

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	8

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

2.2 Liaison à un autre dataset

Chapitre 3

Inférences

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.

On pourrait donc opérer une transformation sur notre dataset, et utiliser le code région comme URI, en le reliant aux valeurs au moyen d'un propriété de type :granted

Chapitre 5

Description VOID

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 ?code_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
21    FROM <file:dataset_rd.csv>
22    WHERE {
23        BIND (URI(?Valeur) as ?URI)
24    }
```

FIGURE 6.1 – Requête construisant le fichier RDF.