

Biosample Database に対応したオープンソースのプロテオーム用 研究情報管理システム改良の取り組み

森澤拓[†]、岩本真知子[†]、津元裕樹[†]、廣田三佳子^{††}、三浦ゆり[†]、戸田年総^{†††}

東京都健康長寿医療センター研究所 老化機構研究チーム[†]、

帝京平成大学 薬学部^{††}、横浜市立大学 先端医科学研究センター^{†††}

プロテオーム研究においては、加速度的に研究情報が增大する中、研究情報の交換、データマイニングの基盤のために、さらには研究の真正性、信頼性、完全性の確保のために研究情報管理システムを活用することは重要である。特に複数の研究者や研究グループが共同でプロテオーム研究を行う際、バイオサンプルに関する文字情報の取り扱い、系統的かつ確実にを行うことが重要になっている。

European Bioinformatics Institute (EBI)、National Center for Biotechnology Information (NCBI)、DNA Data Bank of Japan (DDBJ) 等は、バイオサンプル情報の取り扱いについて、サンプルデータの供給者や疾患名等を用いて、構築している公開データベースの横断的な検索を可能にし、データの透明性の確保するために、公開データベースとして BioSample database の開発、運用の開始を行っている¹⁾²⁾。研究プロジェクト内等で利用される研究情報管理システムにおいても、研究データの第三者による確認をシステムティックに行うために、バイオサンプル情報を多面的に検討できるよう改良することは、とても重要である。本研究では、そのような公開データベース上のバイオサンプル情報の取り扱いを参考に、開発を行っているオープンソースのプロテオーム用研究情報管理システム³⁾におけるバイオサンプル情報の取り扱いを整備し、可能なかぎり EBI, NCBI, DDBJ との整合性も視野に入れ改良を行ったので報告する。本システムでは、セキュリティレベルを維持するとともに、利便性の向上のために、データ入力や取り扱いが煩雑にならないよう検討することも重要である。そのような検討から、本改良において、あわせてユーザーグループの導入等を行いデータの取り扱いについて煩雑にならないよう開発を行った。

Development of the application software for BioSample database in the laboratory information management system for proteomics

[†]Hiraku Morisawa, Machiko Iwamoto, Hiroki Tsumoto, Yuri Miura. Tokyo Metropolitan Institute of Gerontology

^{††}Mikako Hirota. Teikyo Heisei University,

^{†††}Tosifusa Toda. Yokohama City University

材料と方法

オープンソースのプロテオーム用研究情報管理システムは、Web 形式のユーザビリティを実現するため、PostgreSQL-PHP-Apache システムとして開発を行なっている。本改良は、Postgresql 9.2.4、PHP 5.5.3、Apache 2.4.6 を用いて Fedora version 19 上で行った。暗号化は、PHP スクリプトの mcrypt ライブラリを用いた。グラフィック機能をサポートするために Javascript も活用した。本研究情報管理システムにおける、バイオサンプル情報は、サンプルヘッダー情報、サンプル投稿情報、サンプル詳細情報の 3 種類で構成した。サンプル投稿情報、及びサンプル詳細情報は、公開データベースの BioSample database との整合性を視野に入れテーブル項目を作成した。サンプル詳細情報は、東京都健康長寿医療センターで利用していた形式(TMIG original)の他に、12 のテーブルを新たに作成した。(図.1)。サンプルヘッダー情報には、Submission ID、Format ID、Sample ID が付帯情報として含まれ、サンプル投稿情報、サンプル詳細情報にリンクする。サンプルヘッダー情報を各実験手順とリンクし、サンプル情報を実験手順とあわせ管理できるようにした。また、キーワード検索によりバイオサンプル情報に登録している文字情報を検索できるようアプリケーションを作成した。

結果

バイオサンプル情報の取り扱いの改良により、研究情報管理システムにおいて、多様なバイオサンプルに適した形式でデータ管理が行えるようになった。しかし、入力画面数は増加し、入力項目数も増加した。図 2 は、研究情報管理システムにログインした画面とバイオサンプル情報の一覧表示画面である。サンプル情報のグループ名やユーザー名と共にサンプルヘッダー情報を一覧できる。さらに出力ボタンを用意した。そしてサンプルヘッダー情報とサンプル詳細情報の更新画面と閲覧画面にリンクしている。図 3 は、サンプル詳細情報の更新画面と閲覧画面である。ユーザーグループの導入により、バイ

オサンプル情報及び実験プロトコルの入力、電気泳動画像解析、ゲル内消化、質量分析の実験手順において、追加修正モード以外に閲覧モード画面を作成した。バイオサンプル情報の出力については、実際に公開データベースの投稿に有用であるか検討する必要があるが、本改良では検討できなかった。

考察

本改良によって研究情報管理システムにおける、バイオサンプル文字情報を用いた研究データの検索が容易になり、研究データの検討を促進することができるようになったと思われる。しかし、画像解析データ、質量分析データ等の実験手順においては、機器分析より出力されるファイルを変換するアプリケーションの作成などにより、入力作業の自動化を促進することができるが、バイオサンプル情報については、データ入力作業を簡便にするためにさらなる検討、改良が必要と思われた。バイオサンプル情報は、生命科学研究における情報の透明性を高めていくことは、今後さらに重要になっていくと思われる。研究環境における情報管理の必要性は、今後も増大していくと考えられ、オープンソースによる研究情報管理システムをサポートする情報ツールの開発と合わせ、システム改良を検討していきたい。尚、改良したソフトウェアツールは、インターネットを介して公開していきたい。

参考文献

- 1) Barrett T, Clark K, Gevorgyan R, Gorelenkov V, Gribov E, Karsch-Mizrachi I, Kimelman M, Pruitt KD, Resenchuk S, Tatusova T, Yaschenko E, Ostell J. : BioProject and BioSample databases at NCBI: facilitating capture and organization of metadata. *Nucleic Acids Res.* 2012 Jan;40(Database issue):D57-63.
- 2) Faulconbridge A, Burdett T, Brandizi M, Gostev M, Pereira R, Vasant D, Sarkans U, Brazma A, Parkinson H. : Updates to BioSamples database at European Bioinformatics Institute. *Nucleic Acids Res.* 2014 Jan;42(Database issue):D50-2.
- 3) Morisawa H, Hirota M, Toda T. : Development of an open source laboratory information management system for 2-D gel electrophoresis-based proteomics workflow. *BMC Bioinformatics.* 2006 Oct 4;7:430

