第2回Linked Open Data活用ワークショップ in 神戸

2017年9月22日 (金)@理化学研究所融合連携イノベーション推進棟



オープンデータのLOD化

<u>大阪大学産業科学研究所</u>

古崎 晃司

Kozaki@ei.sanken.osaka-u.ac.jp

データソンで行う内容



■ <u>目標</u>

CSV形式で公開されている「オープンデータ」をLOD化 する基本的な方法を身につける

■ 実施内容

- CSVをRDFに変換
 - できるだけ、他のLODと「共通の語彙」を用いる
 - 他LODとのリンクをつける
- 2. RDFファイルを公開
 - SPARQLエンドポイント(API)の利用
 - DBpediaと同じような形(参照解決可能なURI)での公開

既存データをRDF化する方法



- RDFデータを作成するツールを使用
 - Open Refine (http://openrefine.org/) +
 RDF Refine (http://refine.deri.ie/)
 - データを整備・公開するための多様な機能をサポート
 - LinkData (http://linkdata.org/)

LODチャレンジ2011 アプリケーション部門最優秀賞

- テーブルデータをRDFに変換して公開することができるサイト
- SparqIEPCU (http://lodcu.cs.chubu.ac.jp/SparqIEPCU/)
 - LODの作成・活用のための支援サイト

LODチャレンジ2012 アプリケーション部門最優秀賞

- CSVファイルをRDFとしての公開する機能もあり
- StatLD (http://satolab.tiu.ac.jp/statld/)
 - 統計Linked Dataの活用を中心としたツールを公開
 - 一般のLinked Dataの作成にも使用可能
- 独自のプログラム用でRDFデータを作成

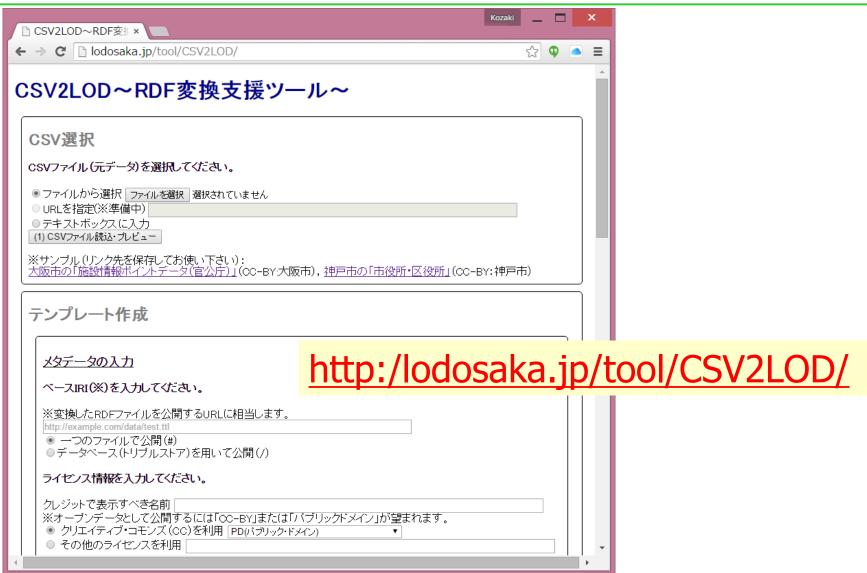
本日、利用するツール



- CSV2LOD~RDF変換支援ツール~
 - http://lodosaka.jp/tool/CSV2LOD/
 - CSV形式のデータをRDFに変換する
 - プロトタイプなので、ご意見大歓迎!
- <u>なんでもリンク</u>
 - http://link.lodosaka.jp/
 - オープンデータに外部LOD(DBpedia)へのリンクを追加する
- Simple LODI
 - https://github.com/uedayou/simplelodi
 - DBpediaのようにLinked Open Data(LOD)を公開する
 - =参照解決可能+コンテンツネゴシエーションに対応した LODを公開する.

使用するツール① CSV2LOD~RDF変換支援ツール~





ツール開発の動機



- Linked Dataの基本原則に沿ったLODを、誰でも、(ある程度) 簡単に作成できるツールを提供したい
- 元データ(CSV)の編集を,極力,減らしたい
 - データが更新されたら、すぐに再変換したい
 - →元データと変換情報のテンプレートは別ファイルに
- 足りない機能があれば、気軽に拡張したい
 - 小規模なプログラムで開発+オープンソース化により改変を自由に
- 初心者から上級者まで使いやすいものを
 - 汎用性の高いテンプレート(Turtleライクな形式)により、複雑なRDFモデルにも対応
 - 初心者向けには、語彙選択支援など、簡単にテンプレートが作れる

■ 動作環境

インストール不要、かつ、ネットがなくても使える→JavaScritptのみで開発

IE8にも、できれば対応したかったが…(現時点は×)

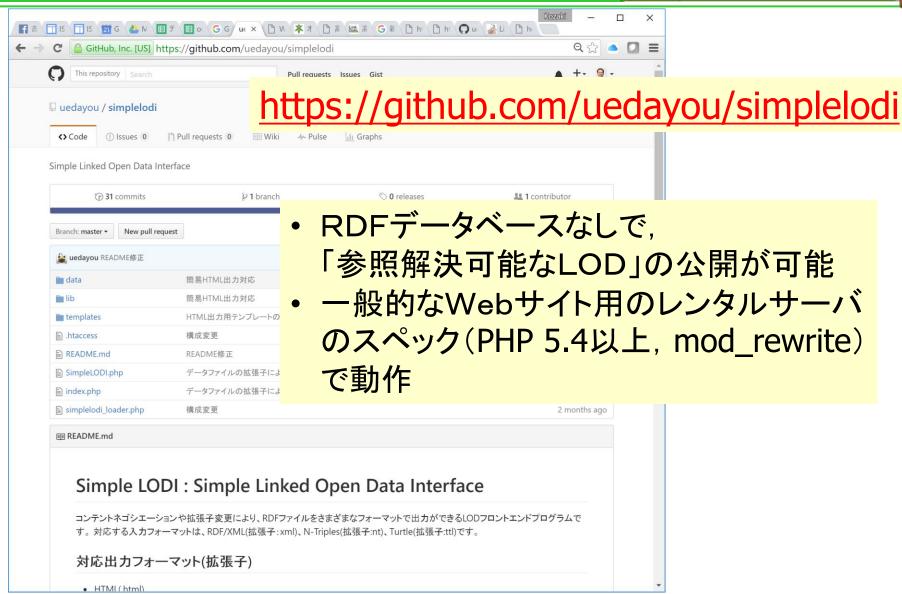
使用するツール
 2
 なんでもリンク





使用するツール③ Simple LODI





利用するオープンデータ



■「CSV形式」のファイルを利用

- 最初の1行目に「データの項目」2行目以降に各データが記入されているもの
- 不要な行など、が入っているものは使えません
- [CSV2LODの既知のバグ]
 "1列目に空白セルが含まれる"と、正しく変換できません。

ハンズオンに利用するオープンデータ

 神戸市オープンデータ: 観光施設情報 https://data.city.kobe.lg.jp/data/dataset/dataset-0000005/resource/369be024-e7c2-46be-98ad-

2017/9/22 a 6 b 2 5 1 9 f 1 2 e 3

利用可能なデータ例



ΧĪ		5	O	<u>.</u>	Ŧ			tourism_o	d2810.csv	- Excel				?	1	_		×
ファ	イル	ホーム	挿入	ページ し	レイアウト	数式 デー	タ 校閲	表示	ACROBAT								ţ	ナインイン
A1		~	X	/	f x 名称													٧
	Α		В		С	D	Е	F	G	Н	I	J	K	L		М	N	
1	名称		名称のル	Ľ	郵便番号	カテゴリ	住所	緯度	経度	電話番号	説明文	営業時間	定休日	URL	バリ	アフリー	-対応	
		えオラン	ノダ館(旧	ヴォルヒ	'6500002	′2401	神戸市中:			078-261-3		9:00~18:0	なし	http://v	wなし			
3	デンマー	-ク館()	日コハン・	フラウベ	′6500002	′2401				078-261-3				http://v	wなし			
			ストリアの			′2401	神戸市中:	34.70241	135.1908	078-261-3	平成4年、	9:00~18:0	なし	http://v	wなし			
5	竹中大	L道具	たけなかフ	だいくど	′6510056	′2401	神戸市中:	34.70686	135.1975	078-242-	竹中大工	火~日 9	月(祝日	の http://d	dougukai	n.jp		
6	風見鶏(か館	かざみどり	りのやか	′6500002	′2401				078-242-								
7	ラインの	館※	工事のた。	め休館	′6500002	′2401	神戸市中:	34.70113	135.1912	078-222-	大正4年(1	9:00~18:0	メンテナ	ンスのたと	か休館中	っです。	H.284.1	\sim
8	うろこの	家・う?	5こ美術館	3	′6500002	′2401				0120-888								
9	異人館		ティン邸_		′6500002	′2401				078-241-								
	萌黄の餌		もえぎのヤ	さかた	′6520002	′2401				078-222-						y.kobe	.lg.jp/cu	ltui
	山手八都				′6500002	′2401				0120-888								
	北野外[′6500002	′2401				0120-888						ikan.ne	et	
13	坂の上の	の異人	館(旧中国	領事館	′6500002	′2401				0120-888								
	ベンの暑				′6500002	′2401				0120-888								
	洋館長月	屋(仏藤	朝西館)		′6500002	′2401				0120-888								
	英国館				′6500002	′2401				0120-888						ikan.ne	et	
17			一ト不思議			′2401				0120-888					wb あり			
18			トン装飾	美術館)		′2401				078-271-					なし			
	神戸北野				′6500002	′2401				078-251-								
	北野工店	_			′6500004	′2401				078-221-						<u>ス、オ</u> ,	ストメイ	
			RVISH MU			′2401				078-291⊸								
			推進センタ	7—	′6500006	′2401				078-351-								ク有
			レージアム		′6500021	′2402				078–327⊸								
			号館24階層			′2402				078-331-8	フラワーロ	平日 8:15	12月29日					
			理センター			′2402	神戸市中							http://v				fet:
					′6500004	′2403				078-351-								
					′6500034	′2404				078-393-				http://v				ım. 🤇
28	神古市	ケ†歯粉!	こうべしれ	つけてふ	16500024	'24 <u>0</u> 4	神戸市山-	24 69727	125 1929	In79-391-4	1777年	ila∙20 – 17∙	4日/如终	Platte://	动状协动	1.70	まま	

ハンズオンの概要



■ オープンデータのLOD化①

- 1.「CSV2LOD」を使いRDFに変換する
- 2. LODの公開
 - ①RDFファイル(Turtle形式)をWebサイトに公開して閲覧
 - ②データをRDFデータベース(Fuseki)に格納する

■ <u>オープンデータのLOD化②【外部リンク付き】</u>

- 3.「なんでもリンク」を使いWikidataへのリンク情報 をCSVファイルに追加する
- 4. リンク情報付きのRDFをRDFデータベース (Fuseki)に格納して、Federateクエリを試す

Simple LODIを用いたLODの公開

■ 5. 「Simple LODI」を使い公開

O. CSVファイルの準備



1. CSVファイルを準備する

 練習では、神戸市オープンデータ: 観光施設情報にWikidata へのリンクを追加したもの https://github.com/KnowledgeGraphJapan/LODws2nd/blob/ master/tourism_od2810%2Bwd.csv を利用

2. Google Sheets(またはExcel等)でCSVファイルを開く

- Google Sheetsでは「File」→「Open」→「Upload」
- エクセル等を利用してもよいが「UTF-8」のファイルを開く際には、 注意が必要

3. 形式に問題がないかを確認する

- 1 1行目が項目, 2行目以降がデータ(値)になっているか?
- 2. 1行目の項目と2行目以降にずれがないか?
- 3. 1列目に空白セルが含まれていないか?
- 4. 不要な行がないか?

1. CSV2LODの作業手順

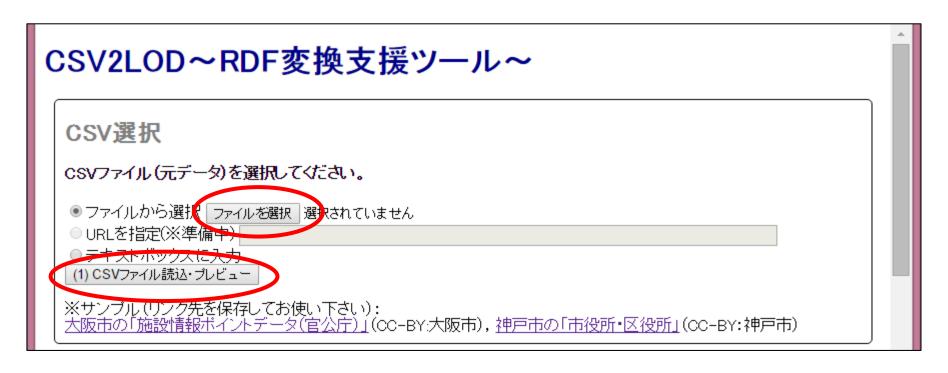


- 作業の手順
- 1. 元データ(CSV形式)を選択して読み込む
- 2. 変換に必要な情報を入力
 - URIに関する基本情報
 - 利用するプロパティ(関係の名称)
- 3. RDF変換用のテンプレートを生成
- 4. CSVをRDFに変換(今回はTurtle形式のみ)
- 5. RDFファイルをダウンロード

CSVファイルの読み込み



- 1. 「ファイルを選択」ボタンで、あらかじめ用意し ておいたCSVファイルを選択する
- 2. 「CSVファイル読み込み・プレビュー」ボタンを 押し、正しく読み込めることを確認する



CSVのプレビュー結果(例)





CSVファイル読み込み時の注意



- 文字コードは自動判定されます
 - サンプルデータについては動作確認済み
- ファイルサイズが大きい(200KB以上?)と読 み込めない場合があります
 - テキストエディタ等で開いて、「テキストボックスに入力」の欄にコピー&ペースとして読み込むと、ある程度のサイズのデータでも読み込めます。
- 「ファイルを選択」がうまく動かないとき

「テキストボックスに入力」を利用して下さい

テンプレート作成① メタデータの入力



■ 変換するデータに関する基本情報を入力する



今日のハンズオンでは



ベースIRI

- http://data.lodosaka.jp/lodws/ファイル名とする(拡張子はつけない)
- ファイル名は、練習では「作業者の名前(例: kozaki)」とする. (実際の公開時には、公開するURLを元に検討する)
 - 日本語ファイル名は避けたほうが無難(?)→環境によっては、上手く認識しない場合あり。
- オプションは「データベース(トリプルストア)を用いて公開(/)」を選択

■ ライセンス

- 神戸市のオープンデータは、すべてCC-BYで公開されているものなの、下記のように選択・入力する
 - クレジットで表示すべき名前:神戸市,作業者
 - ライセンスは, クリエイティブ・コモンズ(CC)の「CC-BY」

ベースIRI(URI)とは、



■ ベースIRIとは、

- RDFにおいてIDとするIRIを決める際に基準とするIRI
- 例) Wikidataの場合は
 http://www.wikidata.org/entity/Q41765
 ↑この部分に相当
- RDFファイルでは、Prefixがついていないリソース (データ)のIRIは、<u>行頭にベースIRIが付加される</u>もの と扱われる
- 例)ベースIRIがhttp://test.data/とすると そのRDFファイルで<1>というリソースのIRIは <http://test.data/1>となる
- ※Turtle形式の@baseには、#のIRIは使えない。

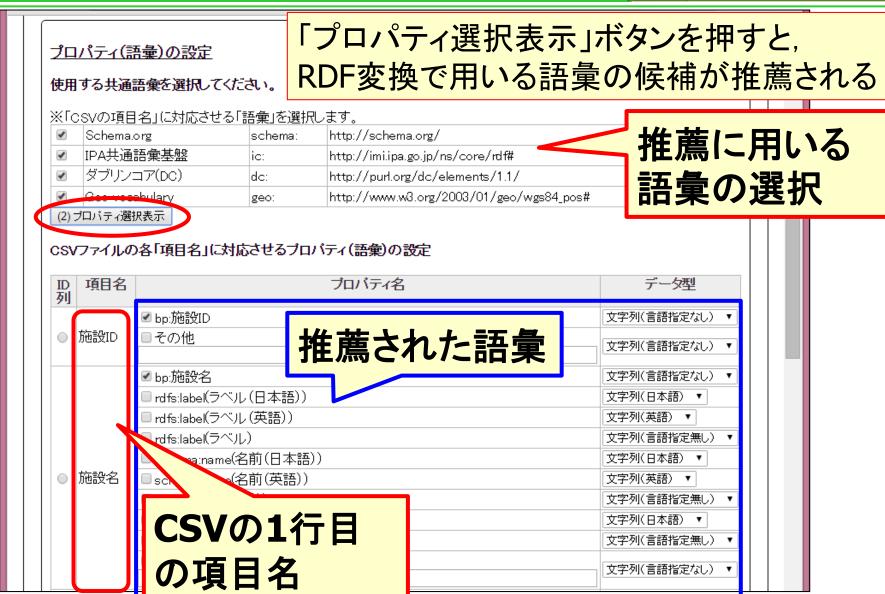
ベースIRI(URI)の決め方



- 各データへはベースIRIによって決まるIRIでアクセスされるので、
 - データの公開者が管理している(できる)IRI
 - データを公開する際に用いるIRI(URL) を用いることが望ましい
- # と / の使い分けは, 一般に,
 - 1つRDFファイルでデータを公開するときは http://test/data#001 のように#を
 - データサイズが大きく、RDFデータベースを用いて 公開するときは http://test/data/001 のように/を使うことが多い

テンプレート作成② プロパティ (語彙)の設定





語彙選択の考え方



■ ID列

- CSVファイル内で「一意のID」となる項目があれば、「ID列」として選択する
- なければ「IDを自動付与」を選択する

■ プロパティ

- 適切な語彙が推薦されていたら選択
- 今回,最低限,下記のプロパティを用いる (他のプロパティが使われる場合もある)

■ 名称 rdfs:label

■ 説明 rdfs:comment

Webサイト foaf:homepage

■ 緯度 geo:lat

■ 経度 geo:long

♣ 外部LODとのマッピング情報 rdfs:seeAlso

バグがあって 正しく動かないため, 選択しないで下さい.

22

語彙選択の考え方



- プロパティ(続き)
 - それ以外の語彙を利用する場合「その他」欄に入力
 - ■「推薦に用いる語彙」の一覧にあるprefixは利用可能
 - それ以外の場合は、語彙のIRIをすべて記載 →よく使われる語彙の一覧は次スライド参照.
 - CSVの項目名を利用するときは「bp:項目名」を選択
 - bp: は「ベースIRI+/property#」を表すprefix
 - データ型は、できれば適切なものを選択する

※プロパティを複数選択することも可

→複数の語彙に対応したいときに利用 (例:共通語彙基盤+Schema.org)

共通語彙



Schema.org http://schema.org/docs/schemas.html

Schema.org(日本語訳サイト) http://schema-ja.appspot.com/

<u>共通語彙基盤(IMI)</u> <u>http://imi.ipa.go.jp/</u>

Linked Open Vocabularies (LOV)

http://lov.okfn.org/dataset/lov/

DBPedia

http://mappings.dbpedia.org/index.php/Main_Page

Dublin Core

http://dublincore.org/documents/dcmi-terms/

→解説 http://www.kanzaki.com/docs/sw/dublin-core.html

Friend of a Friend (FOAF)

http://xmlns.com/foaf/spec/

→解説 http://www.kanzaki.com/docs/sw/foaf.html

語彙の推薦について



- 現状では、推薦できる語彙は「ごく一部」のものに限られています
 - 共通語彙基盤には、ごく一部のプロパティのみ、 試験的に対応。
- 今後, 対応する語彙は順次増やす予定です

RDFファイルへの変換



RDFファイルへの変換

(3) テンプレート生成便新 上記で設定した情報でRDF変換テンプレートを生成します テンプレートのダウンロード 上記のRDF変換テンプレートをダウンロードします (4)「CSV→RDF」の変換実行 下記のRDF変換テンプレートを用いてOSVファイルをRDFファイルに変換します

RDF変換テンプレート(※下記を直接編集することもできます)

```
##baseIRI=http://data.lodosaka.jp/lodc-kobe2015/kobe/institution01_20141128###ID=[[施設ID]]
##cc:attributionName=神戸市,古崎
##c:license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.com/license=http://creativecommons.c
```

- 「テンプレート生成/更新」: 設定した情報に基づいてCSVをRDFに変換するテンプレートを作成する
- 「テンプレートのダウンロード」で、作成したテンプレートをダウンロードして再利用できる
- 「CSV→RDFの変換実行」で、CSVファイルをRDF に変換する

既存のRDF変換テンプレートの読み込み

- ファイルから選択 ファイルを選択 選択されていません
- ●URLを指定

読み込みを実行

RDFファイルへの変換結果



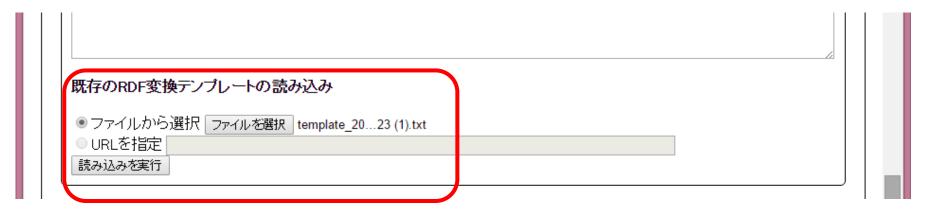
- 変換結果は下記のように表示される
 - 「ファイル名」を入力し「RDFファイルのダウンロード」ボタンで RDFをダウンロードする. →ファイル名は「作業者名(練習時)」 か「元ファイル名」+ Lttlとする
 - うまくダウロードできない場合は、コピー&ペーストを利用する

```
RDFファイルへの変換結果
ファイル名:|institution01_20141128 (1).csv_201509120512.ttl
@base <http://data.lodosaka.jp/lodc-kobe2015/kobe/institution01_20141128#> .
@prefix bp: <http://data.lodosaka.jp/lodc-kobe2015/kobe/institution01 20141128/property#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
@prefix xsd: <http://www.w3.org/2001/XMLSchema>
@prefix cc: <http://creativecommons.org/ns#> .
@prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#> .
<> cc:attributionName "神戸市,古崎";
    cc:license <a href="http://creativecommons.org/licenses/by/4.0/deed.ja">http://creativecommons.org/licenses/by/4.0/deed.ja</a>.
<a href="http://data.lodosaka.jp/lodc-kobe2015/kobe/institution01_20141128#1">hp:施設ID "1";</a>
 NILP://data.lodosaka.jp/lodc-kobe2U
bp:施設名 "神戸市役所";
rdfs:label "神戸市役所";
bp:郵便番号 "650-8570";
bp:電話番号 "078-331-8181(代)";
geo:lat "34.689454"^^xsd:double;
geo:lng "135.195739"^^xsd:double;
bp:行政区コード"中央区";
| bp:カデゴリリスト "市役所・区役所"。
<http://data.lodosaka.jp/lodc-kobe2015/kobe/institution01_20141128#2> | bp:施設ID "2";
 bp:施設名 ″東灘区役所″;
rdfs:lahel ″東灘区役所″
 (5) RDFファイルのダウンロード ※うまくダウンロードできない場合はこちらをご覧ください
```

保存したRDFテンプレートの利用



- ダウンロードした「RDF変換テンプレート」を読み込んで利用することも可能
 - 同じ形式のCSVファイルが複数ある場合に有効
 - 設定画面へは反映されないため「テンプレート生成/更新」ボタンを押すと、読み込んだテンプレートが破棄されるので注意
- テンプレートは、直接、画面上で編集することも可能
 - 読み込んだテンプレートの修正や、設定からは生成できない 複雑なRDF変換用のテンプレートを生成するのに利用できる



2. LOD公開



■ 作成したRDFファイルをWebサイトで公開する

- 今回は、イベント用のWebサイトに公開
- 公開するURLは, http://data.lodosaka.jp/lodws/ファイル名
- アップする前に、RDFのファイル名がベースIRIで指定したもの+.ttlになっているか確認
 例)http://data.lodosaka.jp/lodws/kozakiならば kozaki.ttl

LOD公開(1): Webサイト



1. FTPクライアントへログイン

- <u>https://webftp.heteml.jp/</u> を利用
 - ユーザー名:koujikozaki_lodws
 - パスワード: kobe2017
- FTPクライアントを利用する場合は
 - ホスト名(アドレス) ftp187.heteml.jp
 - ※FTPS 接続 の場合は ssl187.heteml.jp(推奨)

2. FTPでアップロード

「ファイルを選択」ボタン
 →作成したRDF(****.ttl)を選択してアップロード
 ※上書き時には、「上書き」にチェック

LODの閲覧



- ①Webブラウザでの閲覧
 - WebブラウザのURL欄に
 - http://data.lodosaka.jp/lodws/ファイル名
 - のように、IRIを入れるとWebブラウザで閲覧可能
- ②LODブラウザでの閲覧
 - **▶ 下記のLODブラウザのIRIの欄に入力する**
 - Quick and Dirty RDF browser※日本語IRIは文字化けする

http://graphite.ecs.soton.ac.uk/browser/

(Yet Another) Linked Data Browser

※フォーマットでTurtleを選択する必要あり

http://www.kanzaki.com/works/2014/pub/ld-browser

LODの公開2 DBへのRDFデータの登録



■ Fusekiのインストール

「トリプルストアの導入」のハンズオンの方法を参考に、Fusekiをインストール https://goo.gl/fUtv5m

■ Fusekiで新しいDBを作成

- Manage datasets > add dataset
- 作成したDBにRDFファイルをアップロード
 - upload data
- SARQLクエリを試す

Federateクエリの利用



外部のLODと連携してのクエリの利用例

PREFIX rdfs: http://www.wikidata.org/prop/direct/>

```
SELECT ?s ?wd ?l
WHERE {
?s bp:参照 ?wd

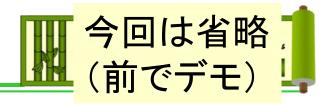
SERVICE <https://query.wikidata.org/sparql>
{
?wd rdfs:label ?l.
FILTER(lang(?l)="en")
}
WikidataのSPARQLエンドポイント(機械用)
```

2017/9/22

て、結果をまとめて返す

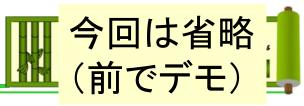
から、英語のラベルを取ってくるクエリを実行し

3. 外部LODとのリンク



- 1. なんでもリンクにアクセス
- 2. オープンデータ(CSV)の「名称」に相当する列をコピー
- 3. なんでもリンクの入力欄にペースト
- 4. 「SPARQLエンドポイント」欄に <u>https://query.wikidata.org/sparql</u> を入力
- 5. 「割当実行」DBpediaのデータと一致するデータの一覧 が表示される(最初の10件)
- 6. 「ダウンロード」ボタンで**結果をダウンロード**
- 7. エクセル等で開き、元のオープンデータの最終列に マッピング情報をコピー&ペースト
 - 項目名は「参照」としておく

マッピング情報に関する注意

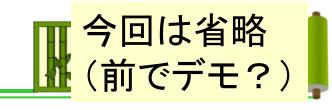


- マッピング情報はUTF8形式となっているので、 以下のように扱う
 - Google Sheetsでは 「File」→「Open」→「Upload」
 - Excelの場合は、 「データ」→「挿入」→「外部データの取り込み」→「テキストファイル」

■ 以下の作業は、作業1.2と同様

マッピング情報に用いる「プロパティ」の選択のみ気 を付ける

5. Simple LODIの設定



- 1. Simple LODIをダウンロード(ZIP形式が簡単)して、解凍する.
 - https://github.com/uedayou/simplelodi から
- 2. 「simplelodi-master」のフォルダ名を、LODを 公開するフォルダ名に変更
 - 例)http://data.lodosaka.jp/lodws/kozaki ならば「kozaki」
- 3.「2のサブフォルダdata」に「5. で分割した. ttlファイルを格納」する
- 4.「2のフォルダー式」を http://data.lodosaka.jp/lodws/ にアップロードする ※FTPソフトを利用する(例: FFFTP, など)

公開したLODの確認



- 公開したLODは,
 - http://data.lodosaka.jp/lodws/フォルダ名/ファイル名.ttl
 - のアドレスでアクセスできる.
- 拡張子を変更することで、様々な形式で取得可
 - .ttl → Turtle形式
 - .nt → N-Triple形式
 - .xml → RDF/XML形式
 - .html → HTML形式
 - .json → JSON形式
 - ___isonId → JSON-LD形式