

○臼田 大輝¹、櫛田 達矢¹、小林 紀郎²、榎屋 啓志¹

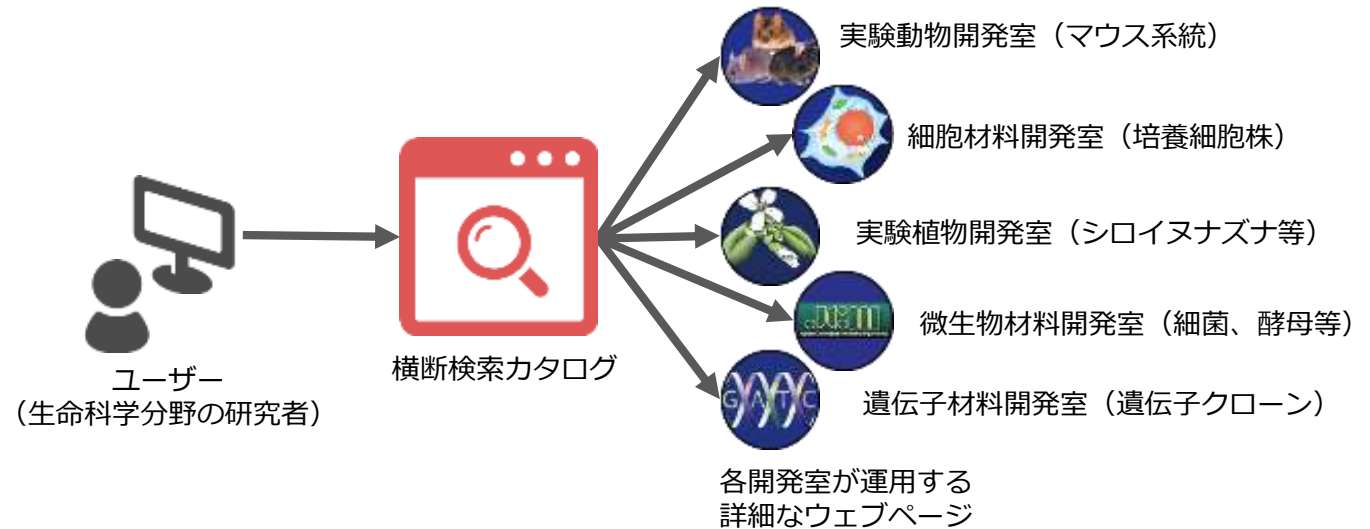
1. 理化学研究所バイオリソース研究センター、2. 理化学研究所情報システム本部



目的

理化学研究所バイオリソース研究センター（RIKEN BRC）では、マウス、植物、培養細胞、遺伝子クローン、微生物株の5種類のバイオリソース（生命科学研究に用いられる材料）を提供している。生命科学では、遺伝子、疾患との関連性、物質産生など、様々な生物特性情報と紐づけてバイオリソースを利用される。

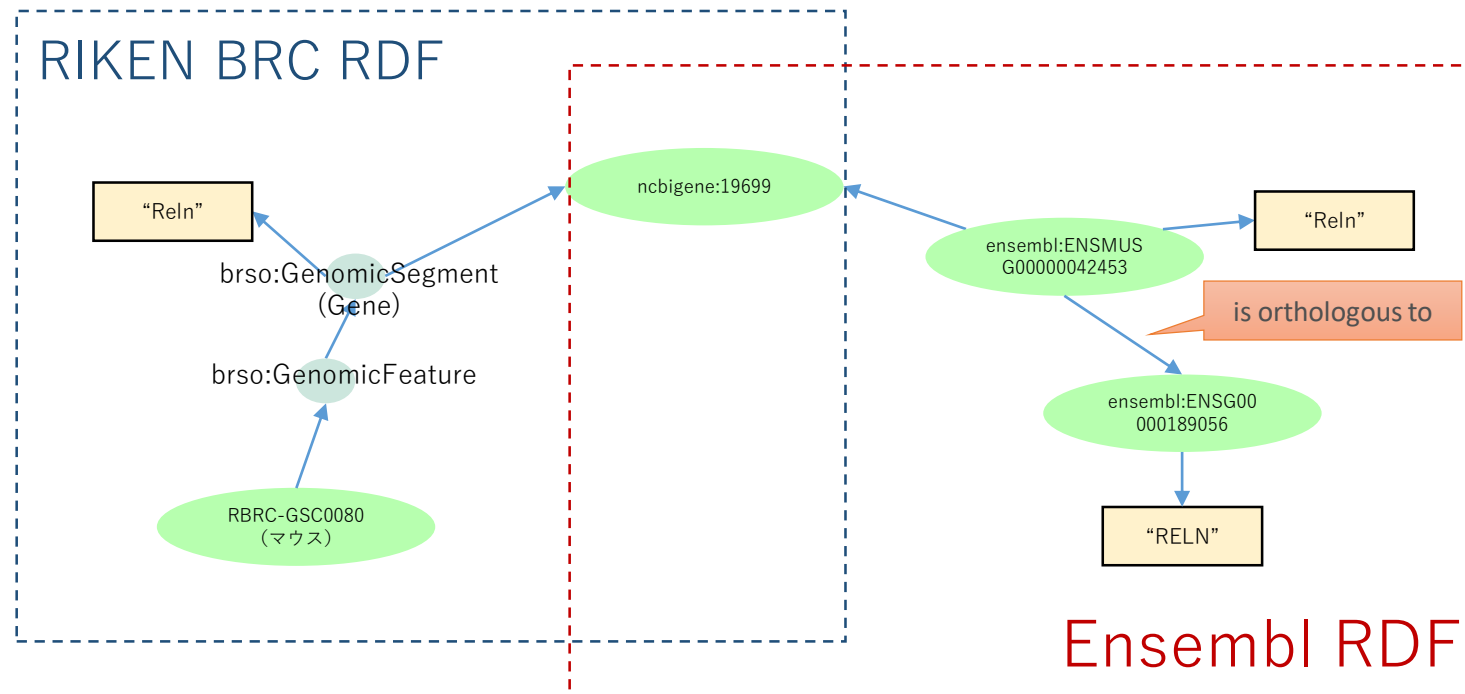
RIKEN BRCにおける従来のバイオリソースの検索方法は、各バイオリソースを担当する部署が独自のウェブサイトで運用していたため、包括的検索が出来なかった。そこで我々は情報統合に適したセマンティックWeb技術を用いることで横断的に検索可能なバイオリソースカタログの開発に取り組んだ。



1つの検索語でRIKEN BRC全体のリソースが検索でき、
詳細情報については各担当が運用しているカタログのウェブページにリンク

RDF形式によるデータ作成

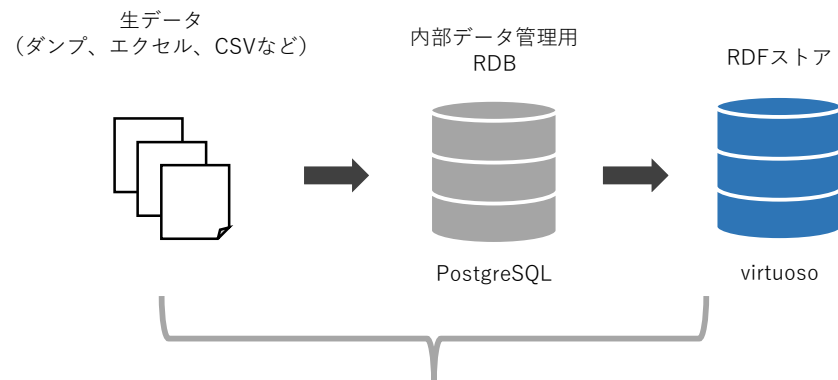
データの形式として情報統合に適しているRDFを採用した。RDFのスキーマ作成にあたり、バイオリソースの基本情報の共通化を目的としたbiological resource schema ontology (brso※)を参考にした。また、遺伝子情報についてはEnsemble RDFと共通のURIを用いることでデータ統合し、オーソログおよびパラログ関係にある遺伝子シンボルでもバイオリソースの検索をできるようにした。



※本会ポスター番号36番: 「BRSO: Biological Resource Schema Ontology の開発」を参照

データ更新業務用アプリケーションの開発

バイオリソースの各担当部署において、データは関係データベース（RDB）、ファイルメーカー、エクセルなど様々なフォーマットで管理されているため、これらの一次データをRDBで一元管理することにした。またデータ更新作業を簡便化するために業務用アプリケーションを独自に開発した。



業務用アプリケーション



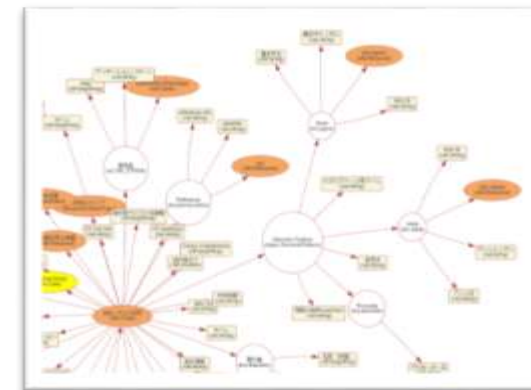
データ取込の設定



更新データの確認



RDF化 前データの設定

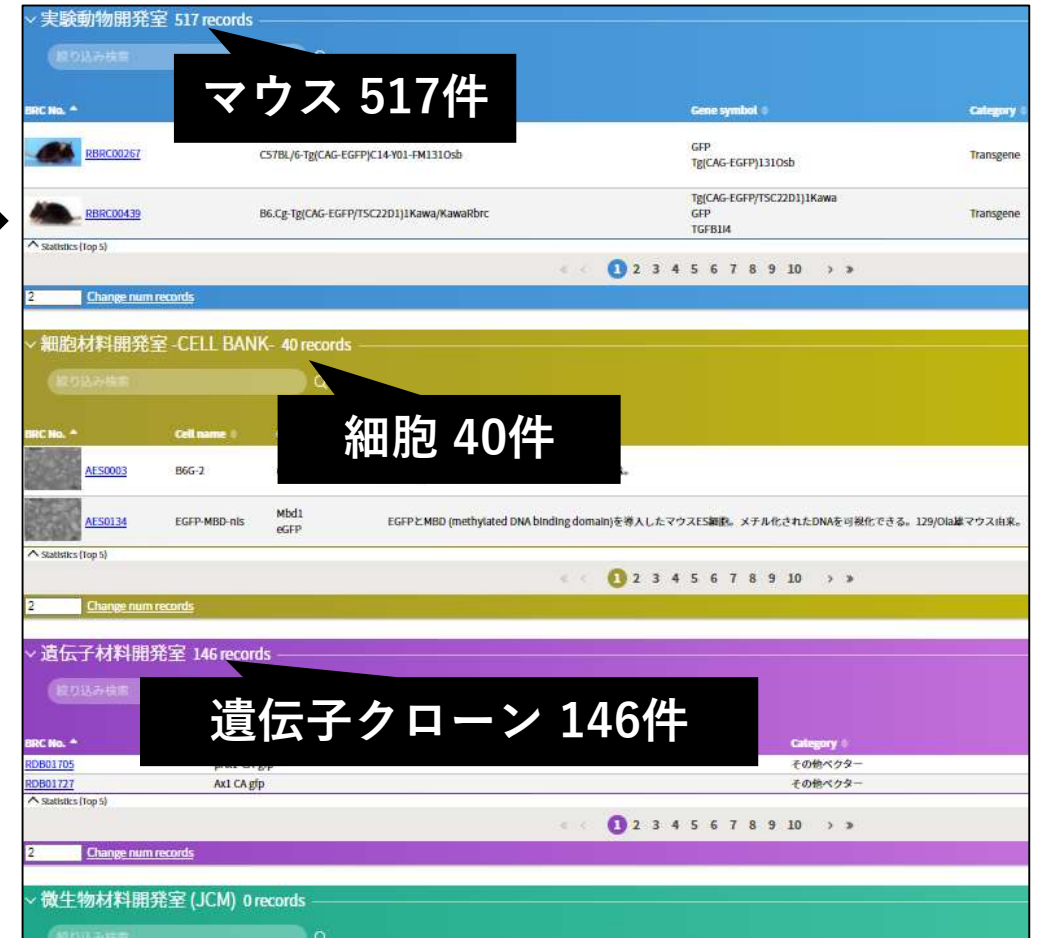


RDF化 前データ 可視化による確認

横断検索カタログシステム



検索実行



マウス 517件

細胞 40件

遺伝子クローン 146件

RIKEN BRCの横断検索カタログシステムは2020年7月より正式に運用開始しており、RIKEN BRCのホームページ (<https://brc.riken.jp>)にあるバイオリソース検索窓から実行できる。

●横断検索カタログへのアクセス実績（運用開始から2か月）

月	ページビュー数	セッション数	ユーザー数
2020/07	1,682	846	668
2020/08	1,782	745	550

“GFP”で検索した結果
リソース番号をクリックすると開発室が運用する詳細ページへと遷移する

RDFデータの公開

作成したRDFデータはBioResource MetaDatabase (<https://knowledge.brc.riken.jp/bioresource>) にて公開している。Bioresource MetaDatabaseは、RIKEN MetaDatabase (Kobayashi 2016) のシステムを利用しており、RDFデータの閲覧、ttlファイルのダウンロード、RDFの問い合わせ言語であるSPARQLによるデータ取得等が実行できる。



BioResource MetaDatabaseで公開している
RIKEN BRCのバイオリソースデータ (2020年9月現在)

バイオリソース種	リソース数	トリプル数
マウス	7,173	903,830
細胞	8,175	265,059
遺伝子クローン	162,187	3,307,228
微生物	18,033	847,474
植物	278,492	5,165,317

[Kobayashi 2016] Kobayashi, N., Lenz, K., Masuya, H.: RIKEN MetaDatabase: a database platform as a microcosm of linked open data cloud in the life sciences. JIST 2016: Semantic Technology, 99-115 (2016)

今後の展開

検索でヒットする共通項目は、ID、名前、生物種名、遺伝子シンボルなど限られている。検索対象となる項目は拡張中であり、マウスについてはMammalian Phenotype Ontology (MP)の哺乳類表現型データの語彙による検索を可能にした。今後は疾患や化合物などでも検索できるようにしていく予定である。また検索履歴について解析を行い検索の質の向上に努めていく。



横断検索カタログのご意見もこちらからお願いします！
(<https://brc.riken.jp>)

謝辞

本研究の一部は、内閣府 戦略的イノベーション創造プログラム (SIP) 「スマートバイオ産業・農業基盤技術」(管理法人: 農研機構生研支援センター) によって実施されました。また本研究を進めるにあたり、BRSO 関係者であるライフサイエンス統合データベースセンターの川島秀一様ならびに高月照江様、国立遺伝学研究所の川本祥子様、農研機構の鐘ヶ江弘美様に感謝いたします。

関連ポスター

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: 36
「BRSO: Biological Resource Schema Ontology の開発」
ポスター番号: 37
「MoG+: 理研BRCマウスゲノム多型データベース」