

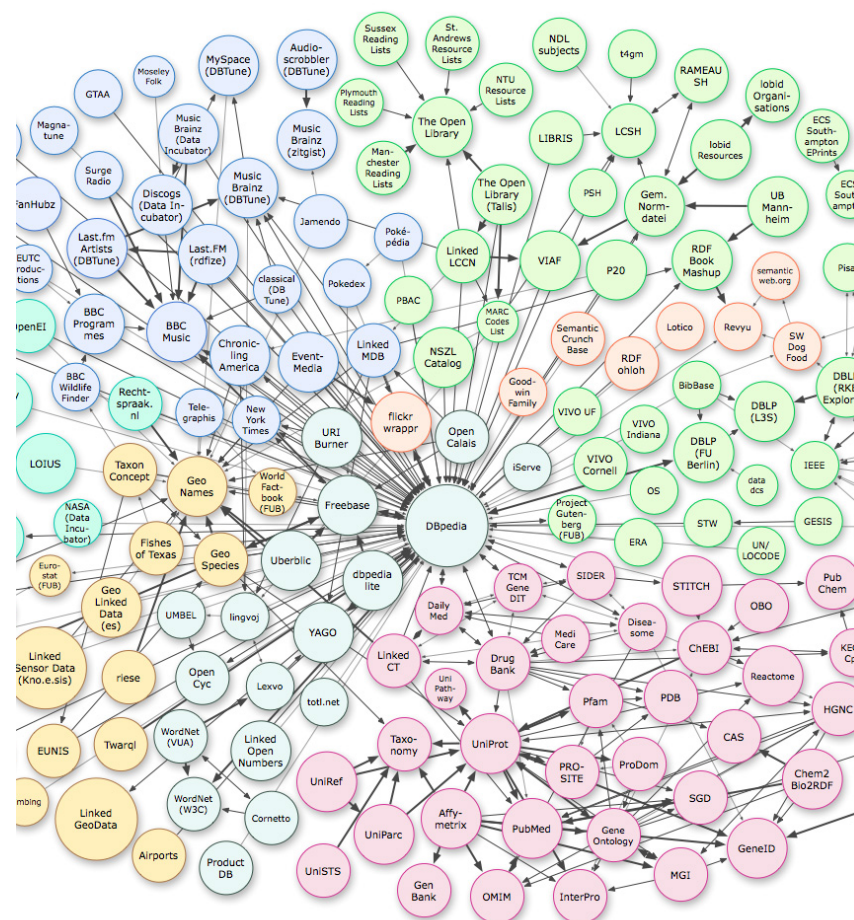
LODPaddle: Linked Open Data for French Region Pays De la Loire

Hala Skaf-Molli
Associate Professor
Nantes University, France
GDD Team, LINA
Hala.skaf@univ-nantes.fr
<http://pagesperso.lina.univ-nantes.fr/~skaf-h>
Nantes 15-mars-2013



LODPaddle Objectives

- Publish regional open data as semantic data interlinked with LOD.
 - Linked Data ¹ provides a set of best practices to publish and interconnect information on the web, creating the so-called "Web of Data".
 - Linked Data simplifies greatly the task of data integration, allowing data consumers to easily combine different datasets to perform complex queries.



¹*Christian Bizer, Tom Heath and Tim Berners-Lee. Linked Data - The Story So Far. International Journal on Semantic Web and Information Systems, 2009*

Problems ?

- Region published many datasets as open data and not linked data
 - 416 datasets
- Datasets need to be cleaned, linked (entity matching), aligned with other linked data before being really useful (i.e. accessed by queries).
- This processing is costly and can be mutualized.




With current datasets, I want:

- Les bibliothèques de Nantes ouvertes Mercredi et Samedi.
- Les galeries d'art qui possèdent un site web.
- Les noms des médiathèques à Rezé.
- La liste de musées de Nantes ouverts pendant les vacances de Toussaint.
- It is possible...
- ... but costly
- All processing required to express these queries has to be done by each open data developer...

Rechercher des données

Mots clés :

 Valider

1 à 10 sur 416 résultats

Trier par : [pertinence](#) | [date de publication / mise à jour](#)

Horaires des Transports Express Régionaux (TER) en Pays de la Loire

Mis à jour le 13/03/2013

[API](#) [GTFS](#)

Horaires des Transports Express Régionaux (TER) en Pays de la Loire. Les horaires des TER sont publiés sur le site Open Data de la SNCF:

[Site Open Data de la SNCF](#)

[Pour plus d'information](#)

Thématiques : Mobilité

Gestionnaire : Région des Pays de la Loire

Propriétaire : SNCF

 [Consulter la fiche détaillée](#)

Services ferroviaires régionaux de voyageurs prévision budgétaire en Région des Pays de la Loire

Mis à jour le 12/03/2013

[CSV](#) [JSON](#)














[XLS](#) [XML](#)

Affiner votre recherche par :

Thématique

- ☐ Citoyenneté / Institution (85)
- ☐ Santé / Social (81)
- ☐ Mobilité (74)
- ☐ Culture / Tourisme (49)
- ☐ Éducation / Formation (38)
- ☐ Environnement (31)
- ☐ Territoires (30)
- ☐ Économie (28)
- ☐ Sport (11)
- ☐ Logement (9)
- ☐ Urbanisme (8)
- ☐ Jeunesse (1)

Type de fichiers

- | | | | |
|---|---|------------------------------------|---|
| <input type="checkbox"/> CSV (328) |  | <input type="checkbox"/> XLS (322) |  |
| <input type="checkbox"/> JSON (249) |  | <input type="checkbox"/> XML (244) |  |
| <input type="checkbox"/> SHP (L93)... |  | <input type="checkbox"/> KMZ (67) |  |
| <input type="checkbox"/> SHP (CC...)... |  | <input type="checkbox"/> KML (17) |  |
| <input type="checkbox"/> PDF (16) |  | <input type="checkbox"/> API (7) |  |
| <input type="checkbox"/> GTFS (3) |  | <input type="checkbox"/> MIF (3) |  |
| <input type="checkbox"/> JPG (3) |  | | |

Rechercher des données

Mots clés :

Valider

1 à 10 sur 416 résultats

Trier par : pertinence | date de publication / mise à jour

Horaires des Transports Express Régionaux (TER) en Pays de la Loire

Mis à jour le 13/03/2013

API GTFS

Horaires des Transports Express Régionaux (TER) en Pays de la Loire. Les horaires des TER sont publiés sur le site Open Data de la SNCF:

[Site Open Data de la SNCF](#)

[Pour plus d'information](#)

Thématiques : Mobilité

Gestionnaire : Région des Pays de la Loire

Propriétaire : SNCF

Consulter la fiche détaillée

Services ferroviaires régionaux de voyageurs prévision budgétaire en Région des Pays de la Loire

Mis à jour le 12/03/2013

CSV JSON

XLS XML

Eléments du budget prévisionnel concernant les services régionaux de voyageurs en Région des Pays de la Loire. Source Budget Primitif 2012 et 2013.

Affiner votre recherche par :

Thématique

- ☐ Citoyenneté / Institution (85)
- ☐ Santé / Social (81)
- ☐ Mobilité (74)
- ☐ Culture / Tourisme (49)
- ☐ Éducation / Formation (38)
- ☐ Environnement (31)
- ☐ Territoires (30)
- ☐ Économie (28)
- ☐ Sport (11)
- ☐ Logement (9)
- ☐ Urbanisme (8)
- ☐ Jeunesse (1)

Type de fichiers

- | | |
|---------------------------------------|------------------------------------|
| <input type="checkbox"/> CSV (328) | <input type="checkbox"/> XLS (322) |
| <input type="checkbox"/> JSON (249) | <input type="checkbox"/> XML (244) |
| <input type="checkbox"/> SHP (L93)... | <input type="checkbox"/> KMZ (67) |
| <input type="checkbox"/> SHP (CC...) | <input type="checkbox"/> KML (17) |
| <input type="checkbox"/> PDF (16) | <input type="checkbox"/> API (7) |
| <input type="checkbox"/> GTFS (3) | <input type="checkbox"/> MIF (3) |
| <input type="checkbox"/> JPG (3) | <input type="checkbox"/> RDF(200) |

Les galeries d'art qui possèdent un site web.

```
prefix dbpedia: <http://dbpedia.org/resource/>
prefix foaf: <http://xmlns.com/foaf/0.1/>
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?name ?homepage
WHERE {
    ?place rdfs:label ?name .
    ?place rdf:type dbpedia:Museum .
    ?place foaf:homepage ?homepage
}
```

Les galeries d'art qui possèdent un site web.

name	homepage
"Musée Naval Maillé Brézé"	< http://www.nwt.fr/escorteur-maille-breze >
"Musée des Beaux Arts"	< http://www.nantes.fr/mairie/services/responsabilites/dgc/musees >
"Musée Jules Verne"	< http://www.nantes.fr/mairie/services/responsabilites/dgc/musees >
"Musée Dobrée"	< http://www.cg44.fr/cg44/Special/Dobree/index.html >
"Planétarium Nantes"	< http://www.nantes.fr/mairie/services/responsabilites/dgc/musees >
"Muséum d'Histoire Naturelle"	< http://www.museum.nantes.fr >
"Musée d'Histoire de Nantes"	< http://www.chateau-nantes.fr >
"Musée de l'Imprimerie"	< http://musee-imprimerie.com >

Steps to achieve this query

- Input data set:
 - *Equipements publics culturels (CSV)*
- Vocabulary alignment
 - NOM_COMPLET -> rdfs:label
 - WEB -> foaf:homepage
- Toolkits: googleRefine¹, DataLift²
- Data cleaning:
 - “Musée de l’histoire de Nantes” -> “Musée de l’histoire de Nantes”
- *Output data set:*
 - *Equipements publics culturels (rdf)*

¹ <https://code.google.com/p/google-refine/>

² <http://datalift.org/>

Libellé	Description	Exemple(s)
IDOBJ	Identifiant numérique attribué par Nantes Métropole à chaque équipement public	« 4095,000000000000000 »
NOM_COMPLET	Nom complet de l'équipement culturel	« Musée de la Machine à Coudre »
THEME	Code d'identification du thème de l'équipement public	« 1,000000000000000 »
LIBTHEME	Libellé du thème de l'équipement public	« CULTURE »
CATEGORIE	Code d'identification de la catégorie de l'équipement culturel	« 108,000000000000000 »
LIBCATEGORIE	Libellé de la catégorie de l'équipement culturel	« Musées & Châteaux »
TYPE	Code d'identification du type (sous-catégorie) de l'équipement culturel	« 10301,000000000000000 »
LIBTYPE	Libellé du type de l'équipement culturel	« Ludothèque »
STATUT	Statut de l'équipement public	« Public »
COMMUNE	Nom de la commune de l'équipement de public	« THOUARE-SUR-LOIRE »
ADRESSE	Adresse de l'équipement public	« Avenue des Argonautes »
TELEPHONE	Numéro de téléphone de l'équipement public	« 02 28 44 98 91 »
WEB	Adresse du site internet de l'équipement public	« www.pathe.fr »
CODE_POSTAL	Code postal de la commune de l'équipement public	« 44700 »
LONG_WGS84	Longitude de l'équipement public dans le système WGS 84	« -1,504202418358 »
LAT_WGS84	Latitude de l'équipement public dans le système WGS 84	« 47,287959592088 »

TAB. 2 – Champs du fichier CSV des équipements culturels.

Vocabularies alignment

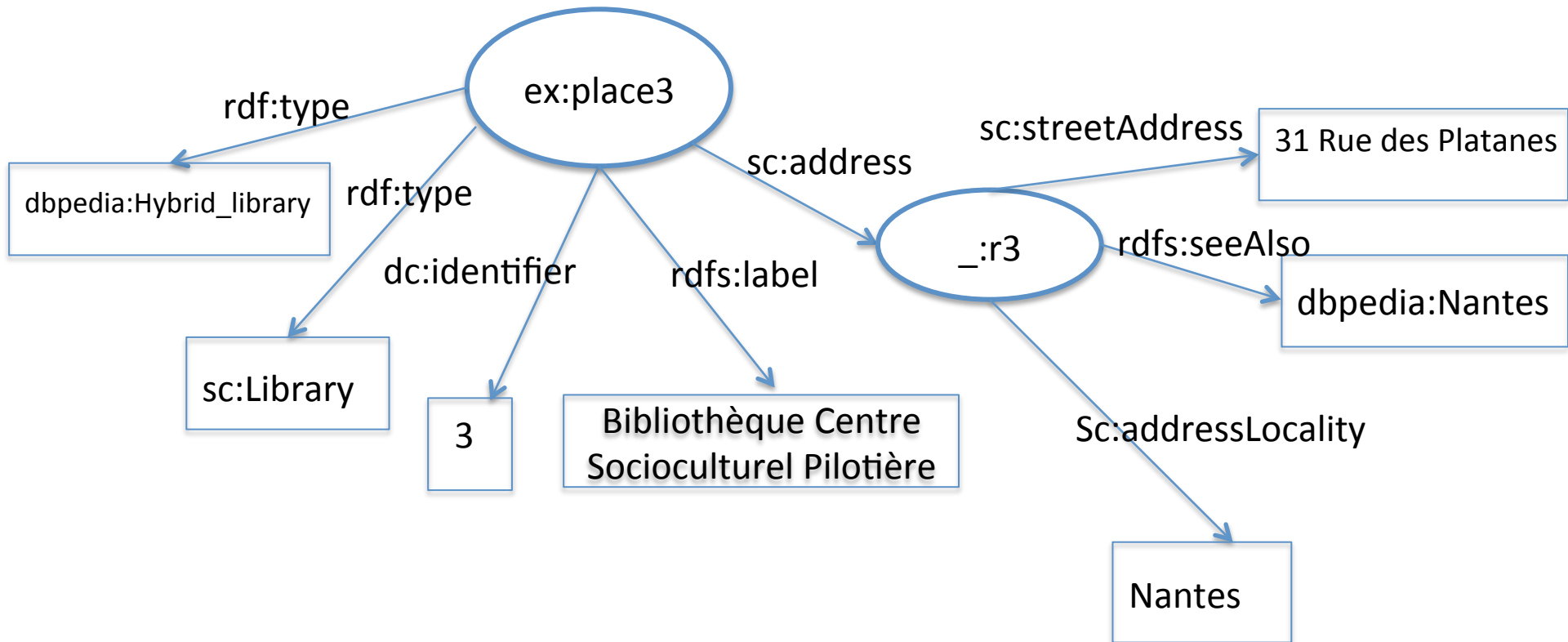
Champs	Prédicat
IDOBJ	dc:identifier
NOM_COMPLET	rdfs:label
LIBTHEME	dcterms:subject
LIBCATEGORIE	rdf:type
LIBTYPE	rdf:type
COMMUNE	sc:addressLocality
TELEPHONE	foaf:phone, sc:telephone
WEB	foaf:homepage
LONG_WGS84	geo:long
LAT_WGS84	geo:lat

foaf: <http://xmlns.com/foaf/0.1/>
 sc: <http://schema.org/>

TAB. 3 – Prédicats RDF utilisés lors de l'étape 1

Champs	Valeur	Classe
Thème	Culture	dbpedia:Culture
Catégorie	Cinéma	dbpedia:Movie_theater
Catégorie	Médiathèque	dbpedia:Hybrid_library
Catégorie	Monument	dbpedia:Monument
Catégorie	Musées & Châteaux	dbpedia:Museum
Catégorie	Salle d'exposition	dbpedia:Art_museum
Catégorie	Salle de spectacle	dbpedia:Auditorium
Type	Bibliothèque	dbpedia:Library
Type	Château	dbpedia:Castle
Type	Ludothèque	dbpedia:Toy_library

Extract of RDF graph of “équipement culturels”



Steps to achieve this query

- Input data set:
 - *Equipements publics culturels (CSV)*
- Vocabulary alignment
 - NOM_COMPLET -> rdfs:label
 - WEB -> foaf:homepage
- Toolkits: googleRefine¹, DataLift²
- Data cleaning:
 - “Musée de l’histoire de Nantes” -> “Musée de l’histoire de Nantes”
- *Output data set:*
 - *Equipements publics culturels (rdf)*

¹ <https://code.google.com/p/google-refine/>

² <http://datalift.org/>

Les bibliothèques de Nantes ouvertes

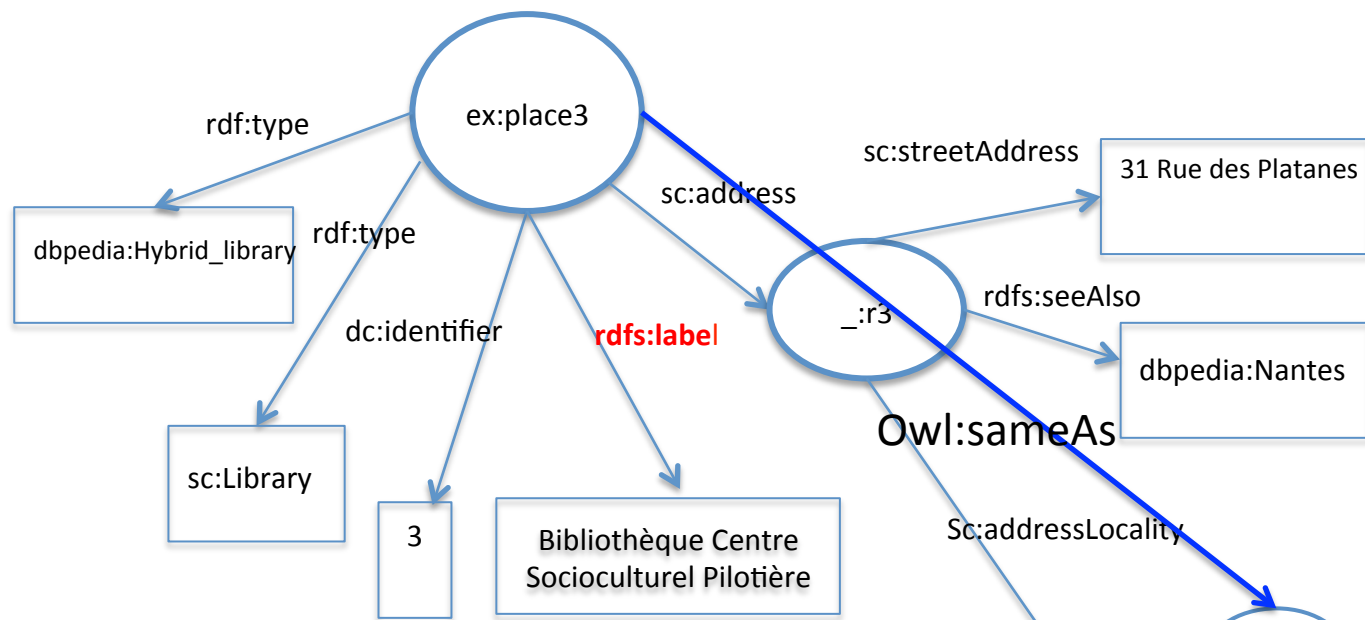
Mercredi et Samedi

```
SELECT ?nom ?adresse
WHERE {
  GRAPH <http://example.org/equipements> {
    ?place rdfs:label ?nom .
    ?place sc:address ?address .
    ?address sc:streetAddress ?adresse .
    ?place rdf:type dbpedia:Library .
    ?address sc:addressLocality "NANTES"
  }
  GRAPH <http://example.org/horaires> {
    ?place2 foaf:name ?nom .
    ?place2 ical:Vcalendar ?cal .
    ?cal ical:Vevent ?event .
    ?event ical:rrule ?rule .
    ?rule ical:byday ?days .
    FILTER fn:contains(?days, "WE")
    FILTER fn:contains(?days, "SA")
  }
}
```

nom	adresse
"Bibliothèque Chantenay"	"8 Rue de la Constitution"
"Bibliothèque La Halvêque"	"23 Rue Léon Serpollet"
"Bibliothèque Manufacture"	"8 Cour Jules Durand"
"Bibliothèque Breil Malville"	"37B Rue des Plantes"

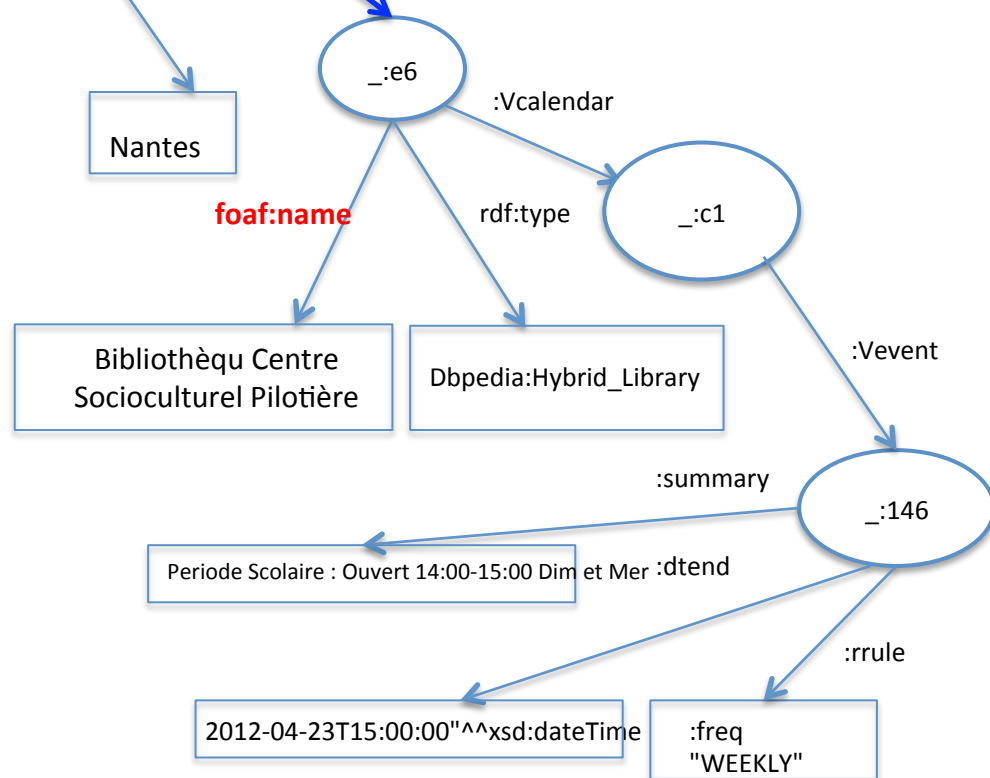
Main problems to achieve this query

- Input data sets:
 - *Equipements publics culturels et horaires des équipement culturels*
- Same as before...
- Entity matching between datasets.
 - `ex:place3 owl:sameAs _:e2` (place3 in Equipement and e2 in horaires)



équipement culturel

Owl:sameAs



Horaire équipement culturel

Impacts

- Any web developer can now build a linked data application without the burden of open data cleaning/matching/aligning.
- Alignment with LOD can enlarge the scope of queries
 - i.e. Horaires des lieux culturels référencés dans Wikipedia? (3 datasets involved: 2 extracted from local open data and dbpedia).

To DO

- Continuous semantification of regional open data
 - Include data cleaning, alignment and matching
 - Perennity of queries
- Endpoint for LOD federation
- Support for web developer (example, tutorial)

Conclusion

- Transforming and publishing structured raw datasets into linked data should allow more complex queries
- Do it once for all consumers
- Give new opportunities for innovative applications.
- Give the opportunities for the region to be a leader in this domain.

Open Questions ?

- Best practices to:
 - **continuously** transform all data of <http://data.loire-atlantique.fr/donnees/> into LOD
 - maintain a **reliable** endpoint? **Consistent** RDF datasets?
- Who pay?
 - For infrastructure, continuous transformation, formation of data producers and data consumers.

Acknowledgements

- Thanks to master students “M1 Informatique” at the university of Nantes for their contributions to lodpaddle.

Fin