

Introduction au web sémantique

Yannick Prié

Département Informatique – Faculté des Sciences et Technologies

Université Claude Bernard Lyon 1

2009-2010

Objectifs de ce module

- Comprendre ce qu'est le web sémantique / web de données
- Comprendre ce qu'est RDF
- Comprendre ce qu'est un vocabulaire / une ontologie
- Concevoir une petite ontologie et l'utiliser

Moyens

- TD introductif : RDFa
- Cours magistral (ce cours)
- TD découverte Protege, RDFS et OWL
- TD conception d'ontologie

Plan

- Web sémantique / web de données
- RDF
- RDF schéma
- OWL

Plan

- **Web sémantique / web de données**
- RDF
- RDF schéma
- OWL

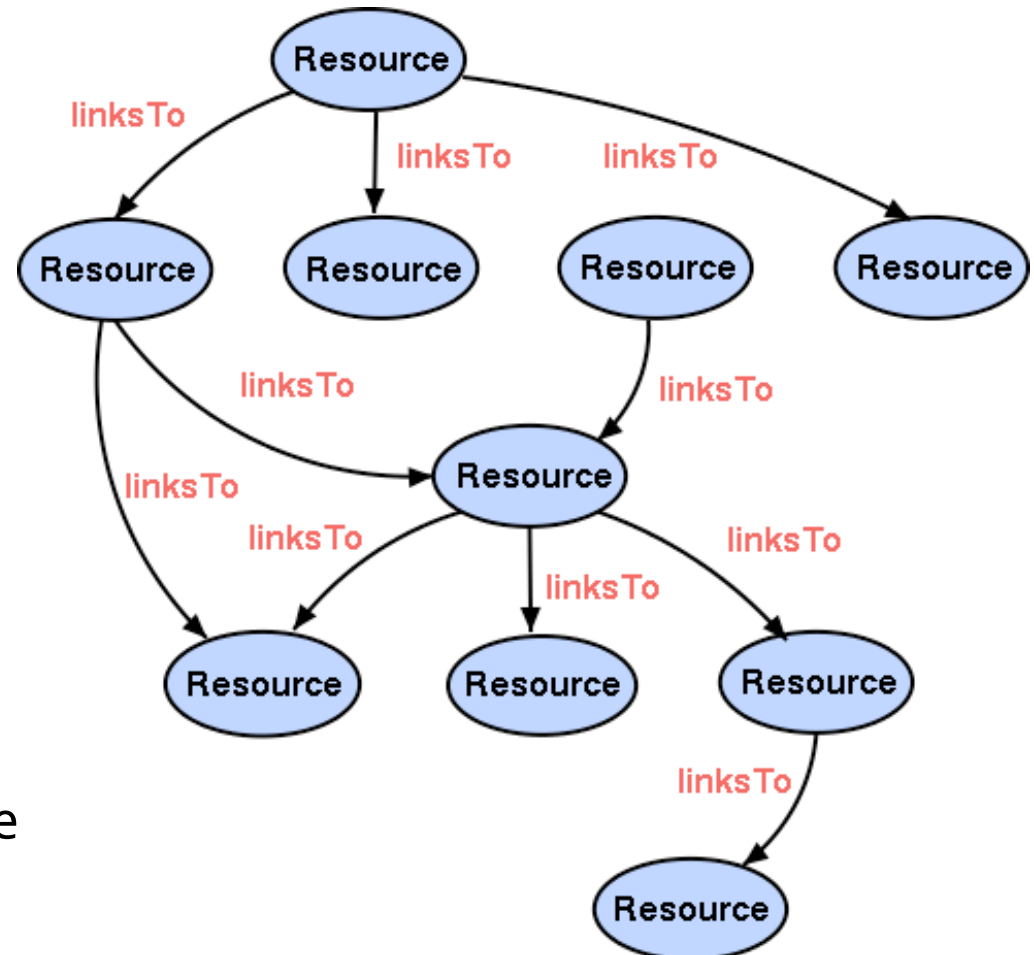
Web « classique » (1/2)

- Essentiellement
 - des pages web
 - décrites en (X)HTML
 - structures documentaires orientées description des documents
 - disponibles à des URLs sur le Web
 - avec des liens entre les pages web
 - `tutu`
- Il n'y a pas de beaucoup de choses pour les machines
 - contenus textuels des pages
 - → indexation / recherche d'informations
 - structures documentaires peu significatives
 - → parties, paragraphes, images
 - très peu de métadonnées
 - un seul type de lien : « la page X est en relation avec la page Y »

Web « classique » (2/2)

<http://www.w3.org/2004/Talks/o120-semweb-umich/>

- Ressources
 - identifiées par des URI's
 - non typées
- Liens
 - href, src...
 - limités, non-descriptifs
- Humain:
 - Exciting world - semantics of the resource, however, gleaned from content
- Machine:
 - Very little information available - significance of the links only evident from the context around the anchor.



Tim Berners-Lee, 1st World Wide Web Conference, Geneva, May 1994

To a computer, the Web is a flat, boring world, devoid of meaning. This is a pity, as in fact documents on the Web describe real objects and imaginary concepts, and give particular relationships between them. For example, a document might describe a person. The title document to a house describes a house and also the ownership relation with a person.... Adding semantics to the Web involves two things: allowing documents which have information in machine-readable forms, and allowing links to be created with relationship values. Only when we have this extra level of semantics will we be able to use computer power to help us exploit the information to a greater extent than our own reading.

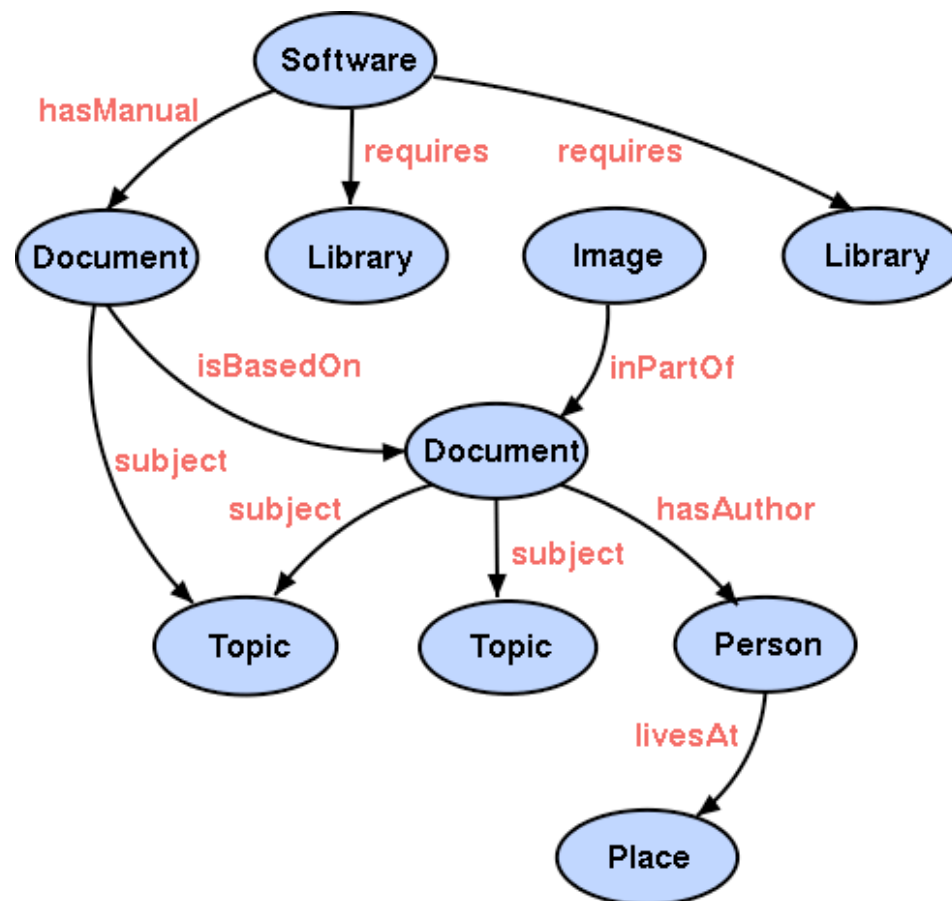
Tim Berners-Lee, 1st World Wide Web Conference, Geneva, May 1994

To a computer, the Web is a flat, boring world, devoid of meaning. This is a pity, as in fact documents on the Web describe real objects and imaginary concepts, and give particular relationships between them. For example, a document might describe a person. The title document to a house describes a house and also the ownership relation with a person.... ***Adding semantics to the Web involves two things: allowing documents which have information in machine-readable forms, and allowing links to be created with relationship values. Only when we have this extra level of semantics will we be able to use computer power to help us exploit the information to a greater extent than our own reading.***

Web sémantique

<http://www.w3.org/2004/Talks/o120-semweb-umich/>

- Resources
 - Globally Identified by URI's or Locally scoped (Blank) Extensible Relational
 - Links: Identified by URI's Extensible Relational
- Humain
 - Even more exciting world, richer user experience
- Machine
 - More processable information is available (Data Web)
- Humaines et machines
 - : Work, learn and exchange knowledge effectively



Web sémantique : définition

- De multiples définitions, à peu près cohérentes
- <http://www.uen.org/core/edtech/glossary.shtml>
 - *The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URIs for naming. A concept proposed by World Wide Web inventor Tim Berners-Lee.*


Plan

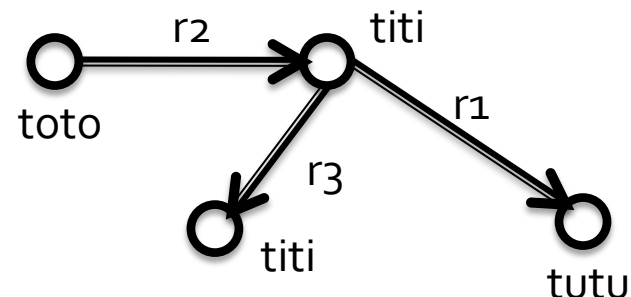
- Web sémantique / web de données
- **RDF**
- RDF schéma
- OWL

Rappel sur les URI

- Uniform Resource Identifier
- Séquence de caractères qui identifie une ressource abstraite ou physique, telle que définie dans la RFC3986.
- Permet d'identifier n'importe quoi de façon unique et non ambiguë
 - pages web, personnes, documents, livres, centres d'intérêts, bâtiments, etc.
- Remarque
 - on peut avoir une URI qui représente une personne, sans que cette URI soit une URL ou un document à propos de cette personne
 - <http://liris.cnrs.fr/yannick.prie> - ma page
 - <http://yannickprie.net/me> - moi
 - tout est question de convention

RDF : carte d'identité

- Ressource Description Framework
- Recommandation W3C depuis 2009
- Modèle de données pour représenter des graphes orientés étiquetés avec des URI
 - orientés : les relations sont orientées 
 - labellisés : les nœuds et les relations ont des étiquettes



RDF : notion de triplet

- < sujet > < prédicat > < objet >
- Trois types de termes
 - URI
 - Nœuds vierges (blank)
 - Littéraux
 - chaînes de caractères, nombres, dates, etc. (cf. XML schema)

RDF : triplet

<http://liris.cnrs.fr/labo>

<http://xmlns.com/foaf/0.1/member>

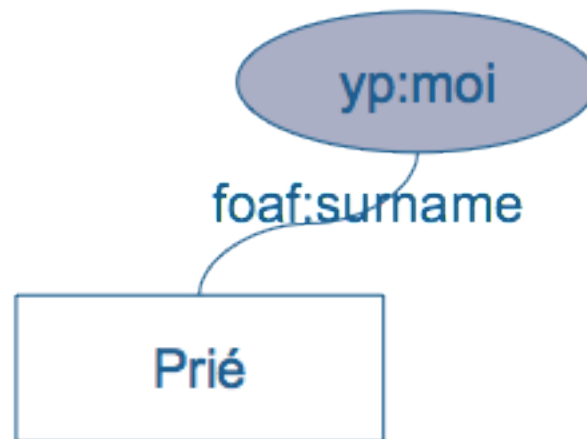
<http://liris.cnrs.fr/yannick.prie/moi>

RDF : espaces de noms



foaf: <http://xmlns.com/foaf/0.1/>
liris: <http://liris.cnrs.fr/>
yp: <http://liris.cnrs.fr/yannick.prie/>

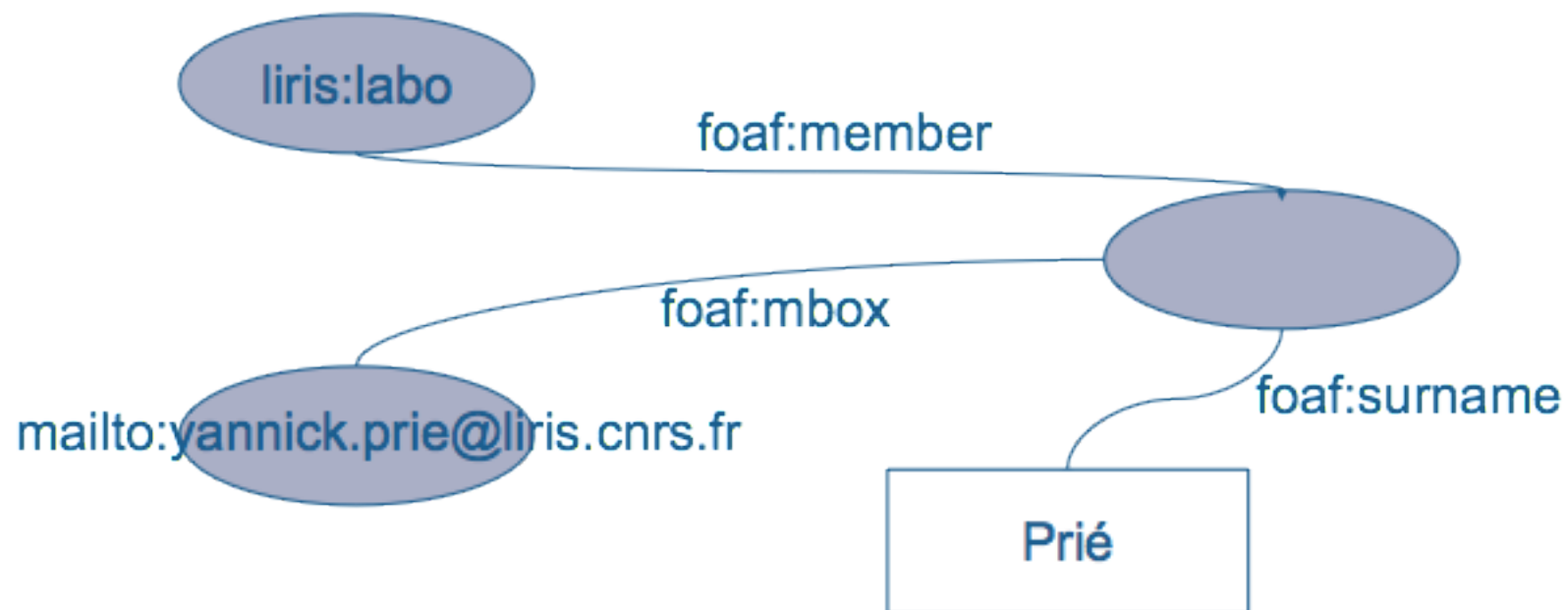
RDF : littéral



RDF : nøed vierge



RDF : exemple



Sérialisation RDF : RDF/XML

- Verbeux, difficile à lire, le plus utilisé
- <http://www.w3.org/TR/rdf-syntax-grammar/>

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:ex="http://example.org/stuff/1.0/">
<rdf:Description rdf:about="http://www.w3.org/TR/rdf-syntax-grammar"
  dc:title="RDF/XML Syntax Specification (Revised)">
  <ex:editor>
    <rdf:Description ex:fullName="Dave Beckett">
      <ex:homePage rdf:resource="http://purl.org/net/dajobe/" />
    </rdf:Description>
  </ex:editor>
</rdf:Description>
</rdf:RDF>
```

Sérialisation RDF : N3/Turtle

- Plus facile à lire pour les humains
- <http://www.w3.org/TeamSubmission/turtle/>

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix ex: <http://example.org/stuff/1.0/> .

<http://www.w3.org/TR/rdf-syntax-grammar>
  dc:title "RDF/XML Syntax Specification (Revised)" ;
  ex:editor [
    ex:fullname "Dave Beckett";
    ex:homePage <http://purl.org/net/dajobe/>
  ] .
```

Sérialisation RDF : RDFa

- Le RDF est inséré dans du XHTML
 - Une page web lisible par les humains et les machines
- <http://www.w3.org/TR/rdfa-syntax/>
- Quelques attributs
 - `about` : URI spécifiant la ressource décrite par les métadonnées ; en son absence, il s'agit de l'élément en cours
 - `href`, `src`, `resource` : spécifiant l'objet d'une propriété
 - `rel`, `rev` : spécifiant une relation avec une autre ressource
 - `property` : spécifiant une propriété pour le contenu d'un élément, également décrit avec `content`
 - `typeof` : attribut optionnel spécifiant le type RDF d'un sujet
- Exemple
 - Voir TD1 !

Langage de requête

- SPARQL
- <http://www.w3.org/TR/rdf-sparql-query/>

```
PREFIX foaf:    <http://xmlns.com/foaf/0.1/>
SELECT ?name ?mbox
WHERE {
    ?x foaf:name ?name .
    ?x foaf:mbox ?mbox .
}
```

- Plutôt réservé aux informaticiens
- Fournit un mode d'accès uniforme aux graphes RDF

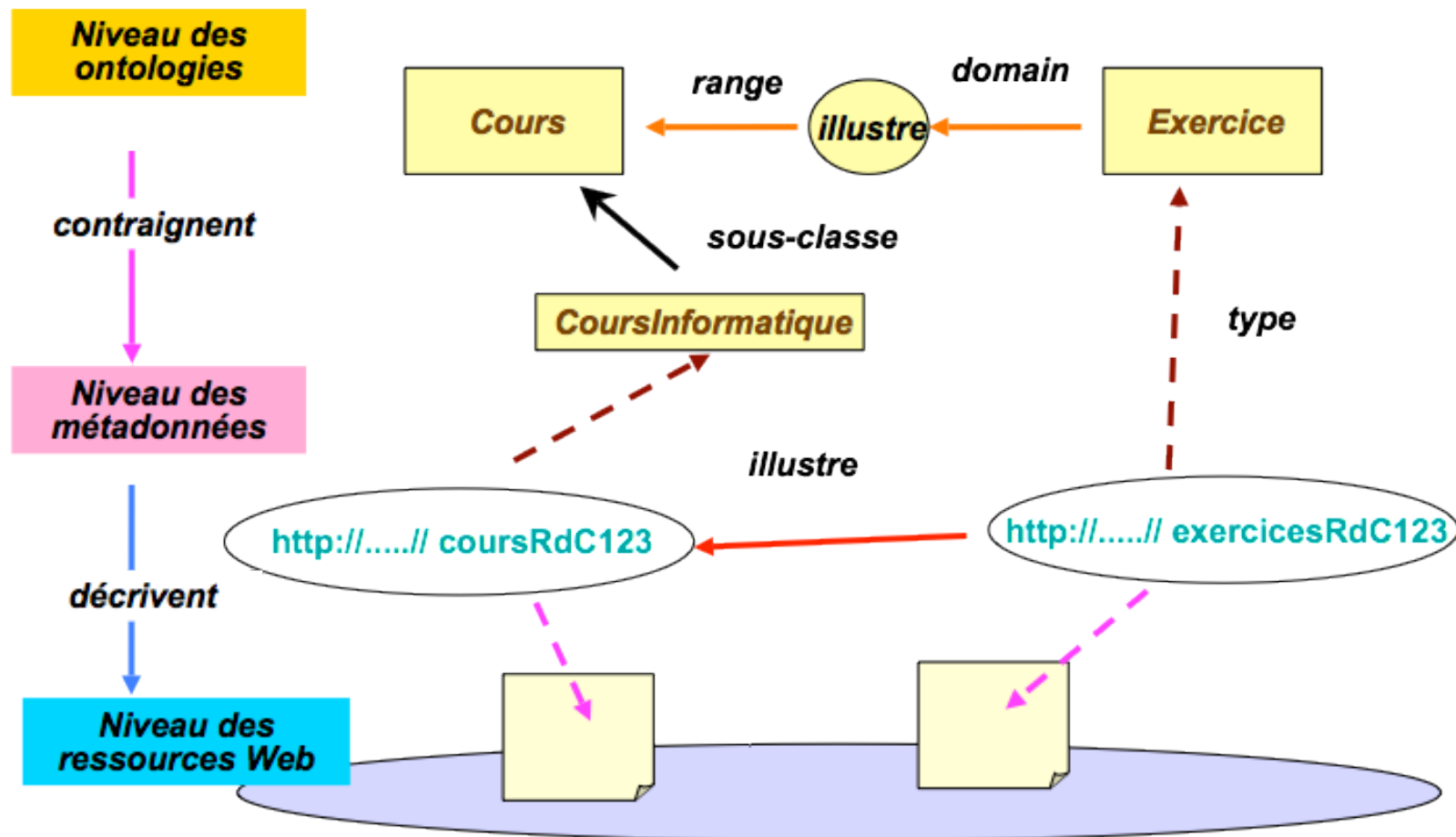
Schémas : vocabulaires, ontologies

- On sait exprimer n'importe quel graphe avec RDF, que manque-t-il ?
 - pouvoir spécifier des manières de décrire des graphes RDF
 - pouvoir spécifier la signification des graphes RDF
 - pouvoir partager ces significations
- Idée
 - donner plus d'informations abstraites et sémantiques sur les nœuds dans les graphes RDF
 - typer les nœuds RDF
 - leur donner des catégories
 - typer les relations RDF
 - comme des propriétés des noeuds
 - attributs
 - relations

Niveau du vocabulaire – schéma

Niveau des assertions

(d'après Philippe Laublet)



Dans la suite

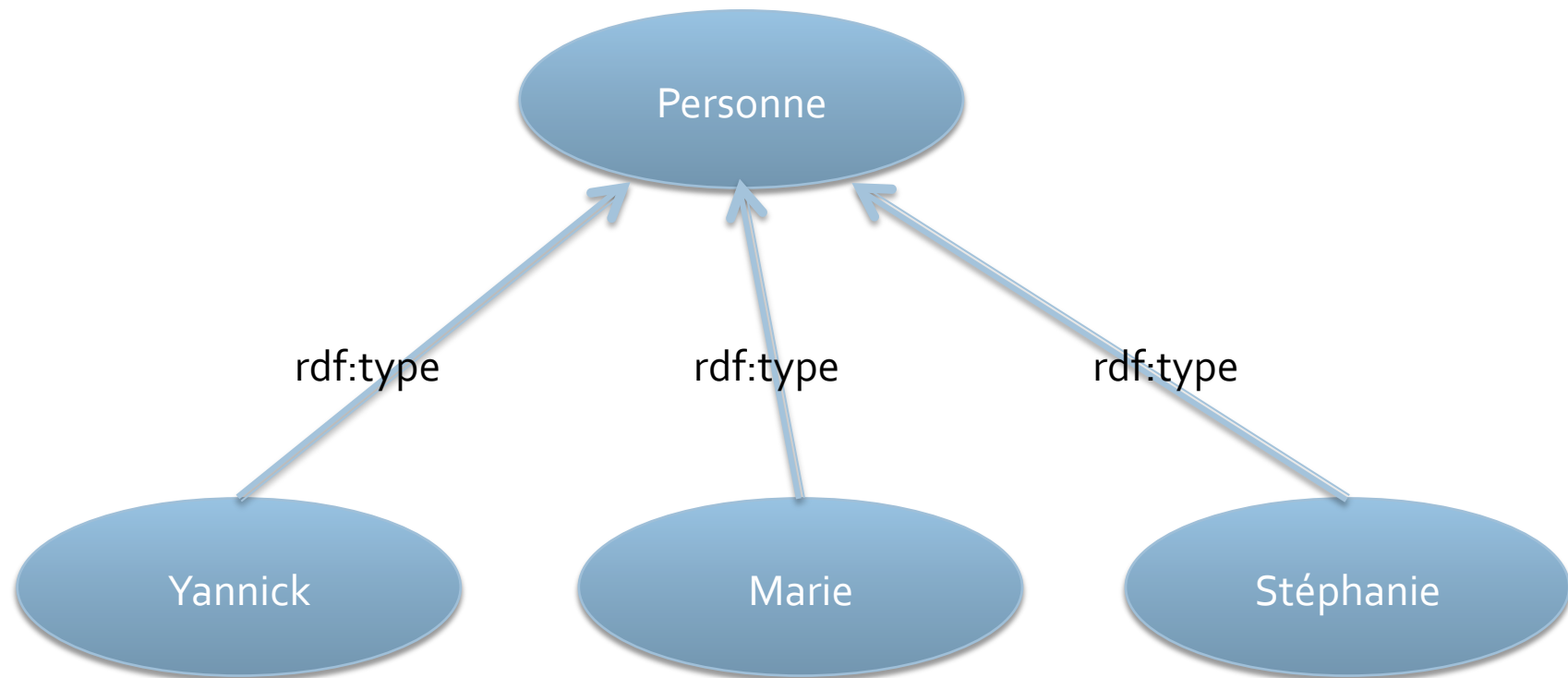
- RDF-schema :
 - pour des vocabulaires basiques
 - pour spécifier des hiérarchies de types et de propriétés
- OWL (Ontology Web Language)
 - pour des vocabulaires plus riches
 - pour spécifier des hiérarchies de concepts, des relations, des axiomes, etc.
- Remarque
 - au sens large du terme, une ontologie est une « spécification partagée d'une conceptualisation » (Gruber)
 - on peut faire des ontologies en RDFS ou OWL

Plan

- Web sémantique / web de données
- RDF
- **RDF schéma**
- OWL

Notion de classe et d'instance

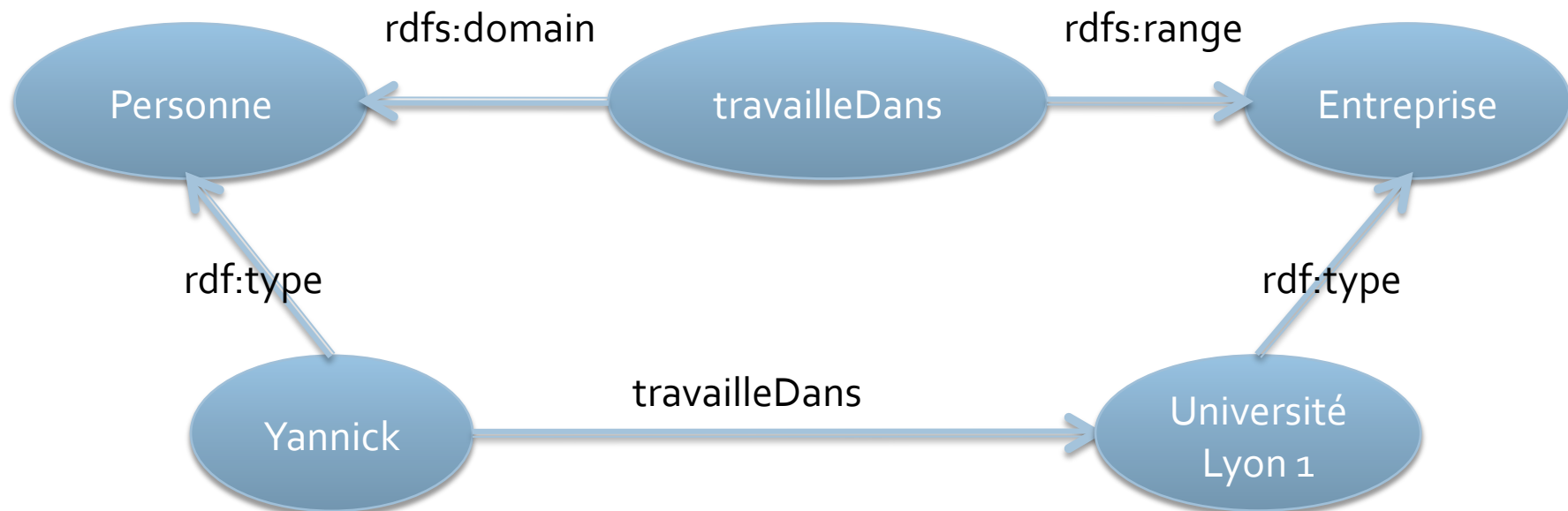
Niveau des classes: connaissances abstraites



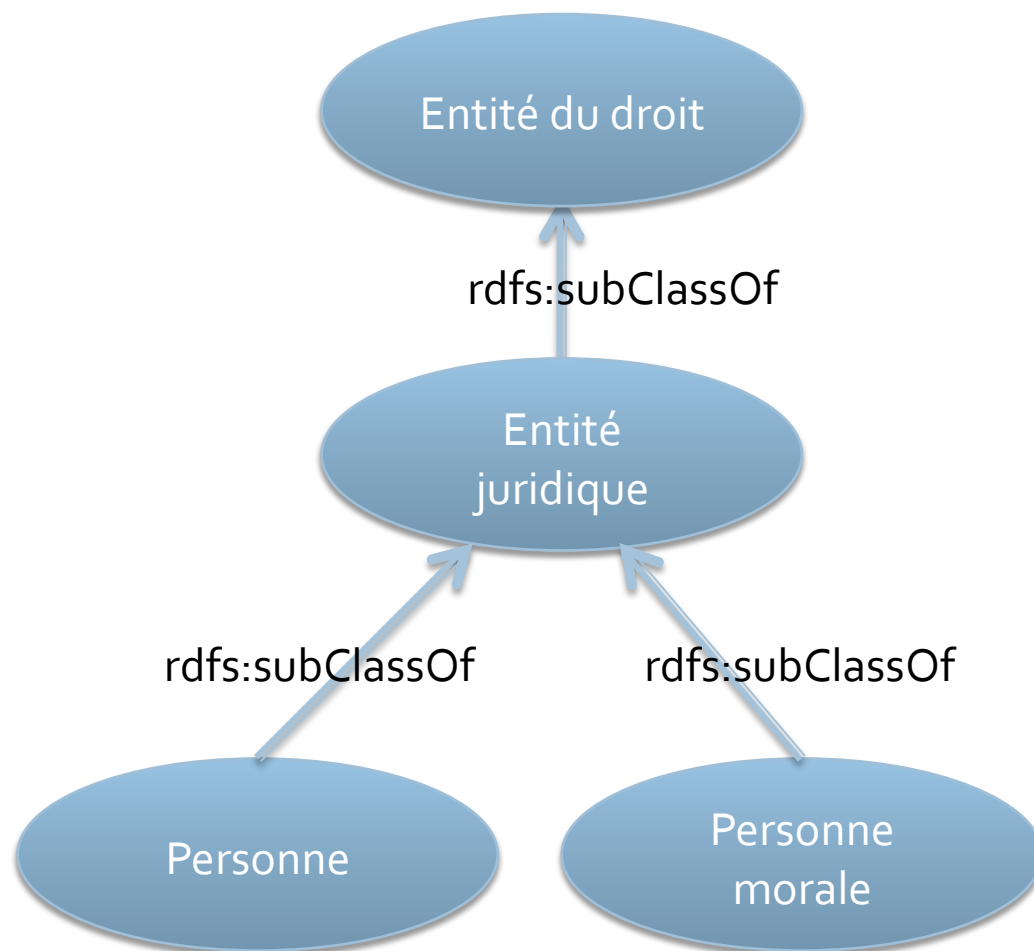
Niveau des instances : connaissances concrètes

Notion de propriété

- Une propriété RDFs à un domaine (*domain*) et une portée(*range*)

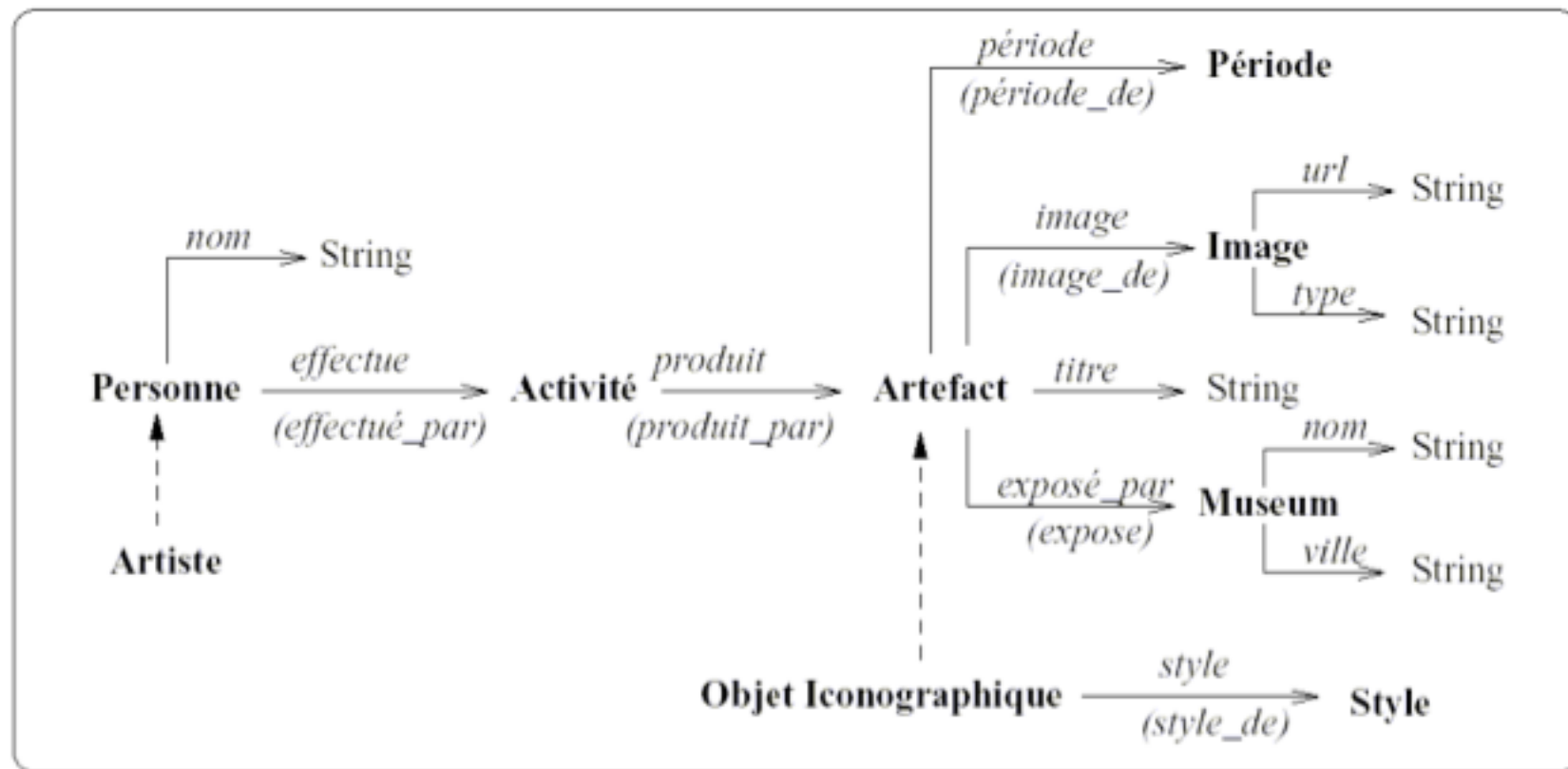


Notion de hiérarchie



- Interprétation
 - Toute instance d'une classe est aussi instance de ses super-classes
 - *Yannick* est une instance de *Entité juridique*
- Relations équivalente :
 - Généralisation (inv. Spécialisation)
- Relations transitives
 - *Personne morale* est une sous classe de *Entité du droit*

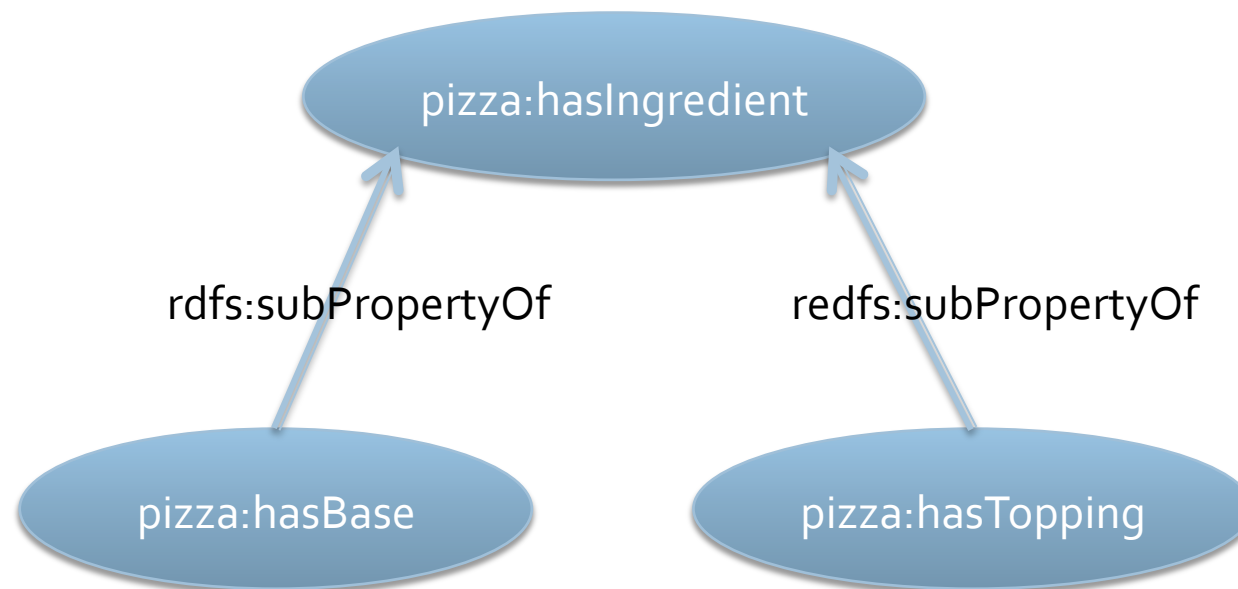
RDF-Schema : hiérarchie de classes



D'après Bernd Amman

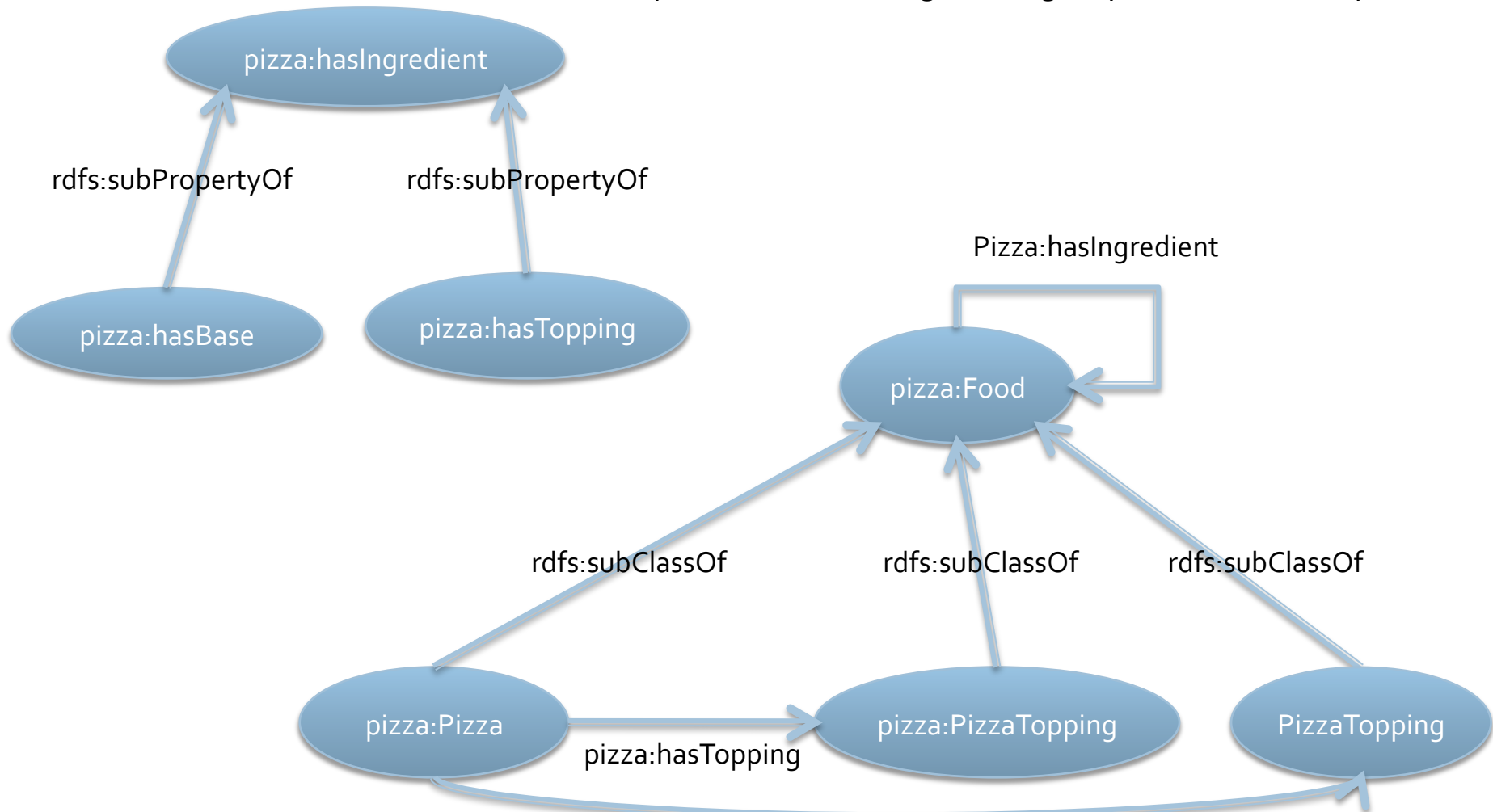
RDF-Schema : hiérarchie de propriétés

<http://www.co-ode.org/ontologies/pizza/2007/02/12/pizza.owl>

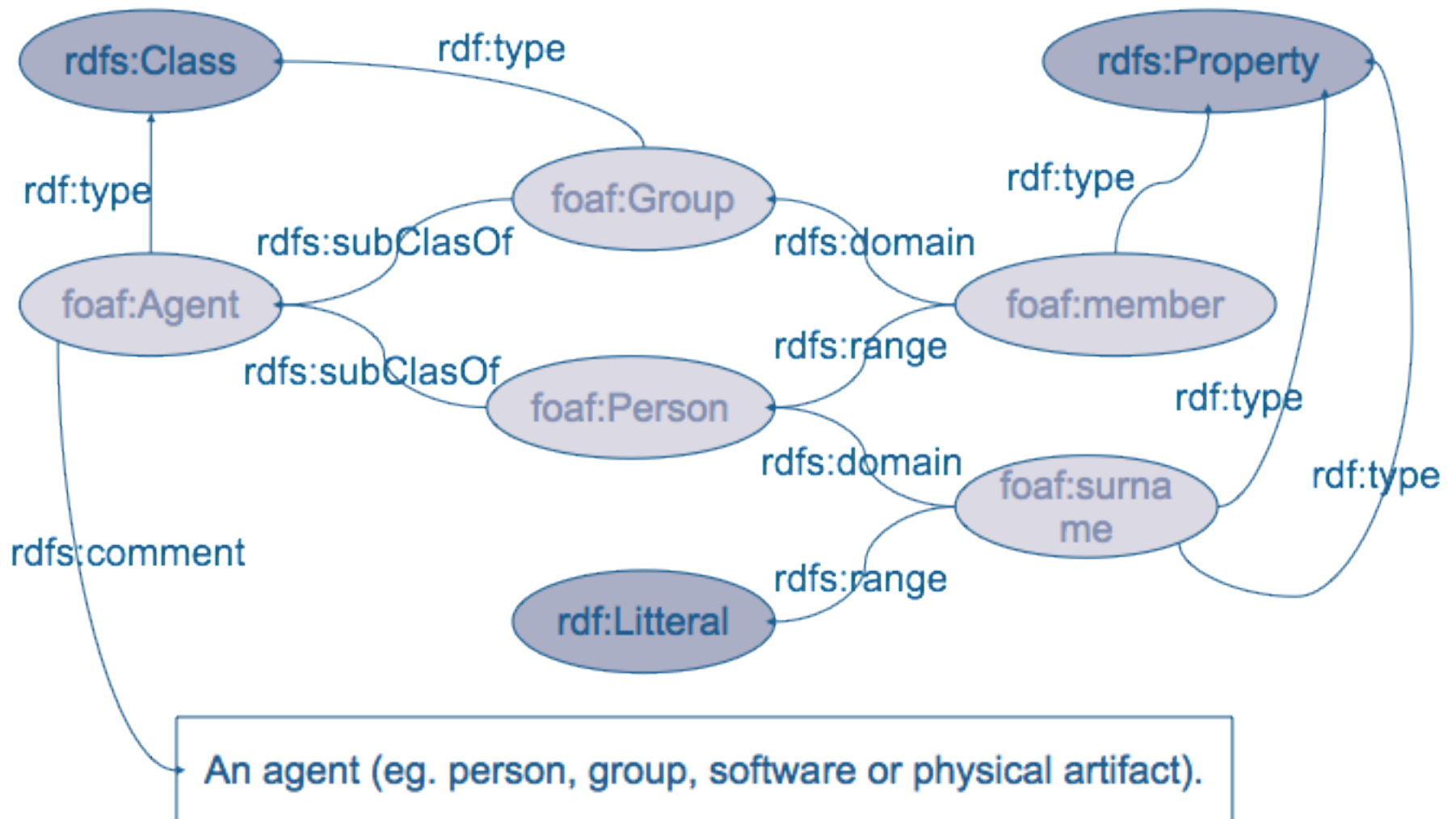


RDF-Schema : hiérarchie de propriétés

<http://www.co-ode.org/ontologies/pizza/2007/02/12/pizza.owl>



Exemple FOAF



RDF Schema : inférences

- Inférence en général
 - déduction automatique à partir d'une structure informatique, effectuée sur la base de la syntaxe du langage utilisé, mais dont les résultats doit correspondre à une sémantique formelle précise

RDF-Schema : inférence de spécialisation

- Inférence principale liée à une hiérarchie, en suivant la sémantique classique de la spécialisation : toute instance de la sous-classe est instance de la super-classe
- Si je cherche toutes les *Entités de droit*, alors j'ai comme réponse *Yannick, Marie, Stéphanie* même si je n'ai pas explicitement demandé des *Personne*
- Si je cherche des choses qui ont des ingrédients (*hasIngredient*), je trouverai des *Pizza*

Exemples RDF-schema

- FOAF : Friend of a Friend
- DOAC : Description of a career
- DOAP : Description of a project
- Dublin Core
- etc.

Plan

- Web sémantique / web de données
- RDF
- RDF schéma
- **OWL**

OWL

- Changement de vocabulaire
 - Classes → concepts
 - Instance → individus
- On garde les hiérarchies de concepts et de propriétés
- On garde range et domain pour les propriétés
- On ajoute des possibilités
 - plus grande expressivité du langage
 - inférences plus riches, mais...
 - plus grande complexité pour faire des inférences
 - en terme de calculs informatiques

OWL : caractériser des propriétés

- Déclarer une propriété comme transitive
 - Si $A p B$ et $B p C$ alors $A p C$
- Déclarer une propriété comme réflexive
 - $A p B$ équivalent à $B p A$
- Déclarer une propriété comme inverse d'une autre
 - $A p B$ équivalent à $B q A$ si $q = \text{inv}(p)$
- etc.

OWL : caractériser des concepts (1/2)

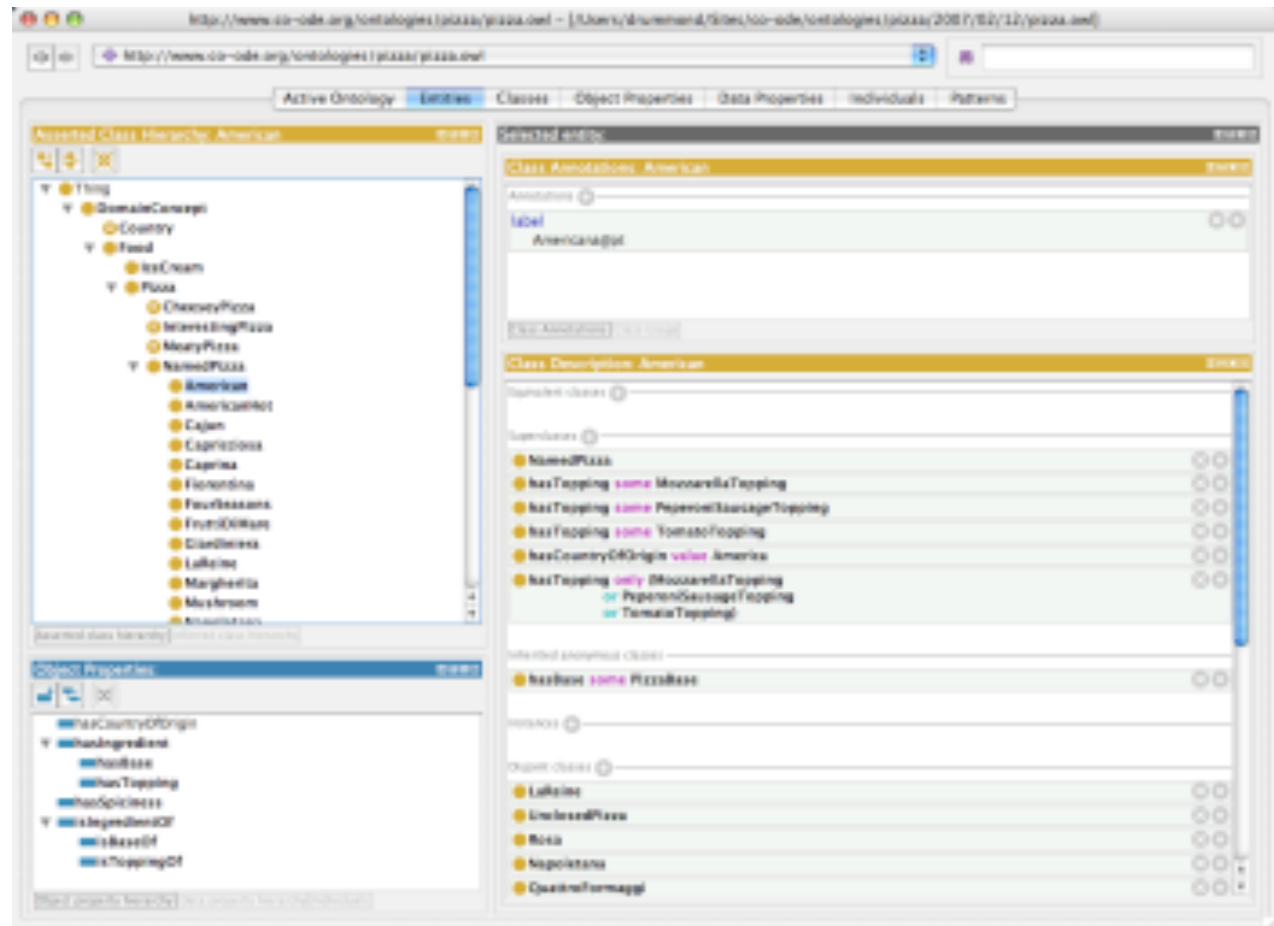
- Ne pas juste définir un concept comme un sous-concept d'un autre (rdfs:subClassOf)
- Mais le caractériser par rapport à d'autres concepts ou à des propriétés
 - Déclarer qu'un concept est disjoint d'un autre
 - A disjoint de B ssi aucun individu ne peut être instance de A et de B à la fois
 - Déclarer qu'un concept est le complémentaire d'un autre
 - Le monde des individus se divise en deux, les instances de A, et les instances de B
 - Déclarer qu'un concept est l'intersection de deux autres
 - Les individus instance de ce concept sont ceux qui sont instance de A et de B

OWL : caractériser des concepts (2/2)

- Déclarer un concept en listant les propriétés qu'il doit avoir
 - Les choses qui ont un époux (une propriété aPour Epoux)
 - Les choses qui sont des Voiture (sous-classe) et qui ont au moins 3 roues
 - etc.

Outil pour gérer des ontologies

- Protege :
outil le plus répandu
- Concepts
- Propriétés
- Individus
- Requêtes
- Raisonneur
 - Fact++

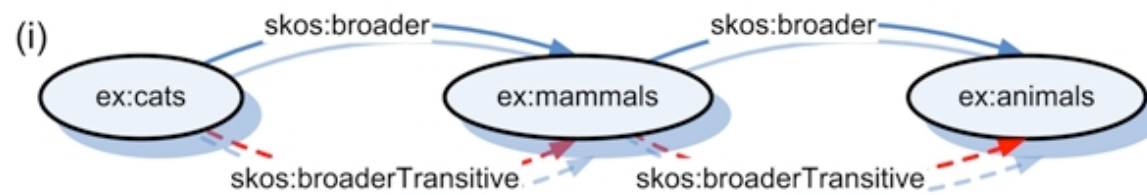
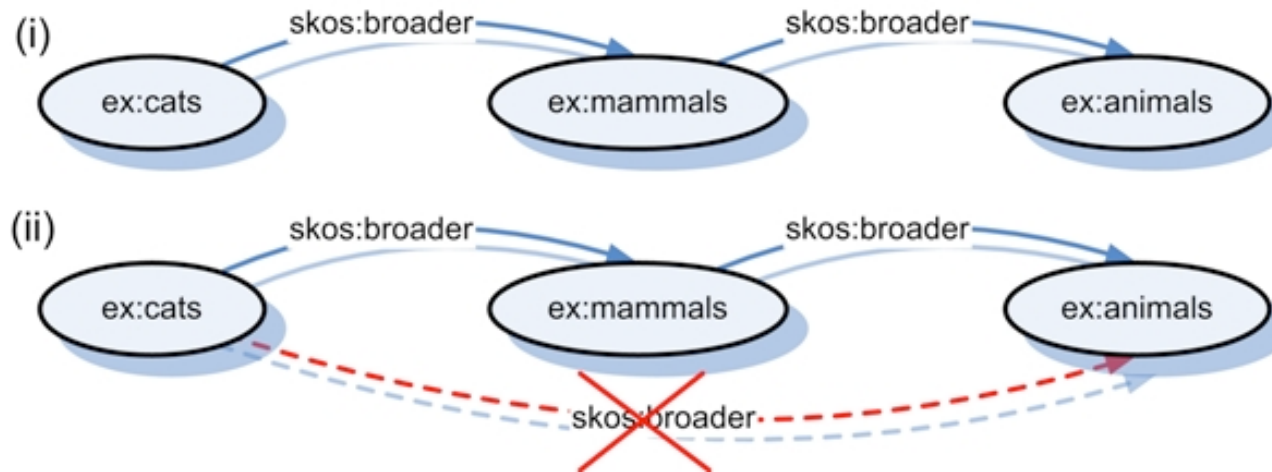


<http://protege.stanford.edu>

Exemples OWL

- Wine, Pizza, Travel
 - ontologies jouet accessibles dans Protege
- SKOS : Simple Knowledge Organisation System
 - <http://www.w3.org/TR/2009/REC-skos-reference-20090818/>
 - Pour représenter des thésaurii existants et les porter sur le web sémantique
- Nombreux autres exemples

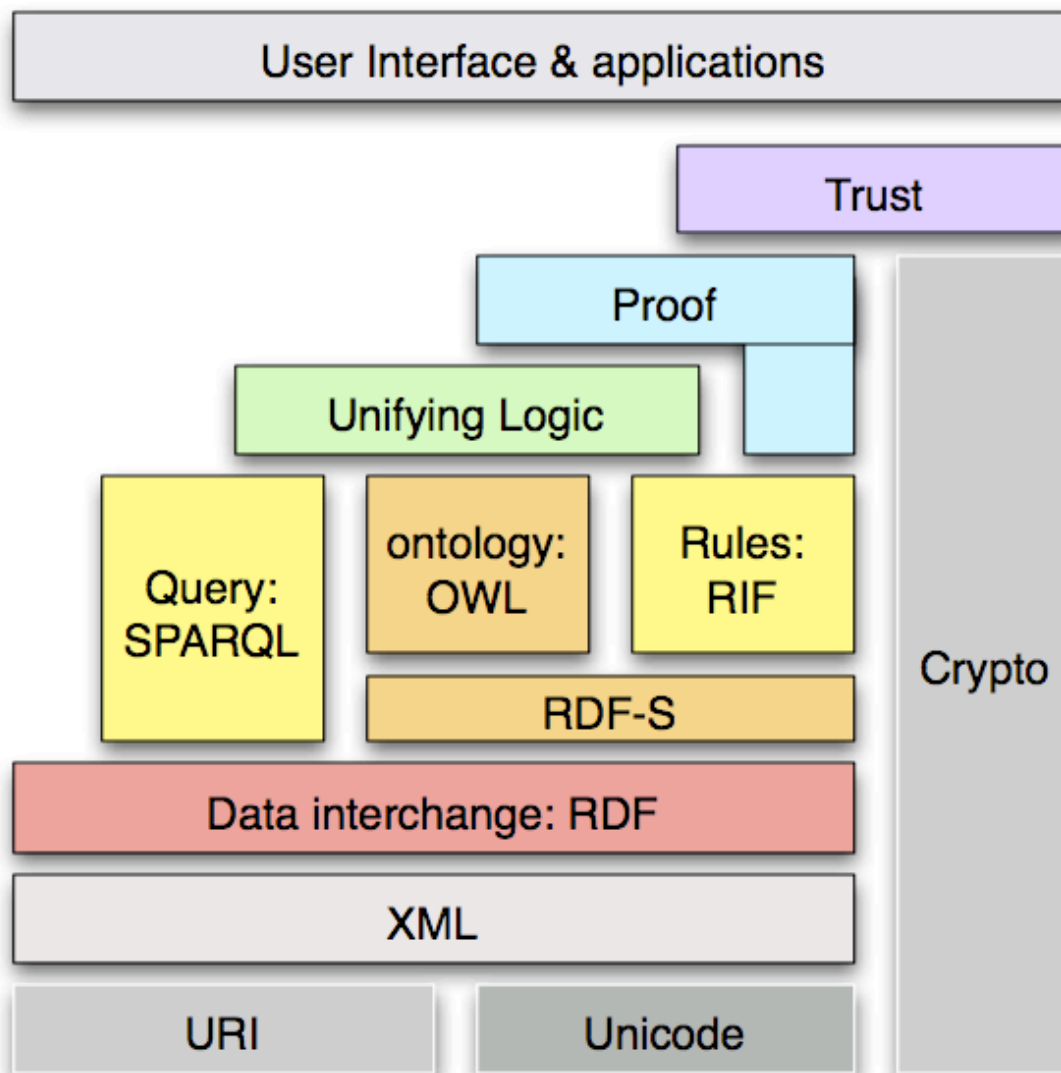
SKOS : remarque sur la transitivité



Plan

- Web sémantique / web de données
- RDF
- RDF schéma
- OWL
- **Pour conclure**

La couche des langages

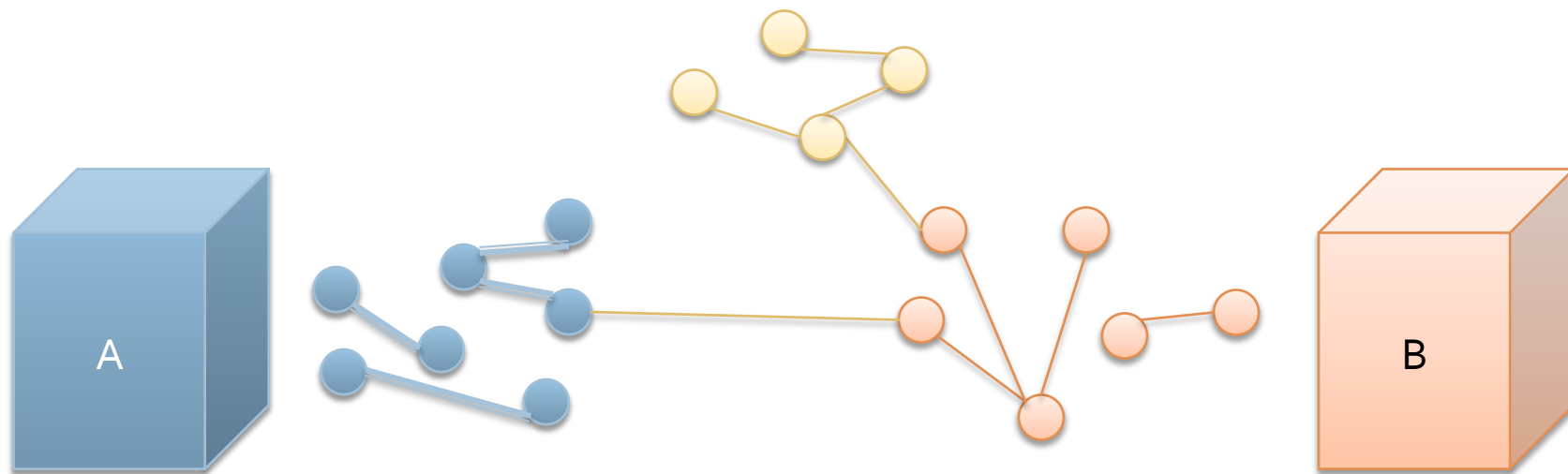


Où on en est

- Plus de 10 ans de travail du W3C
 - technologies qui arrivent à maturité
 - applications industrielles
- Les acteurs basculent petit à petit
 - applications web construites « à base de données sémantiques »
 - bases internes en RDF
 - bases extérieures
 - exposition de données propres sur le web de données
 - Service d'accès à ses données en RDF
 - exposition de données en RDFa
 - etc.

Web of data : principes

- Exposition des données sous forme RDF
- Vision unifiée des données de A, de B, dans le cadre de C
- Construction d'un service par C sur ce graphe unique



Exemple LOD : DBPedia

- Les données structurées de wikipedia disponibles en RDF
- Exemple d'utilisation
 - *All soccer players, who played as goalkeeper for a club that has a stadium with more than 40.000 seats and who are born in a country with more than 10 million inhabitants*

SPARQL Explorer for <http://dbpedia.org/sparql>

SPARQL:

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX : <http://dbpedia.org/resource/>
PREFIX dbpedia2: <http://dbpedia.org/property/>
PREFIX dbpedia: <http://dbpedia.org/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
```

```
SELECT DISTINCT ?page {
  ?s foaf:page ?page.
  ?s rdf:type <http://dbpedia.org/ontology/SoccerPlayer> .
  ?s dbpedia2:position <http://dbpedia.org/resource/Goalkeeper> .
  ?s <http://dbpedia.org/ontology/club> ?club .
  ?club <http://dbpedia.org/ontology/capacity> ?cap .
  ?s <http://dbpedia.org/ontology/birthPlace> ?place .
  ?place ?population ?pop
  Filter (xsd:int(?cap) >40000 ) .
  Filter (xsd:int(?pop) >10000000 ) .
```

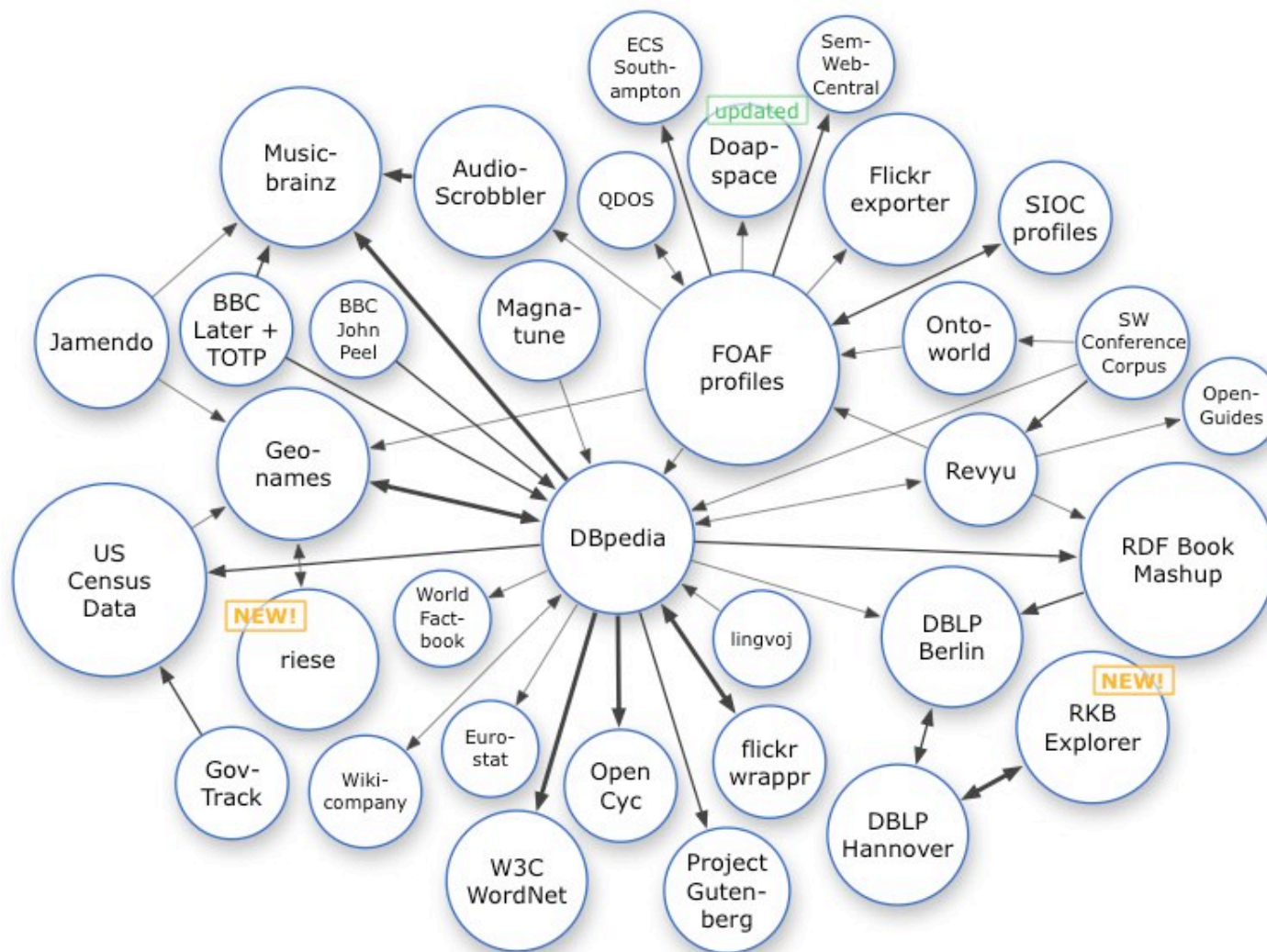
Results:

SPARQL results:

page
http://en.wikipedia.org/wiki/%C3%89dson_Bastos
http://en.wikipedia.org/wiki/Olivier_Renard
http://en.wikipedia.org/wiki/Francisco_Ramos
http://en.wikipedia.org/wiki/Leonardo_Zamora
http://en.wikipedia.org/wiki/Paulo_Garc%C3%A9s
http://en.wikipedia.org/wiki/Ra%C3%BAI_Olivares
http://en.wikipedia.org/wiki/Richard_Leyton
http://en.wikipedia.org/wiki/Joseph-Antoine_Bell
http://en.wikipedia.org/wiki/Cantarele
http://en.wikipedia.org/wiki/Matthew_Nash
http://en.wikipedia.org/wiki/Danny_Milosevic
http://en.wikipedia.org/wiki/Ivan_Necevski

Web of data - Linked Open Data

(2008)



Dans la suite

- TD : découverte de Protege
- TD : conception d'une ontologie

Remerciements

- Alexandre Passant
- Eric Miller
- Pierre-Antoine Champin