ライフサイエンスデータベース統合推進事業
統合化推進プログラム(平成 26 年度採択)
課題名「生命と環境のフェノーム統合データベース」(研究代表者　桝屋　啓志)

表現型情報で生物をつなぐフェノーム統合データベース

桝屋　啓志
(理化学研究所バイオリソースセンター)

生命原理の理解とその応用を加速するには、様々な生物で得られた研究成果情報を、種の違いを踏まえながら統合する必要がある。従来、生物種横断的な情報統合は分子レベルでは広く行われているが、より高度な情報である表現型(フェノタイプ)に関しては、標準化や統合が困難なために公開されても各生物種の研究コミュニティー内で限定的にしか用いられず、新たなイノベーションの障壁となっている。

本課題では、第 1 期課題「生命と環境のフェノーム統合データベース」で開発された表現型情報統合の技術をさらに高度化するとともに、収集・公開するデータの質的・量的な拡充を行う。遺伝因子と環境因子の相互作用によって現れる表現型、およびそこから紐づく遺伝子や系統、疾患、環境などの情報を、モデル動物(マウス、ラット、ゼブラフィッシュ、メダカ)、ゲノム編集研究、ゲノムコホート研究など、幅広い研究コミュニティーから収集し、研究分野の垣根を超えて横断的に利活用できるように標準化した上で、NBDC の他の情報とも連結し、全データをオープンに利用できるかたちで公開する。モデル生物とヒト疾患との関係や、環境への応答など、これまで利用者側で結び付けにくかった異種データ間の新たな結合が容易に可能なるよう提供する。

今年度はデータの共通利用性をさらに向上させるためのデータ基盤の再整備を行なっている。フェデレーション型のデータ統合を実現する Resource Description Framework

(RDF)技術を基盤として、表現型データおよび、表現型とつながる生物系統や株、遺伝子とそのバリエーション（アレル）等の概念を、生命科学の「共通語彙」として上位オントロジーに基づいて定義し、世界的動向を踏まえながら、統合化推進プログラムの各プロジェクトで共通利用できるよう整備を進めている。