

研究開発課題別事後評価結果

1. 研究開発課題名

生命動態システム科学のデータベースの統合化

2. 代表研究者名

理化学研究所 生命システム研究センター チームリーダー 大浪 修一

3. 研究実施概要

本研究開発課題は、新興の生命動態システム科学における、我が国の全てのデータベースを統合する体制、仕組みを先行的に構築し、生命動態システム科学を推進する国内の中心的な研究拠点の一つである理化学研究所 生命システム研究センター(理研 QBiC)をはじめとした有力研究室によって取得されたデータを統合することを目的として実施した。

1) 生命動態システム科学研究におけるデータベース統合の体制と仕組みの構築

1a) プロジェクト・グループの構築

当該分野の研究者コミュニティの合意のもとで進めるため、我が国の当分野の中核的な研究者により構成される「プロジェクト・グループ」を組織し、研究課題の遂行状況や将来計画等について、プロジェクト・グループ委員と頻繁に意見交換を行った(全体会議 1 回、個別の打合せ 18 回)。プロジェクト・グループの委員には、理研 QBiC の PI、JST 戦略的創造研究推進事業 CREST「生命動態の理解と制御のための基盤技術の創出」研究領域(CREST「生命動態」)、さきがけ「細胞機能の構成的な理解と制御」研究領域(さきがけ「細胞構成」)の研究代表者、研究総括の他、本課題開始後に設置された文部科学省「生命動態システム科学推進拠点」(「推進拠点」)に参画する研究者からも委嘱した。

1b) データベース統合の仕組みの構築

生命現象の動態の時空間定量データのほぼ全てを記述できるデータ形式として BDML (Biological Dynamics Markup Language)を開発した。開発に際して、理研 QBiC の研究室全 21 室への聞き取り調査を実施し、生命現象の動態に関する時空間定量データが 14 のタイプに分類される事が分かり、BDML の改良の結果、うち 12 件までを記述できるまでとなった。

また、時空間情報を持つオミクス・データ(例えば特定の組織で経時的に変化する遺伝子発現量)を記述するデータ形式として、OmicsBDML を開発した。既存のオミクス・データ用の形式と異なり、時空間情報を定量的に表現できるものとなっており、既存の形式から変換することも可能である。

さらに BDML あるいは OmicsBDML で記述されたデータと、その計測に使用した動画データ、および BDML/OmicsBDML に対応したデータ活用ツール群を管理・公開するための統

合データベース SSBD (Systems Science of Biological Dynamics; <http://ssbd.qbic.riken.jp>)を開発し、2013年9月に公開開始した。

2) 理研 QBiC 内のデータの統合

大浪研究室で収集を続けている *C. elegans* 初期胚の細胞分裂ダイナミクスの4次元計測データ1,887件をSSBDに統合し、うち286件を公開した。論文投稿中の未発表データは論文公開後すみやかに公開予定である。また、理研 QBiC の全研究室を調査し、論文公開済みの全データ11件をSSBDに統合し、うち9件を公開した。さらに、論文発表準備中の未発表データもSSBDへ統合した。これらは論文発表後すみやかに公開する予定である。

3) 理研 QBiC 外のデータの統合

CREST「生命動態」の平成24年度採択の5課題、および「推進拠点」の4課題に聞き取り調査を行い、論文公開済みのデータ2件をSSBDに統合し、うち1件を公開した。また、今後公開される論文のデータについては、SSBDに統合して公開する内諾を得ている。

また、SSBDの取組みを国外研究者に紹介し賛意を受けたことから、将来の生命動態システム科学のデータベースの全世界的な統合に向けての基盤作りを目的として、当分野における国外の主要なデータベースの統合を、当初研究計画に追加して実施した。当分野を代表する主要なデータベースであるかどうか等の基準に基づいて6つのデータベース(データ数296件)を選択し、各データ生産者との議論を経てSSBDへ統合した。課題終了時点で4つのデータベースのデータ22件を公開し、残り2つのデータベースは公開に向けて調整中である。

4) 統合データベースを活用するツール群の開発

当初研究計画に追加して実施した。

SSBDの登録データを直接利用するクライアント・アプリケーションを、ユーザー自身が多様な環境下で開発できるようにするため、RESTful APIを設計し実装した。

また、ユーザーのプログラミング能力如何に関わらずSSBDを活用できるようにするため、BDMLで記述された定量データを生命科学分野で最も使用されている画像解析ソフトウェアImageJ上で可視化するためのプラグインBDML4DViewerと、表現型の特徴を抽出するためのツールphenocharを開発し、ソースコードも含めて公開した。

さらに、SSBDに登録されたデータを、ImageJに直接読み込むためのプラグインSSBD-OMERO.insight-ijを開発し公開した。

4. 事後評価結果

4-1. 当初計画の達成度

当初の研究計画を概ね達成した。

生命動態システム科学の統合データベースが構築、公開され、当該分野の国内のデータの多くが SSBD に統合された。当初の研究開発計画通りに BDML の開発や SSBD の公開が行われただけでなく、SSBD を活用するツール群の開発にまで研究開発を広げた点も評価でき、これら SSBD や活用ツール群は、「生命動態システム科学」の基盤として、有効に使われることが期待できる。一方、利便性の向上、データ登録数や公開数、アクセス数などには課題が残された。

今後、SSBD が、生命動態システム科学分野のみならず、ライフサイエンス分野全体における時空間データのデータベースとして一般化することが望まれる。

4-2. 研究開発成果の公開および利用の状況等

SSBD および、ImageJ 用プラグイン等の活用ツール群が開発、公開された事は評価できる。SSBD へのデータ統合にあたっては、データ生産者と慎重な議論を重ねながら進めており、このことは、本研究開発課題の国際的な認知度を高め、今後国際標準を定めるうえでも有効であろうと思われる。一方、SSBD へのデータ登録数や公開数、アクセス数などに課題が残された。

4-3. 研究開発成果によるライフサイエンス分野のデータ活用への波及効果

我が国の生命動態システム科学の代表的研究者を結ぶ、ハブ的な機能を果たした点は評価できる。しかし、研究開発終了時点では、ライフサイエンス分野全体への目に見える影響は明確ではない。

今後、SSBD の利用をライフサイエンス分野全体に広めるため、積極的な働きかけが必要である。

4-4. 広報・アウトリーチ活動等

生命動態システム科学の研究者との打合せは充分行われた。

今後、ライフサイエンス分野全体のデータベースとして普及させることにも注力し、生命動態システム科学分野外の研究者へ向けた講演会の開催などの積極的な広報活動にも取り組むことが期待される。

4-5. 総合評価

これから研究データの数も重要性も増すと予想される、生命動態システム科学のデータを、分野の成熟に先行して取り組んだ研究開発課題である。概ね研究開発計画に沿った研究開発が実施され BDML 開発、SSBD 公開がなされた。加えて、活用ツール群も公開された。国外との連携も行われ、生命動態システム科学における基盤データベースが構築されたといえる。

終了時点で、研究データの登録数、公開数、アクセス数が少ないという状況を踏まえ、今後、生命動態システム科学分野以外の研究者との連携を進め、分野を超えた統合化や標準化などへ取り組むことや、SSBD の国際標準化に向け、さらに国際連携を強化することが期待される。

以上