

---

# Linked Open Data

Le Web de données  
*Réseau, usages, perspectives*  
Eric Charton

---

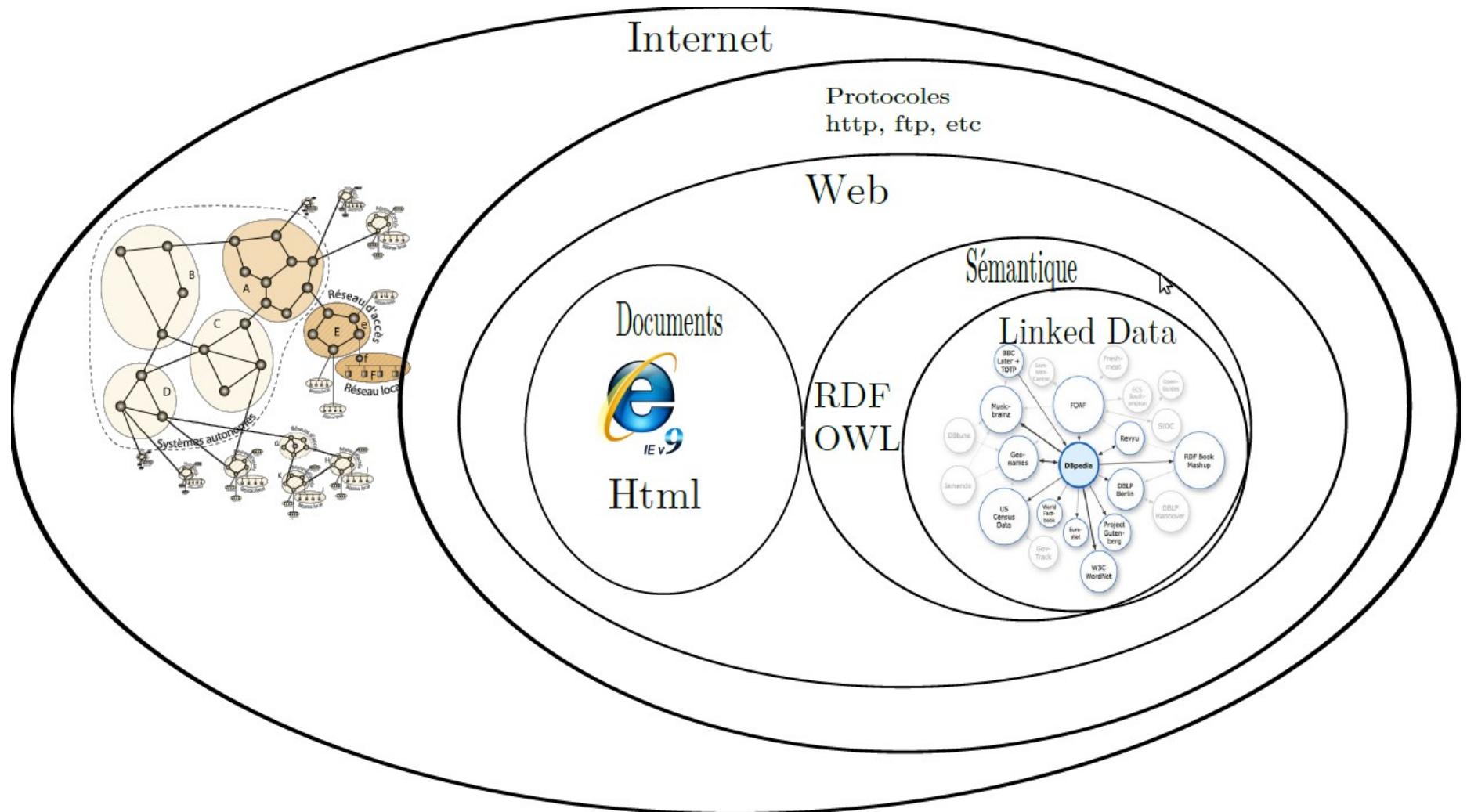
# Sommaire

- Histoire du Linked Open Data
- Structure et évolution du réseau
- Utilisations du Linked Open Data
- Présence sur le réseau LOD
- Futurs contenus
- Futurs usages

---

# Histoire du Linked Open Data

# La structure du Web Sémantique



# Linked Open Data

## Une communauté du W3C

The screenshot shows the homepage of linkeddata.org. The browser address bar displays "linkeddata.org". The main heading is "Linked Data - Connect Distributed Data across the Web". On the left, there is a navigation menu with items like Home, Guides and Tutorials, Frequently Asked Questions, Glossary, Images and Posters, Presentations, Data Sets, Tools, Events, Calls for Papers, Research, News and Blogs, Domains, and See Also. The main content area has a section titled "Linked Data" with a paragraph explaining that it is about using the Web to connect related data that wasn't previously linked, using URIs and RDF. Below this is a network diagram with numerous nodes representing different data sources and services, such as DBpedia, GeoNames, Eurostat, Project Gutenberg, flickr wrapper, Open Calais, Virtuoso Sparger, RKB ECS Southampton, DBLP Hannover, CiteSeer, and many others. On the right side, there is a login section for authorized users, including fields for Username and Password, and a "Log in" button. There are also links for "Log in using OpenID" and "Request new password".



---

# Historique

- Janvier 2007
  - Chris Bizer (DBPedia) et Richard Cyganiak proposent le project au W3C SWEO  
*(Semantic Web Education and Outreach Group)*
  - *Premier pas : contributions de chercheurs et de développeurs.*
  - *Entrée de grands acteurs (BBC, Thomson Reuters) à partir de 2008*

**Linked Data - The Story So Far, Bizer, Heath, Berners-Lee (2009)**



---

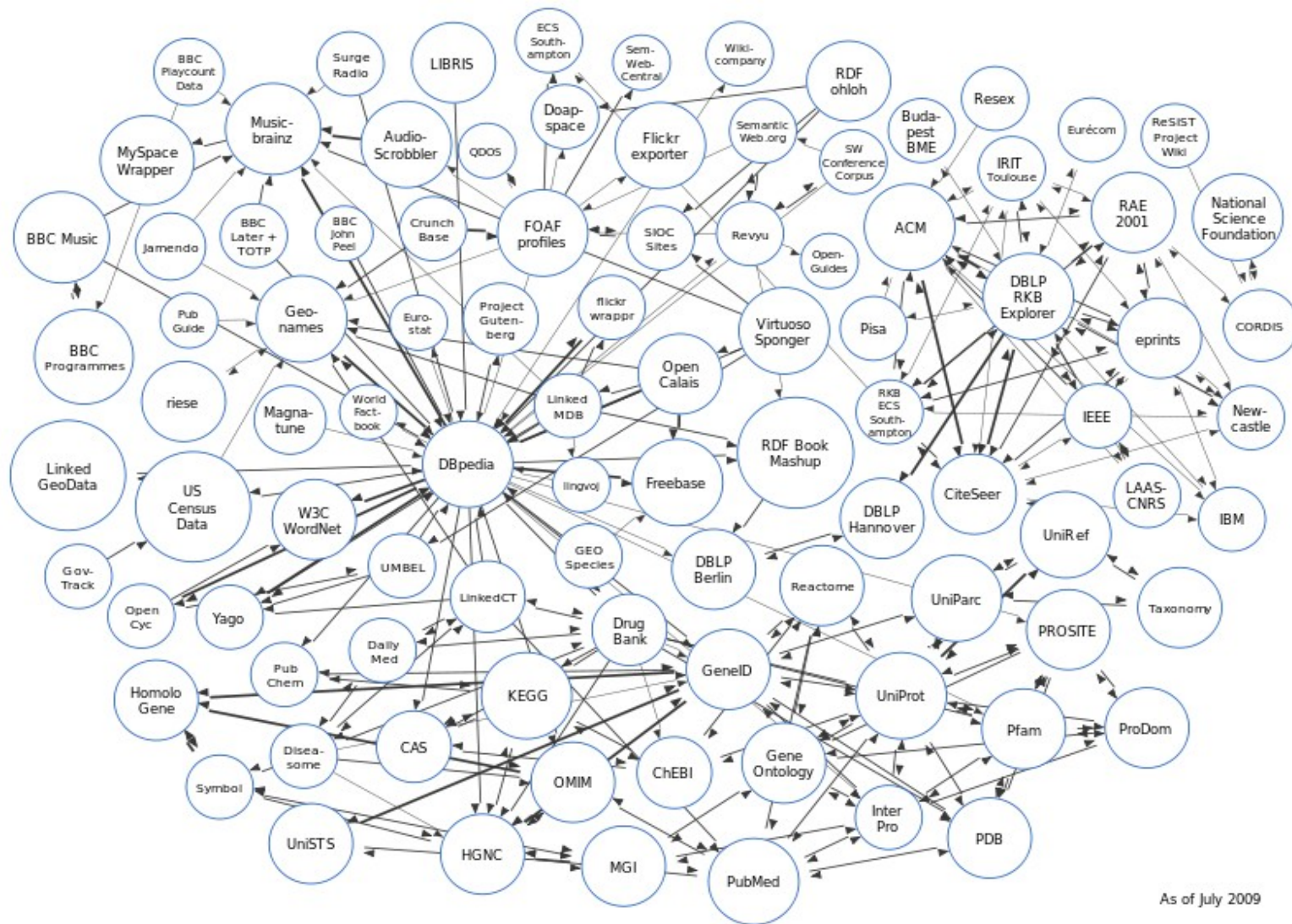
# Linked Open Data (LOD)

- Quelle est sa relation avec le Web sémantique?
  - Les opinions diffèrent :
    - Le Web Sémantique est une branche de recherche et développement, le LOD, un réseau
    - Pour Tim Berner Lee, « *le LOD est le Web sémantique tel qu'il doit être mis en oeuvre* »
  - Mais le concept de Web sémantique dépasse le seul réseau:
    - Il recouvre aussi les algorithmes de production et d'accès aux ressources qui dépassent la seule question du réseau.

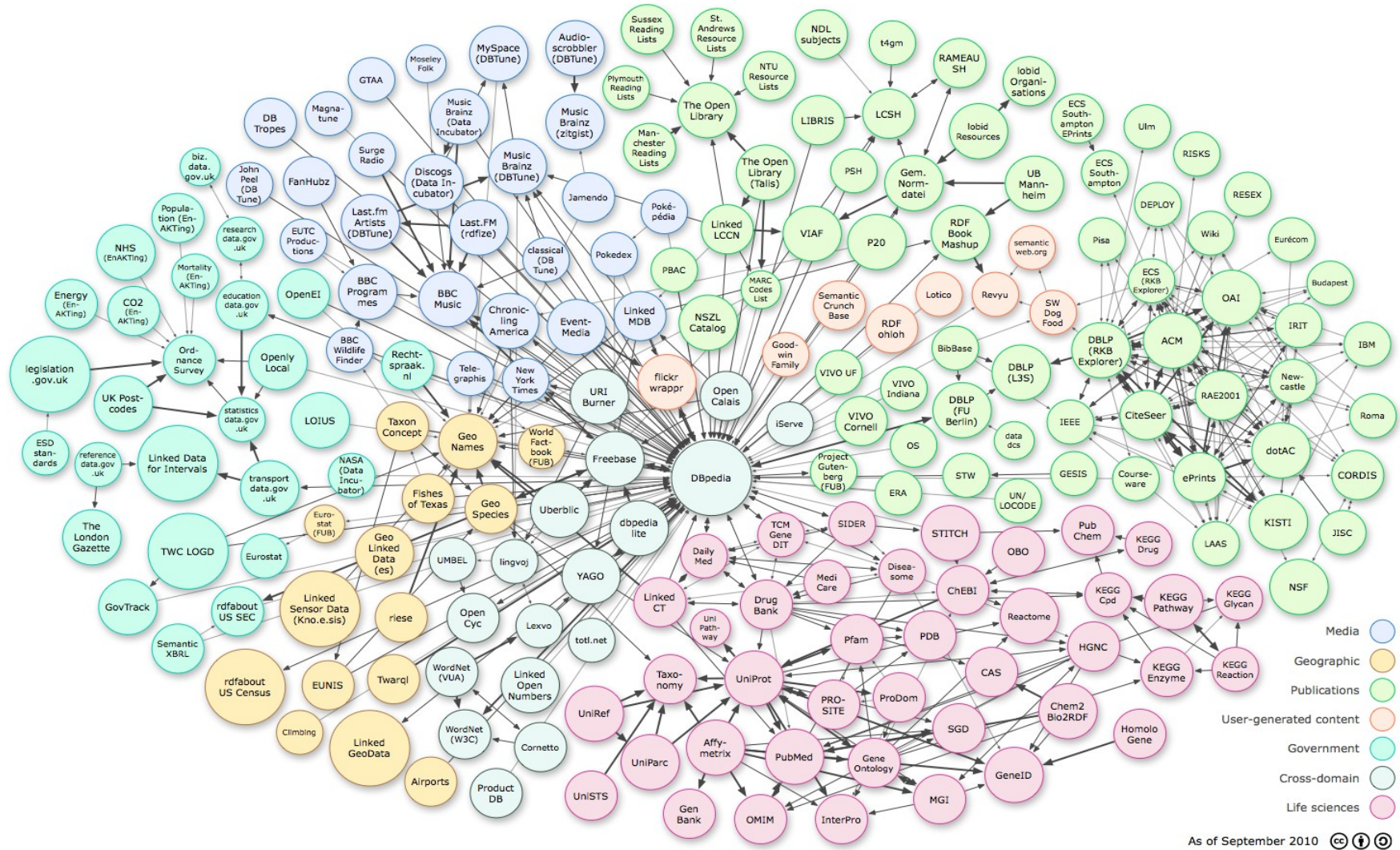




# 2009

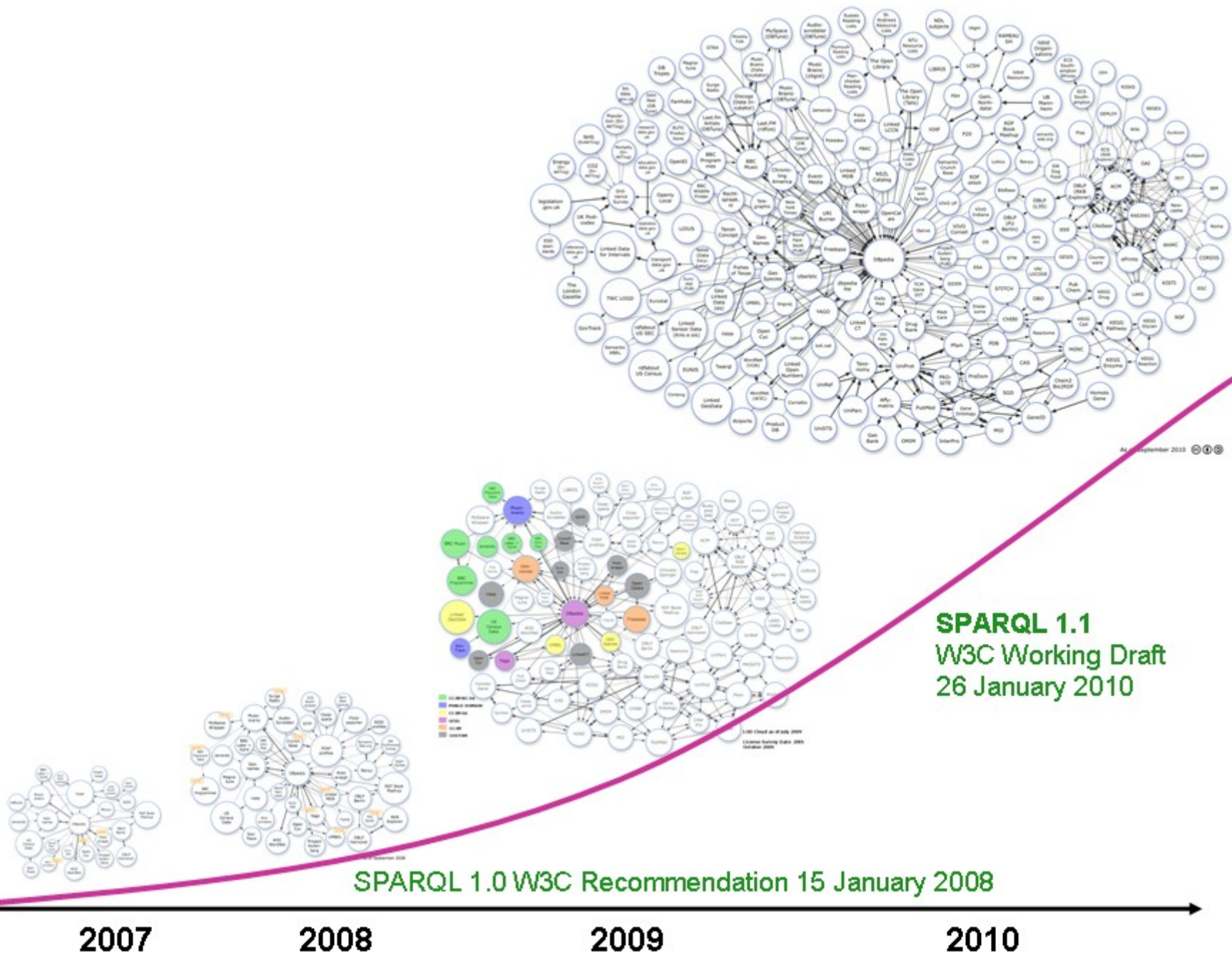


# 2010





Nombre de points d'accès interoperables



# LOD en septembre 2011

Domaine	Sources	Triplets	%	Liens RDF	Percent
Multi-domaines	41	4 185 M	13,2	63 M	12,5
Géographique	31	6 146 M	19,4	36 M	7,1
Gouvernement	49	13 315 M	42,1	19 M	3,8
Médias	25	1 842 M	5,8	50 M	10,1
Publications	87	2 951 M	9,3	140 M	27,8
Sciences de la vie	41	3 036 M	9,6	192 M	38,1
Autres	20	134 M	0,4	3 M	0,7
	294	31 634 M		504 M	

---

# LOD : 4 principes de bases

- Définition par Tim Berners-Lee de quatre principes fondateurs (2006) :
  - 1 Identifier les «choses» avec des URI
  - 2 Les URI sont compatibles avec le web des documents (http)
  - 3 Fournir à travers l'URI des renseignements lisibles par les humains et par les machines.
  - 4 Les URI donnent accès à d'autres URI pour stimuler l'interconnexion

---

# Exemple pratique

- Une bibliothèque A et une bibliothèque B élaborent des dépôt RDF pour décrire leurs livres

# Exemple pratique

- Le dépôt Da et le dépôt Db sont accessibles sur le Web (http)



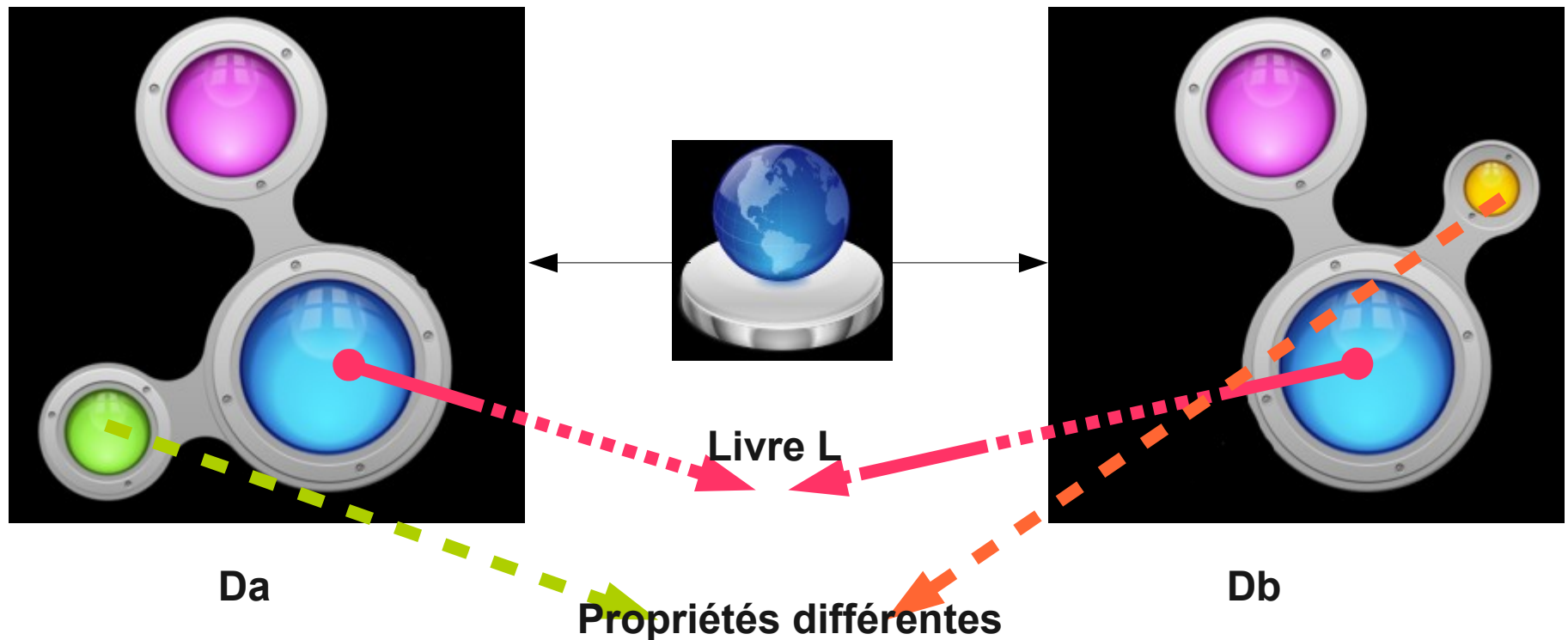
**Db**



**Db**

# Exemple pratique

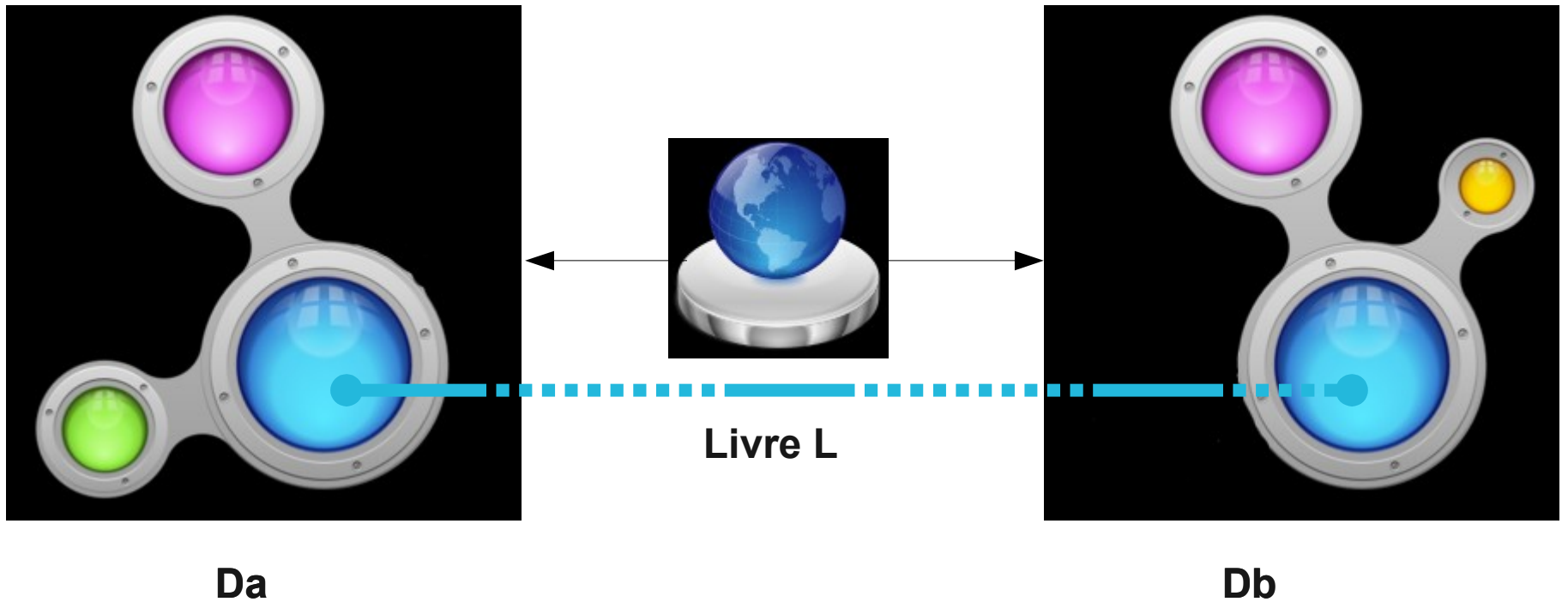
- Pour un livre L, il existe une description dans  $D_a$  et dans  $D_b$ , mais  $D_a$  décrit des propriétés de livres différentes de  $D_b$  (ex nationalité de l'éditeur ou traductions existantes)





# Exemple pratique

- A relie chacun de ses RDF de livres de Db à Da à travers le web via http



---

# Exemple pratique

- Pour l'utilisateur ou le logiciel , le livre L et ses propriétés sont potentiellement devenus une seule entité du Web sémantique



---

# Définir les triples

Les vocabulaires



---

# Les vocabulaires du LOD

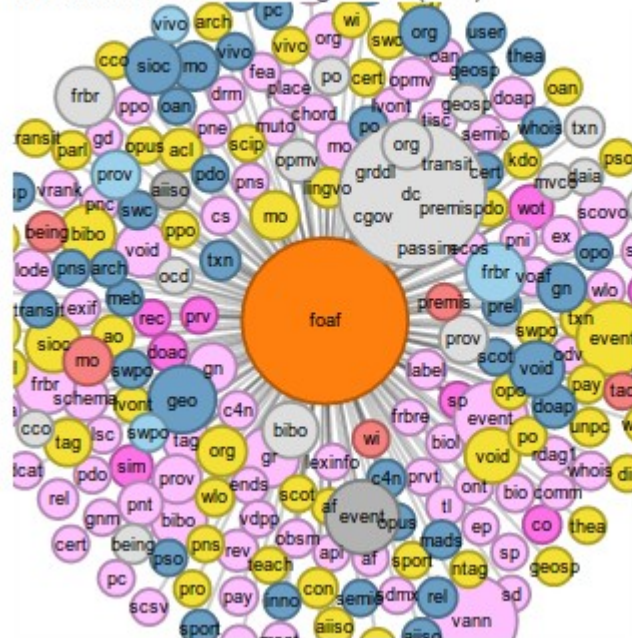
- Des ontologies RDF et OWL
  - Elles constituent des définitions sémantiques des liens entre les « choses » du web sémantique
  - Elles sont une première source de relations standardisées entre les objets des différents dépôts :
    - Une personne est toujours une personne
    - Une ville est toujours une ville
    - Etc ...
    - L'idée est de standardiser la relation de propriété

# Exemple de vocabulaire

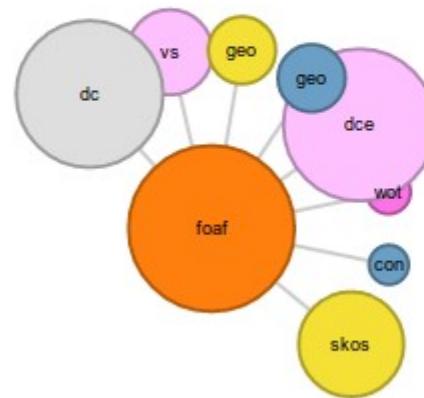
- Foaf (friend of a friend project)

Vocabulary links:

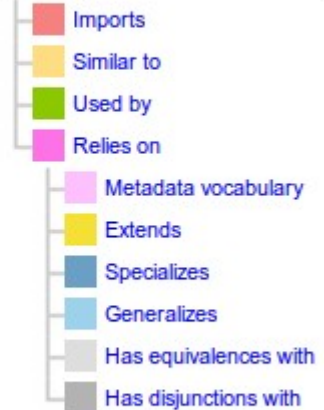
Vocabularies referencing "foaf" (209)



Vocabularies referenced by "foaf" (8)



Link types



# Exemple de vocabulaire *foaf*

FOAF Core	Social Web
<ul style="list-style-type: none"><li>• <ul style="list-style-type: none"><li>◦ <a href="#">Agent</a></li><li>◦ <a href="#">Person</a></li><li>◦ <a href="#">name</a></li><li>◦ <a href="#">title</a></li><li>◦ <a href="#">img</a></li><li>◦ <a href="#">depiction</a> (<a href="#">depicts</a>)</li><li>◦ <a href="#">familyName</a></li><li>◦ <a href="#">givenName</a></li><li>◦ <a href="#">knows</a></li><li>◦ <a href="#">based_near</a></li><li>◦ <a href="#">age</a></li><li>◦ <a href="#">made</a> (<a href="#">maker</a>)</li><li>◦ <a href="#">primaryTopic</a> (<a href="#">primaryTopicOf</a>)</li></ul></li><li>• <ul style="list-style-type: none"><li>◦ <a href="#">Project</a></li><li>◦ <a href="#">Organization</a></li><li>◦ <a href="#">Group</a></li><li>◦ <a href="#">member</a></li></ul></li><li>• <ul style="list-style-type: none"><li>◦ <a href="#">Document</a></li><li>◦ <a href="#">Image</a></li></ul></li></ul>	<ul style="list-style-type: none"><li>• <a href="#">nick</a></li><li>• <a href="#">mbox</a></li><li>• <a href="#">homepage</a></li><li>• <a href="#">weblog</a></li><li>• <a href="#">openid</a></li><li>• <a href="#">jabberID</a></li><li>• <a href="#">mbox_sha1sum</a></li><li>• <a href="#">interest</a></li><li>• <a href="#">topic_interest</a></li><li>• <a href="#">topic</a> (<a href="#">page</a>)</li><li>• <a href="#">workplaceHomepage</a></li><li>• <a href="#">workInfoHomepage</a></li><li>• <a href="#">schoolHomepage</a></li><li>• <a href="#">publications</a></li><li>• <a href="#">currentProject</a></li><li>• <a href="#">pastProject</a></li><li>• <a href="#">account</a></li><li>• <a href="#">OnlineAccount</a></li><li>• <a href="#">accountName</a></li><li>• <a href="#">accountServiceHomepage</a></li><li>• <a href="#">PersonalProfileDocument</a></li><li>• <a href="#">tipjar</a></li><li>• <a href="#">sha1</a></li><li>• <a href="#">thumbnail</a></li><li>• <a href="#">logo</a></li></ul>

---

# Utilisation de vocabulaire *foaf*

- Exemple foaf

- Définition foaf: <http://xmlns.com/foaf/0.1/>
- Vocabulaire utilisée pour concevoir un document ou des triples décrivant une personne:

```
<foaf:Person rdf:about="#michelgagnon" xmlns:foaf="http://xmlns.com/foaf/0.1/">  
  <foaf:name>Michel Gagnon</foaf:name>  
  <foaf:homepage rdf:resource="http://www.professeurs.polymtl.ca/michel.gagnon/" />  
  <foaf:openid rdf:resource="http://www.professeurs.polymtl.ca/michel.gagnon/rdf/" />  
  <foaf:img rdf:resource="http://www.professeurs.polymtl.ca/michel.gagnon/Images/mg.jpg" />  
</foaf:Person>
```



---

# Quelques vocabulaires

- Dublin Core Metadata Element Set (descriptions de ressources bibliographiques)  
<http://dublincore.org/documents/dces/>
- Europeana Data Model (EDM) (collections culturelles)  
<http://www.europeana.eu/schemas/edm/>
- Les vocabulaires existants peuvent être intégrés dans les nouveaux dépôts (ex DBPedia utilise son propre vocabulaire mais aussi Yago, Foaf, etc)

# LOD - Exemple tiré de BNF

@prefix foaf: <http://xmlns.com/foaf/0.1/> .

@prefix rdagroup2elements:

<http://RDVocab.info/ElementsGr2/> .

@prefix xfoaf: <http://www.foafrealm.org/xfoaf/0.1/> .

@prefix owl: <http://www.w3.org/2002/07/owl#> .

<http://data.bnf.fr/ark:/12148/cb11928016k#foaf:Person>

a foaf:Person;

rdagroup2elements:dateOfBirth "08-02-1828";

rdagroup2elements:dateOfDeath "24-03-1905";

rdagroup2elements:placeOfBirth "Nantes";

rdagroup2elements:placeOfDeath "Amiens";

xfoaf:nationality

<http://id.loc.gov/vocabulary/countries/fr>;

owl:sameAs <http://dbpedia.org/resource/Jules\_Verne>;

foaf:name "Jules Verne";

foaf:page <http://data.bnf.fr/11928016/jules\_verne/> .

# LOD - Exemple tiré de BNF

```
@prefix foaf: <http://xmlns.com/foaf/0.1/> .  
@prefix rdagroup2elements:  
<http://RDVocab.info/ElementsGr2/> .  
@prefix xfoaf: <http://www.foafrealm.org/xfoaf/0.1/> .  
@prefix owl: <http://www.w3.org/2002/07/owl#> .
```

```
<http://data.bnf.fr/ark:/12148/cb11928016k#foaf:Person>
```

URI désignant  
Jules Verne

```
  a foaf:Person;  
  rdagroup2elements:dateOfBirth "08-02-1828";  
  rdagroup2elements:dateOfDeath "24-03-1905";  
  rdagroup2elements:placeOfBirth "Nantes";  
  rdagroup2elements:placeOfDeath "Amiens";  
  xfoaf:nationality  
<http://id.loc.gov/vocabulary/countries/fr>;  
  owl:sameAs <http://dbpedia.org/resource/Jules_Verne>;  
  foaf:name "Jules Verne";  
  foaf:page <http://data.bnf.fr/11928016/jules_verne/> .
```

# LOD - Exemple tiré de BNF

@prefix foaf: <http://xmlns.com/foaf/0.1/> .

@prefix rdagroup2elements:

<http://RDVocab.info/ElementsGr2/> .

@prefix xfoaf: <http://www.foafrealm.org/xfoaf/0.1/>Son type

@prefix owl: <http://www.w3.org/2002/07/owl#> .

<http://data.bnf.fr/ark:/12148/cb11928016k#foaf:Person>

**a foaf:Person;**

rdagroup2elements:dateOfBirth "08-02-1828";

rdagroup2elements:dateOfDeath "24-03-1905";

rdagroup2elements:placeOfBirth "Nantes";

rdagroup2elements:placeOfDeath "Amiens";

xfoaf:nationality

<http://id.loc.gov/vocabulary/countries/fr>;

owl:sameAs <http://dbpedia.org/resource/Jules\_Verne>;

foaf:name "Jules Verne";

foaf:page <http://data.bnf.fr/11928016/jules\_verne/> .

# LOD - Exemple tiré de BNF

@prefix foaf: <http://xmlns.com/foaf/0.1/> .

@prefix rdagroup2elements:

<http://RDVocab.info/ElementsGr2/> .

@prefix xfoaf: <http://www.foafrealm.org/xfoaf/0.1/> .

@prefix owl: <http://www.w3.org/2002/07/owl#> .

<http://data.bnf.fr/ark:/12148/cb11928016k#foaf:Person>

a foaf:Person;

rdagroup2elements:dateOfBirth "08-02-1828";

rdagroup2elements:dateOfDeath "24-03-1905";

rdagroup2elements:placeOfBirth "Nantes";

rdagroup2elements:placeOfDeath "Amiens";

xfoaf:nationality

<http://id.loc.gov/vocabulary/countries/fr>;

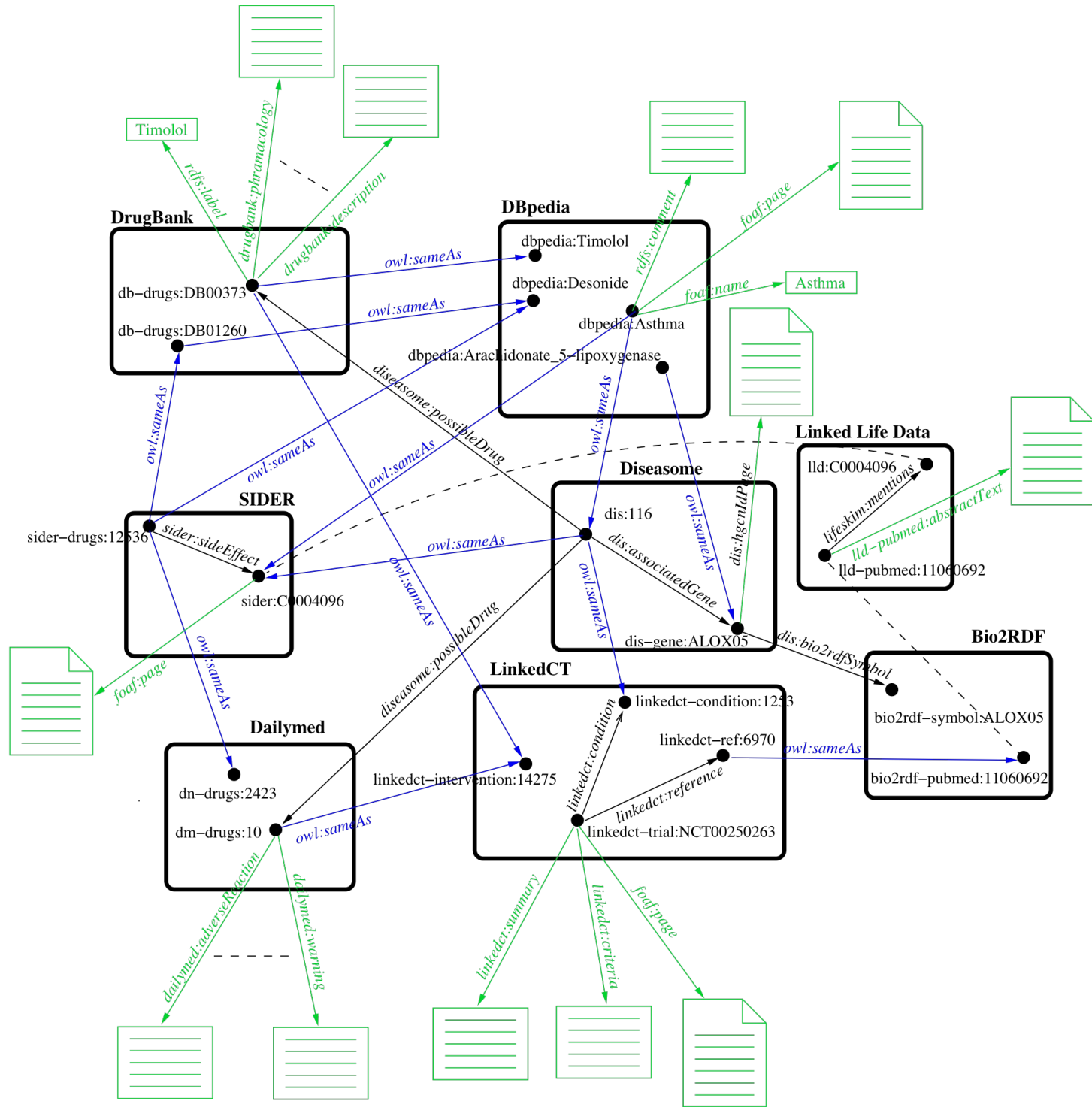
owl:sameAs <http://dbpedia.org/resource/Jules\_Verne>;

foaf:name "Jules Verne";

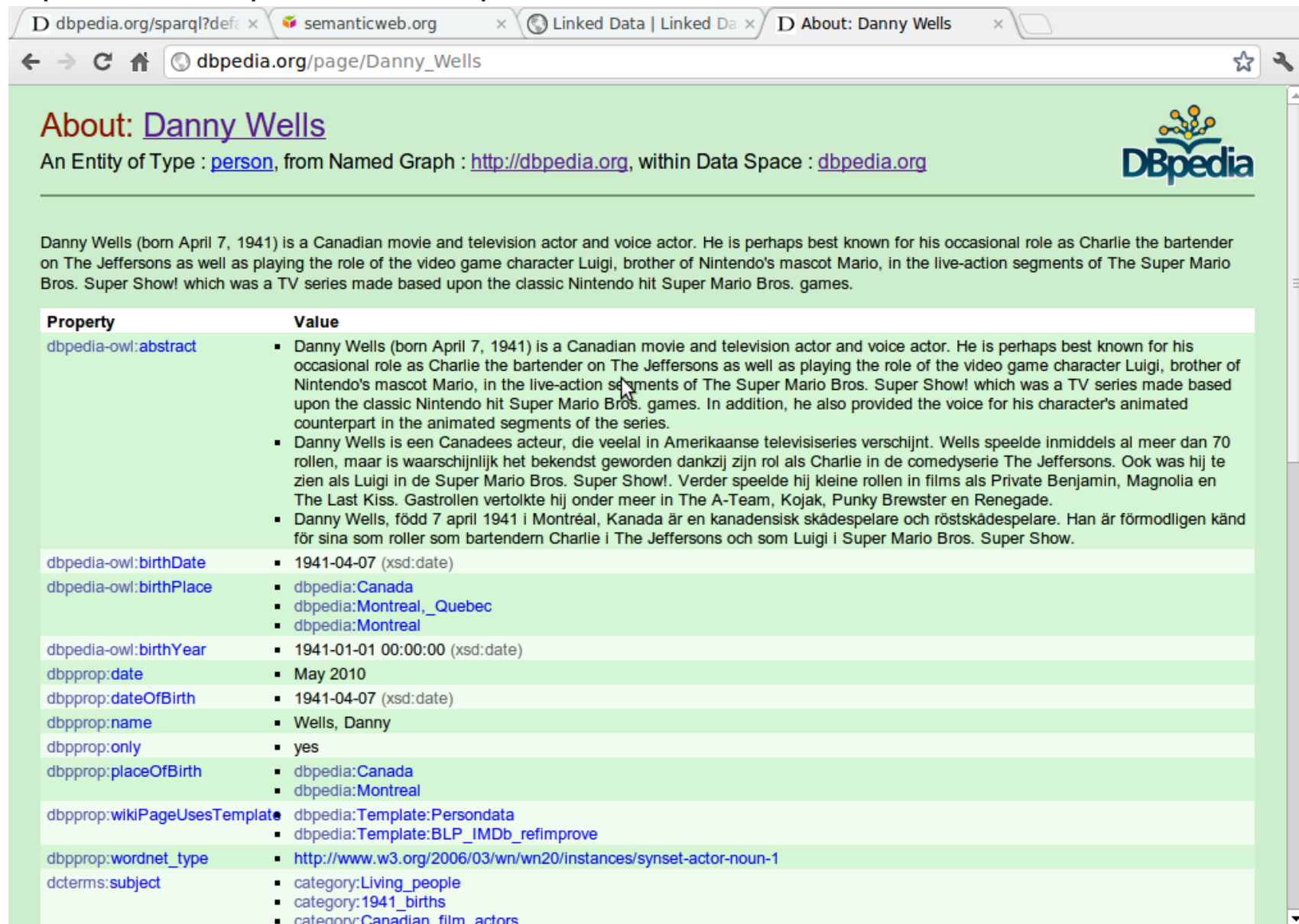
foaf:page <http://data.bnf.fr/11928016/jules\_verne/> .

Liens avec d'autres données RDF

# LOD Domain médical



## Exemple de description dans DBpedia:



DBpedia.org/sparql?def: x semanticweb.org x Linked Data | Linked Da x About: Danny Wells

dbpedia.org/page/Danny\_Wells

### About: [Danny Wells](#)

An Entity of Type : [person](#), from Named Graph : <http://dbpedia.org>, within Data Space : [dbpedia.org](#)

Danny Wells (born April 7, 1941) is a Canadian movie and television actor and voice actor. He is perhaps best known for his occasional role as Charlie the bartender on The Jeffersons as well as playing the role of the video game character Luigi, brother of Nintendo's mascot Mario, in the live-action segments of The Super Mario Bros. Super Show! which was a TV series made based upon the classic Nintendo hit Super Mario Bros. games.

Property	Value
<a href="#">dbpedia-owl:abstract</a>	<ul style="list-style-type: none"><li>Danny Wells (born April 7, 1941) is a Canadian movie and television actor and voice actor. He is perhaps best known for his occasional role as Charlie the bartender on The Jeffersons as well as playing the role of the video game character Luigi, brother of Nintendo's mascot Mario, in the live-action segments of The Super Mario Bros. Super Show! which was a TV series made based upon the classic Nintendo hit Super Mario Bros. games. In addition, he also provided the voice for his character's animated counterpart in the animated segments of the series.</li><li>Danny Wells is een Canadees acteur, die veelal in Amerikaanse televisieseries verschijnt. Wells speelde inmiddels al meer dan 70 rollen, maar is waarschijnlijk het bekendst geworden dankzij zijn rol als Charlie in de comedyserie The Jeffersons. Ook was hij te zien als Luigi in de Super Mario Bros. Super Show!. Verder speelde hij kleine rollen in films als Private Benjamin, Magnolia en The Last Kiss. Gastrollen vertolkte hij onder meer in The A-Team, Kojak, Punky Brewster en Renegade.</li><li>Danny Wells, född 7 april 1941 i Montréal, Kanada är en kanadensisk skådespelare och röstskådespelare. Han är förmodligen känd för sina som roller som bartendern Charlie i The Jeffersons och som Luigi i Super Mario Bros. Super Show.</li></ul>
<a href="#">dbpedia-owl:birthDate</a>	<ul style="list-style-type: none"><li>1941-04-07 (xsd:date)</li></ul>
<a href="#">dbpedia-owl:birthPlace</a>	<ul style="list-style-type: none"><li><a href="#">dbpedia:Canada</a></li><li><a href="#">dbpedia:Montreal,_Quebec</a></li><li><a href="#">dbpedia:Montreal</a></li></ul>
<a href="#">dbpedia-owl:birthYear</a>	<ul style="list-style-type: none"><li>1941-01-01 00:00:00 (xsd:date)</li></ul>
<a href="#">dbpprop:date</a>	<ul style="list-style-type: none"><li>May 2010</li></ul>
<a href="#">dbpprop:dateOfBirth</a>	<ul style="list-style-type: none"><li>1941-04-07 (xsd:date)</li></ul>
<a href="#">dbpprop:name</a>	<ul style="list-style-type: none"><li>Wells, Danny</li></ul>
<a href="#">dbpprop:only</a>	<ul style="list-style-type: none"><li>yes</li></ul>
<a href="#">dbpprop:placeOfBirth</a>	<ul style="list-style-type: none"><li><a href="#">dbpedia:Canada</a></li><li><a href="#">dbpedia:Montreal</a></li></ul>
<a href="#">dbpprop:wikiPageUsesTemplate</a>	<ul style="list-style-type: none"><li><a href="#">dbpedia:Template:Persondata</a></li><li><a href="#">dbpedia:Template:BLP_IMDb_refimprove</a></li></ul>
<a href="#">dbpprop:wordnet_type</a>	<ul style="list-style-type: none"><li><a href="http://www.w3.org/2006/03/wn/wn20/instances/synset-actor-noun-1">http://www.w3.org/2006/03/wn/wn20/instances/synset-actor-noun-1</a></li></ul>
<a href="#">dcterms:subject</a>	<ul style="list-style-type: none"><li><a href="#">category:Living_people</a></li><li><a href="#">category:1941_births</a></li><li><a href="#">category:Canadian film actors</a></li></ul>



<http://dbpedia.org/resource/Danny\_Wells> <http://dbpedia.org/ontology/birthPlace> <http://dbpedia.org/resource/Montreal> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://dbpedia.org/ontology/birthPlace> <http://dbpedia.org/resource/Canada> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://dbpedia.org/property/placeOfBirth> <http://dbpedia.org/resource/Montreal> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://dbpedia.org/class/yago/Person100007846> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://dbpedia.org/ontology/abstract> "Danny Wells (born April 7, 1941) is a Canadian movie and television actor and voice actor. He is perhaps best known for his occasional role as Charlie the bartender on The Jeffersons as well as playing the role of the video game character Luigi, brother of Nintendo's mascot Mario, in the live-action segments of The Super Mario Bros. Super Show! which was a TV series made based upon the classic Nintendo hit Super Mario Bros. games. In addition, he also provided the voice for his character's animated counterpart in the animated segments of the series."@en .  
<http://dbpedia.org/resource/Danny\_Wells> <http://dbpedia.org/ontology/birthYear> "1941-01-01T00:00:00-04:00"^^<http://www.w3.org/2001/XMLSchema#gYear> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://www.w3.org/2000/01/rdf-schema#comment> "Danny Wells is een Canadees acteur, die veelal in Amerikaanse televisieseries verschijnt. Wells speelde inmiddels al meer dan 70 rollen, maar is waarschijnlijk het bekendst geworden dankzij zijn rol als Charlie in de comedyserie The Jeffersons. Ook was hij te zien als Luigi in de Super Mario Bros. Super Show!. Verder speelde hij kleine rollen in films als Private Benjamin, Magnolia en The Last Kiss. Gastrollen vertolkte hij onder meer in The A-Team, Kojak, Punky Brewster en Renegade."@nl .  
<http://dbpedia.org/resource/Danny\_Wells> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://dbpedia.org/class/yago/CanadianFilmActors> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://purl.org/dc/terms/subject> <http://dbpedia.org/resource/Category:People\_from\_Montreal> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://www.w3.org/2000/01/rdf-schema#comment> "Danny Wells (born April 7, 1941) is a Canadian movie and television actor and voice actor. He is perhaps best known for his occasional role as Charlie the bartender on The Jeffersons as well as playing the role of the video game character Luigi, brother of Nintendo's mascot Mario, in the live-action segments of The Super Mario Bros. Super Show! which was a TV series made based upon the classic Nintendo hit Super Mario Bros. games."@en .  
<http://dbpedia.org/resource/Danny\_Wells> <http://dbpedia.org/property/wikiPageUsesTemplate> <http://dbpedia.org/resource/Template:BLP\_IMDb\_refimprove> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://xmlns.com/foaf/0.1/page> <http://en.wikipedia.org/wiki/Danny\_Wells> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://purl.org/dc/terms/subject> <http://dbpedia.org/resource/Category:Canadian\_voice\_actors> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://dbpedia.org/ontology/property/name> "Wells, Danny"@en .  
<http://dbpedia.org/resource/Danny\_Wells> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://dbpedia.org/class/yago/CanadianVoiceActors> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://dbpedia.org/ontology/birthDate> "1941-04-07"^^<http://www.w3.org/2001/XMLSchema#date> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://www.w3.org/2002/07/owl#sameAs> <http://rdf.freebase.com/ns/m/090mls> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://dbpedia.org/property/date> "May 2010"@en .  
<http://dbpedia.org/resource/Danny\_Wells> <http://purl.org/dc/terms/subject> <http://dbpedia.org/resource/Category:Living\_people> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://dbpedia.org/ontology/abstract> "Danny Wells is een Canadees acteur, die veelal in Amerikaanse televisieseries verschijnt. Wells speelde inmiddels al meer dan 70 rollen, maar is waarschijnlijk het bekendst geworden dankzij zijn rol als Charlie in de comedyserie The Jeffersons. Ook was hij te zien als Luigi in de Super Mario Bros. Super Show!. Verder speelde hij kleine rollen in films als Private Benjamin, Magnolia en The Last Kiss. Gastrollen vertolkte hij onder meer in The A-Team, Kojak, Punky Brewster en Renegade."@nl .  
<http://dbpedia.org/resource/Danny\_Wells> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://schema.org/Person> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://dbpedia.org/ontology/property/only> "yes"@en .  
<http://dbpedia.org/resource/Danny\_Wells> <http://www.w3.org/2000/01/rdf-schema#comment> "Danny Wells, f\u00F6dd 7 april 1941 i Montr\u00E9al, Kanada \u00E4r en kanadensisk sk\u00E5despelare och r\u00F6stsk\u00E5despelare. Han \u00E4r f\u00F6rmodligen k\u00E4nd f\u00F6r sina som roller som bartendern Charlie i The Jeffersons och som Luigi i Super Mario Bros. Super Show."@sv .  
<http://dbpedia.org/resource/Danny\_Wells> <http://dbpedia.org/property/dateOfBirth> "1941-04-07"^^<http://www.w3.org/2001/XMLSchema#date> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://purl.org/dc/terms/subject> <http://dbpedia.org/resource/Category:Canadian\_film\_actors> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://purl.org/dc/terms/subject> <http://dbpedia.org/resource/Category:Canadian\_television\_actors> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://dbpedia.org/property/wikiPageUsesTemplate> <http://dbpedia.org/resource/Template:Persondata> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://www.w3.org/2000/01/rdf-schema#label> "Danny Wells"@sv .  
<http://dbpedia.org/resource/Danny\_Wells> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://dbpedia.org/class/yago/CanadianTelevisionActors> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://xmlns.com/foaf/0.1/givenName> "Danny"@en .  
<http://dbpedia.org/resource/Danny\_Wells> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2002/07/owl#Thing> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://www.w3.org/2000/01/rdf-schema#label> "Danny Wells"@nl .  
<http://dbpedia.org/resource/Danny\_Wells> <http://dbpedia.org/property/wordnet\_type> <http://www.w3.org/2006/03/wn/wn20/instances/synset-actor-noun-1> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://dbpedia.org/ontology/Person> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://purl.org/dc/terms/subject> <http://dbpedia.org/resource/Category:1941\_births> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://dbpedia.org/class/yago/LivingPeople> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://dbpedia.org/class/yago/PeopleFromMontreal> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://www.w3.org/2000/01/rdf-schema#label> "Danny Wells"@en .  
<http://dbpedia.org/resource/Danny\_Wells> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://dbpedia.org/class/yago/Actor109765278> .  
<http://dbpedia.org/resource/Danny\_Wells> <http://xmlns.com/foaf/0.1/name> "Danny Wells"@en .



---

# Produire des données sémantiques et présence sur le réseau



---

# Méthodes de production de données

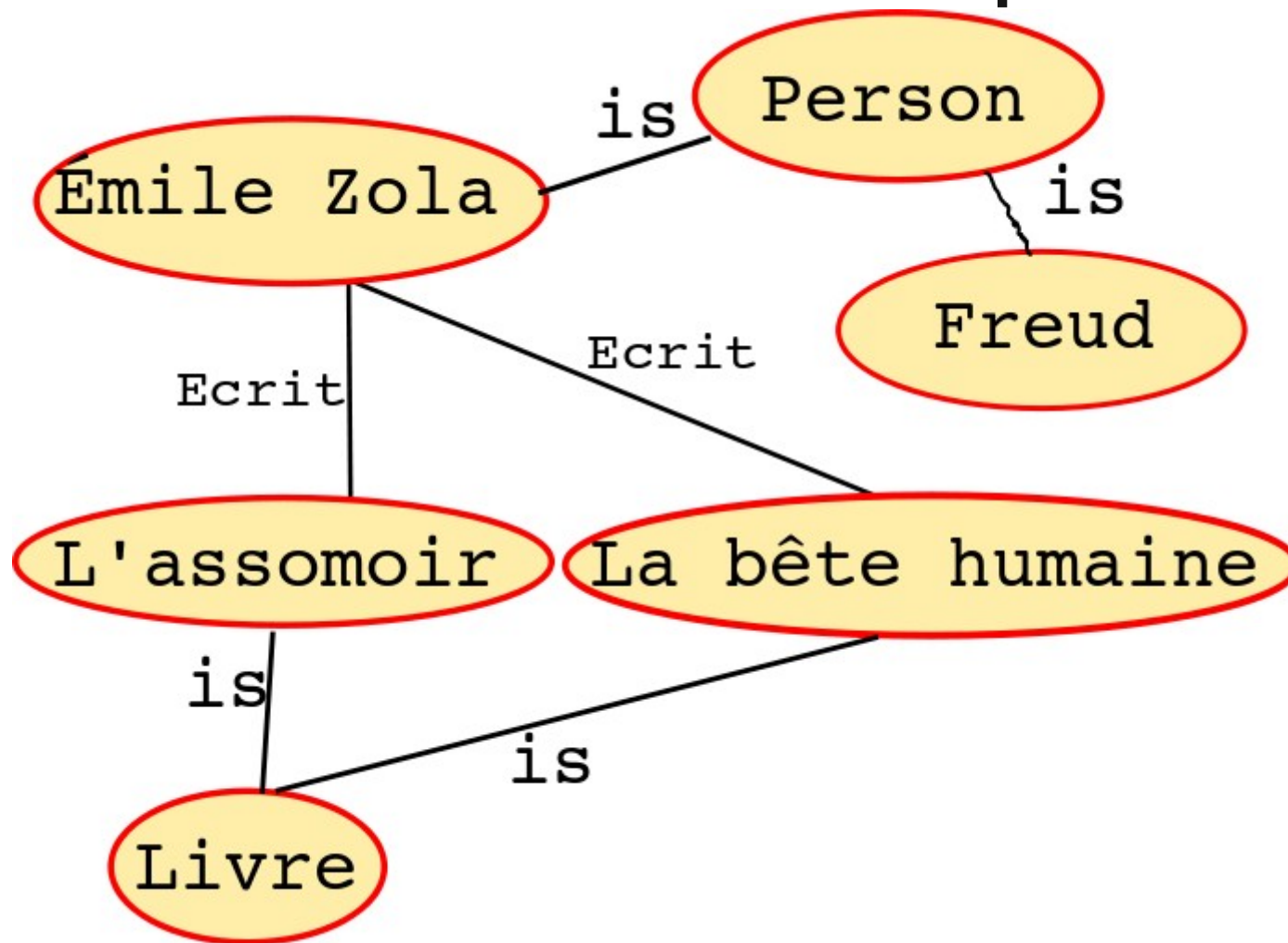
- Écriture manuelle ou édition:
  - La plus fastidieuse
- Profiter des méthodes de production automatique :
  - Fonctions des logiciels (exemple Drupal)
  - Transformation de bases existantes (exemple SQL, CSV avec Virtuoso)
  - Conceptions de modules spécialisés intégrés à des plateformes existantes

# Le défi, passer des tables aux triples

Auteur	œuvre	Catégorie	Nationalité
Émile Zola	L'assomoir	Littérature	Française
Jules Verne	2000 lieues sous les mers	Littérature	Française
Sigmund Freud	Totem et Tabou	Essai	Allemande
Émile Zola	La bête humaine	Littérature	Française

- Phase d'ingénierie des connaissances
- Phase de modélisation
- Choix des bons vocabulaires

# Des tables aux triples



---

# Intégration au LOD






- Comprendre les principes du Web Sémantique
  - Pas d'autorité centrale pour réglementer la diffusion de données
  - Une organisation à but non lucratif pour les recommandations techniques (W3C)
  - Les recommandations ne sont pas des standards (la liberté de publication est totale sur le web)
    - Mais si les recommandations ne sont pas suivies, les autres acteurs du réseau considèrent le dépôt comme non standard
    - Finalement, le dépôt non conforme est marginalisé de fait !

---

# Défi du déploiement

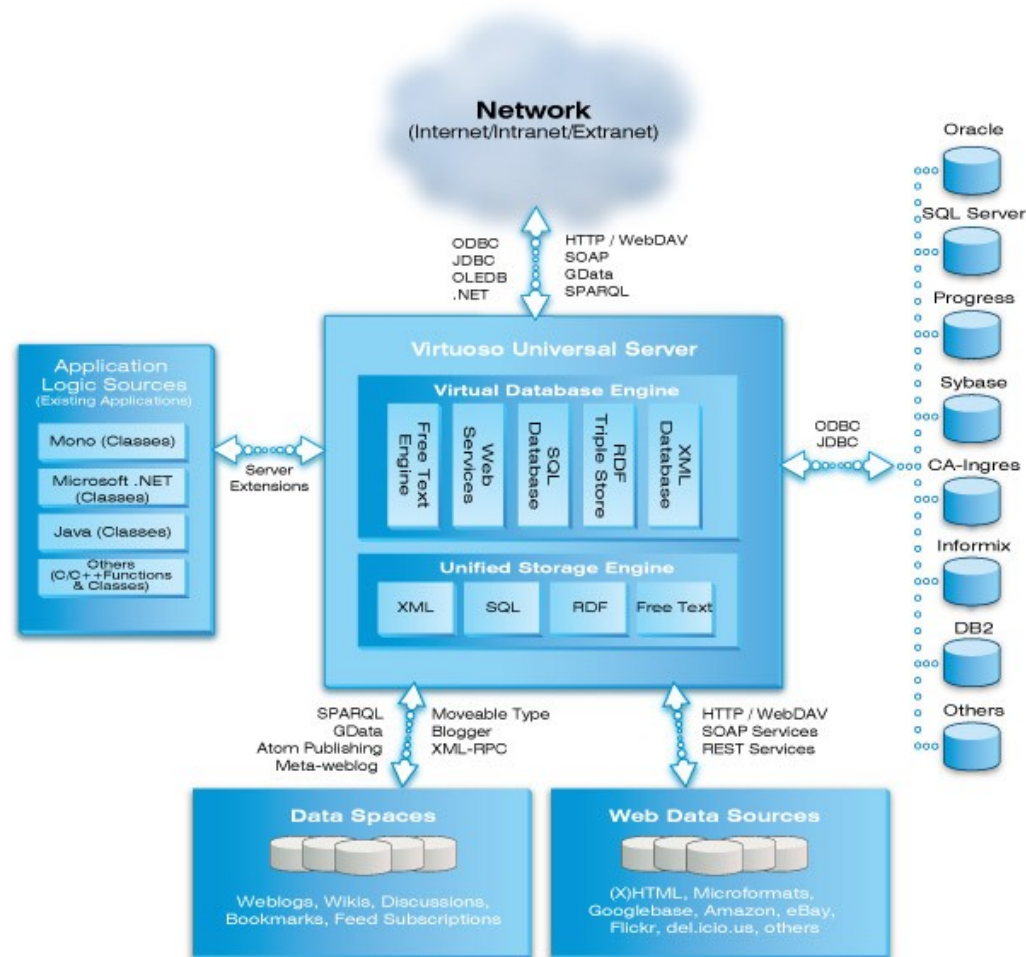
- Définir un ensemble de données à publier
- Suivre les règles de Linked Data
- Diffuser les URI sur le Web
  - Format de base RDF/XML mais aussi N3 ou Turtle.
  - Le point d'accès SPARQL est facultatif
- Relier le dépôt à d'autres

# Les dépôts de données « 5 star »

- Un système d'évaluation de « Linked Open Data » proposé par Tim Berner-Lee en 2010
  - Disponible sur le web avec une licence libre 
  - + Disponible en format lisible par une machine 
  - + Disponible en format libre 
  - + Utilise les standards du W3C (RDF, SparQL) 
  - + Les données sont reliées à d'autres 

En pratique, un dépôt du LOD est 

# Architecture d'un dépôt





---

# Dépôts à explorer

- <http://www.dbpedia.org>
- <http://www.nlgbase.org>
- <http://www.data.gouv.fr/>
- <http://www.freebase.com/>
- <http://www.gutenberg.org/wiki/Gutenberg:Feeds>

---

# Futurs usages du LOD

*Ce n'est pas de la science fiction !*

---

# Futurs usages

- L'étiquetage sémantique (prêt)
- La génération automatique de texte (quasi prêt)
- Le raisonnement (en cours)

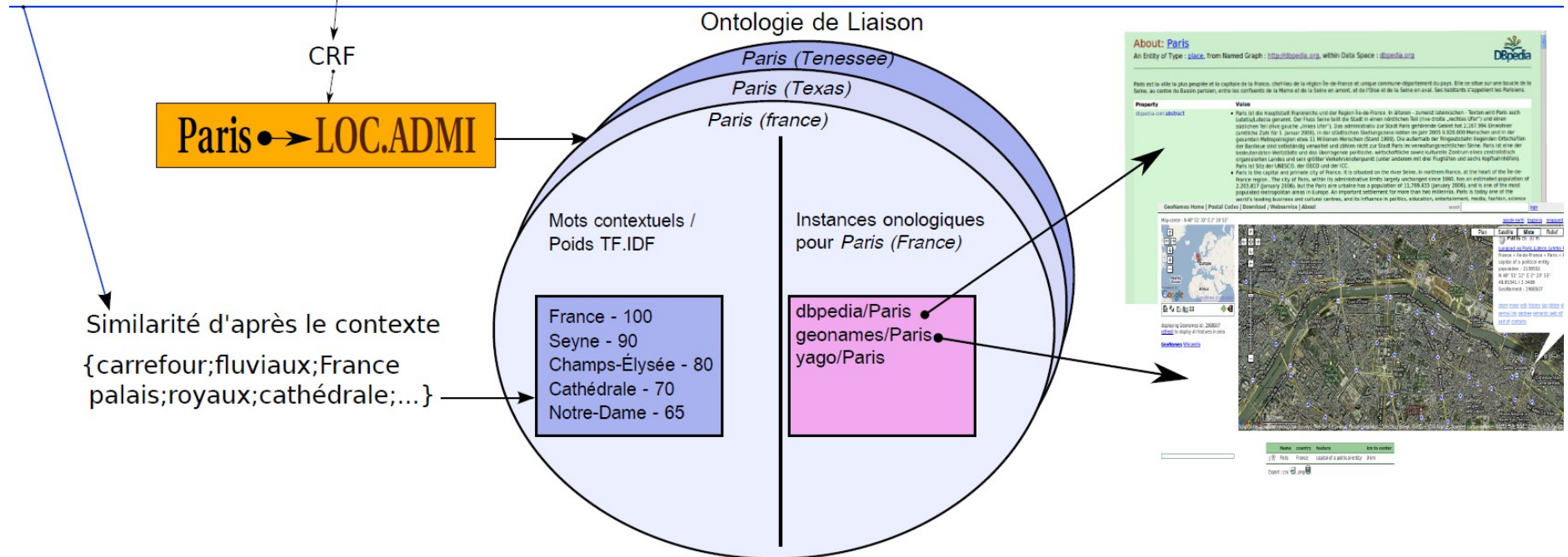
---

# Étiquetage sémantique

- Idée :
  - Utiliser les contenus du Linked Open Data pour enrichir des documents
  - Le Linked Open Data n'est plus seulement un Web de données, il devient une ressource « culturelle »
- Le Web devient véritablement une ressource sémantique

# Étiquetage sémantique

La position de **Paris** à un carrefour entre les itinéraires commerciaux terrestres et fluviaux au cœur d'une riche région agricole en a fait une des principales villes de **France** au cours du Xe siècle, avec des palais royaux, de riches abbayes et une cathédrale



# Étiquetage sémantique

## About: [Paris](#)

An Entity of Type : [place](#), from Named Graph : <http://dbpedia.org>, within Data Space : [dbpedia.org](#)



Paris est la ville la plus peuplée et la capitale de la France, chef-lieu de la région Île-de-France et unique commune-département du pays. Elle se situe sur une boucle de la Seine, au centre du Bassin parisien, entre les confluent de la Marne et de la Seine en amont, et de l'Oise et de la Seine en aval. Ses habitants s'appellent les Parisiens.

### Property

[dbpedia-owl:abstract](#)

### Value

- Paris ist die Hauptstadt Frankreichs und der Region Île-de-France. In älteren - zumeist lateinischen - Texten wird Paris auch Lutetia/Lutezia genannt. Der Fluss Seine teilt die Stadt in einen nördlichen Teil (rive droite „rechtes Ufer“) und einen südlichen Teil (rive gauche „linkes Ufer“). Das administrativ zur Stadt Paris gehörende Gebiet hat 2.167.994 Einwohner (amtliche Zahl für 1. Januar 2006), in der städtischen Siedlungszone lebten im Jahr 2005 9.928.000 Menschen und in der gesamten Metropolregion etwa 11 Millionen Menschen (Stand 1999). Die außerhalb der Ringautobahn liegenden Ortschaften der Banlieue sind selbständig verwaltet und zählen nicht zur Stadt Paris im verwaltungsrechtlichen Sinne. Paris ist eine der bedeutendsten Weltstädte und das überragende politische, wirtschaftliche sowie kulturelle Zentrum eines zentralistisch organisierten Landes und sein größter Verkehrsknotenpunkt (unter anderem mit drei Flughäfen und sechs Kopfbahnhöfen). Paris ist Sitz der UNESCO, der OECD und der ICC.
- Paris is the capital and primate city of France. It is situated on the river Seine, in northern France, at the heart of the Île-de-France region . The city of Paris, within its administrative limits largely unchanged since 1860, has an estimated population of 2,203,817 (January 2006), but the Paris aire urbaine has a population of 11,769,433 (January 2006), and is one of the most populated metropolitan areas in Europe. An important settlement for more than two millennia, Paris is today one of the world's leading business and cultural centres, and its influence in politics, education, entertainment, media, fashion, science and the arts all contribute to its status as one of the world's major global cities. Paris and the Paris Region, with €552.7 billion (US\$813.4 billion) in 2008, produces more than a quarter of the gross domestic product (GDP) of France. According to 2007 estimates, the Paris urban agglomeration is Europe's biggest city economy and the sixth largest in the world. The Paris Region hosts 38 of the Fortune Global 500 companies in several business districts, notably La Défense, the largest purpose-built business district in Europe. Paris also hosts many international organizations such as UNESCO, the Organisation for Economic Co-operation and Development (OECD), the International Chamber of Commerce (ICC) and the informal Paris Club. According to the latest survey from Economist Intelligence Unit in 2010, Paris is the world's most expensive city to live in. Paris is one of the most popular tourist destinations in the world. The Paris region receives 45 million tourists annually, 60% of whom are foreign visitors. The city and region contain numerous iconic landmarks, world-famous institutions and popular parks.
- París es la capital de Francia y de la región de Isla de Francia. Constituida en la única comuna unidepartamental del país, está situada a ambos márgenes de un largo meandro del río Sena, en el centro de la Cuenca parisina, entre la confluencia del río Marne y el Sena aguas arriba, y el Oise y el Sena aguas abajo. La ciudad de París dentro de sus límites administrativos tiene una población estimada de 2 215 197 habitantes (2007). Sin embargo, durante el siglo XX, el área metropolitana de París se expandió más allá de los límites de la comuna. Su área urbana, la segunda más grande de Europa, tiene una población de 11.800.000 habitantes (2006). La región de París es, junto con Londres, el centro económico más importante de Europa. Con

---

# Étiquetage sémantique

- Exemples d'application
  - Contrôles de contenus
  - Analyses de corpus documentaires
  - Livres électroniques interactifs
  - Extraction automatique de nouveaux RDF
  - Analyses de documents
  - ...



# Étiquetage sémantique

**Wikimeta**

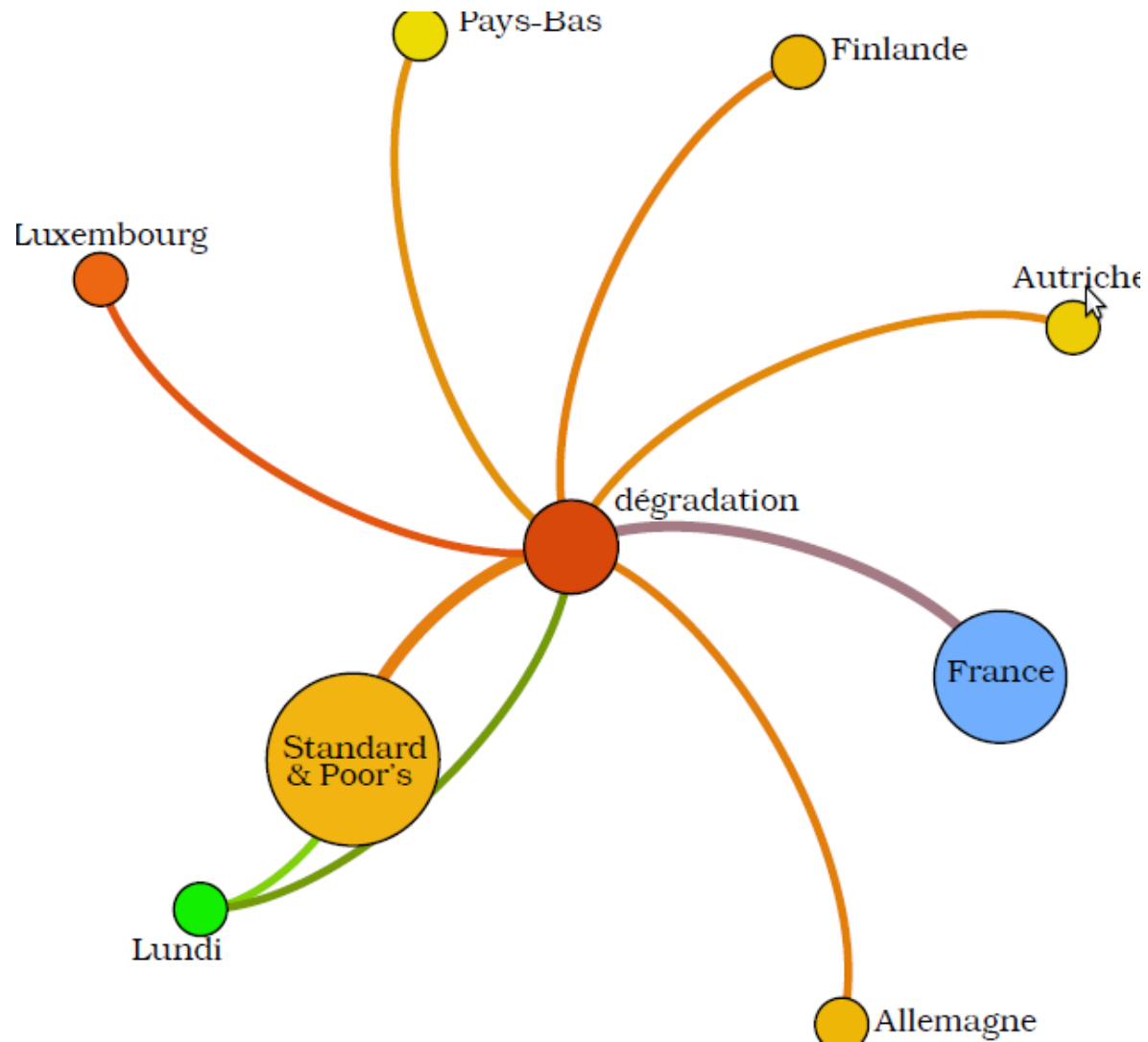
[Semantic Tagger](#)

[Account  
and API](#)

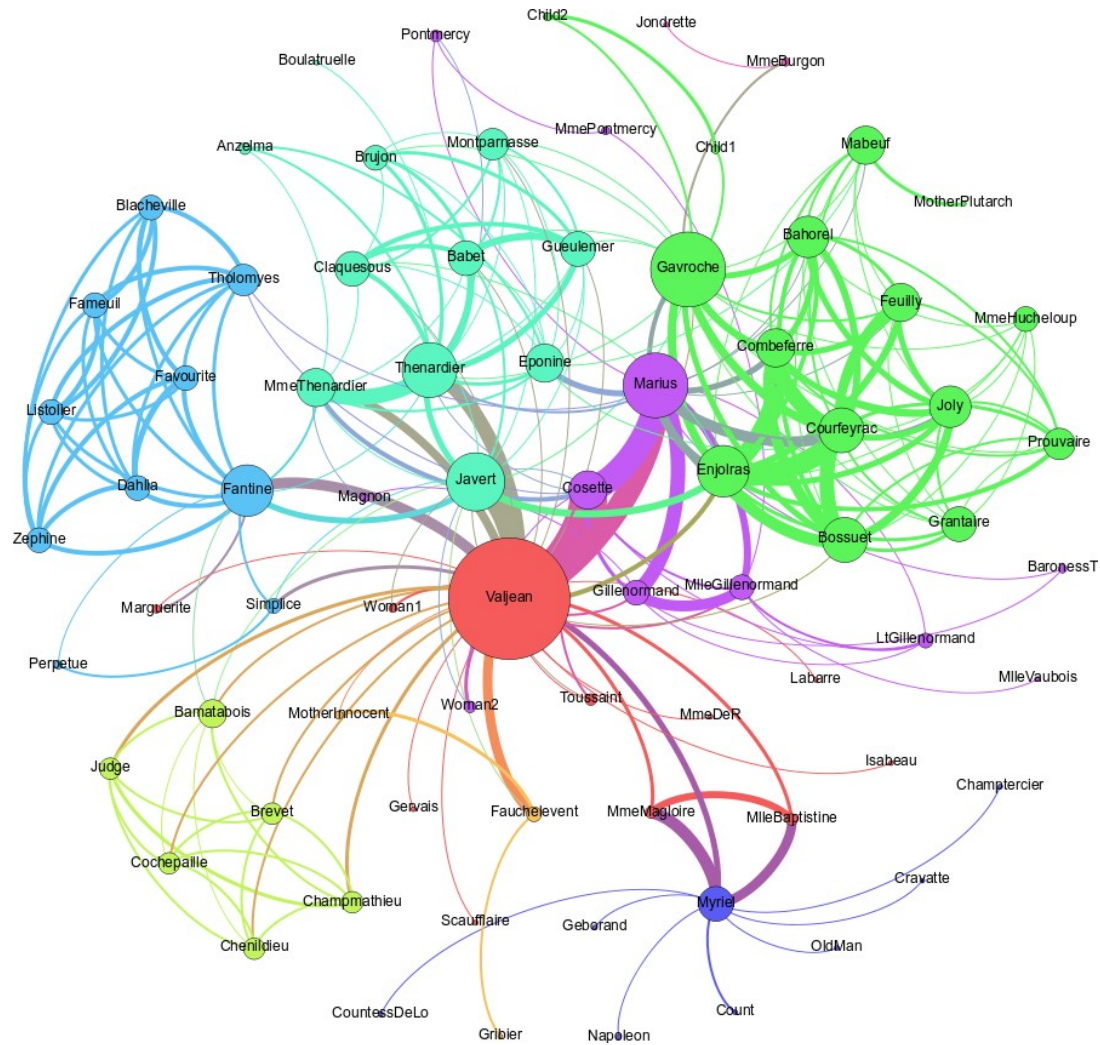
[Standard & Poor \[ORG\]'s \[PROD\]](#) a lancé [lundi \[TIME\]](#) une menace de dégradation du triple A à la [France \[LOC\]](#), qui n'est pas la seule visée. L'agence de notation est sur le point de « mettre sous surveillance avec une implication négative » pas moins de six pays dont l'[Autriche \[LOC\]](#), la [Finlande \[LOC\]](#), le [Luxembourg \[ORG\]](#), les [Pays-Bas \[LOC\]](#) mais aussi l'[Allemagne \[LOC\]](#). [SUR LE MÊME SUJET Accord \[ORG\]](#) Sarkozy-Merkel: l'opposition dénonce « perte de souveraineté » et « austérité » [Hollande \[PERS\]](#) attaque la politique de [Sarkozy \[PERS\]](#), un « échec » [Standard & Poor \[ORG\]](#) 's menace le triple A français et allemandEuro: [Merkel \[PERS\]](#) et [Sarkozy \[PERS\]](#) veulent imposer leur compromis

Les [Etats \[LOC\]](#) auraient été prévenus qu'ils risquaient de perdre leur AAA dans les [quatre-vingt \[TIME\]dix jours \[AMOUNT\]](#) et basculer dans la catégorie [AA \[ORG\]](#) +. Pour la [France \[LOC\]](#), [S & P \[ORG\]](#) envisage même une dégradation plus sévère de deux crans d'un coup.

# Étiquetage sémantique



# Étiquetage sémantique



---

# Extraction des relations

- L'étiquetage des entités n'est qu'une première étape
- On veut aussi extraire des relations sémantiques
- Par exemple:

```
dbpedia:Mozart rel:composerOf dbpedia:Don_Giovanni .
```

---

# Extraction des relations

- Méthode purement statistique:
  - On cherche les co-occurrences des entités dans les phrases
- Par apprentissage machine:
  - Approche supervisée
  - Approche semi-supervisée
- Par règles:
  - Patrons syntaxiques
  -

---

# Approche supervisée

- Corpus de textes annoté manuellement qui indique quelles phrases expriment la relation qu'on veut identifier
- On extrait les caractéristiques intéressantes de ces phrases (mots autour des entités, catégories des mots, structure syntaxique, etc.)
- Avec ces attributs on entraîne un classifieur (réseau bayésien, réseau de neurones, arbre de décision, SVM, etc.)

---

# Approche semi-supervisée

- On part avec quelques exemples de paires d'entités qui sont liées par la relation cible (ex.: Mozart – Don Giovanni)
- On cherche les phrases contenant ces entités
- On extrait les attributs pertinents de ces phrases
- On cherche de nouvelles phrases qui possèdent ces attributs
- À partir de ces nouvelles phrases, on extrait de nouvelles paires d'entités
- On recommence le processus jusqu'à convergence



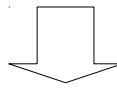
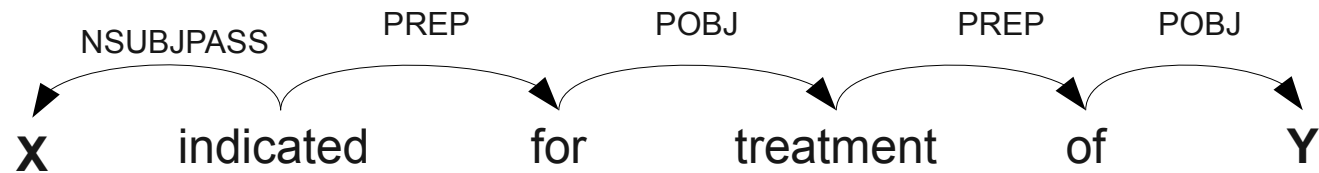
---

# Approche par règles

- On crée une base de patrons syntaxiques qui correspondent aux formes utilisées pour exprimer la relation cible
- Pour chaque patron instancié dans un texte, on construit la représentation RDF qui y correspond

# Approche par règles

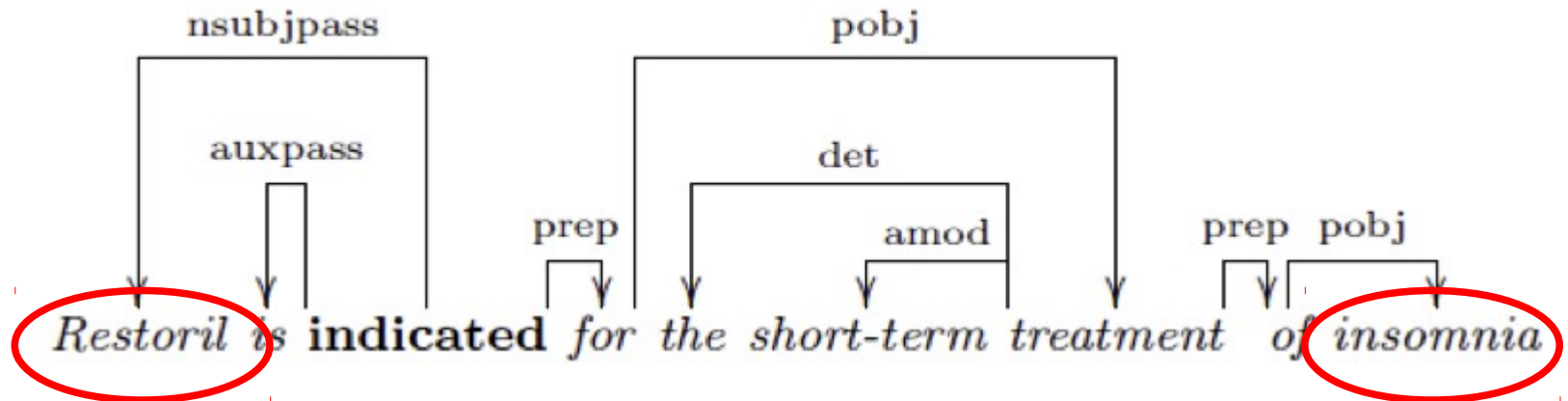
Exemple de patron syntaxique:



```
URI_1 rdfs:label X .  
URI_2 :drugInvolved URI_1 ;  
      rdf:type :Treatment ;  
      :target Y
```

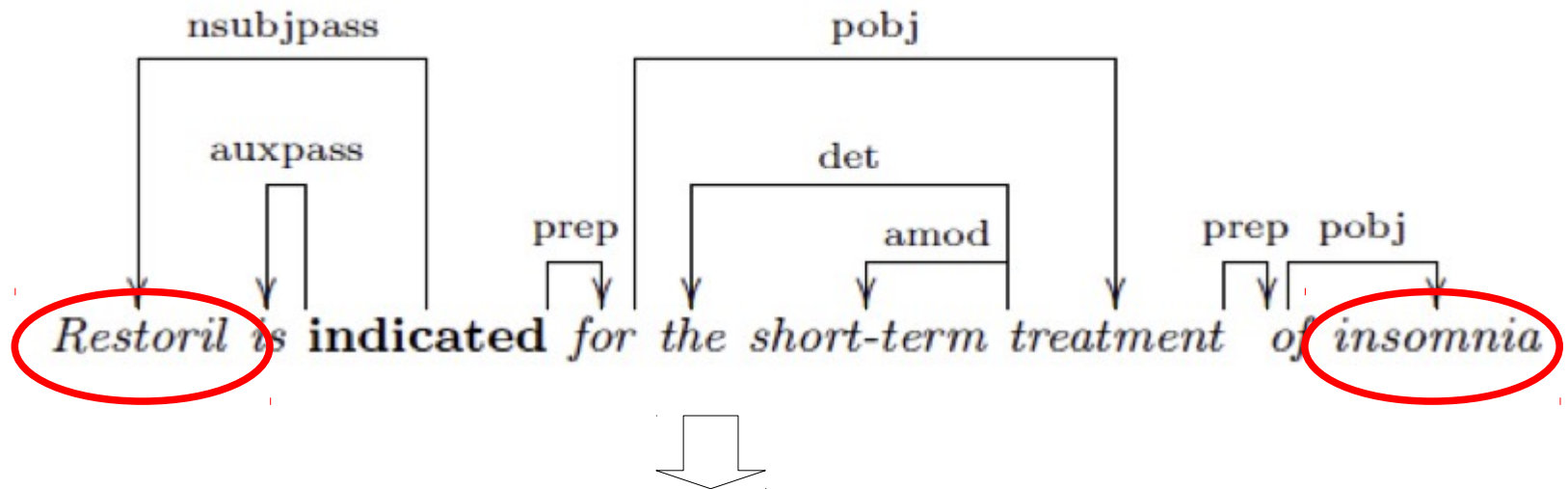
# Approche par règles

Exemple de phrase où on retrouve le patron:



# Approche par règles

On génère les données RDF correspondantes:



```
drug:3239 rdfs:label "Restoril" .
:T234 :drugInvolved drug:3239 ;
      rdf:type :ShortTermTreatment ;
      :target "insomnia"
```

---

# Génération automatique de texte

- Principe

- Utiliser les triplets RDF pour générer automatiquement des documents

- Exemples d'applications

- Produire automatiquement des notices biographiques
- Produire des documentations techniques
- Automatiser la production de contenus répétitifs

# Raisonnements

- Raisonnement

Utiliser les graphes de triples pour extrapoler par des méthodes automatiques de nouveaux liens

