。これを踏まえ、平成21年度案を作成した(資料3-1)。基本的には、提案の増減額を反映したもののだが、本プロジェクト終了後の平成23年度に向けて、JSTの運営費交付金でこのプロジェクトの内容を継承していく体制の前倒しとして、来年度の予算総額11億円のうち本プロジェクトの査定分は8.5億円で、残る2.5億円分がJSTに積まれている。2.5億円分のうち、国立遺伝学研究所の分(0.3億円)をJSTの他の遺伝研のプロジェクトの中に入れて、残る2.2億円分が情報・システム研究機構分となる。

予算関連の手続きとして、本委員会の承認に先立ち、1月7日に松原委員長に配分方針を相談し、1月9日に分担機関、補完課題の方々に業務計画の作成を依頼した。本日、本案をご承認いただければ、21~30日の作業部会分科会で内容を詰めて確定していきたい。ご承認をお願いする。

松原委員長から、予算案の承認について出席委員に意見を求めたが、特に異議がなかったため、本案は承認された。

4. プロジェクト成果公開に関する考え方について

松原委員長から、「プロジェクトの成果公開に関する考え方について説明をお願いしたい」旨の発言があり、資料4に基づき、高木委員から説明がなされた。

このテーマは、今日結論を出すというより、ご意見を伺ったり、次回までに考えていただきたいというものである。

データベースの「統合化」の意味については、平成18年5月の報告書にも書かれており、前回か前々回の委員会で中村桂子委員から3つのフェーズ(第1段階:所在を明らかにし、すべてのものを見通して使えるようにする、第2段階:それらがシームレスに使える、第3段階:それから新しい発見がある)をまとめていただき、プロジェクトでもそれと対応する形(ポータルサイト:どこに何があるか、使えるか、横断検索:それがシームレスに使える[まだ不十分だが]、統合化:新しい発見につながるもの[これから])を考えている。そのためにいろいろなデータベースを集めて我々からまた提供する、あるいは我々自身でコンテンツをつくることもあるが、その際に、本プロジェクトとしてどのように成果公開を考えればいいのか問題になってくる。

プロジェクトのアウトプットの共有が必要だと我々は訴えているが、その際に我々自身でつくったり、加工したデータの共有方針を明確にしないといけない。内部的に議論を重ねてきたが、この段階で皆様にご意見を伺ったほうがいいと考えた。例えばゲノムプロジェクトで言えば、アウトプットは配列断片やアッセムブリ、アプリケーションとしてはアノテーション付きデータがあって、それが医療の応用などのアウトカムにつながっていく。統合データベースの場合は、アウトプットは我々が作った辞書やデータで、アプリケーションとしてはこれらを利用していく検索方法やアルゴリズムで、それが皆さんに広く使えるようになることがアウトカムとなる。プロジェクト一般に関してもアウトプットとアプリケーションとアウトカムが大まかに定義できる。

アプリケーションやアウトカムはプロジェクト開始時には未定だが、アウトプットはプロジェクトの計画時に定義でき、このアウトプットの扱いが社会に対するプロジェクトの価値を左右するものである。一方で、アウトプットは利用条件を明示しないと正式には利用できない。

我々のところでつくった辞書の公開方針(だれにどう渡すか)については幾つかの考え方がある。パブリックドメインにしてだれが使ってもいいとか、引用さえすれば使ってもいいとか、あるいは営利目的は禁止するとか、その背景には営利目的はいけないといった考え方もある。

また、プロジェクトの場合はNEDOとの関係や、省との関係、複雑な契約関係が幾つも絡むケースもある。特に大学等の法人化により、営利の概念が難しくなっている。

これまで我々は科学の引用を原則とする精神で、非営利には自由に使ってもらい、営利目的の場合、使用目的がわからないので相談してもらってきた。今後データベースを受け入れてサービスする場合にも、明確な表示が必要になってくると考え、その意思表示の方法として最近普及しているクリエイティブ・コモンズ(CC)のライセンスを採用しようと考えている。CCのネットワーク上でのデジタル情報の共有に向けた考え方は我々の考えていることとマッチすると考えている。その基本的な考え方は「著作物の自由な利用と共有の仕組みを創造することを目的として、作品の利用を許諾する人と許諾を受けて使う人の間の約束ごと、つまり、利用許諾の条件の雛型であり、著作者が自由に設定することができる」というもの。著者が自由な利用を許可したくとも、現行の著作権法では規制が厳しいので、自由なやりとりを促進する幾つかのライセンスのひな形(クレジット要求、非営利使用可、改変禁止など)を作り、これを我々がつくったデータ等に付け、あるいはデータを受け入れた場合に付与をお願いするという基本精神でやってきたが、既存プロジェクトの参加企業や怪しげなベンチャー企業から申し出があり、このままでほんとうにいいのか、どういう基本精神でいくべきかが問題になってきている。そのために皆さんのご意見をいただきたい。

参考までに、米国では公的資金を用いた基礎データはすべて他の研究者も利用可能な形で納入され、納税者は自由にそのデータを利用することができるが、NCBIなどでやり方はもう少し具体的に調査する必要がある。また、公開されたソフトウェアをできるだけ普及させるためのコピーレフトというCCと近い概念(ソフトウェアの業界ではかなり標準化)もあり、CCの中にもサイエンスの中での特殊な条件を加味したサイエンス・コモンズ(サイエンスの領域での特許や技術移転なども含めて流通させるための活動)があるが、これらについてももう少し調査が必要である。

こういう幾つかの問題があるので、こういう考え方がいいのではないかなど、ご意見をお願いしたい。その意見を踏まえて、次回あたりに具体的なもう少し練った考えを出したいと考えている。

●調査してみて、例えばアカデミックフリーという概念が誤解されていると気がついた。古い固定概念では、営利は悪・富み・私で、非営利は善・貧・公なのだが、実際は営利が納税者・プロジェクトのスポンサー・雇用をつくる人で、非営利が官営で税を消費する人なので、誤解があると思う。また、我々が公開したつもりでも、利用制限のはっきり宣言されていないデータはコンプライアンスの高い企業から見ると危なくて使えないということを認知すべき。だから、CCなどを使って積極的に「だれが何に使ってもいい」と言わないと、怪しげなデータとして使われない。これは本プロジェクトのみが最初に経験できる貴重な体験だと思うので、皆さんと共有しながら合意できる答えを出していきたい。まず本プロジェクトが範となり、他のプロジェクトにまねてもらわないといけない。

●統合化が「シームレスに使えて、新しい発見ができる」という定義だとすると、新しい発見をするためにだれがどのデータを使ってよいかということがはっきりしないと、現実問題として一歩も先に進めない。こういう点をきちんと押さえないと具体的な統合化ができないということで取り組んでいる。逆にこういう点から見ると、統合化とはどういうことかというのは少し見えてくるのではないかな。ぜひこういうことも議論していただきたい。

●営利禁止について、理研などもそうなのだが、最近は任期制で雇用されているところが多く、その職を失ってベンチャーをつくる必要がある場合に、その成果で得た利益や特許を発明者に還元する知財のルールがある場合もある。営利利用を必ず認めないとだめだとすると、発明者がベンチャーや産業を興す際にできる範囲が制限される可能性があるのではないかな。

→おっしゃっているのは、知財がとれる、いわゆるアイデアの部分。共有して欲しいのは、所有権の主張の不可能な、知財の取れないアウトプットで、その扱いを明示する必要があるということである。

●知財の公開時にデータも売れる可能性があるのでは。

→現在はデータそのものを囲い込むことが許されているので、データを誰もが使える状態に比べて発見機会が減り、データの価値を損なっている。それこそが問題の本質。

●公開はするが一定の期間は独占を許すといった条件があるからこそ公開が進むということもある。データを出す段階でなんらかの保証がないと難しい。

→だれがやっても同じ内容のものが出るものがデータであり、契約時にプロジェクト実施期間は提出しなくていいが、終了時には出すという条件で始めたら、それだけでも一定期間独占的にデータを見ることができるわけはかなり有利なのだから、参加を希望する人はいるのではないかな。

●すでにこの件は議論されていて、公的資金で出されたデータは抱え込んではいけないという基本姿

勢はある。しかし、具体的なやり方に関しては一部の科学研究費については決まっているはずだが、それ以外の場合は強要できないということだろうか。理研にはいろいろな経験がはず。

●CCについては、元のデータに付けられている条件を勝手にはずしたり、加工したものを無料で配布するのは困るという声がある。なるべく改変しないという条件ならいいが、ある程度自由な記述を許すときには誰が責任を持つかが不安。CCをそのまま受け入れるのは少し難しい。

→我々もCCを基本にしなが、現実の問題をパターン化したいとは考えている。

●最初は難しいところを厳しくスタートさせ、だんだん緩くするのが同意を得られやすく、とにかく動き出せるのではないか。

●データベースあるいはデータごとにいろいろな事情があるので、CCのこのレベルで統一してやろうというのは、必ずしもうまくいくとは思わない。

→もともとCCの役目は、利用規約や契約条件をわかりやすい典型的なパターンに落とし込むこと。その後、例えば本プロジェクトで引き受けたデータリストにも公開度に応じてグレードを付け、公開度の高いところが評価されればいい。

●先ほどの意見はもう少し統一化したいということでは。

→推奨案は出したい。ただし、あくまで本プロジェクトの範囲での話。

●国の資金で出てきたデータやアウトプット、アウトカムなどはだれのものかという点について皆さんの合意はあるのか。民間企業の立場ではどれも出資者のものであるのは当然だが、研究者間ではアウトプットについても自分のものという人とやっぱり国のものという人と理解がちぐはぐではないか。

→そのちぐはぐなところに問題があるので、解消できるかできないのかを含めて、統合データベースタスクフォースで今後ご議論いただくのだと私は理解している。

●資料4についてももう少しまとめていただけると議論の要点がわかり、非常に意思疎通がよくなる。

●アプリケーションの例として引用があり、そこに「契約必要」とあるが、一々契約しなければいけないとなると非常に大変なので、明示で何とかできないか。同様に営利利用禁止も契約必要とあるのも明示でCCのマークをつければいいので、この「契約必要」という条件を少しご検討いただきたい

●予算申請時に条件付けるのが今後の流れになっていくのか。

→文科省ライフ課のプロジェクトに関しては公募要領に文言を盛り込んでいただいている。ただ、実際にどう担保されるかや契約書への反映方法については、紳士協定的なものにとどまっている。

●知的所有権や公開に関してもこれだけ議論がある。全般に議論すれば何日もかかるだろうが、今日はこのくらいで。

5. 中間評価結果への対応・平成20年度プロジェクト進捗状況について

松原委員長から、「中間評価への対応と20年度プロジェクトの進捗状況について報告をお願いしたい」旨の発言があり、資料5に基づき、高木委員から簡単に説明がなされた。

資料5-1が中間評価指摘された事項で、その右側に関係者や作業部会等で相談した対応策が書かれている。ご意見があれば後ほどでもいいのでいただきたい。中核機関に関して簡単に申し上げると、①風通しのいい運営体制のために作業部会・運営委員会を見直し、②人材育成に関しては、数値目標的に明確にし、③医学、医療関係者の参加については連携を進行中、その他もご指摘事項に関して対応している。これを踏まえて来年度どうするかを資料5-2に記載。基本的には現行の対策をより発展・拡大していく。未公開データは早急に公開という方針で進行中。

●中間評価では国際的に重要な分野にも目配りをして対応しなさいという意見もあった。

●日本で必要なものを取りあえずモデル的に苦労してつくりかけているが、国際的な状況、評価のされ方、国際的な対応についてはいかがか。

→欧米で幾つか本プロジェクトに似た取り組みが始まっている。欧州全体のデータベースを統合しようとか、米国ではそのためのサイバーインフラをつくる動きがある。それらとの連携が確立していないが、EBIとの連携なども具体的な案件の芽は出てきている。

●連携に対応できる人材は育ってきているか。プロジェクトの重要な部分だったはずだが。

→データベースを作る、使う人材は育っているが、国際連携対応ではない。ただ、以前の運営委員会で、国際連携の芽が出てきたときにはサポートするミッションが中核にはあると指摘されており、国際的な動きに加わる人材は個別の案件の中から育っていくと考えている。例えば、データベースを扱うプログラムをモジュール化して標準化を図る国際会議BioHackathon2008を昨年2月に一度東京で開催し、その第2回を今度は3月に沖縄で開く予定。

松原委員長から「現実的な新しい年度に関する審議と本年度の実施内容の概要についての議論をとりあえずこれで終了したい」旨の発言があり、議論を終了した。

6. その他

高木委員から前回の議事要旨(資料6)については委員の変更もあったので、事前に了解を得ている旨の報告があった。

引き続き、6月12日に統合データベースプロジェクト主催のシンポジウム案(参考資料)を検討しており、口頭発表やポスター発表への協力の依頼があった。プロジェクト以外のデータベースに関する各省庁での取り組み等も募集し、紹介する予定。

また、今年度の中核機関の予算計画の変更について、予想よりコンピュータが不足したため予算を変更してコンピュータを購入することにしたという報告があった。

事務局から、各委員からの意見の募集については別途メールでご連絡する旨案内があり、本会を終了した。次回の委員会についても事務局から日程調整する。

以上