



L'apport du (Web) sémantique sur le BI

D'une intelligence de contenant à une intelligence de contenu



Jean-François Martin
Chantal Lamarre



Les principes

$$\Delta\chi\Delta\rho \geq \frac{\hbar}{2}$$



Web sémantique



1.0



2.0



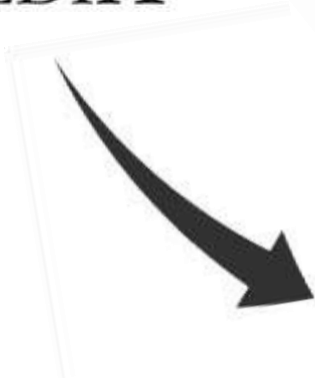
3.0



Web sémantique, un exemple



WIKIPEDIA



Web sémantique, un exemple



Pays	 Canada
Province	 Québec
Région	Montréal
Statut municipal	Métropole
Arrondissements	19 arrondissements (liste)
Fondateur	Paul Chomedey de Maisonneuve
Date de fondation	17 mai 1642
Constitution	1 ^{er} janvier 2002
Maire	Gérald Tremblay
Mandat en cours	(Equipe Tremblay - Union Montréal) 2009 - 2013
Site Web	ville.montreal.qc.ca

Démographie

Population	1 667 700 hab. (2009 ¹)
Densité	4587 hab./km ²
Gentilé	Montréalais, Montréalaise
Langue(s)	Français ²

Géographie

Coordonnées géographiques	 45° 31' 00" Nord 73° 39' 00" Ouest
Altitude	mini. (Pointe-aux-Trembles) 8 m — maxi. (mont Royal) 234 m
Superficie	363.52 km ²
Fuseau horaire	EST (UTC-5)
Indicatif	+1 514 et +1 438
Code géographique	24 66023

dbpedia-owl:areaTotal	<ul style="list-style-type: none"> 140980000.000000 (xsd:double) 365130000.000000 (xsd:double)
dbpedia-owl:areaUrban	<ul style="list-style-type: none"> 1675722307.387392 (xsd:double) 1677000000.000000 (xsd:double)
dbpedia-owl:country	dbpedia:Canada
dbpedia-owl:leaderName	dbpedia:Gérald_Tremblay
dbpedia-owl:leaderTitle	Mayor
dbpedia-owl:maximumElevation	233.000000 (xsd:double)
dbpedia-owl:minimumElevation	6.000000 (xsd:double)
dbpedia-owl:motto	Concordia Salus ("well-being")
dbpedia-owl:populationDensity	<ul style="list-style-type: none"> 4438.630415 (xsd:double) 4439.000000 (xsd:double)
dbpedia-owl:populationMetro	3635571 (xsd:integer)
dbpedia-owl:populationMetroDensity	854.000000 (xsd:double)
dbpedia-owl:populationTotal	1620693 (xsd:integer)
dbpedia-owl:populationUrban	3316615 (xsd:integer)
dbpedia-owl:postalCode	H
dbpedia-owl:thumbnail	http://upload.wikimedia.org
dbpedia-owl:wikiPageExternalLink	<ul style="list-style-type: none"> http://ville.montreal.qc.ca/ http://www.tourisme-montreal.com http://geonames.nrcan.gc.ca/ http://ville.montreal.qc.ca/p http://www12.statcan.ca/en
dbpprop:areaCode	514, 438, and 450, 579
dbpprop:areaMagnitude	1 E8
dbpprop:areaMetroKm	4259 (xsd:integer)
dbpprop:areaMetroSqMi	1644 (xsd:integer)
dbpprop:areaTotalKm	365 (xsd:integer)
dbpprop:areaTotalSqMi	140 (xsd:integer)
dbpprop:areaUrbanKm	1677 (xsd:integer)
dbpprop:areaUrbanSqMi	647 (xsd:integer)
dbpprop:center	Montreal

Investissements



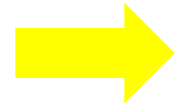
Ouvrir des données

OPEN DATA



Structurer ces données

rdf OWL N-Triples



Relier à d'autres données

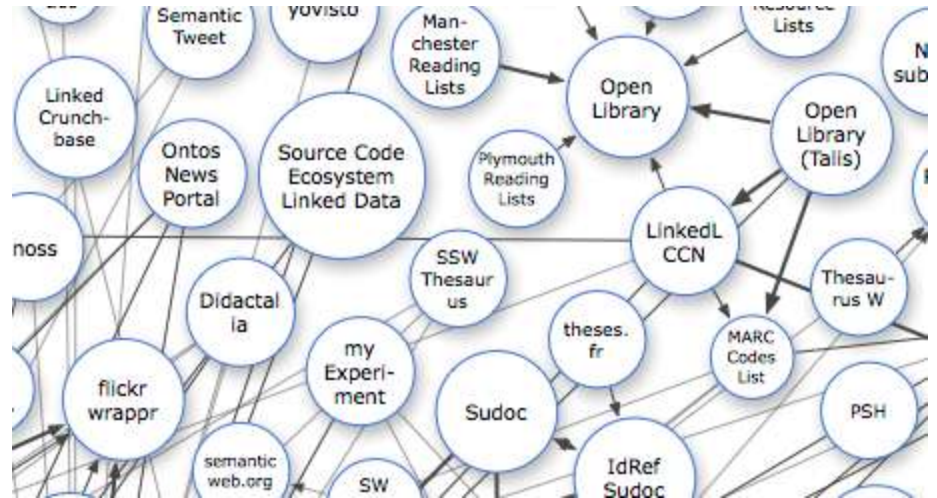


Diffuser — *coopérer sans avoir à coordonner*



Données sémantiques publiques

linkeddata.org



Biodiversité, insectes, plantes

Médicaments et leurs effets secondaires

Disques, artistes, chansons, producteurs et les liens entre eux

Gènes et pseudogènes

Politiciens, partis, agences, fonds, compagnies, lois, etc.



BI sémantique

- Mariage avec les aspects sémantiques de la **gestion de contenu électronique**
- Analyse de **données non structurées** (courriels, microbillets, articles, etc.)
- Interopérabilités avec des sources structurées selon les **standards du domaine sémantiques** (Linked Data, Open Data, etc.)
- Disponibilité de **mécanismes d'inférence**



BI sémantique

Ontology-supported Business Intelligence (OBI)

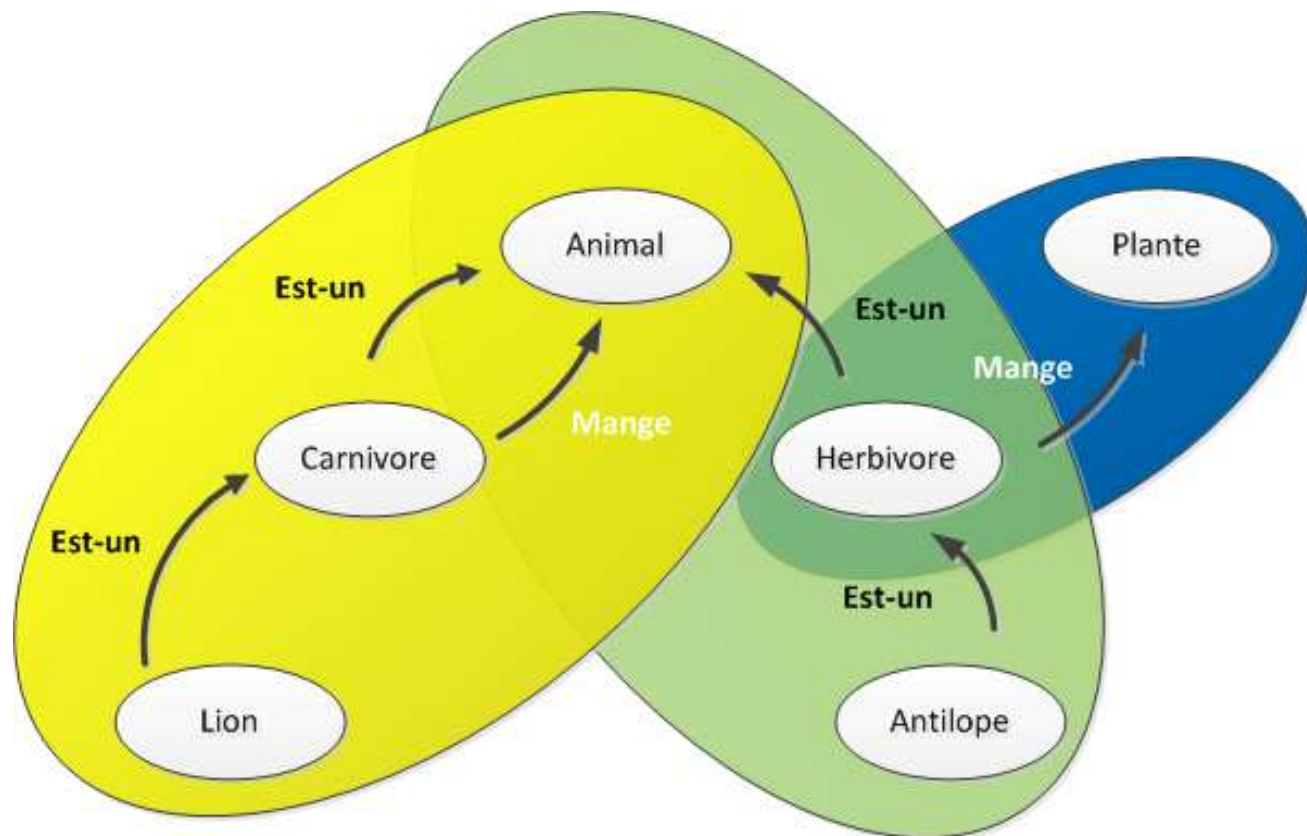


*Semantic Data
Warehouse
(Triplestores)*



Ontology-supported Business Intelligence (OBI)

Une **ontologie** est un **vocabulaire** et ses **liens** (p. ex. : **triplets**) décrivant complètement un **domaine**



Ontology-supported Business Intelligence (OBI)

- Définit le **sens des termes** et ses **relations**
- Permet de **généraliser un problème** afin de le rendre plus compréhensible (évacuer les détails)
- Facilite l'**arrimage** de notions **similaires** ou **interdépendantes**
- Facilite l'**intégration** de sources **externes** et **hétérogènes**
- Participe à la résolution des problèmes de « **big data** »



Ontology-supported Business Intelligence (OBI)

- Ontologies horizontales : modèles objets, taxinomies, vocabulaires XML, etc.
- Ontologies verticales : langage naturel spécifique au domaine
- OWL : langage permettant de partager les ontologies sur le Web



Semantic Data Warehouse (Triplestores)

Un « triplestore » est une base de données spécialisée pour la gestion de données sémantiques (triplets RDF)

L'article traite de sports

Le scarabée est un insecte

Le virus A23s est cancérigène

Jean connaît Paul

	Triplestore	OLAP	NoSQL
Vitesse de mise à jour	+/-	+	++
Requêtes complexes	+	++	-
Inférence	++	+/-	-
Données semi/non structurées	++	-	+/-
Modèle dynamique de données	++	-	-



Semantic Data Warehouse (Triplestores)

- Facilite l'intégration de sources **multiples, hétérogènes et ouvertes**
- Permet la **découverte de connaissances implicites ou cachées** via les mécanismes d'inférence
- Facilite la **mise en relation de données structurées et non structurées**

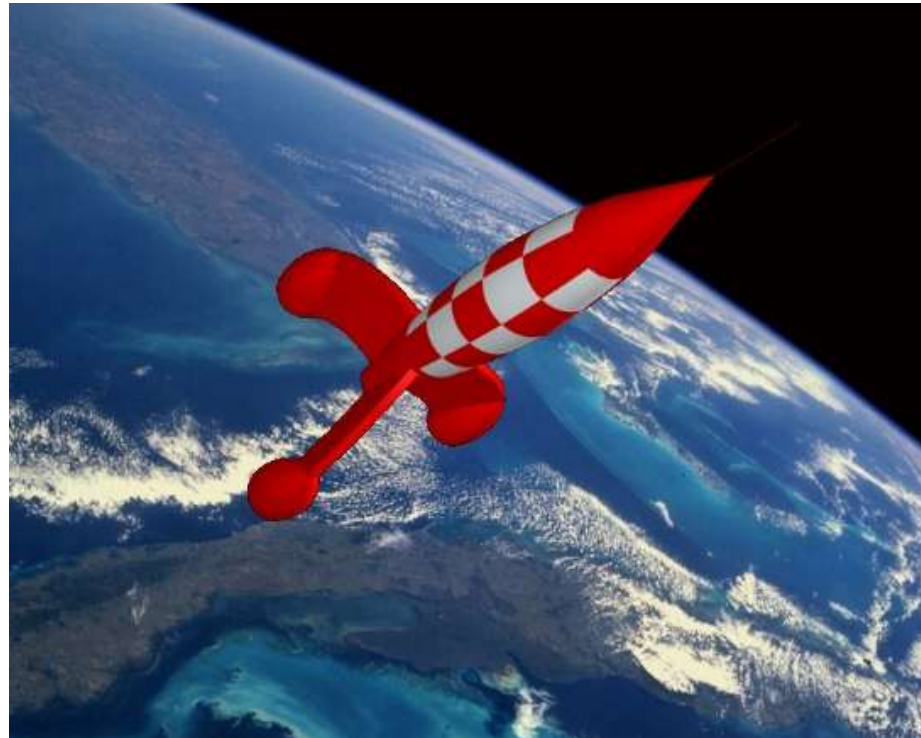


La pratique

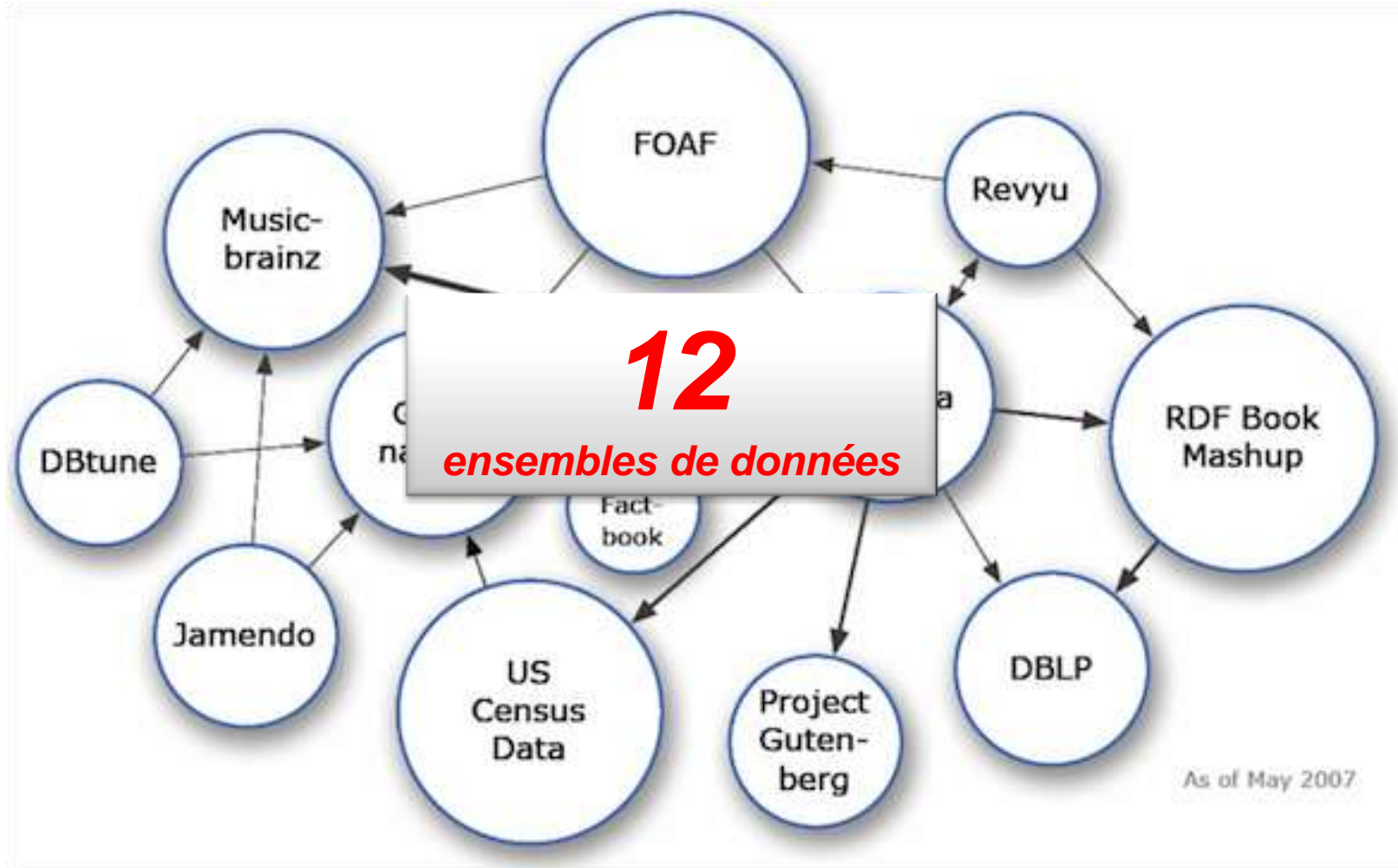


Linked Open Data (LOD)

Une montée fulgurante



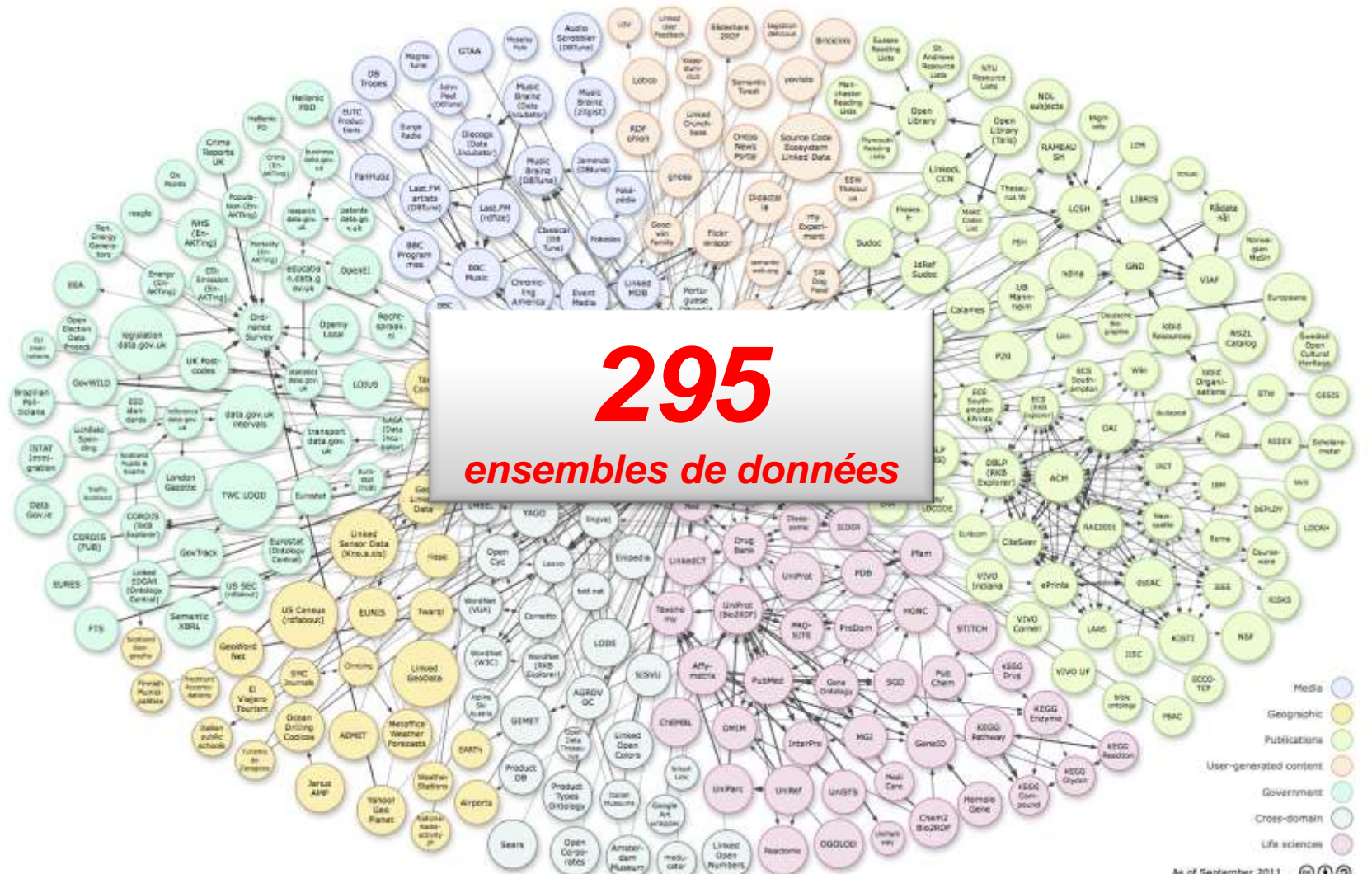
Linked Open Data (LOD) – mai 2007



<http://richard.cyganiak.de/2007/10/lod/>



Linked Open Data (LOD) – aujourd'hui



As of September 2011

<http://richard.cyganiak.de/2007/10/lo/>



Quelques cas pratiques

- Domaine de l'édition juridique
- Domaine de la pratique du droit
- Domaine biomédical



Édition juridique

- Élaboration d'ontologies juridiques
- Utilisation des ontologies pour permettre une gestion efficace de l'information juridique sans cesse croissante
- Acquisition des connaissances, de la rédaction semi-automatique de résumés, de l'indexation semi-automatique



Édition juridique

Contenu

XML

- Acquisition
 - Contenus originaux (de la source) nettoyés et surveillés
- Structuration
 - Contenus structurés
 - Métadonnées
- Annotation (Valeur ajoutée)
 - Ajout de l'enrichissement au résultat de la structuration
- Publication
 - Une fois l'étape d'annotation complétée

Connaissances

- Structuration
 - Ressources terminologiques
 - Ressources ontologiques
- Annotation (Valeur ajoutée)
 - Base de connaissances du domaine
- Publication
 - Taxonomie de publication
 - Base de connaissances publiée

RDF

OWL

SKOS

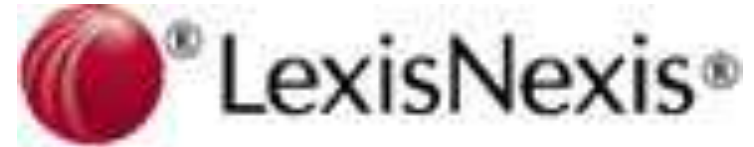


Pratique du droit

- La recherche juridique ne peut plus se limiter à la recherche de mots/termes
 - « dumb data » c. « smart data »
- Les données gouvernementales doivent converger vers des standards ouverts
 - Décisions judiciaires et quasi-judiciaires
 - Législation
 - Documents d'analyse (commentaires du ministre, politique de mise en œuvre, etc.)



Pratique du droit – un exemple



TotalPatent™

The TotalPatent™ service, featuring semantic search technology, is a **patent research, retrieval and analysis solution** designed with the broadest and deepest collection of online first-level patent information available anywhere. With the fast-paced changes and challenges of innovation and patent management, it can be difficult to stay fully informed.

Our TotalPatent service can help you protect, develop and maintain patent assets.

<http://corporate.lexisnexis.com/totalpatent/>



Biomédical

- **Masse** d'information impressionnante
- Information provenant de **différentes sources**
- **Plusieurs disciplines** et domaines d'expertise doivent être explorés
- Information **multilingue**



Biomédical – un exemple



Large Knowledge Collider (LarKC)

- Project end date: September 30th 2011
- The project deals with *large scale ontology mediated Web integration of heterogeneous, evolving and noisy or inconsistent data sources*.
- The Large Knowledge Collider is employing *probabilistic techniques* for selection (e.g. spreading activation techniques), for abstraction (machine learning techniques), and for reasoning (e.g. weighted inference models).
- *Approximate and incomplete reasoning* was one of the main objectives for research to be performed on the Large Knowledge Collider platform.
- The Large Knowledge Collider moves well *beyond existing formalisms* such as RDF, RDF Schema and OWL, without ignoring these achievements. Instead, the Large Knowledge Collider will enable approximate inference on top of these existing formalisms.
- The Large Knowledge Collider is a *reference implementation that can be used as a pluggable experimental platform by other researchers*.

http://www.youtube.com/watch?v=hjUbb14cnAE&feature=player_embedded

