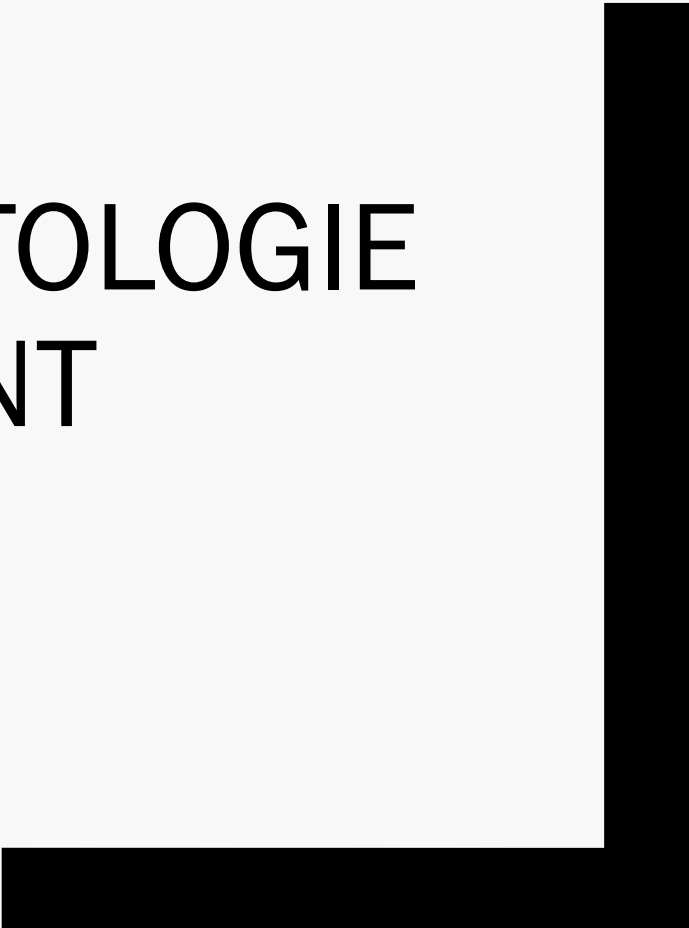


# WEB SÉMANTIQUE, ONTOLOGIE ET RAISONNEMENT

Sana Sellami  
sana.sellami@univ-amu.fr



# Objectif

- Introduire les formalismes de représentation des connaissances dans le Web sémantique et le Web des données
- Interroger et sélectionner les données
- Initier à la programmation logique pour valider sémantiquement et inférer de l'information.

# Organisation

- Intervenants: Belaid Benhamou et Sana Sellami
- *Volume horaire: 9 H cours, 12 h TD, 6h TP*
- Modalités de contrôle des connaissances
  - $NF=0,7*ET+0,3*CC$

# Contenu

- **Partie 1 : Web sémantique et Web des données**
  - *Le web d'aujourd'hui et le web sémantique*
  - *Différentes Technologies du web sémantique*
- **Partie 2: Modèle des données et langage de requêtes**
  - *Modèle de données RDF*
  - *Vocabulaire RDF: RDFS*
  - *Le langage de requêtes SPARQL*
- **Partie 3: Logique de description et ontologies**

# WEB SÉMANTIQUE ET WEB DES DONNÉES

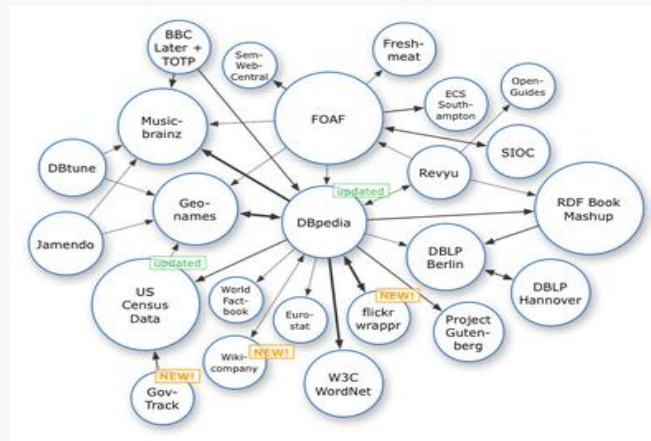
# Le Web

- Qu'est ce que le Web?
- **Quelques rappels :**
- Web  $\neq$  Internet
- Internet, le réseau des réseaux est constitué de plusieurs protocoles, applications, langages...
- le Web en est une des applications, basée notamment sur le protocole HTTP et les liens hypertextes (ou hyperliens).

# Le Web

## ■ Quel Web connaissez vous?

- *Le Web documentaire*
- *Le web applicatif qui surgit des premiers moteurs de recherche*
- *Le web social qui vit naître les premiers wikis?*
- *Le web des services*
- *Le web mobile qui vous suit comme votre ombre?*
- *Ou encore le Web des données liées?*



# Les trois composants de l'architecture Web

## 1. identification & adressage (URL)

ex. <https://www.univ-amu.fr/>

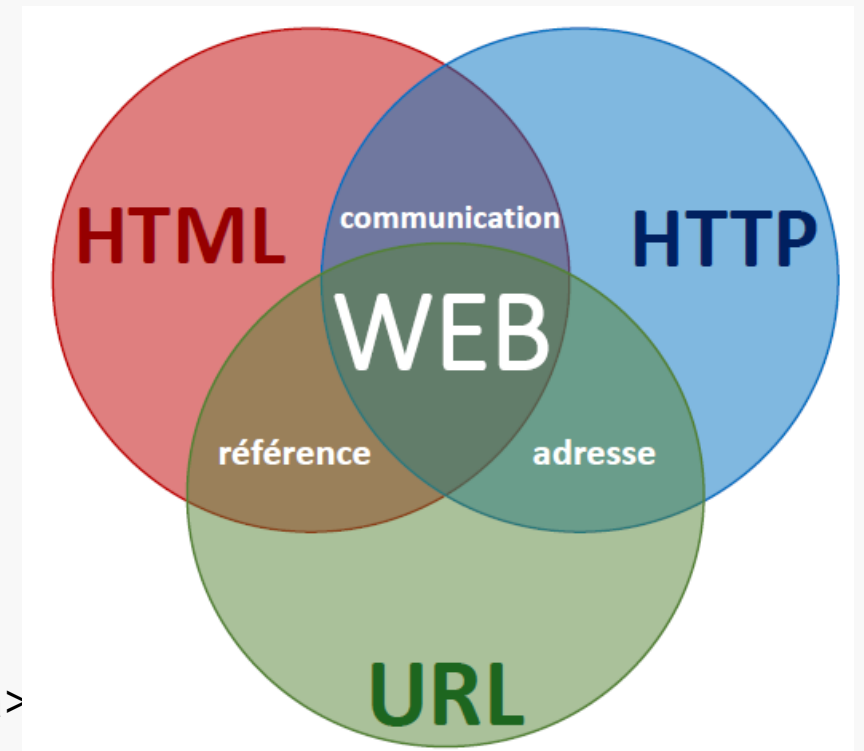
## 2. communication / protocole (HTTP)

GET /masterinfo/SID

Host: <https://www.univ-amu.fr/>

## 3. langage de représentation (HTML)

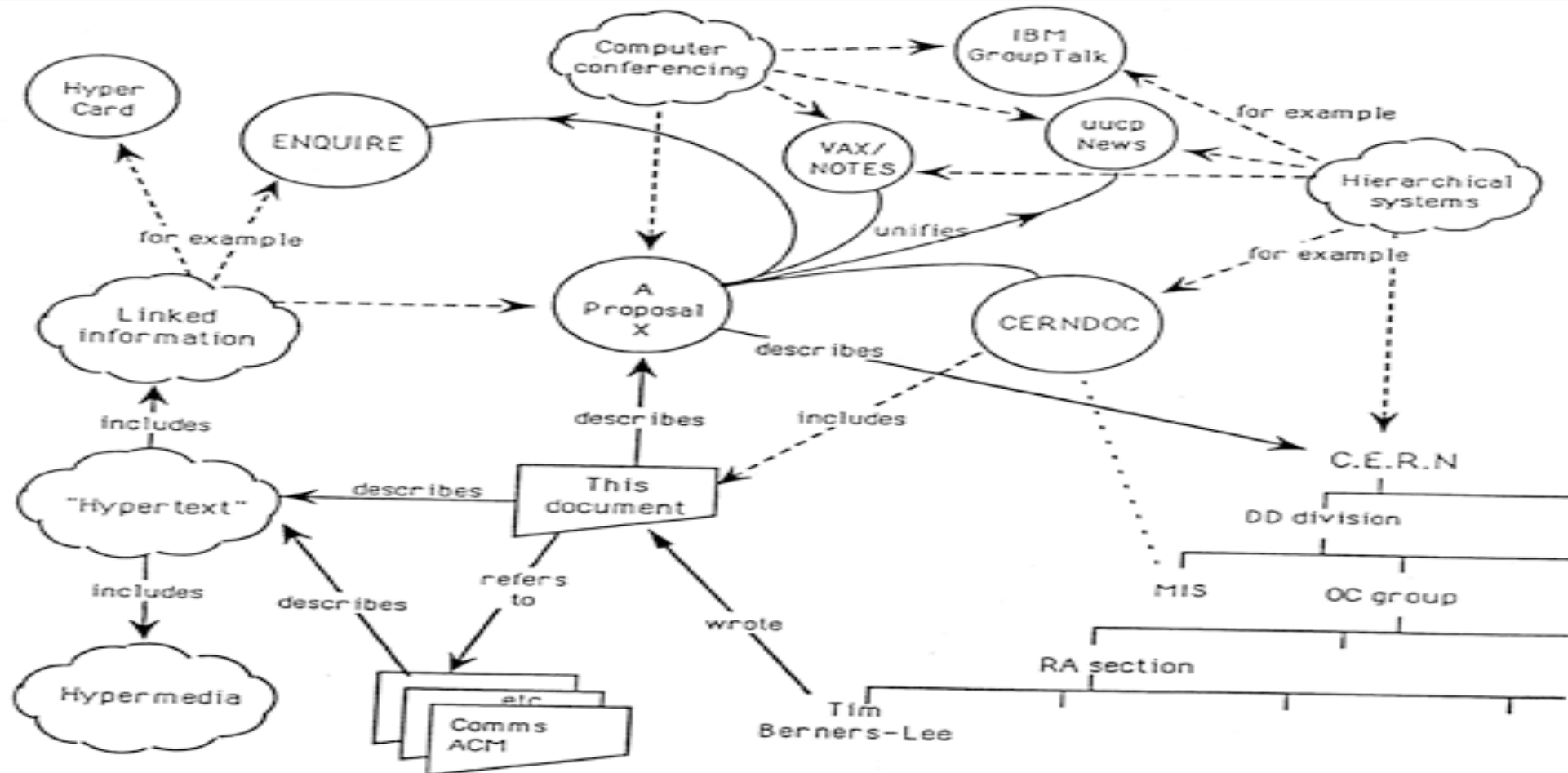
Dupond travaille chez<a href="https://www.univ-amu.fr/">AMU</a>





# Le Web

## ■ Le Web vu par Tim Berners Lee (1989)



# Mais le Web

- Information sur le Web essentiellement prévue pour être affichée (écran, imprimante) et lue par des humains : pages Web
- Seuls les humains peuvent interpréter le contenu des pages Web
- L'interaction entre un utilisateur et le Web passe surtout par un moteur de recherche.
- Ces moteurs font essentiellement une recherche syntaxique de mots-clés
- Usage de « wrappers » pour une extraction automatique à partir de pages Web (par exemple pour indexer les documents selon les mots)
- Le Web actuel est essentiellement **syntaxique** : contenu quasi inaccessible aux traitements machines

# Mais le Web

## ■ INFORMATIONS « cachées » dans le code HTML :

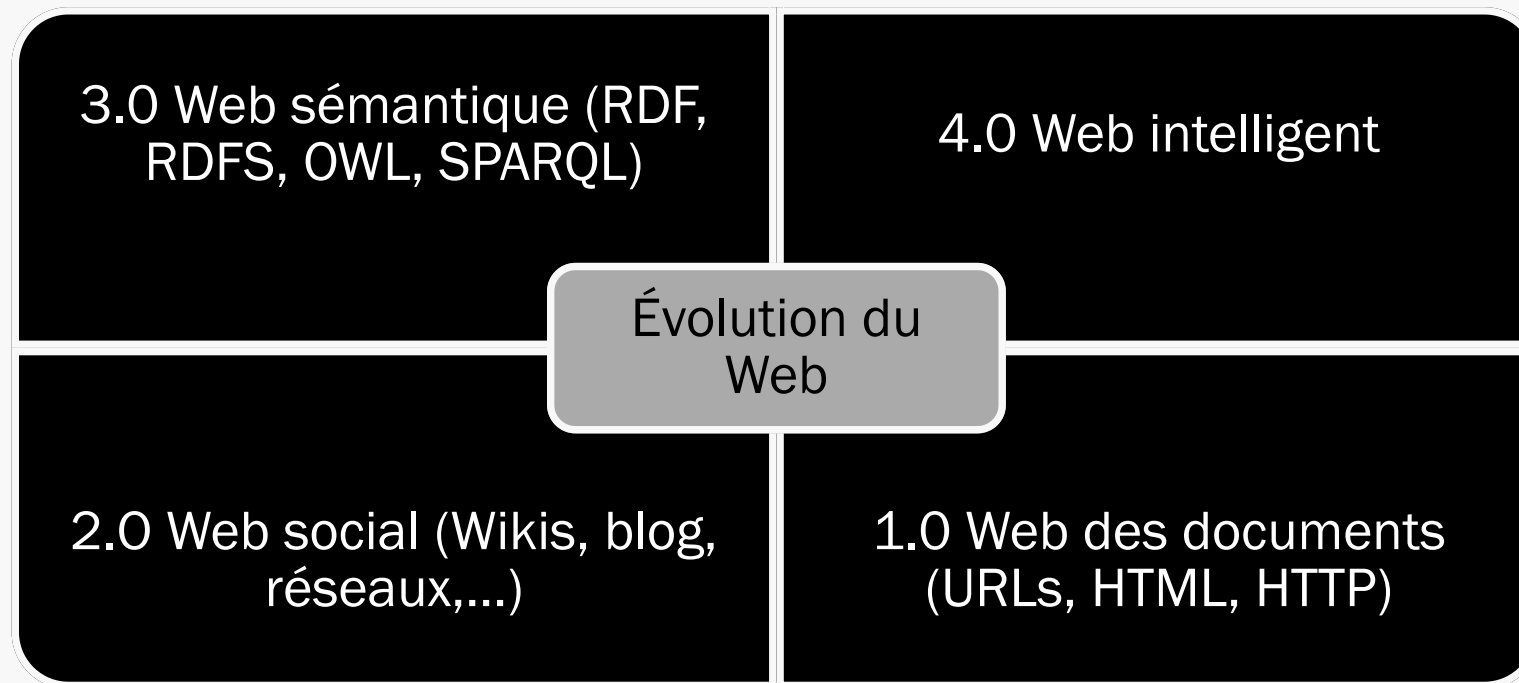
- *code HTML contient l'expression dans une **langue naturelle** (LN) des **informations**, images, fichiers sonores, vidéos, ...*
- ***moteurs de recherche** (sur le texte), pour des raisons de performance (et taille du Web) ne **font aucun traitement sophistiqué** (TALN) des textes :*
  - **recherche de mots**, ce qui est très différent d'une **recherche d'informations**

## ■ SERVICES « cachés » dans le code HTML :

- *Comment **connaître** ce que **propose un service** ?*
- *Comment **utiliser conjointement plusieurs services** ?*

Ex: Achat de billets de trains, validateur html, Web mail, ...

# L'évolution du Web



Le web sémantique propose d'utiliser les métadonnées pour annoter les ressources du web et d'exploiter la sémantique des schémas de ces annotations pour les traiter avec intelligence

# Le Web sémantique

- Un rêve et un projet :

*« J'ai fait un rêve pour le Web [dans lequel les ordinateurs] deviennent capables d'analyser toutes les données sur le Web — le contenu, les liens, et les transactions entre les personnes et les ordinateurs. Un « Web Sémantique », qui devrait rendre cela possible, n'a pas encore émergé, mais quand ce sera fait, les mécanismes d'échange au jour le jour, de bureaucratie et de nos vies quotidiennes seront traités par des machines dialoguant avec d'autres machines. Les « agents intelligents » qu'on nous promet depuis longtemps vont enfin se concrétiser. »* **Tim Berners-Lee, Weaving the Web, 1999**

- **Web sémantique** : permettre aux machines de comprendre la sémantique, la signification de l'information sur le Web

# Le web sémantique: Exemple

- Les Rich Snippets de Google permettent au moteur de recherche d'ajouter des informations, en plus d'un simple lien vers un site, à partir d'informations sémantiques trouvées sur les pages web

Google recettes frites

**Evaluations**

Papilles et Pupilles > fr > 2013/10 > le-secret-des-frites-belges-enfin-re... ▾  
**Le secret des frites Belges enfin révélé (et la recette aussi ...**  
 ★★★★★ Note : 5 - 6 avis **Le nombre d'avis et la durée de préparation**  
2 oct. 2013 - Vous vous demandiez mais pourquoi les **frites** belges sont aussi bonnes ? Ne cherchez plus et venez découvrir leurs secrets de casseroles.

CA Cuisine Actuelle > fr > recettes > frites-maison-277907  
**Frites maison facile : découvrez les recettes de Cuisine Actuelle**  
 ★★★★★ Note : 3,8 - 90 votes - 55 min  
**Recette Frites** maison. Ingrédients (4 personnes) : 1 kg de pommes de terre bintje, Huile de friture, Sel... - Découvrez toutes nos idées de repas et recettes sur ...

Le Journal des Femmes > fr > cuisine > Recettes > Légumes & gratins ▾  
**Recette de Frites parfaites - Cuisine - Le Journal des Femmes**  
 ★★★★★ Note : 4 - 8 avis - 15 min  
Des **frites** croustillantes et fondantes à la fois ! A voir aussi : la **recette** en vidéo. Pelez les pommes de terre, lavez-les une première fois. Détaillez-les ensuite en ...

Donc, certains acteurs du web fournissent des données, d'autres les utilisent.

# Le web sémantique: Exemple

Entreprises et Organisations

## Profil des personnalités

Tim Berners-Lee

Informaticien



Timothy John Berners-Lee, né le 8 juin 1955 à Londres, est un informaticien britannique, principal inventeur du World Wide Web au tournant des années 1990 lors de ses travaux au CERN. [Wikipédia](#)

**Date/Lieu de naissance** : 8 juin 1955 (Âge: 66 ans), Londres, Royaume-Uni

**Épouse** : Rosemary Leith (m. 2014), Nancy Carlson (m. 1990–2011)

**Distinctions** : Bourse MacArthur, Prix Turing, PLUS

**Enfants** : Ben Berners-Lee, Alice Berners-Lee

**Enseignement** : The Queen's College (1973–1976), Emanuel School (1969–1973), Sheen Mount Primary School

**Frères et sœurs** : Mike Berners-Lee

CMA  
CGM

Compagnie



La Compagnie maritime d'affrètement - Compagnie générale maritime est un armateur de porte-conteneurs français. Son offre globale de transport intègre le transport maritime, la manutention portuaire et la logistique terrestre. Il est un leader mondial du transport maritime en conteneurs et le premier français. [Wikipédia](#)

**PDG** : Rodolphe Saadé (9 févr. 2017–)

**Fondateur** : Jacques Saadé

**Création** : 1978

**Siège social** : Marseille

**Nombre d'employés** : 110 000

**Revenus** : 31,45 milliards USD (2020)

**Filiales** : American President Lines, CEVA, Containerships Oy, PLUS

# Le web sémantique

- Le « Semantic Web » est une extension du Web qui favorise les échanges à travers des formats de données communs:
    - *Des liens sémantiques (relations) entre ressources*
    - *Compréhensible par l'humain*
    - *Exploitable par les machines*
    - *Actuellement plusieurs millions de sites Web contiennent des balises propres au web sémantique.*
- Web intelligent → avoir un accès plus facile aux informations: c'est le Web qui viendra à nous



# Le web sémantique Vs Web actuel

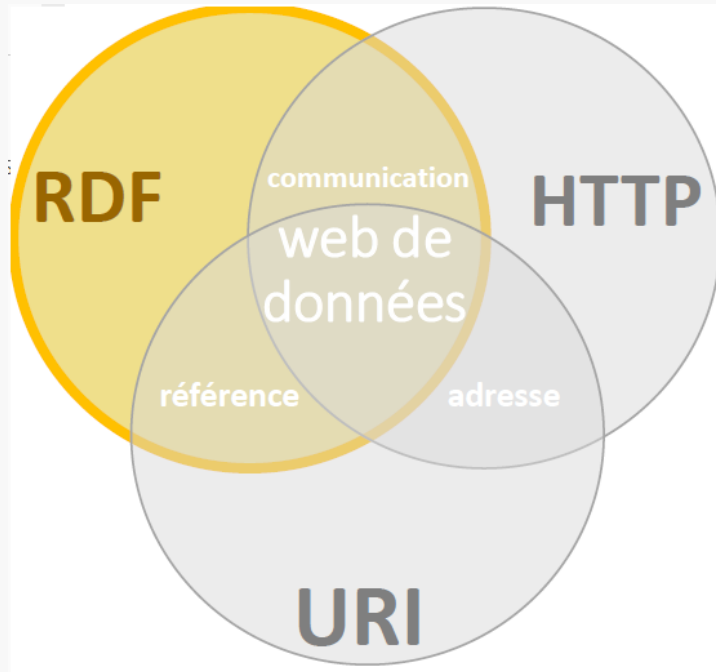
## Web Actuel

- Ensemble de documents
- Basé essentiellement sur HTML
- Recherche par mots clé
- Utilisable par l'humain

## Web Sémantique

- Ensemble de connaissances
- Basé sur XML et RDF(S)
- Recherche par concepts
- Utilisable par la machine

# Le web sémantique: Standards



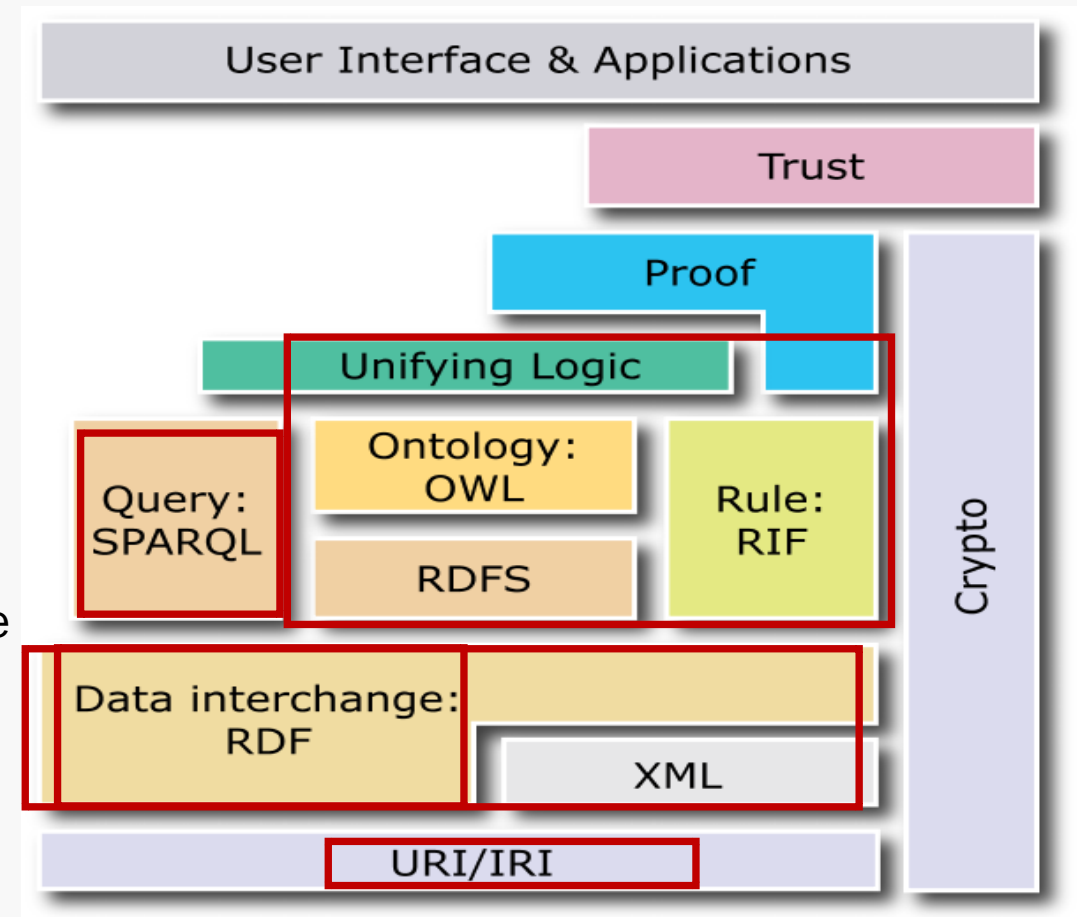
Ontologie = modèle de données représentatif d'un ensemble de concepts dans un domaine, ainsi que des relations entre ces concepts

Raisonnement

Requêtes

La représentation

identification



**Standards du Web sémantique**

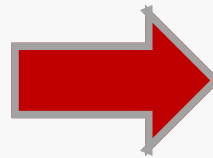
[https://www.w3.org/2006/Talks/0718-aaai-tbl/#\(14\)18](https://www.w3.org/2006/Talks/0718-aaai-tbl/#(14)18)

# De la page à la ressource

## URL

Uniform Resource  
Locator= Identifier ce  
qui existe, sur le Web

L'URL permet  
d'identifier une  
ressource du Web  
(page, ressource,...)



## URI

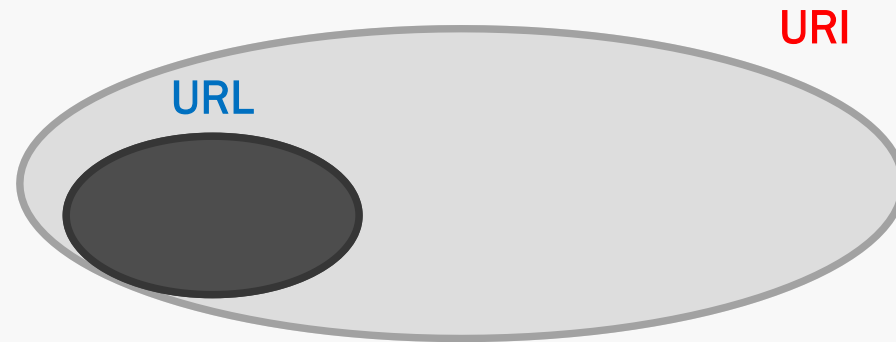
Uniform Resource  
Identifier= Identifier sur  
le Web tout ce qui existe

Une URI est attribuée à  
toute ressource,  
l'insérant dans le Web  
des données

# De la page à la ressource

## Exemple

- On peut fournir des informations sur l'auteur de ce cours mais la personne n'est pas accessible par le Web.
- On trouve des **ressources**: sa page personnelle, une photo, etc. → informations obtenues à partir de leur **URL**
- Les ressources sont reliées à cette personne, mais ne sont pas cette personne.
- Pour désigner la personne, on utilisera une **URI**: un nom unique qui ressemble à une URL mais n'est pas nécessairement accessible sur le Web.



# La ressource

- Toute information qui peut être nommée est une ressource
- Une ressource est identifiée par un identificateur (URI)



# URI (Uniform Resource Identifier)

- Chaque URI identifie une ressource `http://api.domain.com/users/hugo`
- Un URI ne contient pas de verbe
- Un même URI peut être appelé avec différentes **méthodes** HTTP (GET, POST, PUT, DELETE)
- GET `http://api.domain.com/users`  
→ récupère les utilisateurs
- POST `http://api.domain.com/users`  
→ crée un nouvel utilisateur
- DELETE `http://api.domain.com/users/hugo`  
→ supprime Hugo

# Autres couches pour le Web sémantique

- Couche XML
  - *Langage pour représenter les données du Web*
  
- Couche RDF:
  - *RDF: modèle de données*
  - *RDF Schéma: langage pour les ontologies*
  
- Couche Ontologie:
  - *Langage plus expressif que RDF Schéma*
  - *Standard courant pour le Web: OWL*

# Exemples d'applications

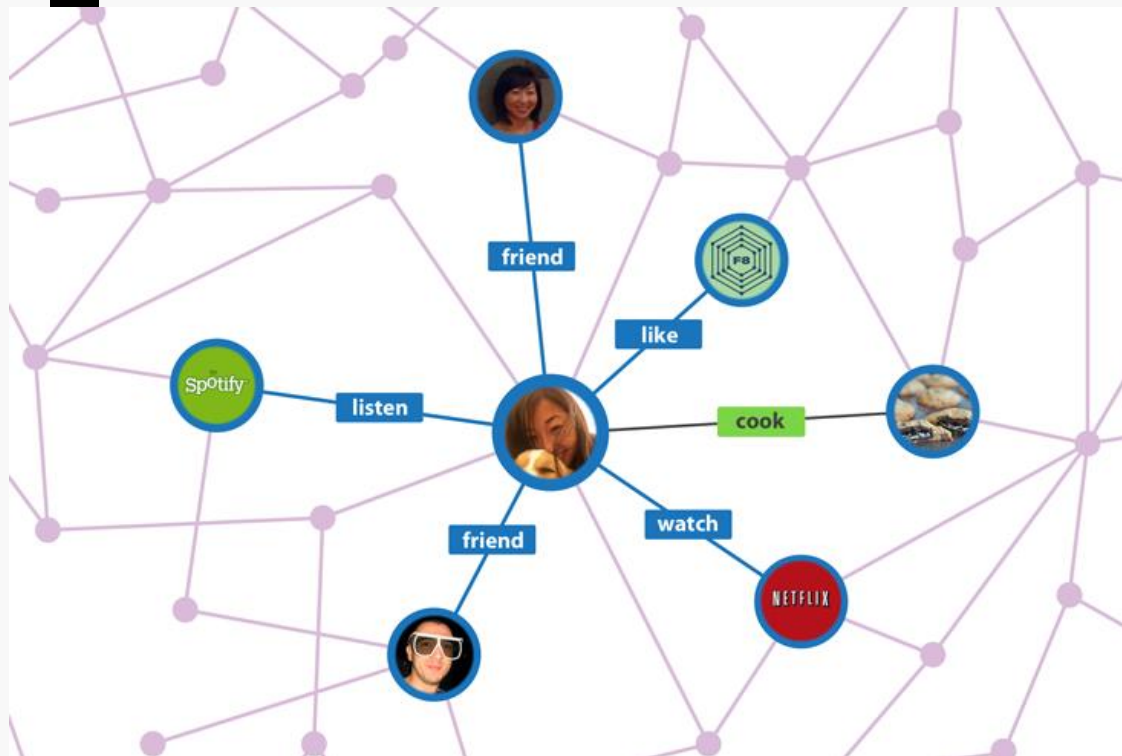
- **Facebook: Open Graph Protocol:** un vocabulaire RDF à intégrer dans une page Web avec RDFa pour enrichir dans le profil Facebook l'affichage de la page Web sélectionné avec le bouton « Like/J'aime ».
- **Google:** (avec la société de gestion de base de données Metaweb): utilise un système de métadonnées pour permettre aux internautes d'arriver aux bonnes informations à partir de requêtes complexes.
- Le site BBC
- ...

<http://ogp.me/>





# Open Graph



<https://developers.facebook.com/docs/opengraph>



Facebook Open Graph Search results for "Music liked by me and my friends who live in Caracas, Venezuela".

Search results list items with their location and the number of likes:

- Music liked by me and my friends who live in **Caracas, Venezuela** (City - 407,193 like this)
- Music liked by me and my friends who live in **Karlsruhe, Germany** (City)
- Music liked by me and my friends who live in **San Antonio De Los Altos, Miranda, Venezuela** (City)
- Music liked by me and my friends who live nearby and live in **Caracas, Venezuela** (City)
- Music liked by me and my friends who live near **Venezuela** (Country)
- Music liked by me and my friends who live nearby and live in **Karlsruhe, Germany** (City)

# Qu'est ce que le web des données?

## ■ Le Web?

Un ensemble de documents interconnectés et accessibles par Internet, à destination des utilisateurs humains qui allaient naviguer d'un document à l'autre pour en traiter eux-mêmes le contenu informatif.

## ■ Le Web des données ?

- *Les données?*

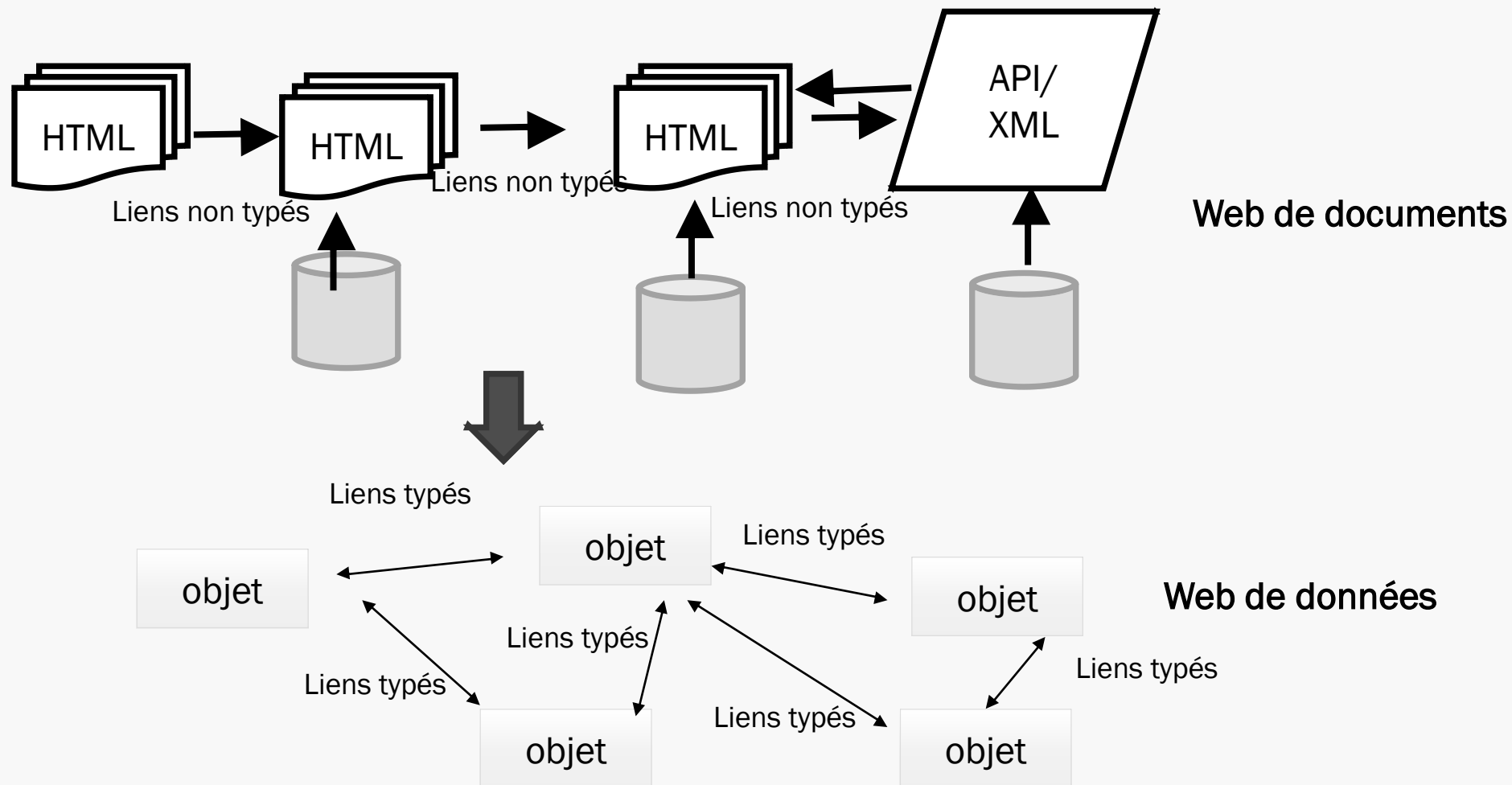
# Web sémantique et Web des données

- Repose sur des technologies existantes préalablement
- + des technologies spécifiquement conçues pour les données
- Le web sémantique met en œuvre le web de données

→ lier et structurer l'information pour accéder simplement à la connaissance qu'elle contient déjà

- Définition de Tim Berners Lee : « Web de données qui peuvent être traitées directement ou indirectement par des machines pour aider leurs utilisateurs à créer de nouvelles connaissances »

# WEB DE DOCUMENTS VERSUS WEB DES DONNEES



# Le web des données

- Une série de **ressources de données** étroitement liées immédiatement lisibles et compréhensibles par l'humain ou la machine.
- Repose sur un standard du Web: **URI** (Uniform Resource Identifier)
- Chaque URI identifie une ressource `http://api.domain.com/users/hugo`
- Ressource:
  - *Documents du Web (identifiés par leur URL)*
  - *Toutes autres choses matérielles ou abstraites dont on parle sur le Web (e.g table, voiture, humain, etc)*

# Le web des données

## ■ Objectif :

- *Mettre à disposition des données en utilisant des techniques standardisées qui garantissent l'interopérabilité*

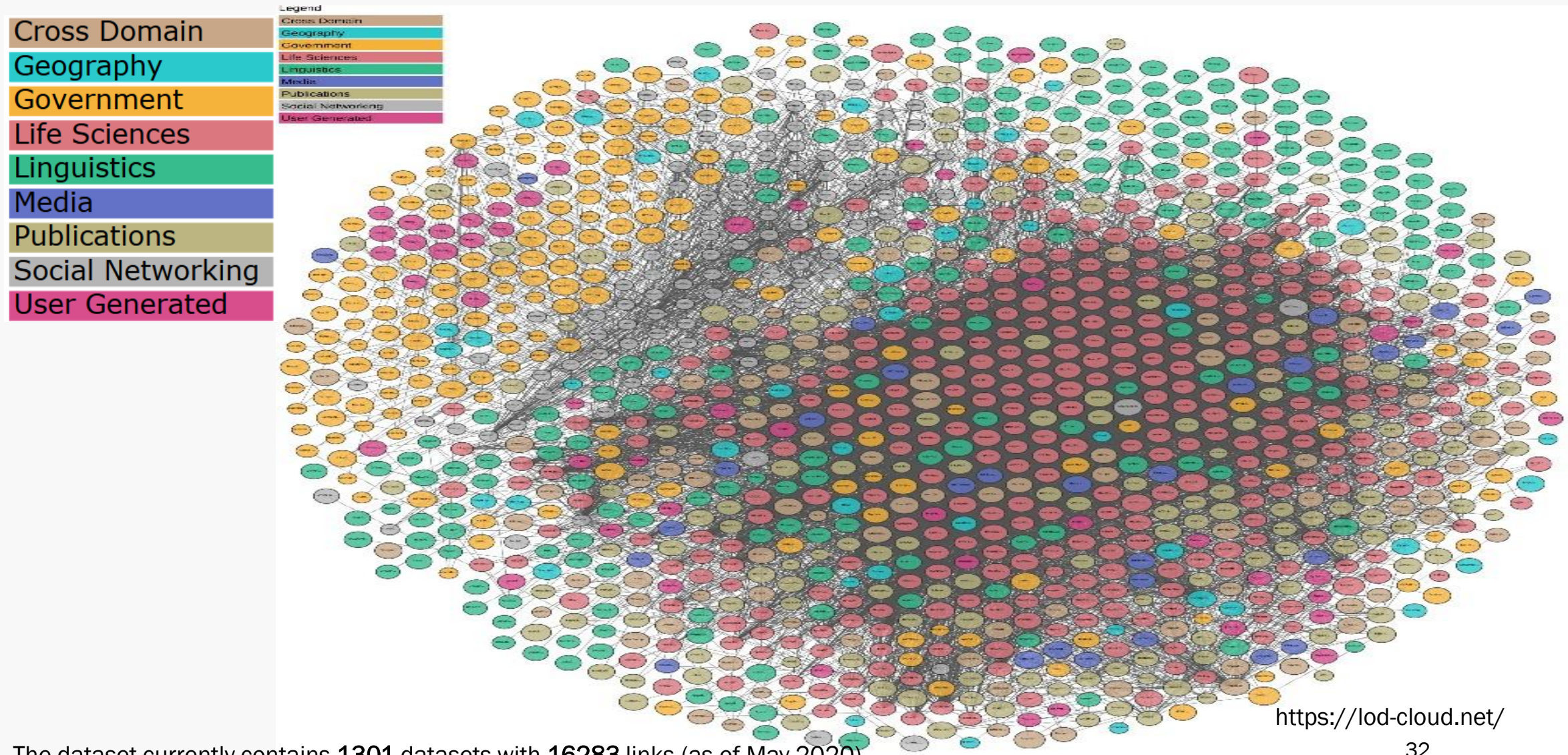
- Relier les données elles-mêmes et les rendre interprétables par les machines
  - « *Permettre aux données d'être partagées et réutilisées au-delà des limites applicatives, organisationnelles ou communautaires* » (W3C)

# Lier le Web des données

- « Le Linked Open Data ou «données ouvertes liées» consiste à rendre libre de toute licence des bases de données préalablement liées entre elles selon le modèle du Linked Data de Tim Berners-Lee. » selon Wikipédia
  
  - **Les quatre principes de Linked Data**
    - *Utiliser des IRIs pour nommer les choses (= ressources).*
    - *Utiliser des IRIs HTTP pour pouvoir obtenir des représentations de ces ressources.*
    - *Fournir ces représentations en utilisant des langages et des protocoles standards (RDF, SPARQL).*
    - *Inclure des liens pour permettre de découvrir de nouvelles ressources.*
- d'après Tim Berners-Lee, <http://www.w3.org/DesignIssues/LinkedData.html>*



# Linking open data



The dataset currently contains **1301** datasets with **16283** links (as of May 2020)



# DBPedia

- dataset de LOD.
- annotation RDF des pages de wikipedia.
- Automatisation de la tâche d'extraction des triplets RDF à partir des pages Wikipédia.
- Interrogation par SparQL
- Stockage des triplets RDF dans une base de données (triple store).

# Exemple Steve Jobs sur Wikipédia



[Accueil](#)  
[Portails thématiques](#)  
[Article au hasard](#)  
[Contact](#)

[Contribuer](#)  
[Débuter sur Wikipédia](#)  
[Aide](#)  
[Communauté](#)  
[Modifications récentes](#)  
[Faire un don](#)

[Outils](#)  
[Pages liées](#)  
[Suivi des pages liées](#)  
[Téléverser un fichier](#)  
[Pages spéciales](#)  
[Lien permanent](#)  
[Informations sur la page](#)

Article [Discussion](#)

Lire [Voir le texte source](#) [Voir l'historique](#)

## Steve Jobs

*Pour les articles homonymes, voir [Jobs](#).*

Vous lisez un « [article de qualité](#) ».

**Steven Paul Jobs**, dit **Steve Jobs**, né à [San Francisco \(Californie\)](#) le **24 février 1955** et mort à [Palo Alto \(Californie\)](#) le **5 octobre 2011**, est un [entrepreneur](#) et [inventeur américain](#), souvent qualifié de visionnaire<sup>1</sup>, et une figure majeure de l'[électronique grand public](#), notamment pionnier de l'avènement de l'[ordinateur personnel](#), du [baladeur numérique](#), du [smartphone](#) et de la [tablette tactile](#). Cofondateur, [directeur général](#) et président du [conseil d'administration d'Apple Inc.](#), il dirige aussi les [studios Pixar](#) et devient membre du conseil d'administration de [Disney](#) lors du rachat en 2006 de Pixar par Disney.

Steve Jobs, [Steve Wozniak](#) et [Ronald Wayne](#) créent [Apple](#) le 1<sup>er</sup> avril 1976 à [Cupertino](#). Au début des [années 1980](#), Steve Jobs saisit le potentiel commercial des travaux du [Xerox Parc](#) sur le couple [interface graphique/souris](#), ce qui conduit à la conception du [Lisa](#), puis du [Macintosh](#) en 1984, les premiers ordinateurs grand public à profiter de ces innovations. Après avoir perdu sa lutte de pouvoir à la tête d'Apple avec [John Sculley](#), le directeur général qu'il avait pourtant recruté, il quitte l'entreprise en septembre 1985 pour fonder [NeXT](#).

En 1986, il rachète la division *Graphics Group* de [Lucasfilm](#), la transforme en [Pixar Animation Studios](#) et rencontre le succès commercial en 1995 avec *[Toy Story](#)*, un film dont il est le [producteur exécutif](#). Il reste directeur général propriétaire de la société (à 50,1 %) jusqu'à son acquisition par la [Walt Disney Company](#) en 2006.

Début 1997, Apple, alors au bord de la faillite, rachète NeXT. L'opération permet à Steve Jobs de revenir à la tête de la firme

### Steve Jobs



Steve Jobs présentant l'iPhone 4 lors du discours d'ouverture

# Exemple Steve Jobs sur Dbpedia

 Browse using  Formats 

Faceted Browser  Sparql Endpoint 

# Construire le Web des données

Deux éléments sont nécessaires :

- Des données que l'on peut référencer sur le Web
- Des gens qui fournissent des liens vers ces données à partir de leurs propres données.
  
- 2 recommandations fondamentales:
  - *RDF pour la représentation des données liées sur le Web.*
  - *SPARQL pour l'interrogation de ces données*

# Modèle de graphe

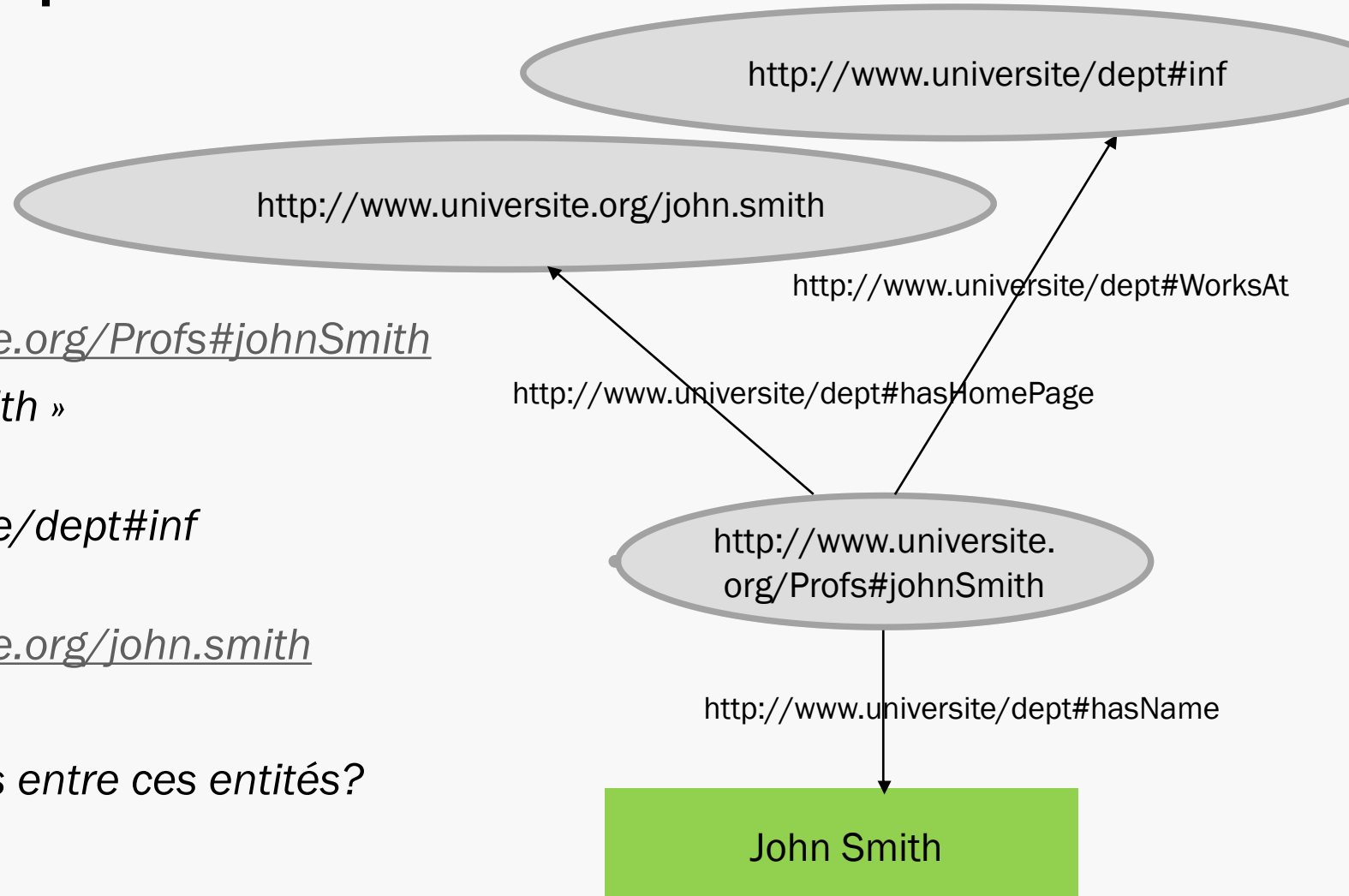
- **Exemple:** Décrire un enseignant John Smith qui travaille au département informatique de l'université et dont la page personnelle se trouve à l'URL suivante:  
<http://www.universite.org/john.smith>
- Combien d'entités?
- 4 entités :
  - *L'enseignant en question*
  - *Son nom*
  - *L'endroit où il travaille*
  - *Sa page personnelle*

# Modèle de graphe

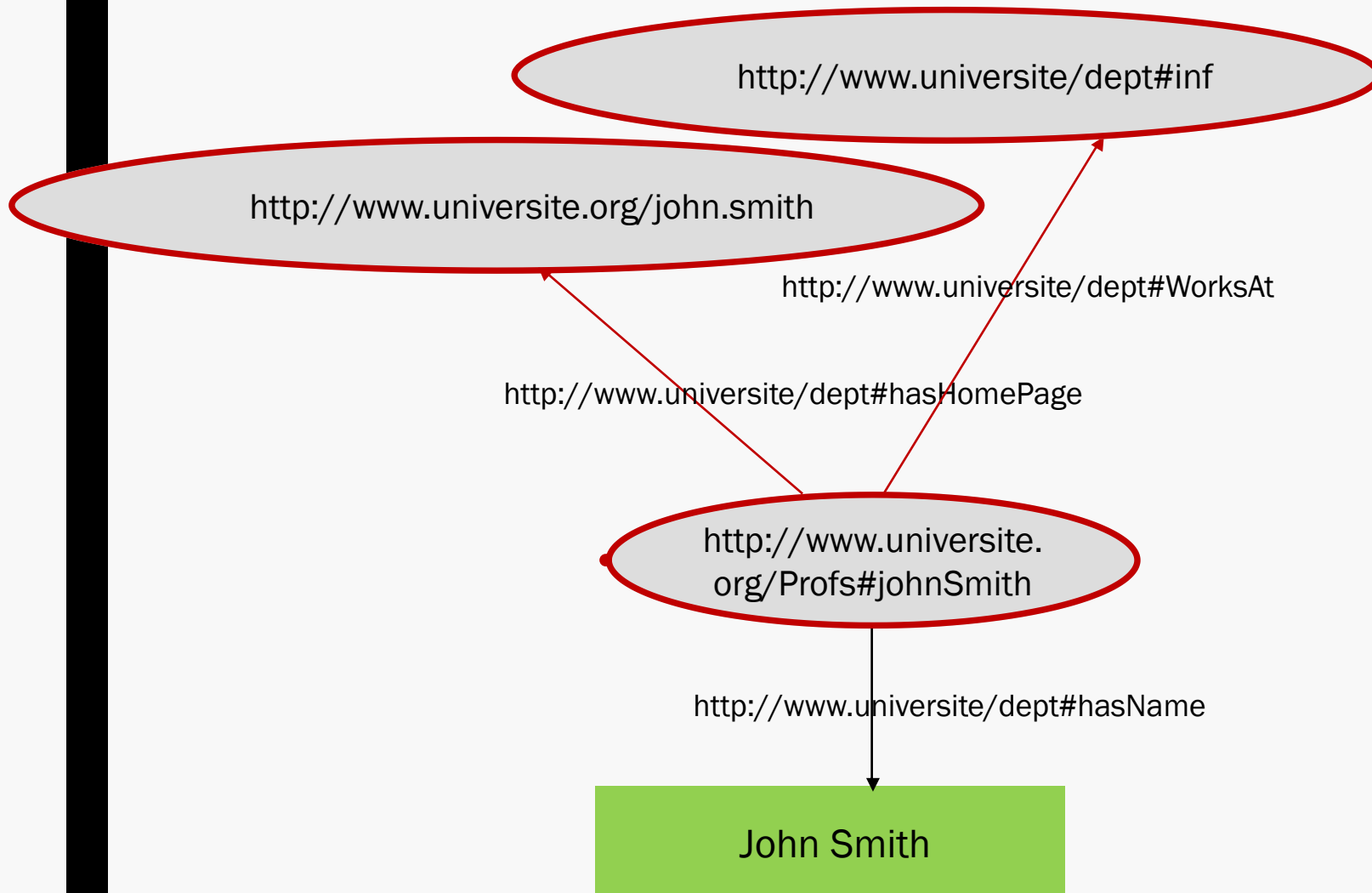
■ 4 entités :

- *L'enseignant décrit:*  
<http://www.universite.org/Profs#johnSmith>
- *Son nom: « John Smith »*
- *Le lieu de travail:*  
<http://www.universite/dept#inf>
- *Sa page personnelle:*  
<http://www.universite.org/john.smith>

*Représenter les relations entre ces entités?*



# Modèle de graphe



Interroger: Motifs de graphe

# Modèle de graphe

1. comment définir des identifiants de ressource à l'échelle du web et sans conflit de nom ?

Utilisation d'URI et d'espaces de noms

2. comment interpréter un littéral : quel est son type, son unité, ... ?

Information de typage (xs:integer, fil:m1, ...)

3. comment savoir quels prédicats utiliser : quels sont les noms de prédicats existants ?

Définis dans des vocabulaires appelés ontologies - RDFS

4. comment publier les triplets sur le web ?

Différents formats de sérialisation (N Triples , xml/RDF, N3, ...)

5. comment interroger en pratique ?

Langage de requêtes SPARQL



# Résumé

- Ce qu'il faut retenir : Il n'y a qu'un seul Web. Le Web de données est une extension du Web. Il permet de relier non pas des documents (pages HTML), mais les données elles mêmes et de les rendre exploitables par des machines.
  
- Le Web de données s'appuient sur des standards informatiques, ceux du Web sémantique:
  - *RDF, qui sera bientôt détaillé ;*
  - *HTTP pour accéder aux données via le Web ;*
  - *URI :*
    - des identifiants uniques pour les entités sur le Web,
    - des pointeurs sur les données ;
  - *les hyperliens.*