

○山本 泰智<sup>1</sup>、山口 敦子<sup>1 2</sup>

1. 大学共同利用法人情報・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 (DBCLS)
2. 東京都市大学



## うまかスイートは、下記のサービスの総称

1. RDFデータを取得する問い合わせ先を検索するうまかサーチ (Umaka Search)
2. 問い合わせ先のデータ構造を俯瞰するうまかビューア (Umaka Viewer)

<https://umaka-viewer.dbcls.jp/>

3. 問い合わせ先の信頼性を測り、適宜関連する情報を共有できるうまかデータ (Umaka-YummyData)

<https://yummydata.org/>

うまかサーチを除き、既に公開済みで誰でも自由にアクセスできる。

うまかサーチについては今年度中の公開を目指して試験運用を続けている。



# うまかサーチ

**課題：** RDFデータに含まれるURIが得られても、当該URIに関連するデータを取得するためのSPARQLエンドポイントの所在は陽には分からない。

また、同じURIを含むエンドポイントが複数ある場合、それらを簡単に知りたい。

**入力：** URI

**出力：** 当該URIを検索対象RDFデータにもつSPARQLエンドポイントリスト

## 例

入力：<http://purl.uniprot.org/uniprot/B1XGB0>

出力：

0:reactome:1.3619678:Subject

1:reactome:1.3619678:Object

2:biomodels:1.3610995:Subject

3:opentggates:1.3497965:Object

4:nextprot:1.3446534:Object

5:jpost:1.3426566:Object

# 今後の予定

- 一般公開に向けた開発を継続
- URIだけではなく、特定の述語についてはリテラルも索引対象にするか検討

# ✓ うまかビューア

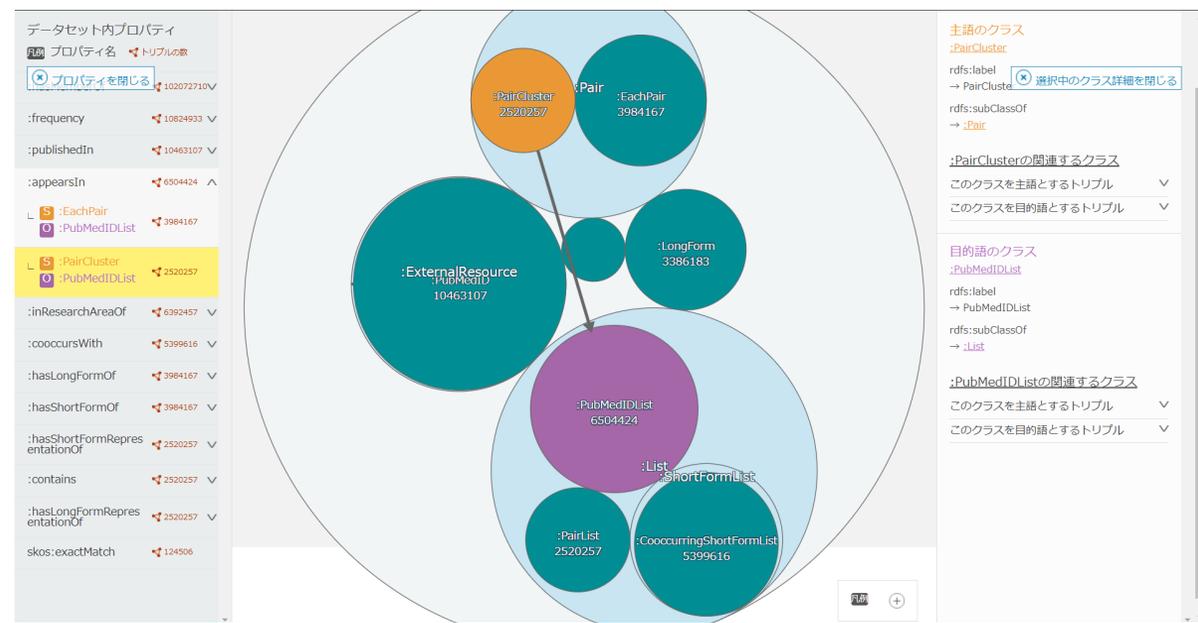
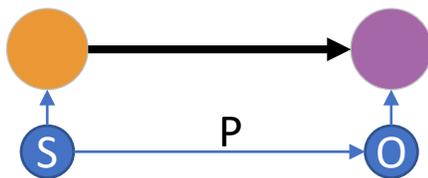
**課題**：SPARQLエンドポイントに含まれるデータの構造が分かりにくい。

<https://umaka-viewer.dbcls.jp/>

**入力**：SPARQLエンドポイント名

**出力**：対象エンドポイントに含まれるデータのクラス階層構造とそれらをトリプルで結ぶプロパティの可視化

可視化された図は対話的で、クラス構造の抽象化及び具体化を行えたり、特定のクラスに注目し、当該クラスに属する主語を持つトリプルで結ばれる目的語が属するクラスをハイライトしたりできる。



# うまかビューアで 可視化するまで

- 必要なメタデータを取得する

<https://bitbucket.org/yayamamo/tripledataprofiler/>

典型的な取得方法

```
java -jar TripleDataProfiler.jar -sbm <Endpoint URL> [<Named Graph Name(s)>] > metadata.ttl
```

- 得られたメタデータから可視化用データを生成する

<https://github.com/dbcls/umakaviewer>

1\*. `umakaparser build-index <ontologies> --dist <directory>`

2. `umakaparser build metadata.ttl --assets <directory> --dist viewer.json`

\* オントロジーファイルを可視化に反映させたい場合にのみ行う

- 可視化用データを可視化サイトに投稿する

<https://umava-viewer.dbcls.jp/>

投稿にはGoogleアカウントが必要

最初のメタデータが得られれば、直接投稿することもできる

# 今後の課題



メタデータ生成ツールの頑健化



可視化データ処理の最適化



可視化データ生成ツールの  
メッセージの改善



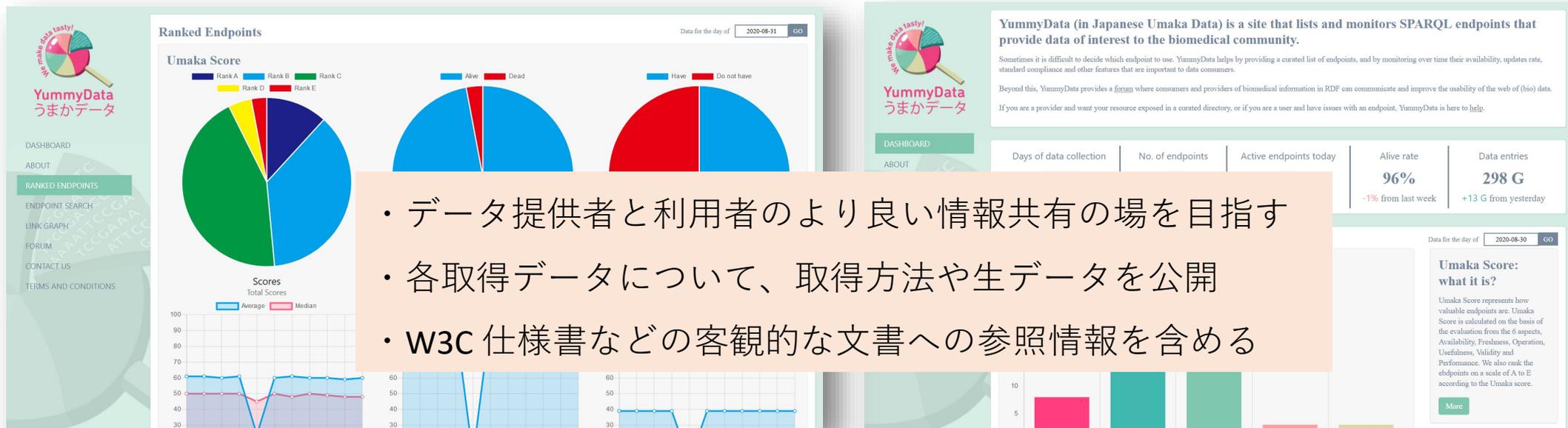
# Umaka-YummyData

課題：SPARQLエンドポイントの信頼性を知りたい

<https://yummydata.org/>

入力：SPARQLエンドポイント名/URI/キーワードなど

出力：対象エンドポイントに対して様々な観点からスコア化した評価結果



- データ提供者と利用者のより良い情報共有の場を目指す
- 各取得データについて、取得方法や生データを公開
- W3C仕様書などの客観的な文書への参照情報を含める



# 評価項目と現状

## 評価項目

- 死活情報
- メタデータ提供情報
- 更新頻度
- **Linked Data** 原則への対応
- 応答速度
- **CORS** 対応状況

など

## 現状

- 2020年9月1日現在、**68**のエンドポイントが対象
- 毎日データを収集
- **1ヶ月以上アクセス不能な場合は削除検討**
- 各エンドポイントについて **github** の **issues** を利用して情報交換可能

# 継続課題



新規エンドポイントの発見方法



スコア方法の改善



データの管理権とURI との関係の定義



サーバー設置場所



フォーラム運営システム

We make data tasty!



終

@yayamamo

yy@dbcls.rois.ac.jp

