

○高月照江¹⁾, 白田大輝²⁾, 櫛田達矢²⁾, 川本祥子^{1) 3)}, 柁屋啓志²⁾, 川島秀一¹⁾

¹⁾ 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, ²⁾ 理化学研究所 バイオリソース研究センター, ³⁾ 国立遺伝学研究所

● 目的

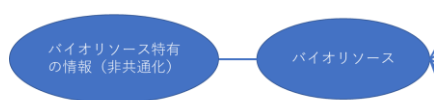
生物資源センター (Biological Resource Centers, BRCs) は、研究用の生物材料を保存・提供する機関であり、維持している生物資源コレクションの様々なメタ情報を整備している。これらのメタ情報には、生物種名、寄託者、配布の可否等の、BRCsや生物種によらず一般的に記載される内容や、表現型情報、繁殖・維持方法、生育環境情報など、特定のコレクションや生物種固有の内容等、有用な情報が多数含まれる。一方で、これらは、コレクション毎に異なる項目や語彙、フォーマットで記載されており、例えばコレクションをまたいだ統合検索を行う為には、事前に語彙やフォーマットの共通化など、多くの労力を必要とする。我々は、生物資源のメタ情報を、共通語彙を用いてRDF化することで、コレクション間のより高い相互運用性が実現できると考え、特に生物資源一般に必要な項目について、RDFとして記述するためのモデルを設計し、そのために必要となるオントロジーであるBRSO: Biological Resource Schema Ontologyを開発してきた。本発表では、BRSOの詳細を紹介すると共に、実際に作成したデータ例を報告する。BRSOを用いて、RDF化することにより、統合的に検索することや、外部のRDFデータとの連携利用も容易に行えるようになること期待される。

BRSO開発の目的

● brso (biological resource schema ontology)

brsoは、バイオリソース情報における共通項目部分をRDF化するために、理研バイオリソース研究センター、遺伝学研究所と協力して、開発を進めているオントロジー。

brsoを用いることで、バイオリソースのデータを共通スキーマで記述することが可能となる。



リソースの情報の中には、共通化できない、リソース特有の情報や、管理施設特有の情報がある。共通化できない情報について、RDF化する場合には、リソース独自のスキーマを検討する必要がある。

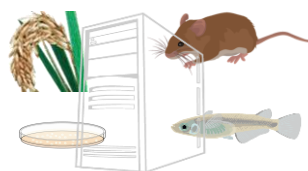
● 国内版バイオハッカソン BH19.7

(<http://wiki.lifesciencedb.jp/mw/BH19.7>) においてバイオリソースとして必要な情報の項目を整理し、スキーマを設計を開始、その後、より使いやすく汎用的に利用できるよう、検討を重ねている。



- 寄託者
- 寄託日
- 寄託機関
- 文献情報 等
- 生物種
- TaxonomyID
- 生物種名 等
- 採取地
- 生息地
- 栽培地 等
- 特徴
- 表現型 等
- 遺伝子名
- 遺伝子番号
- アレル名
- アレル番号 等

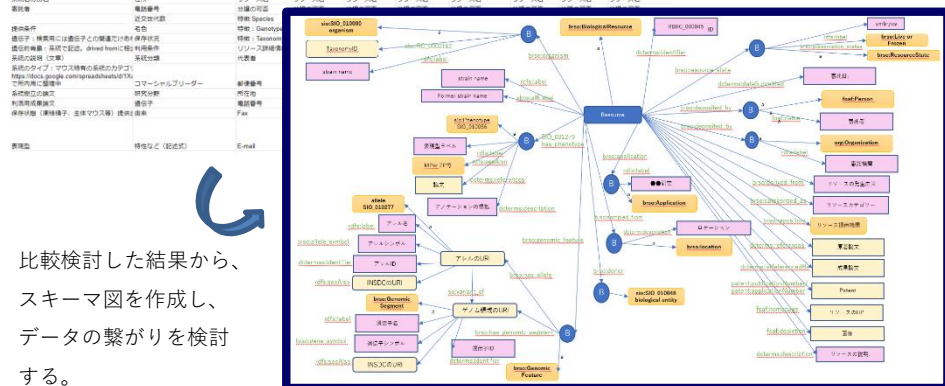
BRSO概要



国内で収集されているバイオリソースの各種データの共通項目を、個々に管理されているデータベースから抜き出し、比較検討を行う。共通情報項目の洗い出しを実施。

国内で一般的に情報が入手できる各バイオリソースの公開データ項目を調査し、比較検討を実施する。

実験動物マウス (種別/性別)	マウス (種別)	シマウナグサ (種別)	雑魚 (種別)	メダカ (種別)	ヒメジョオン (種別)	ヒメジョオン (種別)	ヒメジョオン (種別)	ヒメジョオン (種別)	ヒメジョオン (種別)	ヒメジョオン (種別)
提供機関	提供機関	提供機関	提供機関	提供機関	提供機関	提供機関	提供機関	提供機関	提供機関	提供機関
リソース番号	リソース番号 (DBRP-Rat No.)	リソース番号	リソース番号	リソース番号	リソース番号	リソース番号	リソース番号	リソース番号	リソース番号	リソース番号
寄託者 (姓)	寄託者 (姓)	寄託者 (姓)	寄託者 (姓)	寄託者 (姓)	寄託者 (姓)	寄託者 (姓)	寄託者 (姓)	寄託者 (姓)	寄託者 (姓)	寄託者 (姓)
寄託日 (YYYY-MM-DD)	寄託日 (YYYY-MM-DD)	寄託日 (YYYY-MM-DD)	寄託日 (YYYY-MM-DD)	寄託日 (YYYY-MM-DD)	寄託日 (YYYY-MM-DD)	寄託日 (YYYY-MM-DD)	寄託日 (YYYY-MM-DD)	寄託日 (YYYY-MM-DD)	寄託日 (YYYY-MM-DD)	寄託日 (YYYY-MM-DD)
寄託機関	寄託機関	寄託機関	寄託機関	寄託機関	寄託機関	寄託機関	寄託機関	寄託機関	寄託機関	寄託機関
生物種 (学名)	生物種 (学名)	生物種 (学名)	生物種 (学名)	生物種 (学名)	生物種 (学名)	生物種 (学名)	生物種 (学名)	生物種 (学名)	生物種 (学名)	生物種 (学名)
生物種名	生物種名	生物種名	生物種名	生物種名	生物種名	生物種名	生物種名	生物種名	生物種名	生物種名
採取地	採取地	採取地	採取地	採取地	採取地	採取地	採取地	採取地	採取地	採取地
生息地	生息地	生息地	生息地	生息地	生息地	生息地	生息地	生息地	生息地	生息地
栽培地	栽培地	栽培地	栽培地	栽培地	栽培地	栽培地	栽培地	栽培地	栽培地	栽培地
特徴	特徴	特徴	特徴	特徴	特徴	特徴	特徴	特徴	特徴	特徴
表現型	表現型	表現型	表現型	表現型	表現型	表現型	表現型	表現型	表現型	表現型
遺伝子名	遺伝子名	遺伝子名	遺伝子名	遺伝子名	遺伝子名	遺伝子名	遺伝子名	遺伝子名	遺伝子名	遺伝子名
遺伝子番号	遺伝子番号	遺伝子番号	遺伝子番号	遺伝子番号	遺伝子番号	遺伝子番号	遺伝子番号	遺伝子番号	遺伝子番号	遺伝子番号
アレル名	アレル名	アレル名	アレル名	アレル名	アレル名	アレル名	アレル名	アレル名	アレル名	アレル名
アレル番号	アレル番号	アレル番号	アレル番号	アレル番号	アレル番号	アレル番号	アレル番号	アレル番号	アレル番号	アレル番号

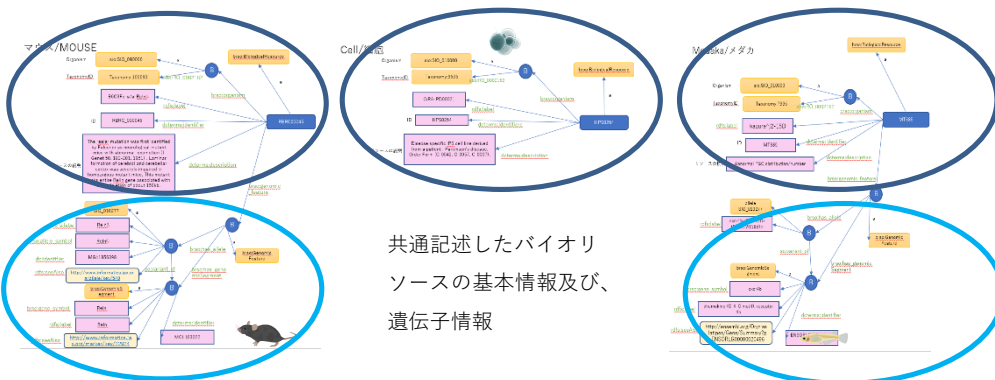


比較検討した結果から、スキーマ図を作成し、データの繋がりを検討する。



BRSOの特徴

BRSOの特徴は、生物種関係なく、汎用的にバイオリソースを記述できる点である。生物が基本的に持つ情報の他に、バイオリソースとしての基本的な共通項目を統一したフォーマットで記述することができる。このことは、生物種を超えて、統計や検索を可能にすることができる。また各種データベースでの項目名の違いを解消するため、国際的なデータベースでの表記の調査も行い、可能な限り国際標準の項目名(表記名)を採用するようにした。



共通記述したバイオリソースの基本情報及び、遺伝子情報

BRSOを用いて作成したスキーマ図。(マウス、細胞、メダカ)生物種は異なるが、項目が整理されている事により、同じ様に記述することができる。

● 今後の展望

多くの生物資源コレクションがRDF化され、公共データとして登録されることにより、それらを統合的に検索することや、ゲノムデータベースなど外部のRDFデータとの統合的な利用が容易に行えるようになると期待される。

各公共データベースでの利活用の促進

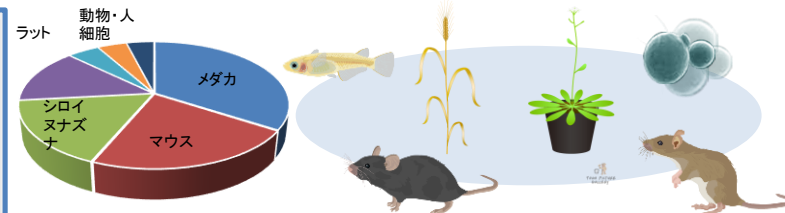
RDF化したバイオリソースデータは、NBDC RDF Portalから公開し、公共データベースにおける、幅広い利活用の促進が期待される。



BRSOを用いたRDF化とその活用

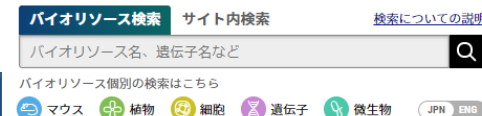
BRSOを用いて、バイオリソースとしての基本情報、生物に関する情報(学術名、タクソノミーID)等の部分を、複数の生物種で実際にRDF化を行う。

Resource_name	count
Medakafish	1299088
Mouse	774238
Arabidopsis(Plant cells, Gene)	589089
Rat	559479
Rice	21421
General Microbes	16113
Animal / Human cells	2378



NBRPに登録されている複数のリソース名と、リソース数をSPARQLでカウントした結果。(7種類 登録リソース数 3,261,806)

● 理研BRCにおける実際の活用事例



BRSOを用いて作成された、理研BRCの実際の検索画面とその結果の表示。遺伝子名"Pax 6"で検索を行うと、関連するマウス、遺伝子材料が表示される。

● 関連ポスター紹介

ポスター番号: 33

「バイオリソースの利活用向上を目指した理研BRC ホーム ページのコンテンツ整備」

ポスター番号: 34

「理研BRCのバイオリソース種横断的に検索可能なカタログ システムの開発」

