

○高月照江1)、臼田大輝2)、北橋優子4)、櫛田達矢2)、市川夏子4)、川本祥子3,1)、梶屋啓志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, ⁴⁾ (独) 製品評価技術基盤機構

- 目的

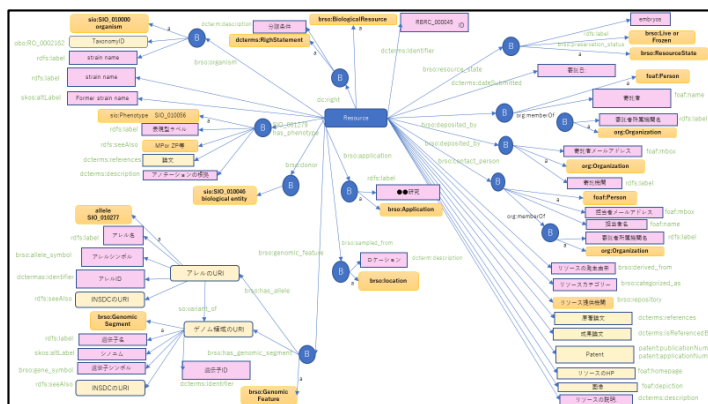
Biological Resource Schema Ontology (BRSO)は、原核生物から動物まで、様々なバイオリソース（生物資源）のメタデータの記述に用いることを目的としたオントロジーである。生物資源のメタデータには、株名、寄託日、寄託者、標品の形態といった基本的な情報や、バイオセーフティーレベルなどの取り扱いに関する情報、表現型やアليل等の生物学的な情報など、多様な内容が含まれているが、BRSOは基本的には、あらゆる生物種に共通して記述される概念を対象として開発されている。一般に、メタデータはバイオリソース保存機関毎に異なる仕様で記載されているが、BRSOを用いたRDFとして記述することで、機関間におけるデータの相互運用性が高まると考えられる。これまでに、BRSOは理研BRC及びナショナルバイオリソースプロジェクト（NBRP）において利用されており、現在、理研のJapan Collection of Microorganisms (JCM)及び製品評価技術基盤機構のBiological Resource Center, NITE (NBRC)が管理する微生物資源についても、BRSOを用いたRDFデータの開発を進めている。本発表では、BRSOの詳細を紹介すると共に、BRSOを利用してRDF化されたデータの実際の利用例や、リソースデータ以外の外部データとの連携等についても報告する。BRSOを用いたRDF化を実現することにより、機関をまたいだデータの横断検索の実現が容易になり、また、ゲノムや疾患等の多様な外部データとの統合による、より高度な生物資源検索につながると考えている。

BRSOについて



● brso (biological resource schema ontology)

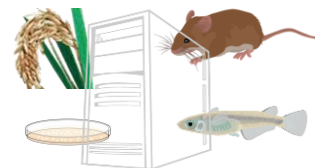
brsolは、バイオリソース情報における共通項目部分をRDF化するために、DBCLSが中心となり、理研バイオリソース研究センター、遺伝学研究所と協力して、開発を進めているオントロジー。必要に応じて拡張を進めている。



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。

BRSOを用いたRDF化とその活用



●理研BRCにおける実際の活用事例

BRCカタログシステム(バイオリソース横断検索)でのBRSOの活用

<https://web.brc.riken.jp/>

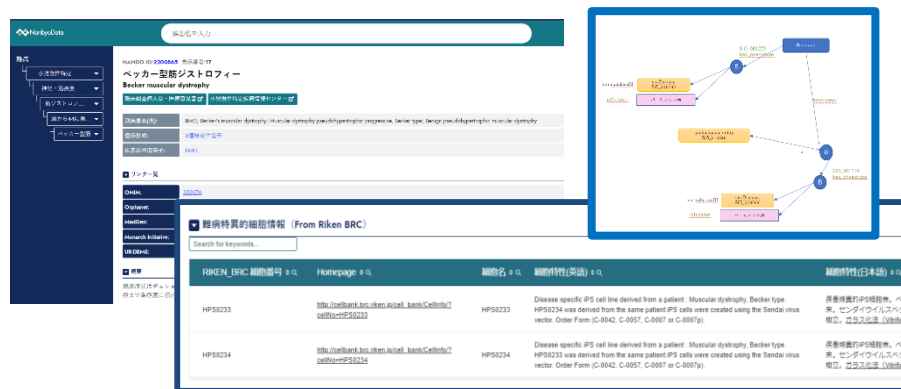
[illegible]

The screenshot displays the Cell Browser database interface. At the top, there's a search bar with the text "Cell Browser (2018-03-01)". Below this, the main content area is divided into several sections. The first section, titled "Cell Browser (2018-03-01)", shows a list of cell lines with columns for "Cell line", "Cell type", "Species", "Disease", "Cell line", "Cell type", "Species", "Disease", and "Cell line". The second section, titled "Cell Browser (2018-03-01)", shows a list of cell lines with columns for "Cell line", "Cell type", "Species", "Disease", "Cell line", "Cell type", "Species", "Disease", and "Cell line". The third section, titled "Cell Browser (2018-03-01)", shows a list of cell lines with columns for "Cell line", "Cell type", "Species", "Disease", "Cell line", "Cell type", "Species", "Disease", and "Cell line". The fourth section, titled "Cell Browser (2018-03-01)", shows a list of cell lines with columns for "Cell line", "Cell type", "Species", "Disease", "Cell line", "Cell type", "Species", "Disease", and "Cell line". The fifth section, titled "Cell Browser (2018-03-01)", shows a list of cell lines with columns for "Cell line", "Cell type", "Species", "Disease", "Cell line", "Cell type", "Species", "Disease", and "Cell line". The sixth section, titled "Cell Browser (2018-03-01)", shows a list of cell lines with columns for "Cell line", "Cell type", "Species", "Disease", "Cell line", "Cell type", "Species", "Disease", and "Cell line". The seventh section, titled "Cell Browser (2018-03-01)", shows a list of cell lines with columns for "Cell line", "Cell type", "Species", "Disease", "Cell line", "Cell type", "Species", "Disease", and "Cell line". The eighth section, titled "Cell Browser (2018-03-01)", shows a list of cell lines with columns for "Cell line", "Cell type", "Species", "Disease", "Cell line", "Cell type", "Species", "Disease", and "Cell line". The ninth section, titled "Cell Browser (2018-03-01)", shows a list of cell lines with columns for "Cell line", "Cell type", "Species", "Disease", "Cell line", "Cell type", "Species", "Disease", and "Cell line". The tenth section, titled "Cell Browser (2018-03-01)", shows a list of cell lines with columns for "Cell line", "Cell type", "Species", "Disease", "Cell line", "Cell type", "Species", "Disease", and "Cell line".

BRSOを用いてバイオリソース情報をRDF化した結果、リソースの横断検索を実現している。

BRSOを利用したRDF化データの活用 外部データとの連携について

日本の難病疾患の詳細情報を提供するNanbyoData (<http://nanbyodata.jp/>) では、理研BRCの取り扱い細胞データを、難病疾患の研究リソースとして表示させている。これは、バイオリソースの特性情報の一つとして、関連疾患名をNANDO (Nanbyo disease ontology) を用いてアノテーションし、それらのデータをBRSOを用いてRDF化した結果、NanbyoDataで表示されているNANDO (Nanbyo disease ontology) との連携が可能になっている。理研のRDFデータは理研のエンドポイントで管理されており、NANDOのデータはDBCLSのエンドポイントで管理されている。RDF化されたデータは、違うエンドポイントに投入されたデータでも連携が可能となる。NANDOのIDに対して、関連ある細胞のIDを、紐づけされたNANDOのIDでSPARQL検索を行い、ID,理研BRCの細胞の詳細ページ、細胞の説明等を表示させているが、これらの表示項目についてもBRSOを用いてRDF化された結果、表示が可能となっている。



実際のNanbyoDataでの細胞の表示例及びスキーマ図

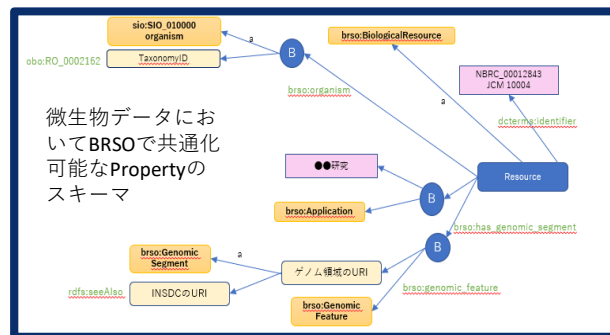
理研BRCでは、細胞以外のリソース（マウス、遺伝子）についても、疾患特性のアノテーションを実施していく予定であり、結果、関連バイオリソースとして、細胞だけではなく、その他のリソースを追加することも可能となる

● 今後の展望

BRSOを用いたRDF化を実現することにより、機関をまたいだデータの横断検索の実現が容易になる。生物種関係なく、ゲノムや疾患等の多様な外部データとの統合が可能になることで、ゲノムからリソース、疾患からリソースと言った、より高度な生物資源検索につながると考えている。

BRSOを用いた新たなデータ連携の取り組み（微生物リソース）

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。今年度の新たな取り組みとして、理研のJapan Collection of Microorganisms (JCM)及び製品評価技術基盤機構のBiological Resource Center, NITE (NBRC)が管理する微生物資源についても、BRSOを用いたRDFデータの開発を進めている。これまでMCCV (Microbial Culture Collection Vocabulary) で記述されていた微生物のRDFデータに、BRSOを用いたRDFデータを加えることで、微生物もバイオリソースの一つとして取り扱いが可能となり、バイオリソースとして共通項目事項については、他の生物種同様のデータ管理が可能となる。



brsoを用いて、生物資源であるという事を記述することにより、生物種関係なく、生物資源としてデータ上扱う事が可能となる。



+

その他のリソース

これまで、MCCVを使ったRDF化により、情報連携を行っているが、更にBRSOを加える事で、その他のバイオリソースとの連携も可能となる

● 関連ポスター紹介

ポスター番号：39「セマンティックウェブ技術を用いたバイオリソース横断 検索システム」

ポスター番号：41「NBRCカイコデータベースのRDF化とその利用」

ポスター番号：42「NBRC藻類リソースのRDFデータ提供およびWebサイト 公開」