

【目的】生命科学分野では、研究成果のデータを適切にオープンにすることが求められ、FAIR原則が提唱されている。生物学データの再利用性を担保するために、異種のデータを統合して共有・活用できる形に整理する必要があるが、その作業には多大な労力が必要である。一方、整理されたデータから、新たな解析目的に応じて柔軟に部分データを再構築できれば、研究データの再利用性を超えた利活用が期待できる。本研究では、TargetMine データウェアハウスの開発経験を生かして、生命科学分野の研究データ再利用性担保及び学習データ自動生成プラットフォームを構築することを目的とした。

【方法】データ（ファイル）とデータを照会するために使用するツール（ソフトウェア）の両方を共有できるプラットフォームを設計した。データベースを使用して共有するすべてのデータを格納し、データ提供者と照会者両方の要求に対して柔軟に対応可能なユーザーインターフェースを持つシステムを開発した。プラットフォーム内の各要素をモジュールとして構成し、それぞれの機能を実装した。

【結果】実装したプラットフォームは、1) 提供者用アプリケーション、2) 照会者用アプリケーション、3) 共有データベース、4) これらを接続するインターフェース、の4つの独立したモジュールから成る。提供者用アプリケーションにより、どのような種類の生物学データを共有するかをユーザーが自由に規定し、データをアップロードできる機能を実装した。照会者用アプリケーションにより、必要な部分データを柔軟に取得することを可能にした。アミロイド前駆体タンパク質の産生メカニズム解明を目的とした網羅的遺伝子ノックダウンとタンパク質間相互作用データとを組み合わせたデータセットデータの共有を例として、これらの機能を実証した。将来的には、本プラットフォームを通して、データ提供者が Docker イメージ作成などの汎用的な方法でデータを公開できる機能の付加を予定している。

実装したプラットフォームのコンセプトとスクリーンショット

