NBDC 共同研究

課題名：「統合データベースにおける基盤技術開発とデータベース運用に係る共同研究」（研究代表者　小原　雄治）

DBCLS におけるデータベース RDF 化への取組みと今後の展望

山口　敦子

（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）

1. はじめに

生命科学研究においては、分野の細分化や計測機器の発展に伴い、多種多様かつ膨大なデータが産出される。これらのデータを集約し公開するデータベースは、生物学研究者に必要不可欠なインフラとして利用されている。しかしながら、従来、プロジェクトや目的ごとなど個別にデータベースが作成、公開されてきたため、データベースの所在や利用方法がわかりにくいなど多くの問題を抱えていた。

DBCLS では、生命科学の多種多様かつ膨大なデータを統合的に扱うために、2011 年 4 月から三年間行われた基盤技術開発プログラムを通じて、RDF などのセマンティックウェブ技術に基づいたデータの記述や公開を推進してきた。その結果、セマンティックウェブ技術を用いたデータベース統合について、一定の成果を得ることができ、技術のポテンシャルの高さを示すことができた。その一方で、これまで DBCLS や国内他機関で作成されてきた RDF 化された DB を基に、セマンティックウェブ技術に基づいた本格的なサービス運用を目指すためには、解決すべき問題が数多く残されている。

これらの問題を解決すべく、DBCLS は 2014 年 4 月より NBDC との共同研究を開始した。本発表では、この共同研究内容、なかでもデータベースの RDF 化について着目し、計画内容および今後の展望について述べる。

2. 計画概要

NBDC との共同研究は、以下の４つの柱から構成される：(1) RDF 統合化のための基盤技術開発、(2) 統合化支援と分散統合 DB 環境の実現、(3) エンドユーザ向け DB 統合利用環境の構築、(4)既存・新規の DB と各種サービスの効率的な運用と拡張。

(1)では、RDF データに対して、一カ所に集めて管理することが難しいデータに対する分散環境構築技術、個人情報などを含むデータに対するセキュリティ管理技術などの研究開発を行う。そして(1)の技術を基盤として、(2)では、統合化推進プログラムで開発された DB の RDF 化を支援する。さらにこれらの DB の RDF を用いた統合化を進めるために、国内外の機関と連携して国際的な標準化を進める。また、セマンティックウェブ技術を扱う敷居の高さを補い、あるいは NGS データなど喫緊の課題に対応するために、エンドユーザ向けのサービス運用および技術開発を(3)において行う。さらに、(4)では、実際にセマンティックウェブ技術に基づいてサービスを運用する際に問題になる安定性やバージョン管理方法などについて技術開発を行う。

これら(1)-(4)の技術開発により、DB 作成者・管理者、エンドユーザなど様々な立場のユーザから安定的かつ簡便にセマンティックウェブ技術を利用する道筋をつけることが可能となる。

3. おわりに

共同研究により開発される技術やサービスを組み合わせることで、DB のより高度な利用が期待できる。共同研究終了時には、さまざまな技術の提供はもちろん、研究開発成果を用いた具体的な高度な利用例を示すことができることを目指している。