

事後評価票（案）

（平成 23 年 3 月現在）

1. 課題名 統合データベースプロジェクト

2. 評価結果

（1）課題の達成状況

【事業の目的】

我が国のライフサイエンス関係のデータベースの利便性の向上を図るため、データベース整備戦略の立案・評価支援、統合化及び利活用のための基盤技術開発、人材育成等を行い、ライフサイエンス関係データベースの統合的活用システムを構築・運用する。

【進捗・達成度】

中間評価までは、分担機関等のコントロールが不十分だった印象が強いが、中間評価以降は、未だ全体が組織的とは言い難いものの組織的構成の改善が見られる。特に中核機関がイニシアチブを発揮して全体を良く掌握し、プロジェクト全体の方向付け、および進捗のチェックのために研究運営委員会、同作業部会が組織されたことによって、全体戦略を浸透させつつ、連携を強化してプロジェクトが遂行される体制が構築され、基本的には重複等がない形で円滑に事業を進めたと考えられる。

また「Open innovation と情報共有」を基本的考え方として、ユーザを増やすための学会などでの啓発活動も継続的に行われており、裾野の広がりなど着実な成果が見られる。

しかし一方で、多くの研究機関が参画して実施している弊害も認められなくはない。同様な機能を持った WWW サービスが構築されたり、出来上がったデータベースやコンテンツが乱立したりしている感がある。ユーザ側に立つと、もう少しサービスをまとめ、画一的なインターフェースで利用できるようにする必要はあるであろう。この問題にはプロジェクトの立ち上げからデータ共有のルールが事前に整備されている必要があること、当初からボトムアップな公募を行ったこと、中核・分担の同時公募などの影響があったことを指摘しておきたい。

（2）成果

【中核機関】

ライフサイエンス統合データベースの実現という、壮大なプロジェクトを実現するに際し、カタログ化／横断検索／受け入れ／統合検索といった現実的な数段階による実現という適切な形にブレークダウンを図り、着実にこれらを実行し、公開という成果に至っている点で、当初目標は達成されていると考えられる。

代表機関である情報・システム研究機構ライフサイエンス統合データベースセンターを中心に、いくつかのデータベースはかなり早い段階から公開されており、有用性と利便性

において評価できる。特に、日本語を用い多数のライフサイエンスのデータベースを検索できる点は非常に有用である。

【分担機関】

京都大学は、ゲノムネット医薬品データベース開発／化合物・医薬品検索ツールの開発という形で医薬品／化合物情報をユーザに提供して行くために必要となる項目をブレイクダウンし、着実に開発を実行しており、所期の成果は得られている。情報解析ツールの開発、化合物・医薬品データベースのための検索システム開発も実施された。一般ユーザの利用が増えた点も評価できる。

東京医科歯科大学および大阪大学は、オントロジー、ターミノロジー、シソーラスの整備と言う点では計画の実施が不十分であるものの、疾患データベースの統合化手法（統合医科学データベース）を開発し、散在する特定の疾患に関する臨床情報などの集約と統合化のモデルが構築された。分子から疾患まで、発散しがちなところをうまくまとめており、疾患データベースという構築困難な物に対するロールモデルとしては非常に有意義な成果である。

東京大学を代表機関とする「疾患解析から医療応用を実現する DB 開発」では、疾患遺伝子の同定における GWAS の重要性を鑑みると、GWAS データを受け入れたデータベース構築という明確な目的のもと着実に成果を挙げている点では評価に値する。ハード的な不足を抱えながら中核機関との連携で当初目的を達成した点は高く評価できる。また、関連学会との連携がとられた点も重要である。

【補完課題実施機関】

理化学研究所は、保有する大量の未公開データを公開したという点では、大いに評価される。機械的なデータ保全・公開の仕組みについては、十分な取り組みがなされ、対象となる3つ（シロイナズナ研究、高等動物由来蛋白質立体構造研究、微生物由来蛋白質立体構造研究）の研究分野においては有用なデータが公開された。特に、標準的オントロジーにより、統一された注釈と構造を持つデータが、植物のみならず哺乳類オミックスデータにも取り入れられ提供されている点は大いに評価できる。

産業技術総合研究所糖鎖医工学研究センターは、「我が国の糖鎖関連データの統合データベースの構築」という目的では、短期間で糖鎖工学の根幹をなす5つの各論的データベースの構築を果たしたことは高く評価される。中核機関との役割分担を明確化にしつつ、糖鎖関連の研究者コミュニティと連携を取りながら省庁の壁を越えた形でデータベースの統合化が進んでおり、本課題の進捗状況及び得られた成果は大変優れていると評価される。

国立遺伝学研究所は、新型シーケンサーの登場に対応した「我が国での配列決定における Trace データの保存と有効利用」という目的に対し、新旧の異なるタイプのデータ保管に対応したデータベースの構築と、データの蓄積、ライフサイエンス統合データベースセンターへのデータ提供が行われており、補完課題として期待された成果は得られたと考えられる。

九州工業大学は、蛋白質の熱力学データと構造データ、蛋白質と核酸の相互作用のデータベース構築、新規情報の収集を行い、「小規模研究室でのユニークなデータベースの構築」という点では、補完課題として期待された成果をあげたと考えられる。

(3) 今後の展望

今回構築したシステムやデータベースの整理については、アップデートし続けることが重要であるため、今後もその維持管理を継続する必要がある、それができる仕組み・体制を構築しなければならない。

また日々、進化するデータ解析手法、解析データの種類や算出量の増加など、今後のデータベース統合は、今回の取り組みがそのまま適用できるとは限らず、時として柔軟にその考え方を考える必要性もあり、今後もこのような取り組みは、継続して、その方法論の開発や具体的なデータベースのカタログ化が進むことを期待する。この際に限られた資金を有効に活用するために、機関の有機的な連携と中核機関の主導性が必要と考えられる。

一方、現在の取り組みのように研究成果として得られたデータを細大漏らさず保存するという考えには一理あるものの、現実には有効なデータは刻々変化する事も事実であり、今後はどのデータを優先的に保存するか、あるいは一次データから抽出した有用データのみを保存するかという考え方の導入も必要である。

ユーザ側の視点とニーズをどのように反映するかという視点が、まだ欠けているように感じられる面もあるので、この点も留意すべきである。

もし今後、後継プロジェクトも同様の体制で進めるのであれば、中核機関決定後に中核機関の意見が反映された形での分担機関の公募が望ましく、また、中核機関が分担機関を評価できる様な仕組みがあってもよいのではないだろうか。

広報活動については、学会への出展等努力を重ねた点は評価できるが、未だ周知が不足しており、地方での講習会等の活動をもっと増やしてもよいのではないかという意見があったことを付け加えておく。

(4) その他特記事項

【今後の課題】

以下はプロジェクト全体の評価であると同時に、このプロジェクトの事後評価から浮き彫りになった国家プロジェクト事業の進め方に関する課題として参考にしていただきたい。

1 公募制度、ミッション等の問題

- ・ゲノム研究を始めとした所謂オミックス研究者だけではなく、個別の研究者に対してどのように「統合された」データベースを提供したいのかという、クリアなビジョンが当初の段階から可視化されていなかった。戦略を策定し、最初に中核、次に分担機関という決定プロセスで与えられた年限で中核のイニシアチブが十分に発揮できるような体制が必要と考えられる。
- ・優秀な人材の確保や、予算の柔軟な執行、全体に投じられた予算に対して中核機関への配分が少なくならざるを得ないなど、「中核と分担を同時に公募した時限プロジェクトの弊害」が多く認められた。事後評価の段階で、以後はデータベース構築業務以外の職種

に異動している人材が多く見受けられた。

- ・ワークフローの構築や、教育など必要と感じられることに取り組んでいる姿勢大いに評価できる。各機関で取り組んでいる課題があまりにも diversity が大きいいため、参画機関が取り組む課題に関しても、最初の段階で中核機関が全体を見渡した強いイニシアチブの下、方向付けを可能とする施策が必要であった。
- ・本プロジェクトのような課題は、年度単位などの期初目標にとらわれることなく、柔軟にその場で必要とされている問題に継続的に取り組んで行くこと、さらには予算の有効利用のためにも隅々までトップの意向が伝わるような一極集中に近いような体制を組むことが可能となるように、予算／体制を整備すべきである。そうでなければ、せっかくの本プロジェクトの成果も時とともに色褪せてしまうことになりかねない。

2 統合データベースの在り方

- ・分子レベルのデータと疾患を含めた個体の表現型レベルのデータを統合するという姿勢が不足している。これは我が国の研究者の特徴として、得意分野における知識は豊富であるが、分野の枠を超えた広範囲の知識と考え方を重視する姿勢が少ない事にも関連していると考えられ、今後の統合データベースの在り方の上で考慮すべき視点である。確かにゲノムワイド関連解析やリシーケンシングのデータなどを取り扱うグループが加えられているが、それがその他の分子データを扱うグループとどのように統合されるか、等「研究階層を超えた連携と統合的理解」をできる人材の育成が今後の問題である。今後、得られるすべてのデータのデータベースを作成し、検索の対象にするという姿勢は不可能になる可能性が高い。統合データベースの在り方を極めて広い立場から展望し、適宜方針を変える姿勢が今後必要である。
- ・計算機環境の整備もこの種の推進には不可欠なことで、その意味でも、今後の維持管理、あるいは発展が期待される。ただし、「統合」がすべてではなく、膨大なデータを無理に統合することで、かえってデータの利用が阻害される可能性もあると思われる。今後は必要な分野の選定を行って、統合化を進めて行くべきと考える。
- ・国内データベースの統合はこれで終わりではなく、継続的に推進されることを期待する。データベースは今後爆発的とは言わないまでも加速度的に件数とデータ容量、多様性が増加すると予想され、決して5年前の当初目標達成に甘んじるわけにはいかない。世界情勢の分析、法整備等々、多面的な早急な取り組みが必要である。特に、今後構築されるデータベースに対し標準化のプロトタイプ提示などが遅ればあとの修正は困難を極めると予想されるので、今後は無駄や過去の執着を省き、必要な整備を精力的に進めるべきである。
- ・本プロジェクトで得られた成果には世界をリードできるユニークなデータベースとして今後も発展できるものが少なくない。それらを最大限に活用し、今後も事業を推進していくための留意点として、「データの提供者が自由に参加できる」「開発を推進する者がメジャーユーザである」「継続的な開発と改良」「プロトタイプの開発から公開までの迅速性」などが評価委員会の議論で挙げられたことを付記したい。

【プロジェクトマネジメント関係】

- ・欧米でのプロジェクトのように中核機関による強力な指導力の発揮は見られなかったが、我が国でのプロジェクトとしては中核機関による全体のマネジメント、更には分担機関、補完課題担当組織についてもゆるいながらも組織的なプロジェクト遂行が認められる点で評価できる。
- ・プロジェクト開始時は中核機関のイニシアチブ、および各機関の連携に不十分な部分もあったが、中間評価後には中核機関が全体を良く掌握し、適切な方向修正も行われた。中核・分担・参画機関の集中的な共同作業によって円滑に事業が進められ、ほぼ当初目的に相当する成果が得られた。比較的短期間の内に統合データベース推進事業で継承し得る成果も上がっている。

統合データベースプロジェクト

予算総額：5,042百万円
 平成22年度予算額：1,009百万円
 平成21年度予算額：1,099百万円
 平成20年度予算額：1,100百万円
 平成19年度予算額：1,584百万円
 平成18年度予算額：250百万円

概要

【目的】

○公的資金による研究成果として作成されたライフサイエンス分野のデータベースの利便性向上を図り、世界トップレベルの研究力を支える基盤を整備する。

【主な成果】

○散在しているデータベースの所在情報等の整理・分類を実施。(912件(国内620件、国外292件))
 ○異なるデータベース同士の横断検索機能の付加。(270件(国内150件、国外120件))

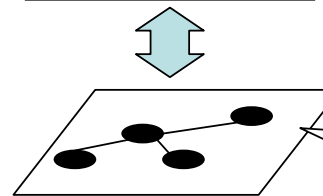
今後の計画

○平成22年度末までに本事業とJSTバイオインフォマティクス推進センター事業の一本化を行う。

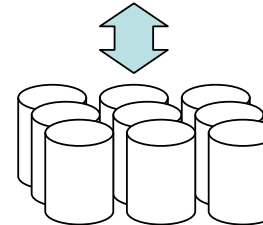
統合データベースとは？ ライフサイエンス関連データベースの統合的活用システム



人体、臓器、細胞モデルを画像化し直感的に分かりやすい、知識情報として整理



データの意味付けやネットワークの構築による、新たな情報を創出。



実験などで得られた生データを統一して運用するためのデータ変換、注釈情報の付加、データ変換のための辞書開発。

