

# GMIN332 : Vision sources de données ouvertes

## 1. Préambule

---

Dans la continuité du premier Tp portant sur RDF, vous allez vous concentrer sur une exploitation mutualisée de jeux de données, référentiels de valeurs et standards de métadonnées. A cet titre, vous aurez à construire un jeu de données portant sur des livres. Ce jeu de données sera enrichi sémantiquement par des termes organisés en thésaurus, au travers de propriétés définies au sein du standard de métadonnées Dublin Core.

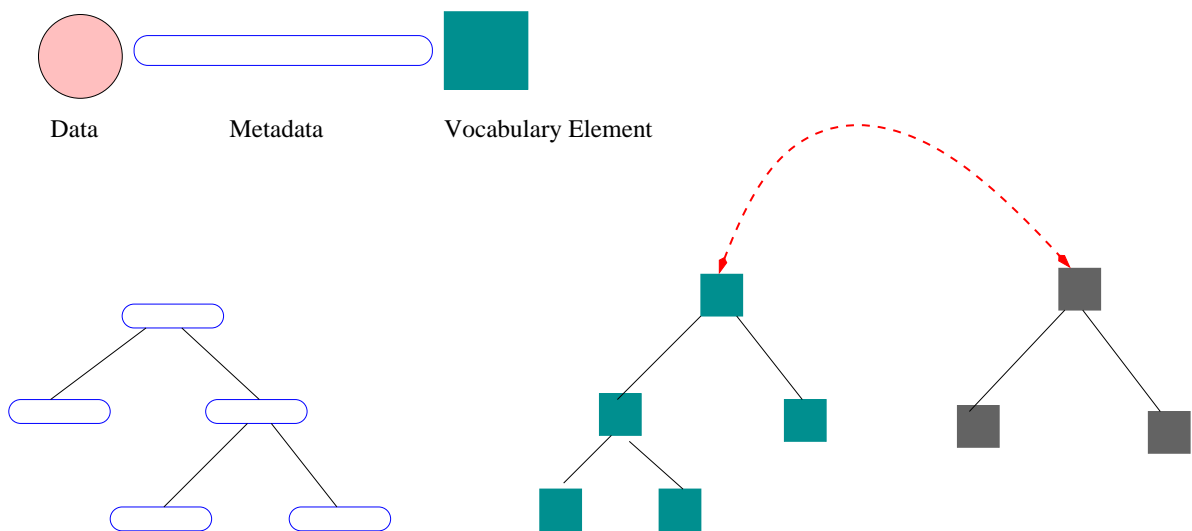


FIGURE 1 – Principes des sources de données ouvertes

## 2. Exercice 1 : Construire le jeu de données

---

### 2.1 Question 1

Construire le jeu de données à partir du fichier xml (sans chercher à construire un analyseur/convertisseur) et à partir de l'exemple donné (Livres\_1.java).

### 2.2 Question 2

Améliorer l'organisation du contenu en proposant deux jeux de données, un pour les livres, un pour les auteurs, et donc deux espaces de noms. Le jeu de données pour les auteurs doit s'adosser au vocabulaire FOAF (Friend Of A Friend) qui facilite la représentation de réseaux sociaux.

### 3. Exercice 2 : Construire le référentiel de valeurs

---

#### 3.1 Question 1

Construire le vocabulaire SKOS à partir de la liste d'items proposée (vous pouvez travailler à partir de votre propre arborescence). La classe permettant de définir les éléments syntaxiques SKOS vous est entièrement donnée. Un exemple de définition d'instances SKOS vous est également fourni. Vous exploiterez cet exemple pour construire un vocabulaire SKOS organisant les descripteurs (concepts SKOS) de l'informatique (nommé ComputerScience ou CS).

- Concept le plus général : ComputerScience
- niveau 1 : System, Language, Software
- niveau 2 pour System : DatabaseSystem, OperatingSystem
- niveau 2 pour Language : ObjectOriented, Descriptive, Functional, QueryBased, ...
- niveau 3 pour Descriptive : KnowledgeBased, RDF
- niveau 3 pour QueryBased : SQL, SPARQL
- niveau 3 pour DatabaseSystem : Relational, Object-Relational, NoSQL

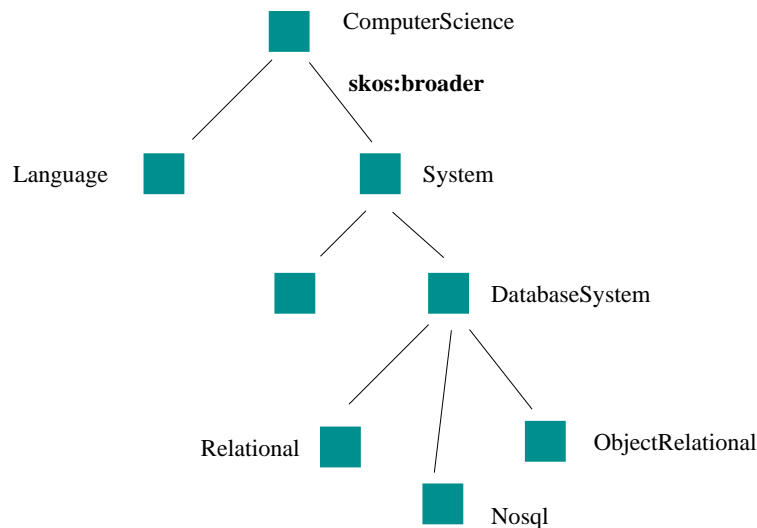


FIGURE 2 – Ossature du vocabulaire SKOS CS

### 4. Exercice 3 : Annoter sémantiquement le jeu de données

---

Construire une classe qui permettra de mettre en relation les ressources de vos jeux de données, notamment Livre avec le vocabulaire SKOS nommé CS au travers d'éléments de métadonnées. Pour ce faire, vous exploiterez le standard Dublin Core (vocabulaire DCTerms) et notamment la propriété dct:Subject qui doit prendre ses valeurs parmi les descripteurs du vocabulaire CS. Pointer pour chaque livre vers les termes informatiques qui leur sont les plus spécifiques. A la fin de l'exercice, vous devez disposer d'un modèle général tirant parti des différents vocabulaires et descriptions réalisées à partir de ces vocabulaires.

### 5. Exercice 4 : Manipuler l'ensemble

---

Il s'agit maintenant de manipuler votre modèle et d'appréhender l'intérêt de coupler standards de métadonnées et ontologies terminologiques pour décrire des jeux de données.

- retourner toutes les ressources de type Livre

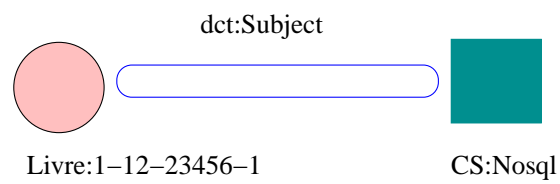


FIGURE 3 – Illustration annotation sémantique

- retourner toutes les ressources de type Livre, qui ont pour sujet le relationnel (étiquette terme relational)
  - retourner toutes les ressources de type Livre qui ont pour sujet de la base de données (DatabaseSystem).
- A votre sens, que manque t'il à votre modèle pour obtenir l'ensemble des bonnes réponses ? Comment pouvez vous y prendre pour remédier à ce manque ?