

BioLOD.org: 動物・植物の枠組みを超えた 新たなフェノーム統合データベースの開発

○土井考爾(理研BASE)、梶屋啓志(理研BRC)、豊田哲郎(理研BASE)

表現型(フェノタイプ)は、生物が環境の中におかれた際に観測される特性である。フェノームはフェノタイプの集合であり、現在の生物科学コミュニティが最も興味を持っている研究対象の一つである。我々はこれまでの事業で培ったノウハウを生かし、マウス・シロイヌナズナなどモデル生物のフェノーム情報を種の壁を越えて収集し、さらに文献データ・分子データとの統合を視野に入れた新しいフェノーム統合データベースの構築に着手した。本データベースでは、これまでばらばらに蓄積されてきた表現型情報を一元化し、Linked Open Data (LOD)と呼ばれる形式で提供する。これにより、異なる発生源のデータを隔離せずに結びつけて解析でき、新しい発見をより容易に導くことが期待できる。

BioLOD.org (Biological Linked Open Databases for Phenomes) からデータ提供を開始

BioLOD.org (フェノーム統合データベース)

我々は新たにBioLODと呼ぶサービスを開始し、様々なフェノームデータを、DBpediaやBio2RDFに代表されるLOD(Linked Open Data)のデータ提供方法に準拠した形式で提供している。LODでは、すべての事柄をURIで参照可能とし、データの外部からの参照や関連付けが容易という特徴がある。BioLODは、研究者に馴染みの深いテーブル形式でのデータ提供にも対応する。

理研サイネスとの連携

理研の研究基盤である「理研サイネス(SciNetS)」は、データベース構築・連携・公開化の共通基盤システムである。多様かつ大量のデータベースを単一のプラットフォームのもとに効率よく集積・公開することができる。

また、セマンティックリンクによるデータの繋がりに対応した推論検索、アクセス権を考慮した統合検索などの機能を備え、検索結果をキーワードとの関連性を統計値で順位付けして提供できる。

統合と共有化の実際

データ統合化の一例としてシロイヌナズナの表現型情報の一つをBioLOD, SciNetSで見てみる(図1-4)。

biolod.orgのトップページからDatabase Listを参照すると、PlantPhenomeデータベースのページが見出される(図1)。ここから全データがダウンロードできる。また、このデータベースにはOntology Directoryとして4種類のデータセットが含まれていることもわかる。そのうちの一つをクリックすると、その詳細が表示される(図2)。このデータセットに含まれる個々の表現型のさらに詳細な情報についてはscinets.orgから閲覧することもできる(図3・4)。

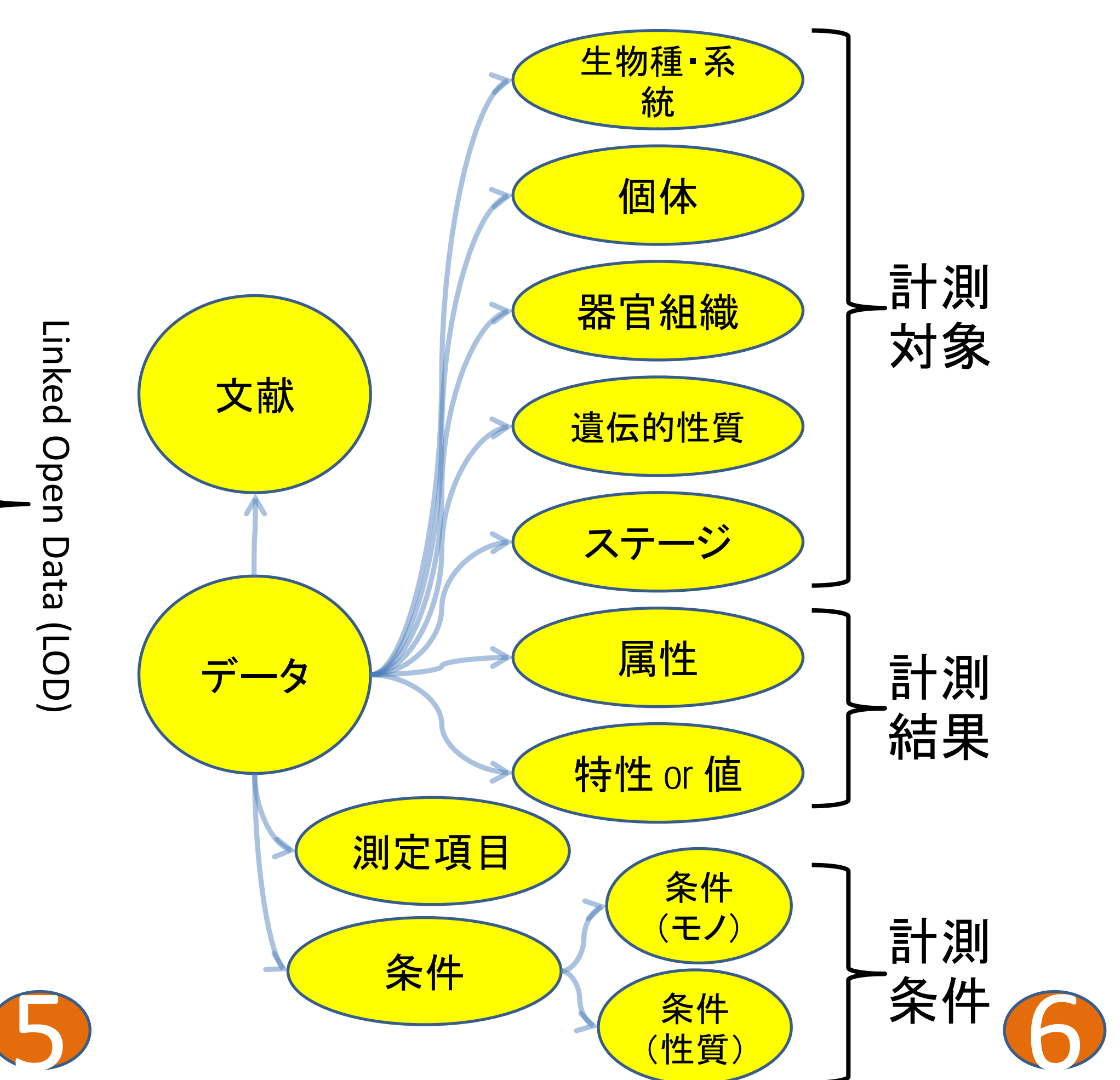
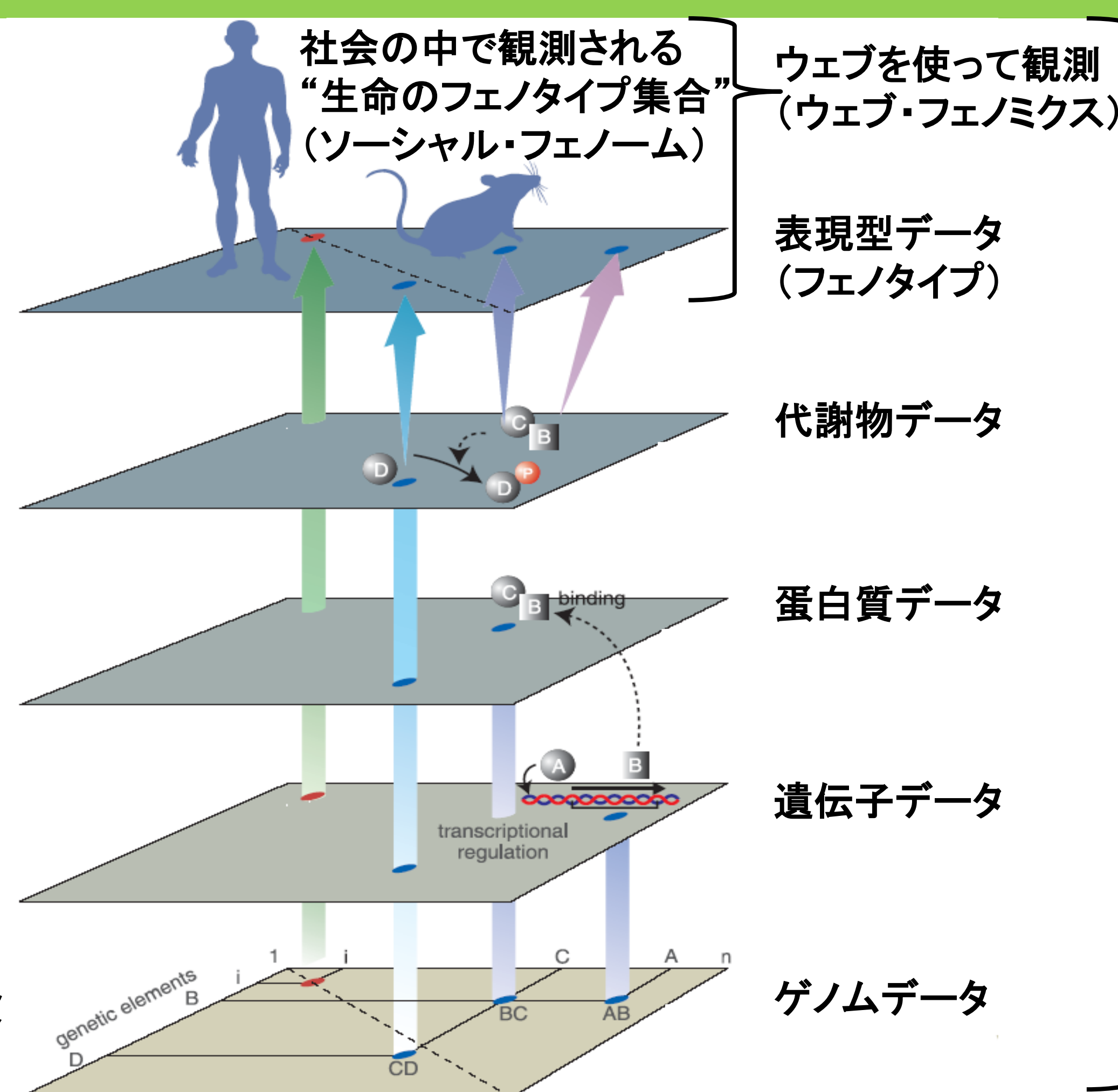
Figure 1-4 consists of four numbered screenshots of the BioLOD.org website.
 1. The 'Database List' page, showing a table of various databases including PlantPhenome.
 2. The 'Ontology Directory' page, showing details for the 'Abnormal in biological quality in response to light stimulus under light regimen' ontology.
 3. A detailed view of an individual phenotype, showing its description and associated data.
 4. A detailed view of a phenotype, showing its description and associated data, including a link to PubMed literature information.

フェノーム情報の統合

各種オミックス情報の中でも、近年は特にフェノームが注目され、積極的に活用して詳細な研究がなされるようになってきている。

フェノームとは表現型(フェノタイプ)の総体であり、表現型とは、遺伝子型と環境要因(実験条件)によって決定される細胞または個体の性質である。表現型の観測値は、遺伝的要因や環境要因に加えて、観測に使用する計測手段の要因、時間的要因、それ以外の生物サンプル固有の要因などに依存して変動する。

また、種を超えて表現型間の関係性を明らかにし、さらにゲノム・分子データと関連付けることで新しい生物情報学的な研究の発展が期待できる(図5)。



このような事情を考慮し、生物種によらず、個体ごとに個々の表現型を標準化する方法を確立したうえで、あらゆるフェノームを統合したデータベースを新たに構築する必要があると考えた。現在、マウスとシロイヌナズナの表現型情報をモデルケースとして、図6に示したように表現型の関連要素を整理し、具体的なフェノームの表現方法について検討を重ねている。これに基づいてデータベース構築を進め、BioLODを通じて公開する予定である。