

Web Sémantique

-

Rapport

·X1I1030·

Hippolyte Dubois, Florent Gaillard, Marvin Jean

Novembre 2017

Table des matières

Introduction	2
1 Choix du Dataset et Sémantification	3
2 Requêtes	4
2.1 Dataset seul	4
2.2 Liaison à un autre dataset	4
3 Inférences	5
4 Liaison au cloud Linked Data	6
4.1 Sélection d'un URI	6
5 Description VOiD	7
6 Annexe	8
6.1 Construct Query	9
6.2 Requêtes	10

Introduction

Dans le cadre du cours X1I1030, il nous a été demandé de choisir un dataset libre répertorié sur le site <https://data.esr.gouv.fr> afin de le sémantifier.

Nous présenterons d'abord la méthode que nous avons utilisé afin de transformer les données au format CSV en données RDF, puis quelques requêtes réalisées sur le dataset, d'abord seul puis lié à celui d'un autre groupe.

Ensuite, nous présenterons les inférences réalisées sur le dataset, nous proposerons des façons de le lier au cloud de [Linked Data](#), et finalement décrirons notre dataset au moyen d'un vocabulaire VOID.

Chapitre 1

Choix du Dataset et Sémantification

Nous avons choisi d'utiliser un dataset concernant [les moyens consacrés à la R&D par les administrations](#).

Ce dataset est le fruit d'une enquête concernant les moyens consacrés à la R&D réalisé par le Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche en Octobre 2015. Les données ont été modifiées pour la dernière fois en octobre 2016.

Ces données sont structurées dans une table avec les headers suivants :

- Le code INSEE de la région de l'organisme (codes des anciennes régions pré réforme)
- Le nom de la région
- L'année du financement
- Le code de l'indicateur
- Le label complet de l'indicateur
- Le code du type d'administration
- Le label complet du type d'administration
- Le code du sexe du bénéficiaire
- Le label complet du sexe du bénéficiaire
- Le code du type de personnel
- Le label complet du type de personnel
- L'état des données
- La valeur de l'indicateur

Pour sémantifier le dataset nous avons dû l'épurer légèrement en ignorant les colonnes redondantes tels que le nom de la région (trouvable en liant notre dataset à celui de l'INSEE).

Le fichier RDF a été construit au moyen de [tarql](#). Pour la requête Construct, se référer à [Figure 6.1](#).

```
1 <2.53> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> dbo:Statistic ;
2     igeo:codeRegion "52" ;
3     gn:name "Pays de la Loire" ;
4     dbo:creator "4" ;
5     dbo:alias "Autres administrations" ;
6     dbo:year "2009" ;
7     dc:type "Depense interieure de R&D" ;
8     dbo:status "definitif" .
```

FIGURE 1.1 – Exemple de tuple produit par la requête.

Chapitre 2

Requêtes

2.1 Dataset seul

La premiere requête a pour but d'afficher le nombre de projet subventionnés groupés par année, qui ont comme chef d'équipe une femme. On peut ainsi constater l'évolution du nombre de projet dirigés par des femmes au court des années. Voir le code ici : [Figure 6.2](#).

Cette seconde requete permet d'etudier le dynamisme des régions au fil des années. Le type de projet sert à differencier les secteurs, pour une étude qui pourrait être plus spécifique. Voir le code de la requete ici : [Figure 6.3](#).

2.2 Liaison à un autre dataset

La liaison avec d'autres datasets se fait sur les codes régions. Nous avons choisi le dataset du groupe Alapetite - Bellot - Boudine, en essayant de voir s'il y a de potentielles corélations entre l'attribution de financements par les administrations et le taux de réussite à ce concours.

Pour exploiter la liaison de notre graph avec un autre possédant aussi des region, nous pouvons regarder les régions ayant une subvention et ayant un participant au concours étoile. Voir le code de la requête ici : [Figure 6.4](#)

Chapitre 3

Inférences

Via apache fuseki nous avons pu inférer des triples et voir la modifications sur notre data set. Le changement est l'apparition de nouveau triple concernant le modèle du data set, par exemple :

```
1 1 rdfs:domain rdfs:domain rdf:Property
2 2 rdfs:domain rdfs:range rdfs:Class
3 3 rdfs:domain rdf:type rdf:Property
4 4 <http://www.w3.org/2002/07/owl\#DatatypeProperty> rdfs:subClassOf rdf:
   Property
5 5 <http://www.w3.org/2002/07/owl\#Property> rdf:type rdfs:Class
```

Ces tuples inférés sont le résultat de l'application des règles OWL sur les ontologies de chaque espèces de nommage (car nous avons tentés au mieux d'utiliser des nommages connus et cohérents à nos données). Les inférences ont donc déduit des propriétés par rapport aux ontologies des propriétés du modèle de notre data set.

Chapitre 4

Liaison au cloud Linked Data

4.1 Sélection d'un URI

La norme Linked Data stipule que tous les URIs doivent être des URIs HTTP. Cela soulève un vrai problème dans le cas de notre dataset, car les données sont anonymisées et que le seul élément permettant de séparer deux tuples est la valeur de l'indicateur.

Chapitre 5

Description VOiD

La description VOiD de notre dataset et de la jointure réalisées avec d'autres datasets, qu'ils appartiennent à un autre groupe ou qu'ils soient fournis par des organismes, permet de spécifier la structure de cette jointure.

```
1      @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns\#> .
2      @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema\#> .
3      @prefix foaf: <http://xmlns.com/foaf/0.1/> .
4      @prefix dcterms: <http://purl.org/dc/terms/> .
5      @prefix void: <http://rdfs.org/ns/void\#> .
6
7      :RFStats rdf:type void:Dataset ;
8          foaf:homepage <http://research-funding.org/> .
9
10     :RFStatsAdministration rdf:type void:Dataset ;
11         void:subset :RFStats ;
12         foaf:homepage <http://research-funding.org/administration/> ;
13         dcterms:subject <http://http://dbpedia.org/ressource/Research> .
14
15     :DBpedia rdf:type void:Dataset ;
16         foaf:homepage <http://dbpedia.org/> .
17
18     :INSEEgeo rdf:type void:Dataset ;
19         foaf:homepage <http://id.insee.fr/geo/region/> .
20
21     :EtoileDataset rdf:type void:Dataset ;
22         foaf:homepage <http://concours-etoiles-europe.org/> .
23
24     :RAnalysis rdf:type void:Linkset ;
25         void:target :RFStatsAdministration ;
26         void:target :DBpedia ;
27         void:target :INSEEgeo ;
28         void:target :EtoileDataset .
```

FIGURE 5.1 – Description VOiD du projet.

Chapitre 6

Annexe

6.1 Construct Query

```
1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#/>
2 PREFIX dbo: <http://dbpedia.org/ontology/>
3 PREFIX gn: <http://www.geonames.org/ontology#>
4 PREFIX dc: <http://purl.org/dc/elements/1.1/>
5 PREFIX igeo: <http://rdf.insee.fr/def/geo>
6
7 # Value
8 CONSTRUCT {
9     ?URI a dbo:Statistic ;
10     igeo:codeRegion ?URI_REGION ;
11     gn:name ?Region ;
12     dbo:sex ?Sexe ;
13     dbo:creator ?code_type_uai ;
14     dbo:alias ?type_administration ;
15     dbo:year ?Annee ;
16     dbo:roleInEvent ?type_personnel ;
17     dc:type ?Indicateur ;
18     dbo:status ?etat_donnees.
19
20     ?URI_REGION owl:sameAs ?URI_REGION_DBPEDIA;
21                 owl:sameAs ?URI_REGION_CONCOURS;
22                 owl:sameAs ?URI_REGION_INSEE.
23
24 }
25 }
26 FROM <file:dataset_rd.csv>
27 WHERE {
28     BIND (URI(?Valeur) as ?URI)
29     BIND (URI(CONCAT('http://research-funding.org/administration/region/', ?
30                     code_region) as ?URI_REGION)
31     BIND (URI(CONCAT('http://fr.dbpedia.org/page/', ?code_region)) AS ?
32             URI_REGION_DBPEDIA)
33     BIND (URI(CONCAT('http://id.insee.fr/geo/region/'), ?code_region) AS ?
34             URI_REGION_INSEE)
35     BIND (URI(CONCAT('http://concours-etoiles-europe.org/ns#', ?code_region)
36             ) as ?URI_REGION_CONCOURS)
37 }
```

FIGURE 6.1 – Requête construisant le fichier RDF.

6.2 Requêtes

```
1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
3 PREFIX ex: <http://example.org/>
4 PREFIX dbo: <http://dbpedia.org/ontology/>
5 PREFIX dc: <http://purl.org/dc/elements/1.1/>
6 select ?year ?type (count(?x) as ?countwomen)
7 where {
8     ?x dbo:sex "Femmes";
9         dbo:year ?year;
10        dc:type ?type.
11 }
12 group by ?year ?type
```

FIGURE 6.2 – Requête donnant le nombre de projet qui ont pour leader une femme et ce par an.

```
1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
3 PREFIX ex: <http://example.org/>
4 PREFIX dbo: <http://dbpedia.org/ontology/>
5 PREFIX dc: <http://purl.org/dc/elements/1.1/>
6 PREFIX gn: <http://www.geonames.org/ontology#>
7
8 select ?year ?region ?type (sum(?x) as ?countbudget)
9 where {
10     ?x gn:name ?region;
11         dbo:year ?year;
12         dc:type ?type.
13 }
14 group by ?region ?year ?type
```

FIGURE 6.3 – Affichage du nombre de projet par région/type/an.

```

1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
3 PREFIX ex: <http://example.org/>
4 PREFIX dbo: <http://dbpedia.org/ontology/>
5 PREFIX dc: <http://purl.org/dc/elements/1.1/>
6 PREFIX gn: <http://www.geonames.org/ontology#>
7 PREFIX dbp: <http://dbpedia.org/property/>
8 PREFIX igeo: <http://rdf.insee.fr/def/geo>
9 PREFIX owl: <http://www.w3.org/2002/07/owl#>
10
11 select ?y ?region
12 where {
13     ?x igeo:codeRegion ?region.
14     ?region owl:sameAs ?region_dataset.
15     GRAPH <http://concours-etoiles-europe.org/ns#> {
16         ?y dbp:region ?region_dataset
17     }
18 }
19 group by ?y ?region

```

FIGURE 6.4 – Affichage des participants au concours de l'étoile dans les régions ayant une subvention.