RIKEN META DATABASE

理研メタデータベースは、理研の成果をより広く研究者の方に活用していただくことを目的に、
理研の研究者が公開するデータベースのメタデータを体系的に整理して公開するサービスです。

http://metadb.riken.jp

SPARQL endpoint

en ja

理研 データベース 電話帳

データベース | オントロジー語彙 | 理研メタデータベースについて

理研メタデータベースの運用とデータ統合の実際

懸津　魁[1], 桝屋 啓志[2,1] , 小林 紀郎[1,2,3]

1. 情報基盤センター　2. BRC バイオリソースセンター
3. ライフサイエンス技術基盤研究センター

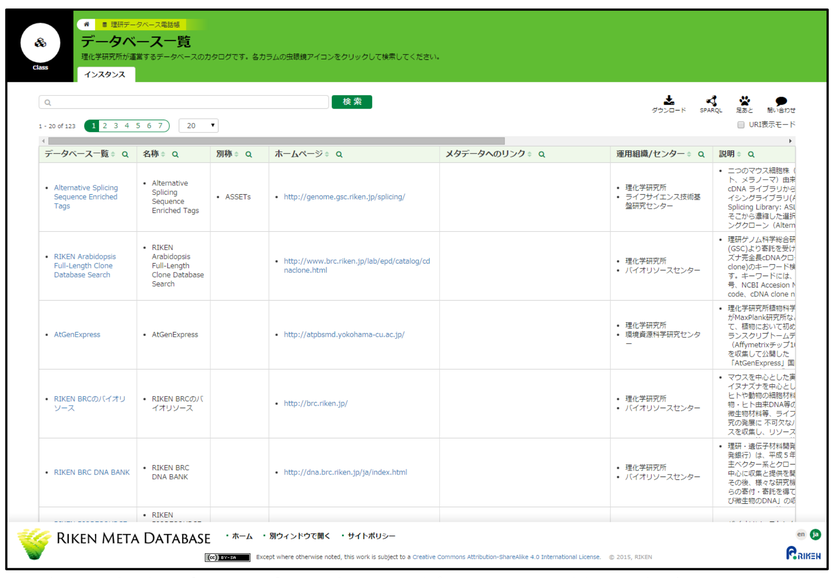
概要

理化学研究所は物理、化学、ライフサイエンスをはじめとするサイエンスの総合研究所であり、多種多様な研究データを産出している。これらのデータを公開するために、理研メタデータベース (http://metadb.riken.jp/) を構築し、公開インフラとして利用している。理研メタデータベースは、メタデータ公開の標準仕様であるResource Description Framework (RDF)に準拠して設計されており、複数のデータベースを統合、公開することができる。1つのデータベースは1つのRDF Graphとして管理され、更にデータレコード（RDFリソース）はすべてクラス概念を付してまとめ上げることで、ライフサイエンス分野でよく使われている互いに関連するテーブル形式のデータ構造を持たせ、表示することができる。RDFの特性を活用し、従来のテーブルデータとは異なり、異なるデータベースのデータも統合しつつテーブルの上に表示することが可能となっている。理研メタデータベースはSPARQLエンドポイント機能も備わっており、ユーザーはテーブル形式で表示されるデータ構造を参考にしながらSPARQLクエリを記述し、より詳細なデータ解析を行うことができる。2016年10月現在、センター横断的に112個のデータベースが統合され、クラス数2,206、トリプル数161,122,245に及ぶデータを公開している。

メタデータの作成・流通を容易に

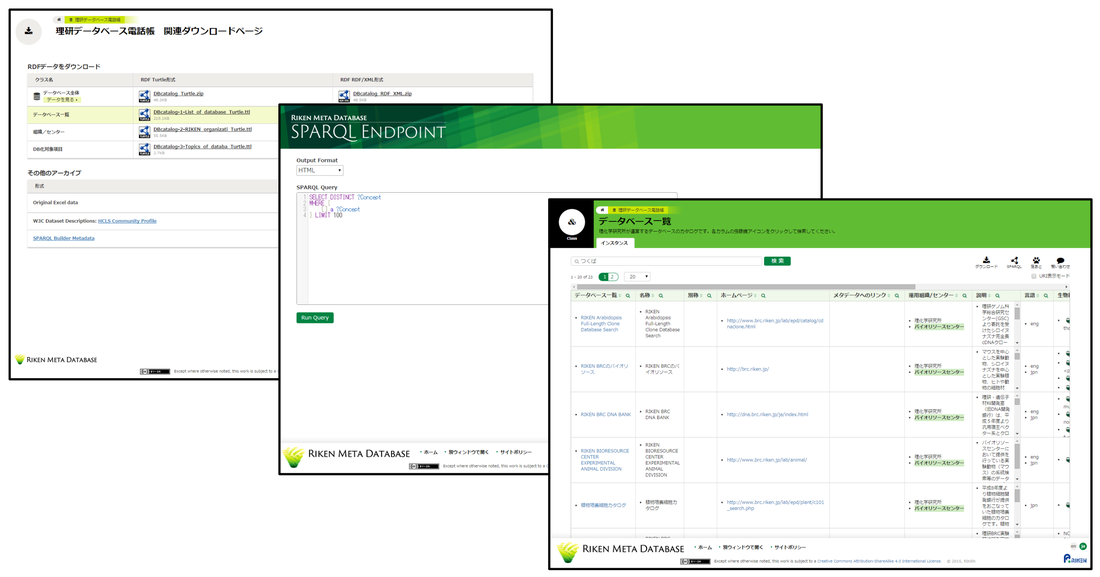
①スプレッドシートに記入 → ②システムでRDFに変換（世界標準仕様） → ③メタデータベースで公開 → ④SPARQL検索 RDFダウンロード

RDFデータのテーブル形式表示



主語を左端列・述語をヘッダー行・目的語を各セルにそれぞれ対応させクラスに関わるトリプル群を表示

その他GUI（ダウンロード・検索）



RDFのダウンロード・SPARQLエンドポイント・テーブル内での検索機能

系統名	BRC ID	Former Common name	系統のタイプ	系統説明	背景系統	寄託機関	オリジナルサイト	種	Linked External class	Linked External B6-ChrNMSMコンソミック系統
B6-Chr6T<MSM>	RBRC02539		Consomic				http://www2.brc.riken.jp/lab/animal/detail.php?brc_no=RBRC02539	Mus musculus	B6-Chr6T<MSM>	B6-Chr6T<MSM>
B6-Chr7C<MSM>	RBRC02541		Consomic				http://www2.brc.riken.jp/lab/animal/detail.php?brc_no=RBRC02541	Mus musculus	B6-Chr7C<MSM>	B6-Chr7C<MSM>
B6-Chr7T<MSM>	RBRC02542		Consomic				http://www2.brc.riken.jp/lab/animal/detail.php?brc_no=RBRC02542	Mus musculus	B6-Chr7T<MSM>	B6-Chr7T<MSM>
B6-Chr8<MSM>	RBRC02543		Consomic				http://www2.brc.riken.jp/lab/animal/detail.php?brc_no=RBRC02543	Mus musculus	B6-Chr8<MSM>	B6-Chr8<MSM>

同一URIのリソースを主語または目的語とするトリプルが他データベースにもあればテーブル右端側に外部リンク列を作成し表示する

統合されたデータベース群

データベースURI	データベース数	オントロジー数
BRC マウスリソース表現型メタデータ	13	14
BRC細胞リソースメタデータ	13	10
遺伝研 マウス表現型データベースメタデーター	13	10
NBRP ラット表現型メタデータ	12	13
JCGGDB: ノックアウトマウスを用いた機能糖鎖科学データベースのメタデータ（ベータ版）	12	10
理研データベース電話帳	12	5
bioresource_schema	10	12
NBRPメダカ表現型メタデータ	8	10
JCM微生物リソースメタデータ	8	9
SSBD:生命動態システム科学の統合データベースに登録されている定量データと顕微鏡画像セットのメタデータ	8	9

同一URIのリソースを持つ外部データベース・オントロジーの数（上位10データベース）

利用オントロジー

- Metagenome and Microbes Environmental Ontology
- Clinical Signs and Symptoms Ontology
- Pathogenic Disease Ontology
- Chemical Entities of Biological Interest (ChEBI)
- Cell Ontology
- Cell Line Ontology (CLO)
- Gene Ontology (GO)
- Left Unit Ontology
- Mouse Adult Gross Anatomy (MA)
- Microbial Culture Collection Vocabulary (MCCV)
- Medaka fish anatomy and development
- Mammalian Phenotype (MP)
- Mouse pathology
- NCBI Organismal Classification (NCBITaxon)
- Orphanet Rare Disease Ontology
- Phenotypic Quality (PATO)
- OBO-relation ontology (RO)
- Rat Strain Ontology (RS)
- Semanticscience Integrated Ontology
- SSBDO
- Uber Anatomy Ontology (UBERON)
- Units of Measurement (UO)
- Zebrafish Anatomy and Development (ZFA)
- Zebrafish Phenotype Ontology

データベース一覧

- 理研データベース電話帳
- FANTOM5 SSTAR
- FANTOM4
- FANTOM eeDB
- BRC マウスリソース表現型メタデータ
- BRC細胞リソースメタデータ
- JCM微生物リソースメタデータ
- NBRPメダカ表現型メタデータ
- NBRP ラット表現型メタデータ
- 遺伝研 マウス表現型データベースメタデーター
- IMPC_RDF
- JCGGDB: ノックアウトマウスを用いた機能糖鎖科学データベースのメタデータ（ベータ版）
- RDF of MGI data
- bioresource_schema
- NIG ゼブラフィッシュ（ベータ版）
- SSBD: 生命動態システム科学の統合データベースに登録されている定量データと顕微鏡画像セットのメタデータ
- FOX Hunting
- RIKEN Arabidopsis Activation tagging line
- DS tag line
- SSBC (Systems and Structural Biology Center)
- RIKEN Protein Database
- RIKEN Harima Heavy-atom Database [Data]
- Bacpedia (Riken Harima SPring8-Center)
- Bacpedia data download project
- The Rice Gene From RAP-DB
- 理研 Plant gene family
- 理研 Ortholog
- シロイヌナズナ変異体フェノーム情報
- 理研 RARGE Promoter
- Gramene O.sativa GO Annotation [Data]
- Promoter collection for genome design
- 理研 RARGE Transposon Mutant
- 理研 RAFL cDNA
- RIKEN Arabidopsis Phenome Information Database
- RAPID: Resource of Primary Immunodeficiency Diseases
- 理研　RARGE Alternative Splicing
- 理研　シロイヌナズナ転写因子データベース (RARTF)
- ARTADE2 予測遺伝子モデル データベース
- Sorghum bicolor gene
- Cerebellar Development Transcriptome Database
- Codon Usage Database
- Selaginella moellendorffii gene
- タンパク質実験・構造統合データベース
- Pfam
- Domain
- 遺伝子組み換え生物等の規制
- SMART
- Bacterial Transcriptional Regulations
- Populus trichocarpa (JGI v1.1)
- Vitis vinifera gene
- 植物フェノーム
- A.thaliana cDNA
- Genome Annotation
- Plant Small RNAs
- フェノームミーティング
- AraCyc Pathway (BioPAX Level2)
- PDFファイルリポジトリ
- 科学誌一覧
- JavaScriptサンプル集
- Experimental condition
- Supp Data
- Evidence Code Description
- 合理的ゲノム設計：パーツ集
- 配列設計プロジェクト
- SciNeS公開仕様集
- SciNeSデータ構造
- A.thaliana locus
- GO Evidence Codes
- Ontology template
- Entrez Human
- TAIR PO Annotation[growth]
- TAIR PO Annotation[structure]
- Human Ensembl Gene
- Protein Data Bank
- Entrez Rat
- RefSeq Human
- HGNC
- Biorel
- MGI gene - Gene Ontology
- TAIR GO Annotation
- MGI Strains and Polymorphisms
- MGI Allele-Mammalian Phenotype
- HGNC-OMIM
- TIGR Rice Gene Models
- Genome Sequence
- A.thaliana gene
- InterPro
- Rat Genome Database
- Rat Ensembl Gene
- Dog Ensembl Gene
- Chimpanzee Ensembl Gene
- Mammalian Ortholog
- miRBase (Sanger Institute)
- Dublin Core Metadata
- MGI gene-MP link
- NAR Categories
- MGI Alleles
- GenoCon 2010
- コンテスト公開プロジェクト（フラグ用）
- Portal Page
- 植物統合データベース
- SemanticTable
- synthetic-biology.jp
- Swiss-Prot
- 植物リン酸化プロテオームデータベース
- Entrez Mouse
- 理研哺乳類統合データベース
- 実験テンプレート
- ダウンロードファイル保存場所
- DNA in RIKEN BioResource Center
- Type project
- TAIR Annotation

個々のデータベースの独立性とRDFによる柔軟な接続・横断的な検索や閲覧を実現