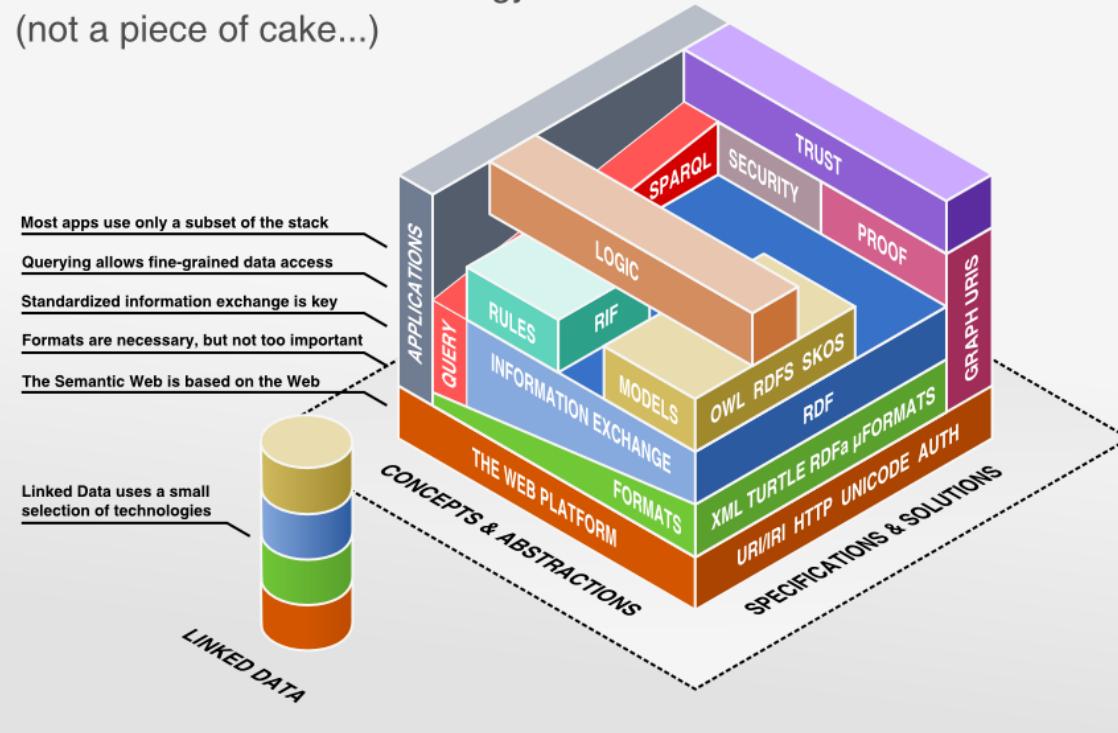


## The Semantic Web Technology Stack (not a piece of cake...)



## Partie 2 : modélisation des données sur le Web

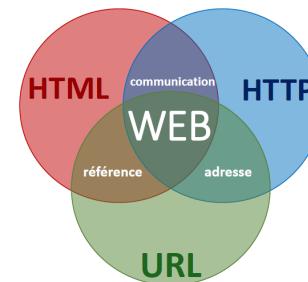
# Résumé de la dernière séance

## Histoire du Web

- Du Web documentaire au Web des données, Web sémantique
- Memex (1945), hypertext/hypermédia (70's), WWW (90's)

## Architecture

- Identification & adressage/localisation (URL/URI/IRI)
- Communication et protocole (HTTP)
- langage de représentation (HTML)

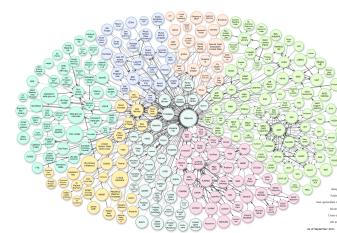


## Séparer le fond de la forme

- CSS
- Echange de données via XML (syntaxe concrète, standard, ouvert) : parseurs, schémas DTD, XPath, Xpointer/Xlink, Xquery, XLS Xproc

## Besoin de données liées

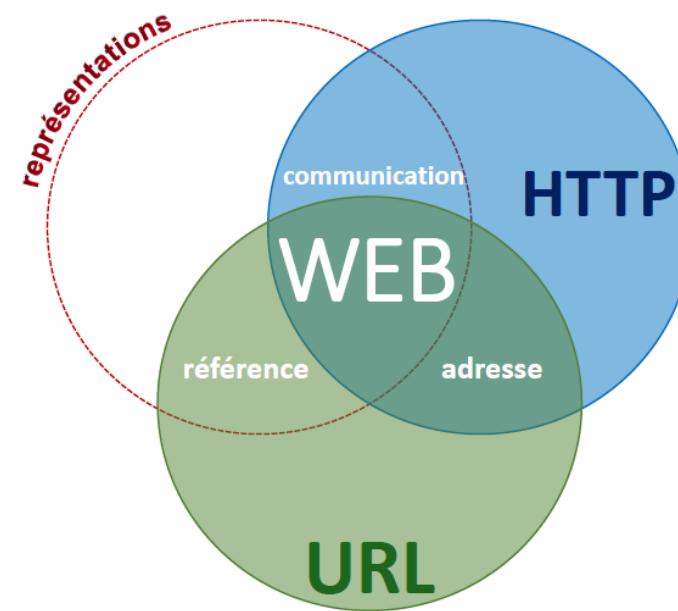
- Graphe de données distribuées



# Aujourd’hui

## Représentation des données du Web - éléments de base

- Recommandation RDF du W3C
- Une recommandation est le fruit du pipeline *atelier, working draft, last call, candidate/proposed recommendation, recommendation*
- W3C suite à la guerre des navigateurs



# Motivations de RDF par des utilisations

- **Métadonnées Web** : afin de fournir des informations sur les ressources Web
- **Modèles d'information ouverts** : annotation de ressources
- **Collaboration entre applications** : combiner les données de plusieurs applications, créer de nouvelles informations.
- **Traitements automatiques de l'information** : d'une information uniquement lisible par les humains vers un état de réseau mondial de processus en coopération.

La valeur de l'information augmente quand elle est échangeable et devient accessible à plus d'applications sur Internet

# Une recommandation fondamentale

## Objectifs de RDF

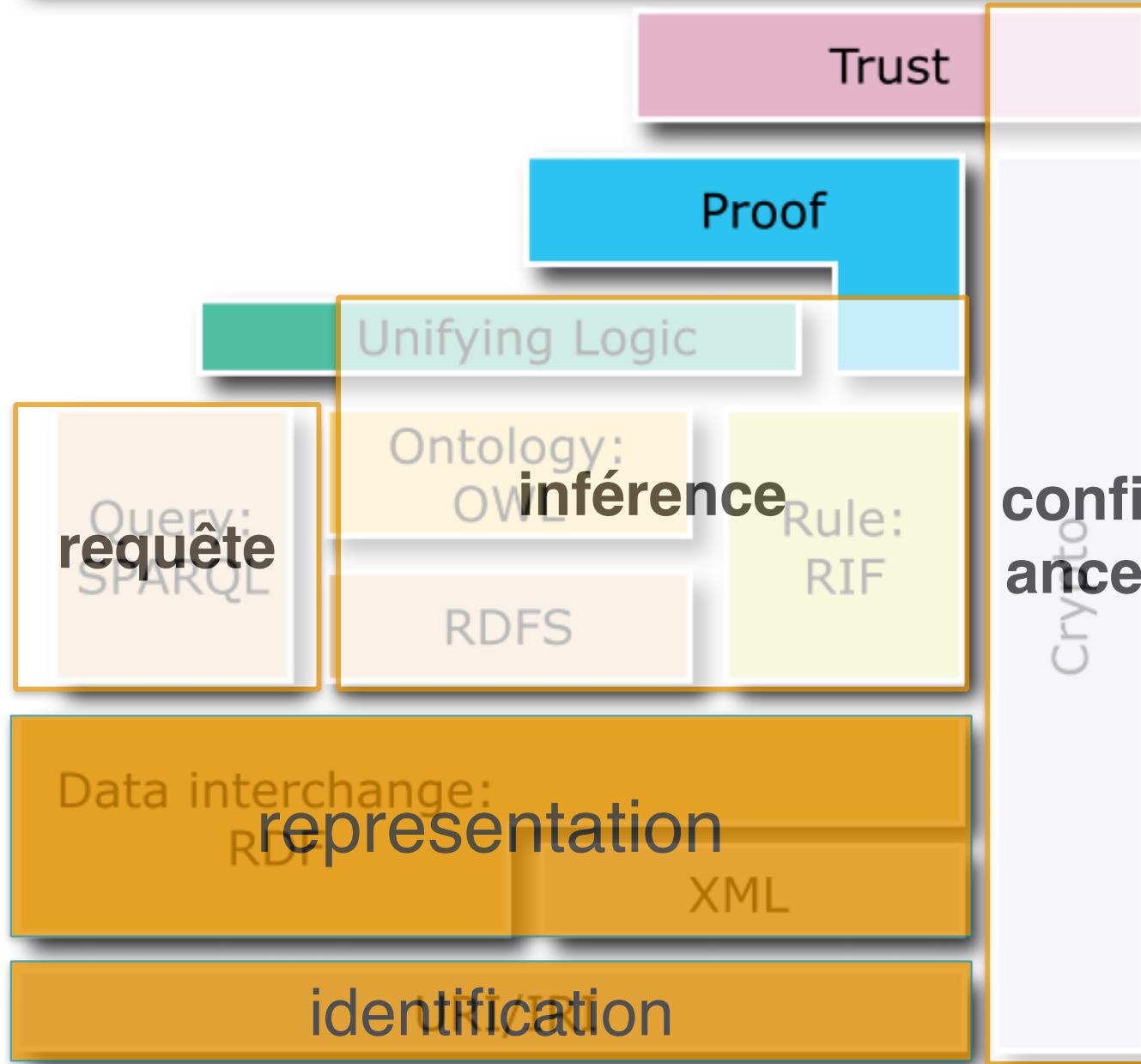
- avoir un modèle de données simple
- avoir une sémantique formelle et une inférence prouvable
- utiliser un vocabulaire extensif
- supporter l'utilisation des types de données de XML
- autoriser quiconque à faire des commentaires sur n'importe quoi

## Basé sur l'architecture classique du Web

- URI pour identifier les ressources décrites et les liens entre elles
- HTTP pour la communication
- XML (et autres) pour la syntaxe

Gérer les données distribuées : représenter l'information de la manière la moins contraignante, la plus flexible possible

# User Interface & Applications



# Plan du cours

*Vers un Web des données*

## 1/ Modélisation

Décrire des données sur Web (triplets et graphes RDF)

Identifier des ressources sur le Web (URI, espace de noms)

Syntaxes concrètes pour un modèle abstrait

## 2/ Publication des données

RDF, RDFa, SGBD2RDF

## 3/ Accès aux données

Déréférencement, négociation, endpoint SPARQL

## 4/ Intégration des données

Fusion, linked data, open data

# 1 / 4 Modélisation

# Resource description framework (RDF)

## Un cadre pour la description de ressources

- **ressource** : « entité », « chose », en fait tout ce à quoi on peut se référer est une ressource (page Web, image, vidéos, personne réelle ou fictive, lieu, dispositif, évènement, organisation, service, ...), c'est à dire, tout ce qui peut être identifié par une URI
- **description** : une ressource peut être décrite par ses attributs et les valeurs que ces derniers prennent mais aussi par les relations (et leur type) qu'elle peut avoir avec d'autres
- **cadre** : ensemble de modèles, langages et syntaxes de ces descriptions

*Une structure de données standard et un modèle pour encoder des données sur n'importe quelle ressource*

<http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>

<http://www.w3.org/TR/rdf-syntax-grammar/>

# La notion de ressource

**Sur le Web, toute chose ‘d’intérêt’ est désigné par une ‘ressource’**

- **identifié par un URI de manière unique**
- **de type ‘information’** : on peut obtenir une représentation via HTTP (une page Web, une image, etc.)
- **de type ‘non-information’** : on ne peut pas obtenir de représentation via HTTP mais on veut pouvoir en parler quand même (le concept de circularité, l’atome de carbone, mon chat)

URI

`http://weather.example.com/oaxaca`

Representation

Metadata:  
Content-type:  
application/xhtml+xml

Data:  
<!DOCTYPE html PUBLIC "...  
" "http://www.w3.org/...">  
<html xmlns="http://www...>  
<head>  
<title>5 Day Forecast for  
Oaxaca</title>  
...  
</html>

Resource  
*Oaxaca Weather Report*

Identifies

Represents

# Quel modèle ?

## Il faut pouvoir

- Distribuer des données sur le Web (distribution)
- Fusionner les données d'information multiples (merging)

## Une représentation classique : les tables

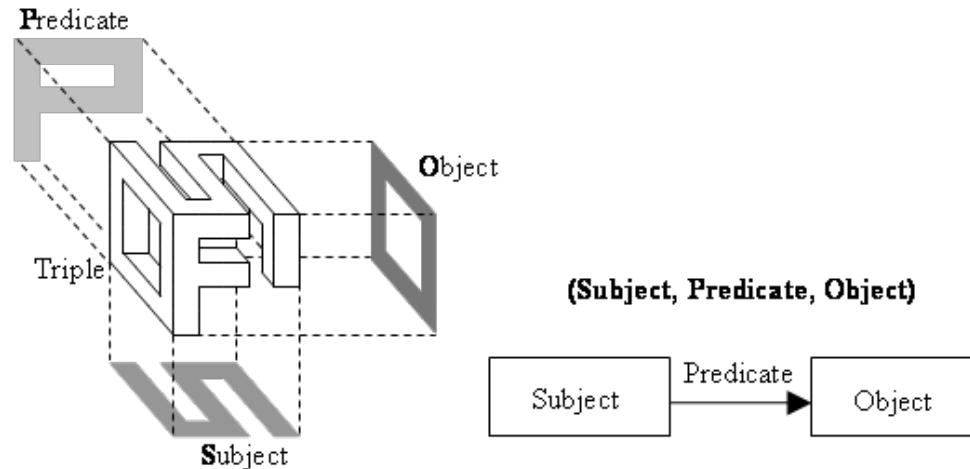
- lignes : sujet
- colonnes : attribut/propriété
- cellules : valeur de la propriété pour le sujet

*Il faut pouvoir identifier les choses et échanger des fragments de connaissances sur eux entre serveurs distants*

# Des atomes de connaissances

## Décrire et connecter des ressources avec des triplets RDF

- < sujet , prédicat , objet > ou ( sujet , prédicat , objet )
- Un énoncé axiomatique, un fait affirmé, une assertion
- On appelle le « prédicat » la « propriété » du triplet



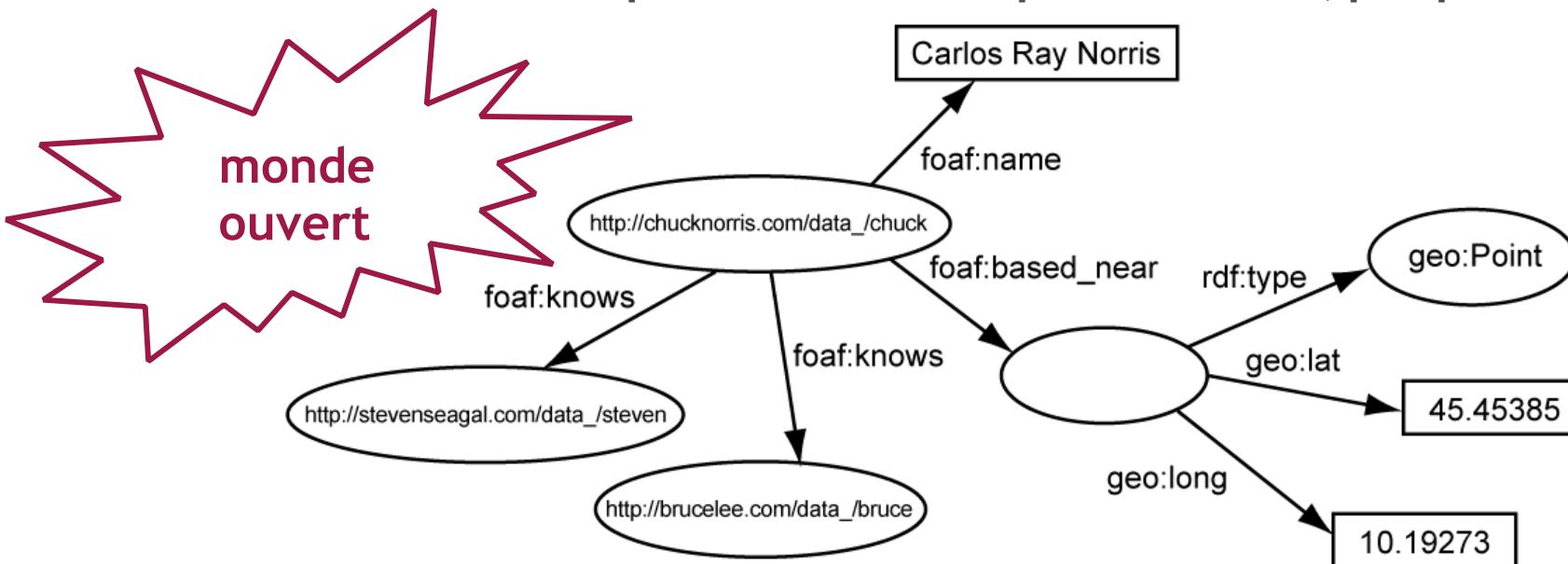
([http://chucknorris.com/data/\\_chuck](http://chucknorris.com/data/_chuck), foaf:name, 'Carlos Ray Norris')

*Un prédicat binaire en logique du premier ordre !  
cf. 4-IF-ALIA-Logique et 4-IF-ALIA-PR1-Prolog*

# Modèle de données générique « graphe »

Un ensemble de triplets RDF se voit comme un graphe RDF

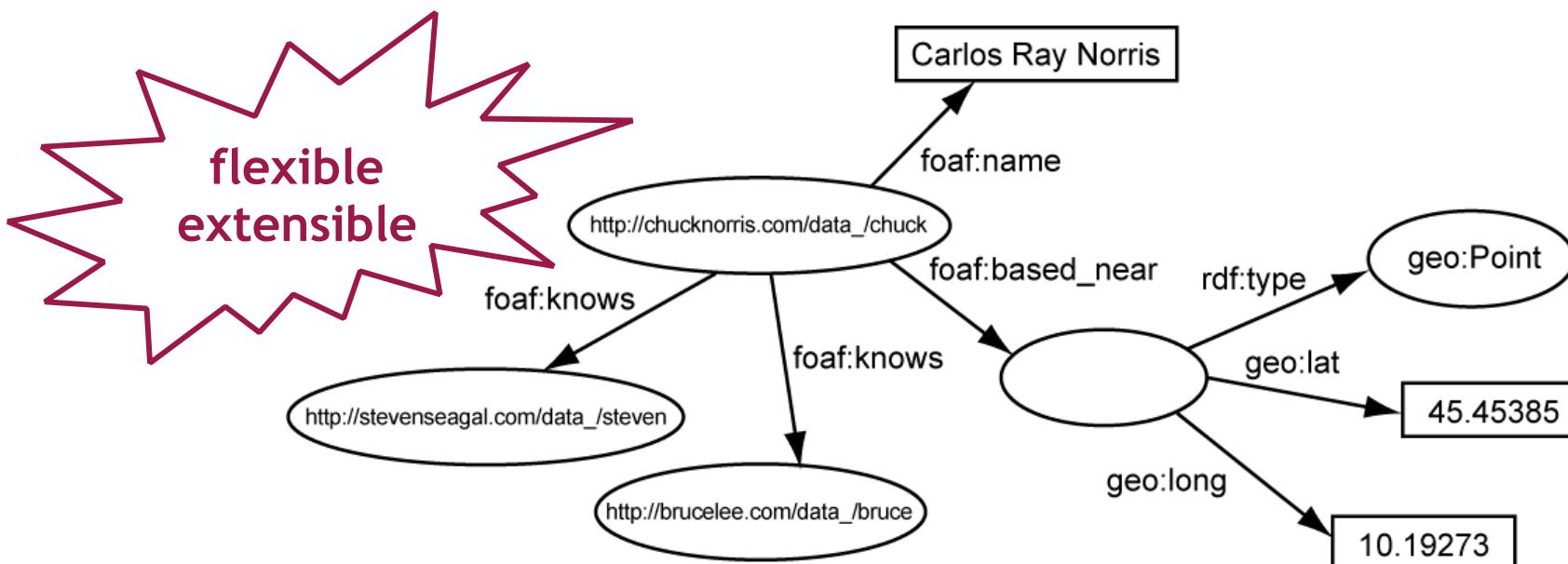
- **nœuds étiquetés** : les objets et sujets des triplets
- **arêtes étiquetés** : les prédictats des triplets, des relations entre les ressources
- **multi-graphe** : différents prédictats peuvent lier un même sujet à un même objet
- **orienté** : du sujet vers l'objet
- Décrire les informations pour être utilisées par la machine, pas par l'humain



# Modèle de données générique « graphe »

## Une syntaxe abstraite

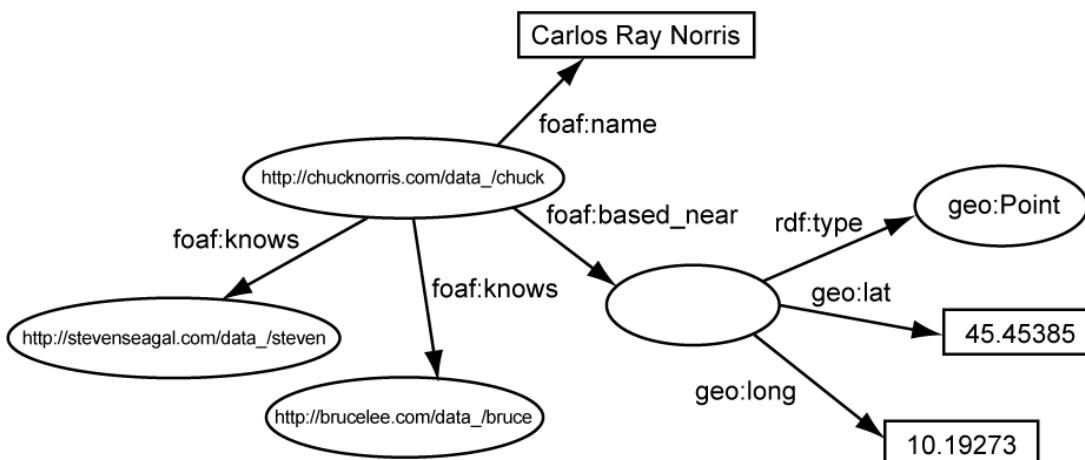
- Syntaxe sur laquelle la sémantique formelle est définie : les implémentations (XML/RDF, JSON/RDF,...) sont libres de représenter les graphes RDF sous toute autre forme équivalente
- L'assertion d'un graphe RDF est équivalente à l'assertion de tous ses triplets, et par conséquence la signification d'un graphe RDF est la conjonction (ET logique) des faits correspondant à tous ses triplets



# Identifier, nommer

## Triplet ( sujet , prédicat , objet )

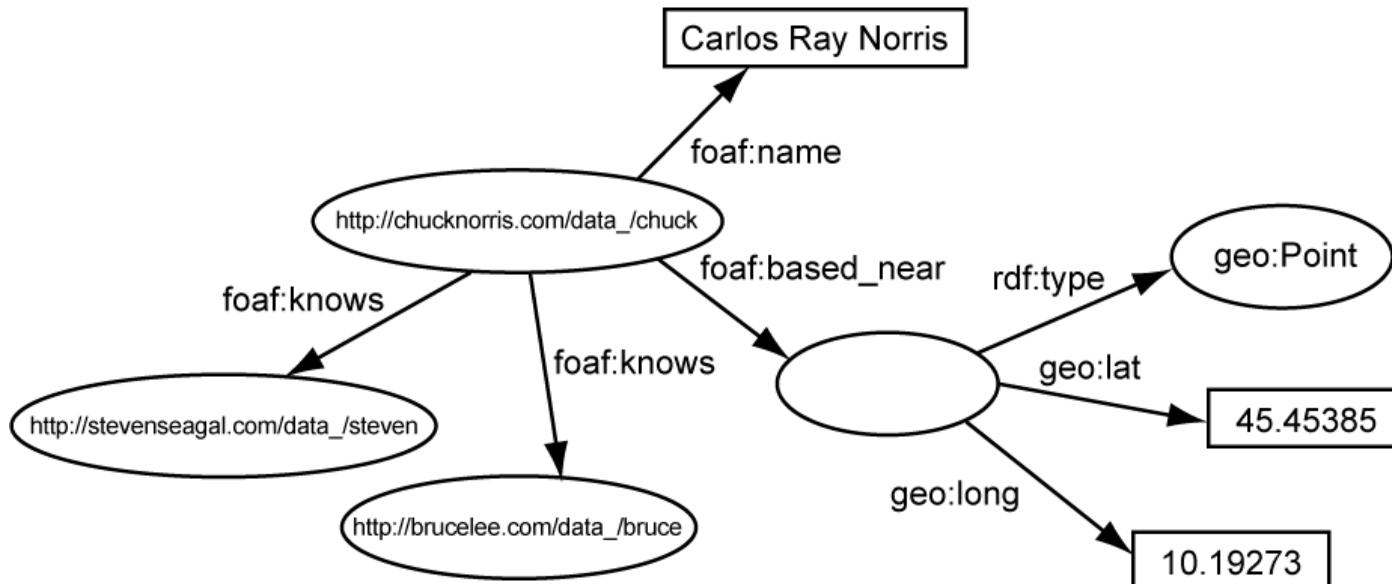
- Un sujet est une ressource identifié par un URI
- Un prédicat dénote une relation, une propriété du sujet (URI)
- L'objet dénote la valeur prise par la propriété : une autre ressource (URI) ou un littéral (valeur, une chaîne, ...)
- On peut connaître l'existence d'une ressource, sans pouvoir la nommer ! (noeuds anonymes)



# Visualisation de graphe RDF

## Représentation usuelle

- Ressource : ellipse étiquetée par son URI
- Littéral : rectangle étiqueté par une chaîne (peut être typée)
- Prédicat : flèche orientée du sujet vers l'objet étiqueté par URI
- Noeud anonyme : ellipse non étiquetée



# Un peu de logique ...

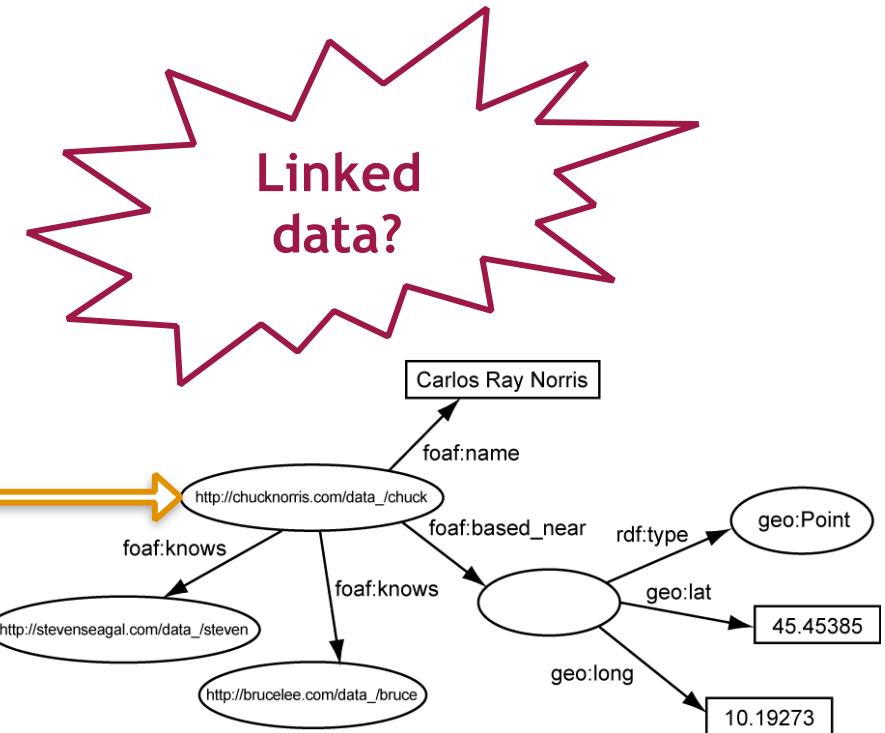
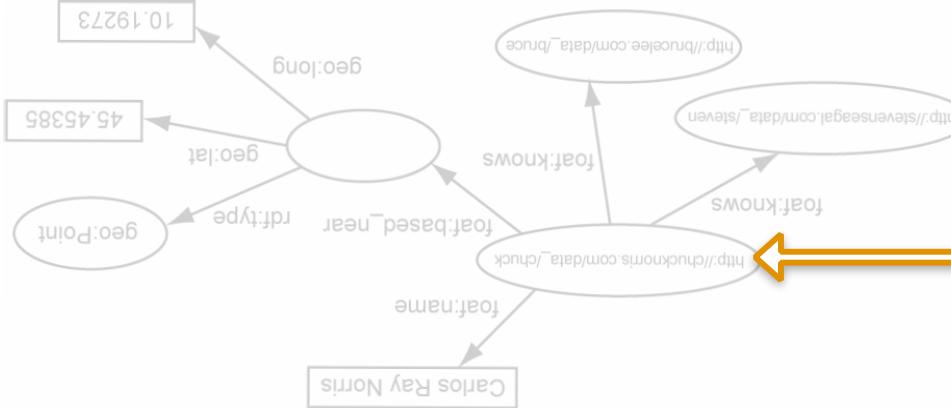
## Sémantique formelle

- RDF est un sous-ensemble de la logique du premier ordre
  - Avec** : prédicats binaires, quantification existentielle ( $\exists$ ), conjonction (^)
  - Sans** : disjonction, négation, quantification universelle ( $\forall$ ), ...
- RDF est monotone : Tout énoncé RDF est considéré comme vrai ce qui est vrai et ce que l'on peut déduire reste vrai si l'on rajoute de nouveaux énoncés.
- Dilemme expressivité/calculabilité  
(languages plus expressifs dans les autres séances)

Pensez-y en 4-if-ALIA (et vice versa) !  
et en PROLOG : assert(...), retract(...) ?

# Fusionner deux graphes RDF ?

Comment répondre à : ‘un noeud dans un graphe est-il le même qu’un noeud dans l’autre graphe ?’

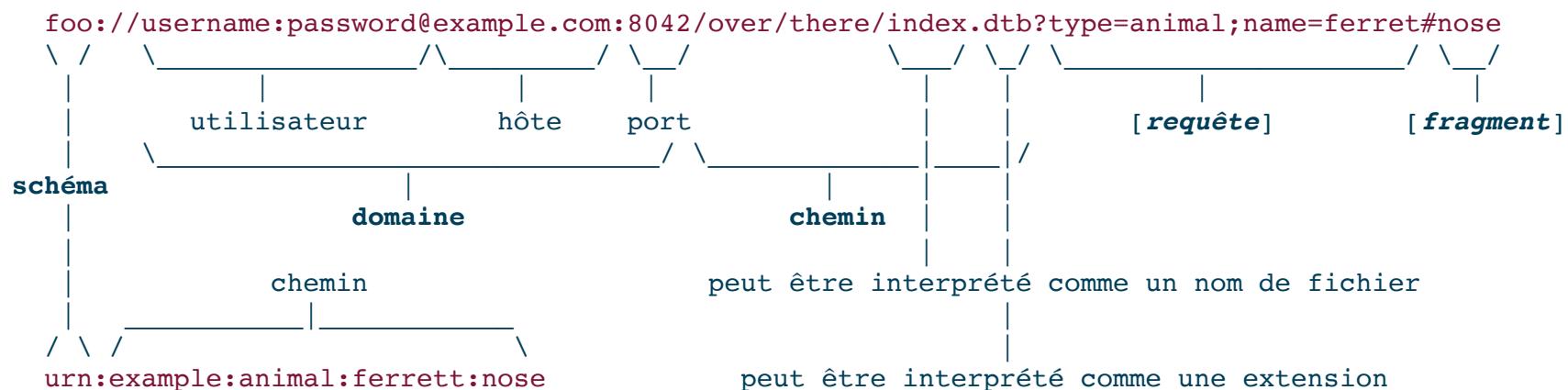


Besoin d’identifiants, et d’un schéma de ses identifiants :  
URI et URI Schema

# Identifier avec des URI

## Uniform Resource Identifier

- RFC 3986 (Berners-Lee et al., 2005) : décrit un schéma simple et extensif pour l'identification d'objets, abstraits ou physiques, à travers le monde
- Séquence compacte de caractères suivant une syntaxe générique  
`<nom du schéma> : <partie hiérarchique> [ ? <requête> ] [ # <fragment> ]`
- Exemples



- Notion similaire existante (ISBN, ISSN, DOI, ...) ne pas confondre avec numéro de série !
- Un identifiant global sur Internet

# Divers schémas d'URI



## Un vaste panel enregistré par l'IANA

Internet Assigned Numbers Authority

- file, ftp, dns, http, https, im, ldap, news, nfs, tel, ...  
(+ de nombreux non-officiels, e.g. chrome, doi, ....).
- Autant de RFC qui décrivent ces schémas
- Un schéma d'URI n'est pas un protocole
- URL : un cas spécial et typique !

telnet://<user>:<password>@<host>[ :<port>/ ]

file://[host]/path    pas de protocole en particulier !

gg:<userid<sub>hasleftthegame</sub>>

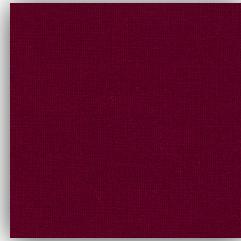
irc://<host>[ :<port> ]/[<channel>[ ?<password> ] ]

lastfm://<radio\_stream>

skype:<username|phonenumber>[ ?[ add|call|chat|sendfile|userinfo ] ]

# URI avec RDF

‘Bordeaux’ : oui, mais dans quel vocabulaire ?



<http://www.insa-lyon.fr/if/ws/couleur#Bordeaux>

<http://www.insa-lyon.fr/if/ws/vin#Bordeaux>

<http://www.insa-lyon.fr/if/ws/geo#Bordeaux>

**URI : Identification globale d'une ressource sur le Web**

- Une URI peut être de-référencée : le nom du serveur, le port, le nom de fichiers... peuvent être utilisés pour localiser
- InterCap convention pour les noms (CamelCase)

# Schémas d'abréviation

## Ecriture d'URI simplifiée

- Qualified name (qname) : namespace + identifier

col pour <http://www.insa-lyon.fr/if/ws/couleur#>

vins pour <http://www.insa-lyon.fr/if/ws/vin#>

geo pour <http://www.insa-lyon.fr/if/ws/geo#>

col:Bordeaux

vins:Bordeaux

geo:Bordeaux

- Quand le contexte est clair...

:Bordeaux

- Un vocabulaire définit ce que les triplets veulent dire,  
et d'identifier le concept spécifique auquel on fait référence

# Espace de nommage (namespace)

**URI particulier pour identifier un vocabulaire, un schéma**

- De nombreux espaces standards
  - xsd:      **Définition de schémas XML**
  - xmnl:     **Définition d'espaces de noms**
  - rdf:       **Pour définir des types et propriétés en RDF**  
<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
  - rdfs:      **RDF Schema language (patience...)**  
<http://www.w3.org/2000/01/rdf-schema#>
  - owl:        **Web Ontology language (patience...)**  
<http://www.w3.org/2002/07/owl#>
- Dé-référencer une URI (pas obligatoire, mais très important pour des applications)

<http://www.w3.org/2000/01/rdf-schema#subClassOf>

# Un exemple de vocabulaire : FOAF



## Décrire les personnes et les relations entre amis

```
<?xml version="1.0" encoding="utf-8" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
           xmlns:foaf="http://xmlns.com/foaf/0.1/"
           xmlns:dc="http://purl.org/dc/terms/">
  <foaf:Person rdf:about="http://njh.me/#http://www.pipm.fr/foaf/foaf.rdf">
    <foaf:name>Julien Plu</foaf:name>
    <foaf:title>M.</foaf:title>
    <foaf:givenname>Julien</foaf:givenname>
    <foaf:family_name>Plu</foaf:family_name>
    <foaf:nick>SnaKe</foaf:nick>
    <foaf:mbox_sha1sum>d14930bd11472c1b9d3b5604aa7c3d231a69a030</foaf:mbox_sha1sum>
    <foaf:dateOfBirth>1987-06-19</foaf:dateOfBirth>
    <foaf:homepage>
      <rdf:Description rdf:about="http://jplu.developpez.com/">
        <dc:title>Page personnelle</dc:title>
      </rdf:Description>
    </foaf:homepage>

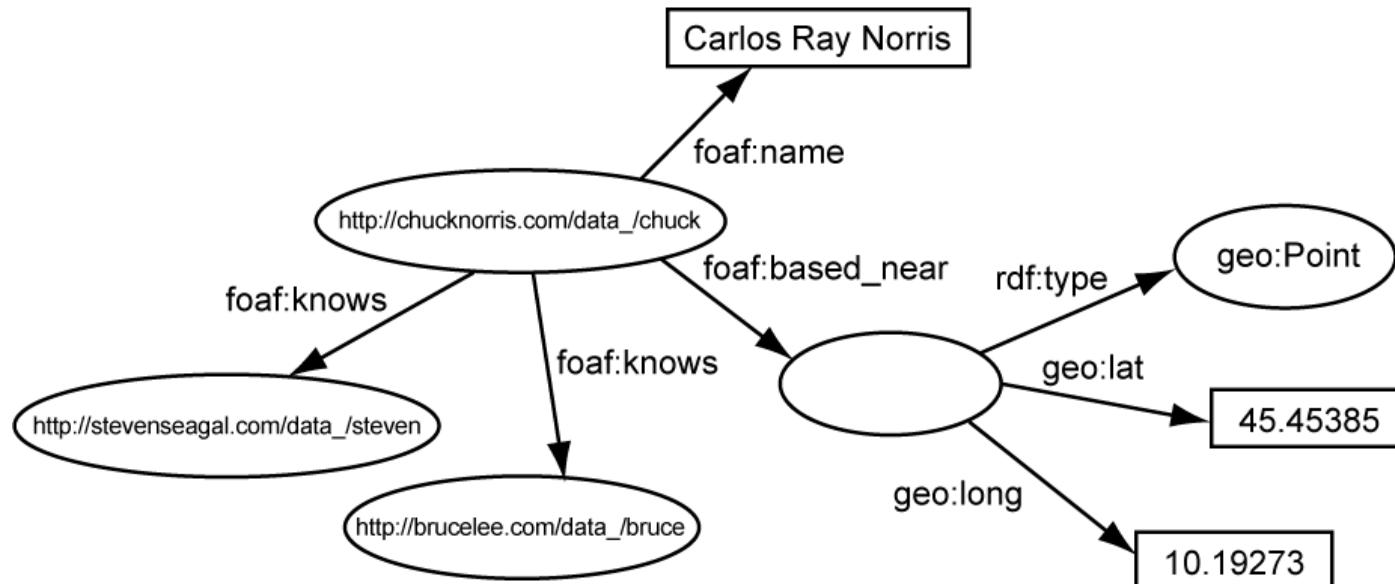
    <foaf:weblog rdf:resource="http://blog.developpez.com/jplu/" />
    <foaf:img rdf:resource="http://www.pipm.fr/foaf/moi.jpg"/>
    <foaf:schoolHomepage rdf:resource="http://www.univ-montp2.fr/" />
    <foaf:gender>male</foaf:gender>
  </foaf:Person>
</rdf:RDF>
```

Jouer avec <http://www.w3.org/RDF/Validator/rdfval>

# Retour vers la visualisation de graphe

## Représentation usuelle

- Ressource : ellipse étiquetée par son URI
- **Littéral** : rectangle étiqueté par une chaîne (peut être typée)
- Prédicat : flèche orientée du sujet vers l'objet étiqueté par URI
- **Noeud anonyme** : ellipse non étiquetée



# Littéral

## Une « chaîne » ...

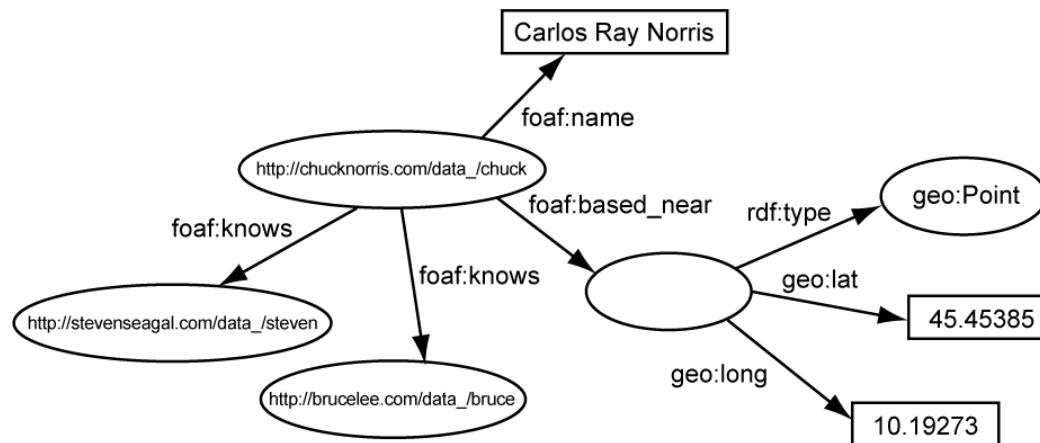
- Chaîne de caractère arbitraire typée ou non typée représentant des valeurs (texte, entiers, flottant, dates, ...)
- Tout ce qui est représenté par un littéral pourrait aussi être représenté par une URI, mais il est généralement plus adéquat ou intuitif d'employer des littéraux
- Un littéral peut être l'objet d'une assertion RDF, mais il ne peut en être ni le sujet ni le prédicat.

```
<http://dbpedia.org/resource/Institut_national_des_sciences_appliqu%C3%A9es>
    <http://dbpedia.org/ontology/wikiPageID>
    "697940"^^<http://www.w3.org/2001/XMLSchema#integer> .
```

# Noeuds vides et quantification existentielle

## Ressources anonymes

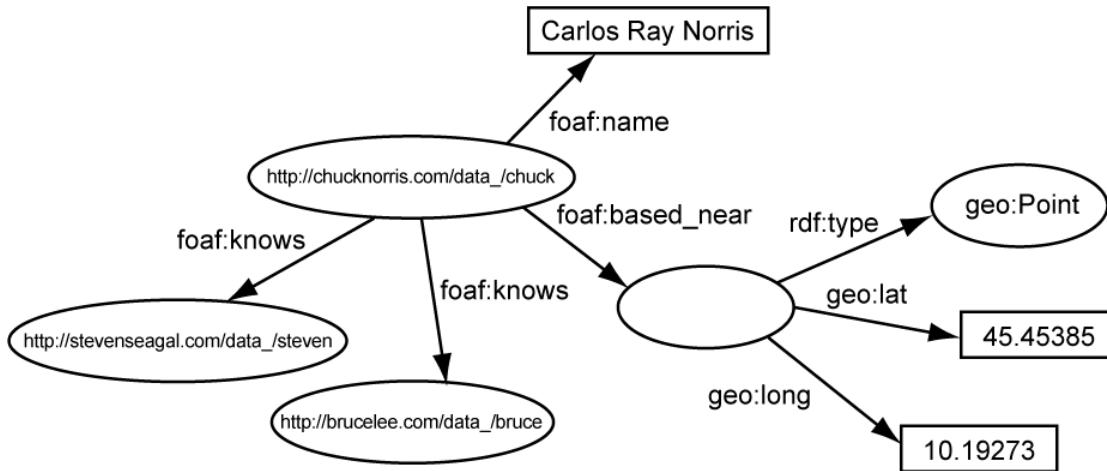
- L'existence d'une ressource peut être connue, sans que l'on puisse la nommer, ou l'identifier par un URI
  - Créer un URI particulier
  - Utiliser un URI existant faisant sens
  - Utiliser le concept de noeud vide (à éviter sauf conteneurs et collections)



# Noeuds vides et quantification existentielle

## Noeuds vides, blank-nodes

- Un nœud vide est un nœud qui ne représente ni un URI, ni un littéral.
- Deux ressources anonymes avec un même identifiant bnode « `_n` » dans deux document différents doivent être considérées, jusqu'à preuve du contraire, comme différentes
- Aucune structure interne : impossible de déterminer si deux noeuds vides donnés sont les mêmes (au final, un problème de syntaxe traité par les syntaxes concrètes)



# Vers l'ontologie ...

## Multi instantiation et typage orienté ontologies

- Primitive `rdf:type` peut être impliquée dans plusieurs triplets avec le même sujet
- RDF est puissant : mécanisme de multi-instantiation (e.g. différent de Java)
- L'idée est de fournir les éléments de base pour définir des ontologies et des vocabulaires destinés à structurer les ressources RDF
- C'est aussi un point majeur afin de pouvoir définir un langage d'interrogation (requêtes SPARQL par exemple)
- Pour le moment, nous nous limiterons à utiliser des vocabulaires simples et existant (`rdf`, `foaf`, `dublin core`)

*Les requêtes SPARQL seront présentées en partie 4  
Les ontologies, RDF-S et OWL seront présentés en partie 3*

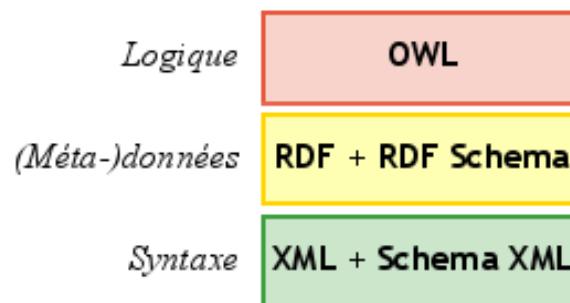
# RDF : plusieurs syntaxes pour un même modèle

**Le graphe RDF est un modèle abstrait**

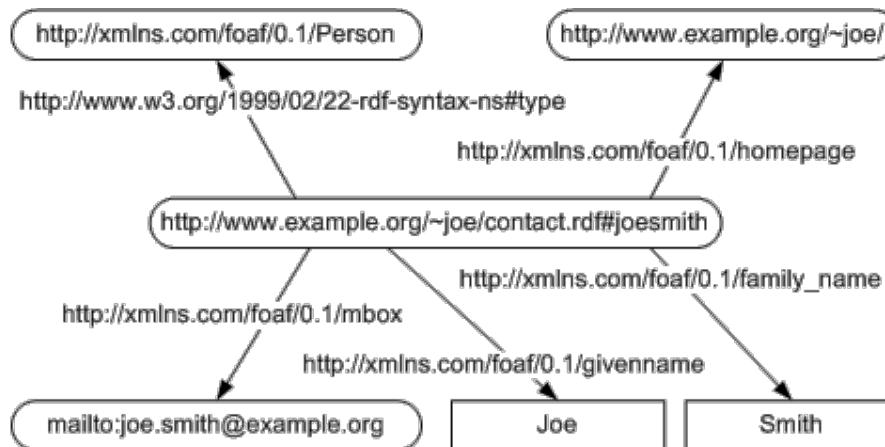
- Visuel et compréhensible facilement par l'homme
- Intuitivement extensif et flexible

**Une représentation machine nécessite une syntaxe concrète (sérialisation)**

- Différentes syntaxes : RDF/XML, N-triples, Turtle, N3
- RDF 1.0 préconise RDF/XML
- Le code XML (arbre) d'un RDF (graphe) n'est pas destinée à être lu par l'utilisateur
- XML permet un format d'échange standardisé et tous les outils qui vont avec



# Graphe RDF



# RDF/XML

```

<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
>
  <foaf:Person rdf:about="http://www.example.org/~joe/contact.rdf#joesmith">
    <foaf:givenname>Joe</foaf:givenname>
    <foaf:homepage rdf:resource="http://www.example.org/~joe/" />
    <foaf:mbox rdf:resource="mailto:joe.smith@example.org" />
    <foaf:family_name>Smith</foaf:family_name>
  </foaf:Person>
</rdf:RDF>
  
```

```

@prefix : <http://www.example.org/~joe/contact.rdf#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

  
```

```

:joesmith a foaf:Person ;
  foaf:family_name "Smith" ;
  foaf:givenname "Joe" ;
  foaf:homepage <http://www.example.org/~joe/> ;
  foaf:mbox <mailto:joe.smith@example.org> .
```

N-3

```

<http://www.example.org/~joe/contact.rdf#joesmith> <http://xmlns.com/foaf/0.1/homepage> <http://www.example.org/~joe/> .
<http://www.example.org/~joe/contact.rdf#joesmith> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://xmlns.com/foaf/0.1/Person>.
<http://www.example.org/~joe/contact.rdf#joesmith> <http://xmlns.com/foaf/0.1/family_name> "Smith" .
<http://www.example.org/~joe/contact.rdf#joesmith> <http://xmlns.com/foaf/0.1/givenname> "Joe" .
<http://www.example.org/~joe/contact.rdf#joesmith> <http://xmlns.com/foaf/0.1/mbox> <mailto:joe.smith@example.org> .
```

# N-Triples

31  
<http://www.easyrdf.org/converter>  
<http://www.w3.org/wiki/ConverterToRdf>

# RDF : conteneurs ouverts

## Conteneurs de ressources/littéraux

- **Sac (rdf:Bag)** : non ordonné, doublons (e.g. contenu d'un colis)
- **Séquence (rdf:Seq)** : ordonné, doublons (e.g. vainqueurs de la FIFA WC depuis 1930)
- **Alternative (rdf:Alt)** : ensemble de valeurs dont une seul peut être tirée selon son contexte d'utilisation (e.g. les titres d'un seul et même livre en différentes langues)

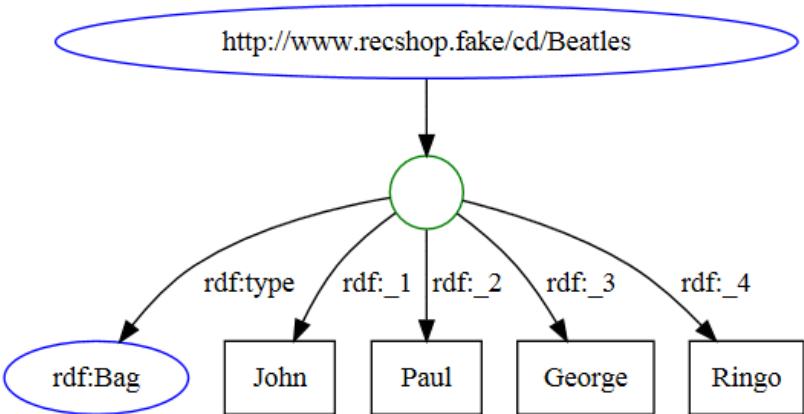
## Utilisation des conteneurs

- Une ressource (souvent anonyme, le conteneur) est créée, et typée comme sac/séquence/alternative
- Les éléments du conteneurs lui sont attachés en utilisant rdf:li
- Chaque occurrence de 'rdf:li' sera transformée en 'rdf\_1', 'rdf\_2', etc. pour un même conteneur pour former le graphe correspondant

## Ces conteneurs sont ouverts

- On ne peut pas affirmer l'ensemble complet, et qu'il n'existe pas d'autre éléments ailleurs

# RDF : conteneurs ouverts (1)



```
@prefix ns0: <http://www.recshop.fake/cd#> .  
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .  
  
<http://www.recshop.fake/cd/Beatles> ns0:artist [  
    a rdf:Bag ;  
    rdf:_1 "John" ;  
    rdf:_2 "Paul" ;  
    rdf:_3 "George" ;  
    rdf:_4 "Ringo"  
] .
```

```
<?xml version="1.0" encoding="utf-8" ?>  
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"  
          xmlns:ns0="http://www.recshop.fake/cd#">  
  
<rdf:Description rdf:about="http://www.recshop.fake/cd/Beatles">  
    <ns0:artist>  
      <rdf:Bag>  
        <rdf:li>John</rdf:li>  
        <rdf:li>Paul</rdf:li>  
        <rdf:li>George</rdf:li>  
        <rdf:li>Ringo</rdf:li>  
      </rdf:Bag>  
    </ns0:artist>  
  
</rdf:Description>  
  
</rdf:RDF>
```

```
<http://www.recshop.fake/cd/Beatles> <http://www.recshop.fake/cd#artist> _:genid1 .  
_:genid1 <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/1999/02/22-rdf-syntax-ns#Bag> .  
_:genid1 <http://www.w3.org/1999/02/22-rdf-syntax-ns#\_1> "John" .  
_:genid1 <http://www.w3.org/1999/02/22-rdf-syntax-ns#\_2> "Paul" .  
_:genid1 <http://www.w3.org/1999/02/22-rdf-syntax-ns#\_3> "George" .  
_:genid1 <http://www.w3.org/1999/02/22-rdf-syntax-ns#\_4> "Ringo" .
```

# RDF : Collections fermées (1)

## Groupes fermés

- Ces groupes ne contiennent que les membres spécifiés lors de leur déclaration

## Principe

- Liste chaînée de membres : rfd:list lie une ressource à sa collection (ressource anonyme) qui est liée à une ressource tête (rdf:first) et aux reste du groupe (rdf:rest)
- Comme en LISP, rdf:nil est utilisé en fin de liste

```
<rdf:Description rdf:about="http://e.org/family/349">
  <s:familyMembers rdf:nodeID="sch1"/>
</rdf:Description>

<rdf:Description rdf:nodeID="sch1">
  <rdf:first rdf:resource="http://e.org/person/Paolo"/>
  <rdf:rest rdf:nodeID="sch2"/>
</rdf:Description>

<rdf:Description rdf:nodeID="sch2">
  <rdf:first rdf:resource="http://e.org/person/Emanuele"/>
  <rdf:rest rdf:nodeID="sch3"/>
</rdf:Description>

<rdf:Description rdf:nodeID="sch3">
  <rdf:first rdf:resource="http://e.org/person/Maria"/>
  <rdf:rest rdf:nodeID="sch4"/>
</rdf:Description>

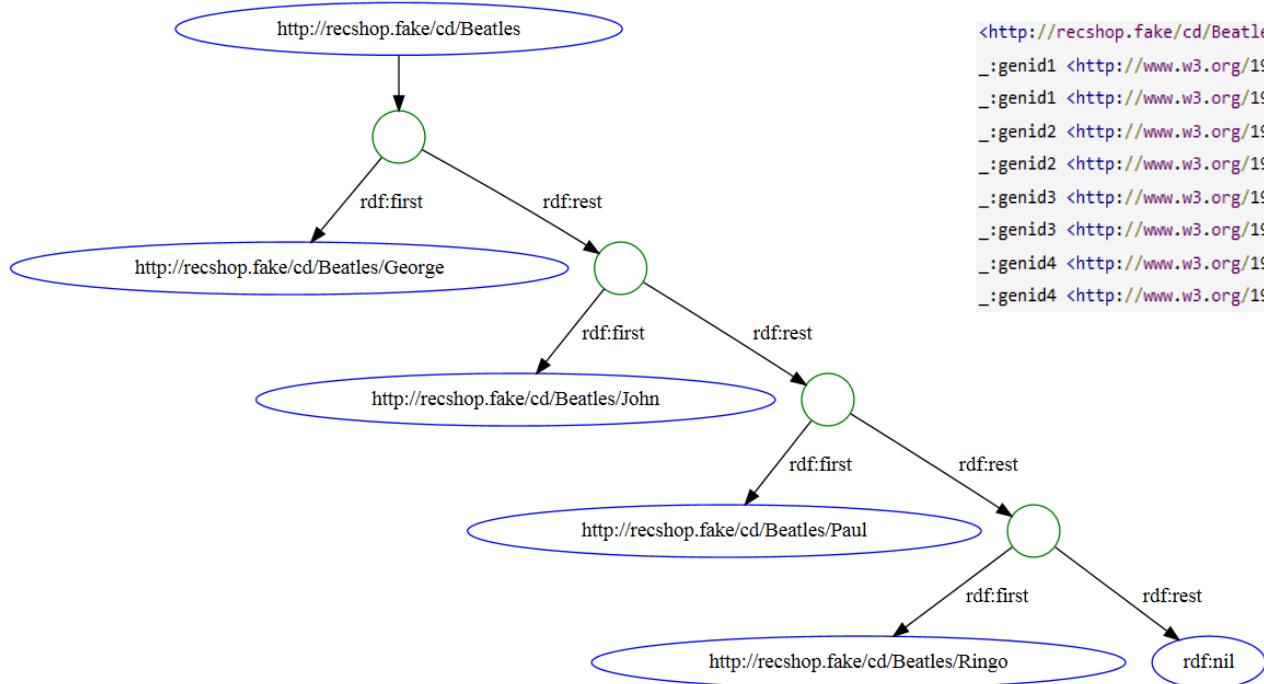
<rdf:Description rdf:nodeID="sch4">
  <rdf:first rdf:resource="http://e.org/person/Franco"/>
  <rdf:rest rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#nil"/>
</rdf:Description>
```

```
<rdf:Description rdf:about="http://e.org/family/349">
  <s:familyMembers rdf:parseType="Collection">
    <rdf:Description rdf:about="http://e.org/person/Paolo"/>
    <rdf:Description rdf:about="http://e.org/person/Emanuele"/>
    <rdf:Description rdf:about="http://e.org/person/Maria"/>
    <rdf:Description rdf:about="http://e.org/person/Franco"/>
  </s:familyMembers>
</rdf:Description>
```

**rdf:parseType="Collection"**

**Vocabulaire  
de collection**

# RDF : Collections fermées (2)



```

<http://recshop.fake/cd/Beatles> <http://recshop.fake/cd#artist> _:genid1 .
_:genid1 <http://www.w3.org/1999/02/22-rdf-syntax-ns#first> <http://recshop.fake/cd/Beatles/George> .
_:genid1 <http://www.w3.org/1999/02/22-rdf-syntax-ns#rest> _:genid2 .
_:genid2 <http://www.w3.org/1999/02/22-rdf-syntax-ns#first> <http://recshop.fake/cd/Beatles/John> .
_:genid2 <http://www.w3.org/1999/02/22-rdf-syntax-ns#rest> _:genid3 .
_:genid3 <http://www.w3.org/1999/02/22-rdf-syntax-ns#first> <http://recshop.fake/cd/Beatles/Paul> .
_:genid3 <http://www.w3.org/1999/02/22-rdf-syntax-ns#rest> _:genid4 .
_:genid4 <http://www.w3.org/1999/02/22-rdf-syntax-ns#first> <http://recshop.fake/cd/Beatles/Ringo> .
_:genid4 <http://www.w3.org/1999/02/22-rdf-syntax-ns#rest> <http://www.w3.org/1999/02/22-rdf-syntax-ns#nil> .
  
```

```

@prefix ns0: <http://recshop.fake/cd#> .

<http://recshop.fake/cd/Beatles> ns0:artist (
  <http://recshop.fake/cd/Beatles/George>
  <http://recshop.fake/cd/Beatles/John>
  <http://recshop.fake/cd/Beatles/Paul>
  <http://recshop.fake/cd/Beatles/Ringo>
) .
  
```

```

<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:cd="http://recshop.fake/cd#">
<rdf:Description
  rdf:about="http://recshop.fake/cd/Beatles">
  <cd:artist rdf:type="Collection">
    <rdf:Description rdf:about="http://recshop.fake/cd/Beatles/George"/>
    <rdf:Description rdf:about="http://recshop.fake/cd/Beatles/John"/>
    <rdf:Description rdf:about="http://recshop.fake/cd/Beatles/Paul"/>
    <rdf:Description rdf:about="http://recshop.fake/cd/Beatles/Ringo"/>
  </cd:artist>
</rdf:Description>
</rdf:RDF>
  
```

```

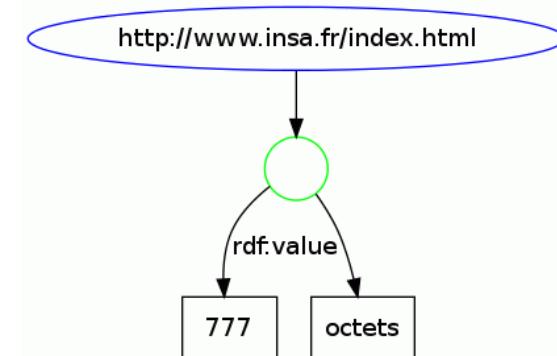
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:ns0="http://recshop.fake/cd#">
<rdf:Description rdf:about="http://recshop.fake/cd/Beatles">
  <ns0:artist>
    <rdf:Description>
      <rdf:first rdf:resource="http://recshop.fake/cd/Beatles/George"/>
      <rdf:rest>
        <rdf:Description>
          <rdf:first rdf:resource="http://recshop.fake/cd/Beatles/John"/>
          <rdf:rest>
            <rdf:Description>
              <rdf:first rdf:resource="http://recshop.fake/cd/Beatles/Paul"/>
              <rdf:rest>
                <rdf:Description>
                  <rdf:first rdf:resource="http://recshop.fake/cd/Beatles/Ringo"/>
                  <rdf:rest rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#nil"/>
                </rdf:Description>
              </rdf:rest>
            </rdf:Description>
          </rdf:rest>
        </rdf:Description>
      </rdf:rest>
    </ns0:artist>
  </rdf:Description>
</rdf:RDF>
  
```

# Les valeurs complexes

## Des bonnes pratiques

- Toute relation n-aire doit être décomposée
- Semantic Web best Practices and Deployment (W3C)  
<http://www.w3.org/2001/sw/BestPractices/>
- Exemple : un document d'une taille de 777 octets

```
@prefix ns0: <http://www.insa.fr/voc#> .  
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .  
@prefix ns1: <www.insa.fr/voc#> .  
  
<http://www.insa.fr/index.html> ns0:taille [  
    rdf:value "777" ;  
    ns1:unit "octets"  
] .
```

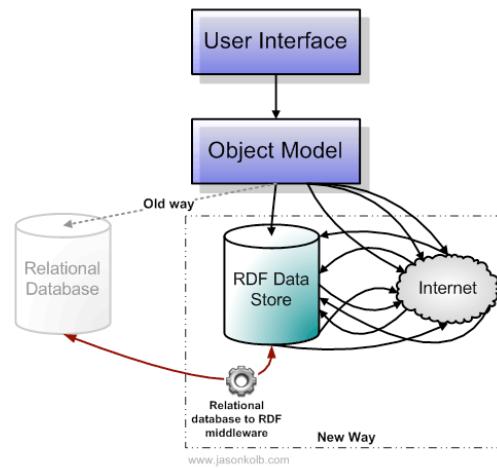


## 2/4 Publication

# Comment publier des données sur le Web?

## Plusieurs moyens

- En publiant un fichier RDF
- En les embarquant dans des pages (X)HTML statiques ou dynamiques : (i) extraire la structure de la page et la convertir en RDF (**GRDDL**), (ii) inclure directement des triples RDF dans le code avec **RDFa**
- A partir d'une BD (D2R Server, Virtuoso)



# RDFa (1)

## « RDF dans des attributs HTML »

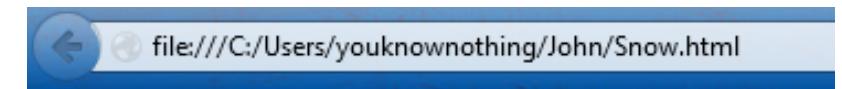
- Une syntaxe qui permet d'indiquer et permettre l'extraction de données présentes dans une page Web, par exemple le titre, l'artiste d'un cd, les horaires d'ouvertures d'un établissement, les horaires/lieux/prix d'une manifestation
- utilise des attributs (X)HTML et d'autres sont ajoutés

<b>rel inv</b>	<i>Liste de propriétés séparées par des espaces Utilisé pour exprimer des relations entre des ressources</i>
<b>content</b>	<i>Valeur littérale d'une propriété</i>
<b>href src</b>	<i>URI d'une ressource utilisée comme objet d'une propriété</i>
<b>about</b>	<i>Le sujet d'une propriété</i>
<b>property</b>	<i>Une liste de propriété ayant une valeur littérale</i>
<b>resource</b>	<i>URI d'une ressource utilisée comme objet d'une propriété (&lt;&gt;&gt; href)</i>
<b>datatype</b>	<i>Permet de spécifier le type de données littérales</i>
<b>typeof</b>	<i>Une liste de type pour une ressource séparés par des espaces</i>

# RDFa (2)

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN"
  "http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd">
<html xmlns="http://www.w3.org/1999/xhtml"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  version="XHTML+RDFa 1.0" xml:lang="en">
<head>
  <title>John's Home Page</title>
  <base href="http://example.org/john-d/" />
  <meta property="dc:creator" content="Jonathan Doe" />
  <link rel="foaf:primaryTopic" href="http://example.org/john-d/#me" />
</head>
<body about="http://example.org/john-d/#me">
  <h1>John's Home Page</h1>
  <p>My name is <span property="foaf:nick">John D</span> and I like
    <a href="http://www.neubauten.org/" rel="foaf:interest"
      xml:lang="de">Einstürzende Neubauten</a>.
  </p>
  <p>
    My <span rel="foaf:interest" resource="urn:ISBN:0752820907">favorite
    book is the inspiring <span about="urn:ISBN:0752820907"><cite
      property="dc:title">Weaving the Web</cite> by
      <span property="dc:creator">Tim Berners-Lee</span></span></span>.
  </p>
</body>
</html>
```

Ici, on utilise les vocabulaires Dublin Core et FOAF



## John's Home Page

My name is John D and I like [Einstürzende Neubauten](#).

My favorite book is the inspiring *Weaving the Web* by Tim Berners-Lee.

A voir : <https://www.youtube.com/watch?v=ldl0m-5zLz4>

A tester : <http://www.w3.org/2012/pyRdfa/> avec  
<http://www.bbc.com/news/world-asia-china-29420802>

A voir : <http://rdfa.info>

# 3/4 Accès aux données

# Déréférencement des ressources

**Etendre le principe du Web des documents au Web des données**

- **Déréférencement d'un URI** : acte de retrouver la représentation de la ressource qu'il identifie
- **Ressource information** (page HTML, image, video, ...) : la ressource est déréférencée directement et on récupère une représentation du document (e.g. un code html à afficher)
- **Ressource non-information** : on récupère une description RDF de la ressource (RDF/XML) à la suite d'une étape de négociation de contenu

# Déréférencement des ressources

## Exemple de déréférencement d'une ressource information

- Page Wikipedia sur John Lennon (HTML)

`curl -I fr.wikipedia.org/wiki/John_Lennon`

Source de : [http://fr.wikipedia.org/wiki/John\\_Lennon](http://fr.wikipedia.org/wiki/John_Lennon)

The screenshot shows the Wikipedia page for John Lennon. The left side displays the raw HTML code of the page, while the right side shows the rendered version. The page includes a sidebar with navigation links like 'Accueil', 'Portails thématiques', and 'Outils'. The main content area features a large image of John Lennon playing a guitar, a bio section, and a 'John Lennon' section with a photo and a quote.

Créer un compte Se connecter

Article Discussion Lire Modifier Modifier le code Historique Rechercher

## John Lennon

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

**John Winston Ono Lennon** (né **John Winston Lennon** le **9 octobre 1940** à Liverpool, et mort assassiné le **8 décembre 1980** à New York) est un **musicien, auteur-compositeur, guitariste, chanteur et écrivain britannique**. Il est le fondateur des **Beatles**, groupe musical anglais au succès planétaire depuis sa formation au début des années 1960. Au sein des Beatles, il forme avec **Paul McCartney l'un des tandems d'auteurs-compositeurs les plus influents et prolifiques de l'histoire du rock**, donnant naissance à plus de deux cents chansons.

Adolescent, influencé par ses idoles américaines du **rock 'n' roll**, il est emporté par la vague de musique **skiffle** qui sévit à Liverpool et fonde en 1956 le groupe des **Quarrymen**, qui évoluent pour devenir, avec **Paul McCartney, George Harrison et Ringo Starr, les Beatles**. De **Please Please Me** en 1963, à **Let It Be** en 1970, les Beatles deviennent un des plus grands phénomènes de l'histoire de l'industrie discographique, introduisant de nombreuses innovations musicales et mélangeant les genres et les influences. Lennon occupe une place centrale dans cette réussite populaire, critique et commerciale, composant une bonne partie des succès du groupe. Les dissensions entre les musiciens, en particulier entre Lennon et McCartney, mettent fin à l'aventure en 1970.

Lorsque les Beatles se séparent, John Lennon se consacre à sa carrière solo, épaulé

**John Lennon**

John Lennon en juin 1969 à Montréal, lors de l'enregistrement de *Give Peace a Chance*.

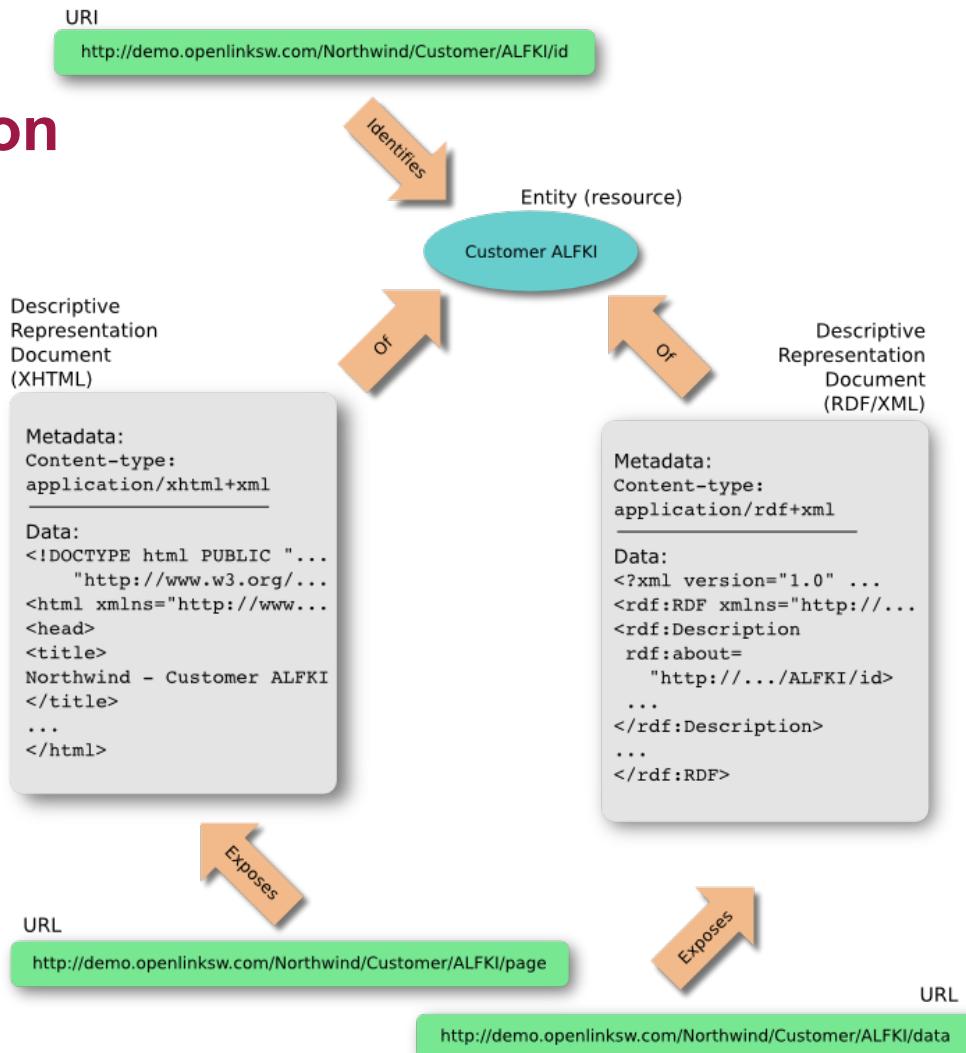
**Informations générales**

<b>Nom de naissance</b>	John Winston Lennon
-------------------------	---------------------

# Déréférencement des ressources

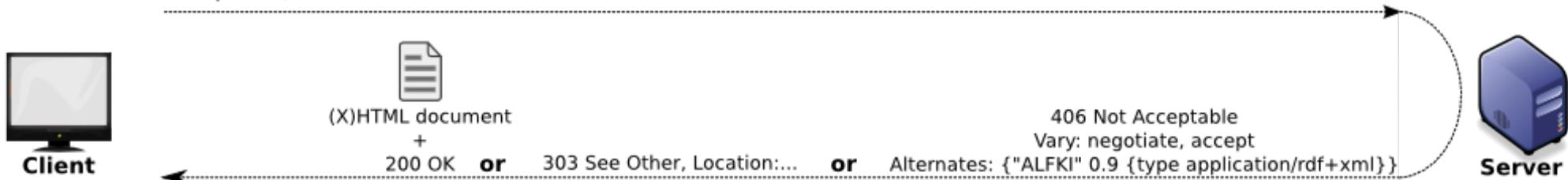
## Exemple de déréférencement d'une ressource non-information

- **URI <> URL!**
- **Il faut négocier le contenu**

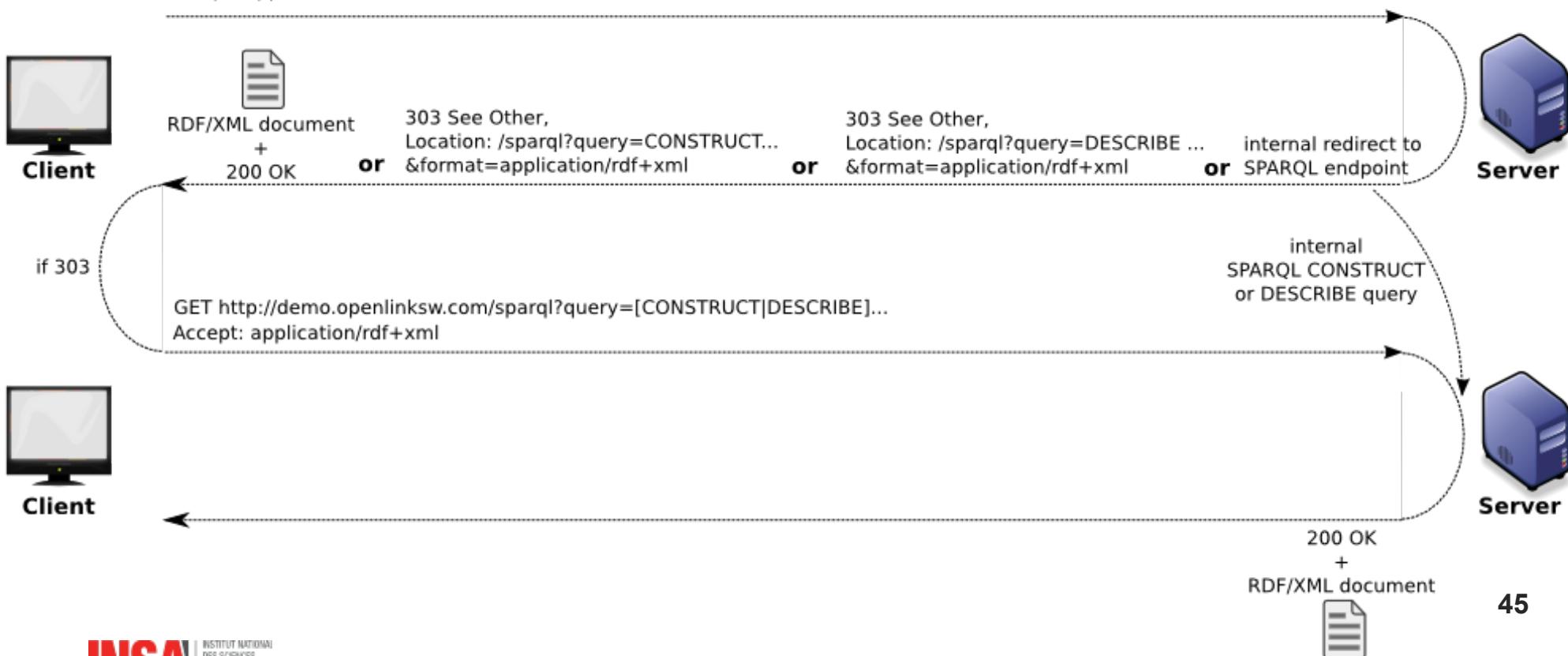


# Négociation de contenu

GET http://demo.openlinksw.com/Northwind/Customer/ALFKI[#this]  
Accept: text/html



GET http://demo.openlinksw.com/Northwind/Customer/ALFKI[#this]  
Accept: application/rdf+xml



# Déréférencement des ressources

Considérons la ressource non information [http://dbpedia.org/resource/John\\_Lennon](http://dbpedia.org/resource/John_Lennon)

- On demande une représentation HTML :

```
$> curl -H "Accept: text/html" -IL http://dbpedia.org/resource/John_Lennon
```

On est redirigé vers la page [http://dbpedia.org/page/John\\_Lennon](http://dbpedia.org/page/John_Lennon) (303 Redirect)

- On demande une représentation RDF/XML

```
$> curl -H "Accept: application/rdf+xml" -IL http://dbpedia.org/resource/John_Lennon
```

On est redirigé vers [http://dbpedia.org/data/John\\_Lennon.xml](http://dbpedia.org/data/John_Lennon.xml) (303 redirect)

# Interroger avec SPARQL

Séance4

Quelles sont les communes d'Île de France >100 000 habitants?

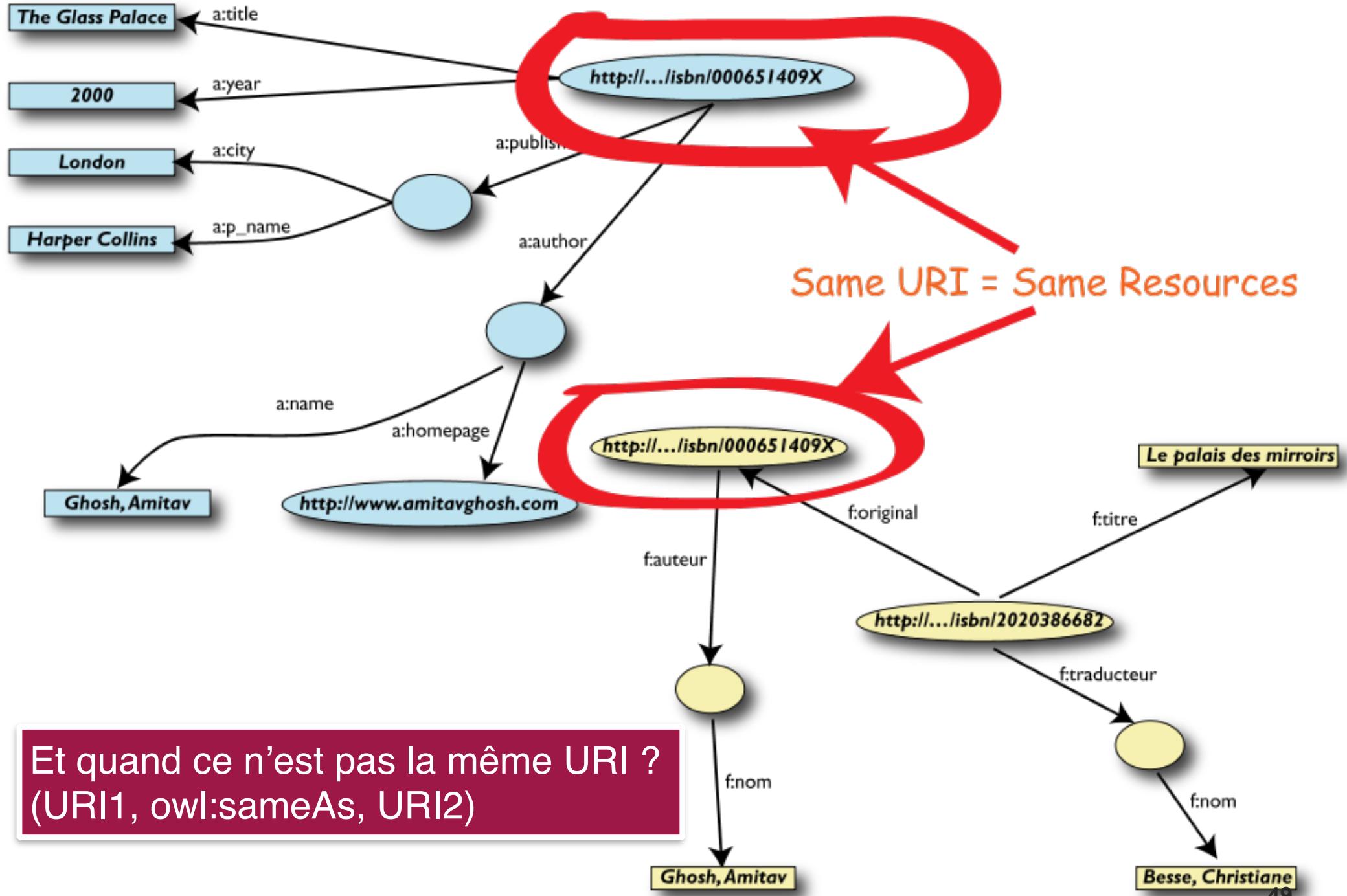
## Requête SPARQL ENDPOINT

```
prefix db-owl: <http://dbpedia.org/ontology/>
select * where {
  ?ville db-owl:region <http://fr.dbpedia.org/resource/Île-de-France> .
  ?ville rdf:type db-owl:Settlement .
  ?ville db-owl:populationTotal ?population .
  FILTER (?population > 100000)
}
```

## Représentation tabulaire de la réponse (en général du JSON)

ville	population
http://fr.dbpedia.org/resource/Saint-Denis_(Seine-Saint-Denis)	106785
http://fr.dbpedia.org/resource/Boulogne-Billancourt	114205
http://fr.dbpedia.org/resource/Montreuil_(Seine-Saint-Denis)	102770
http://fr.dbpedia.org/resource/Argenteuil_(Val-d'Oise)	103125
http://fr.dbpedia.org/resource/Paris	2243833

# 4/4 Intégration de données



# ‘Linking Open Data’

## Divers domaines

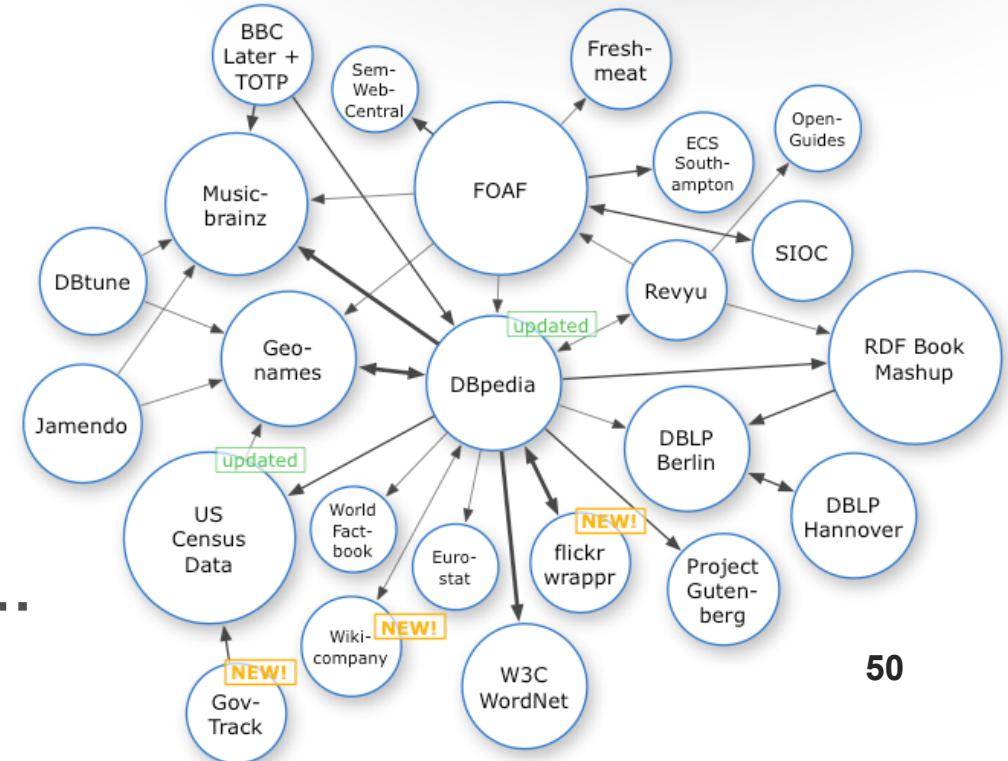
- public (compagnies)
- scientifiques
- gouvernements



## En développement croissant

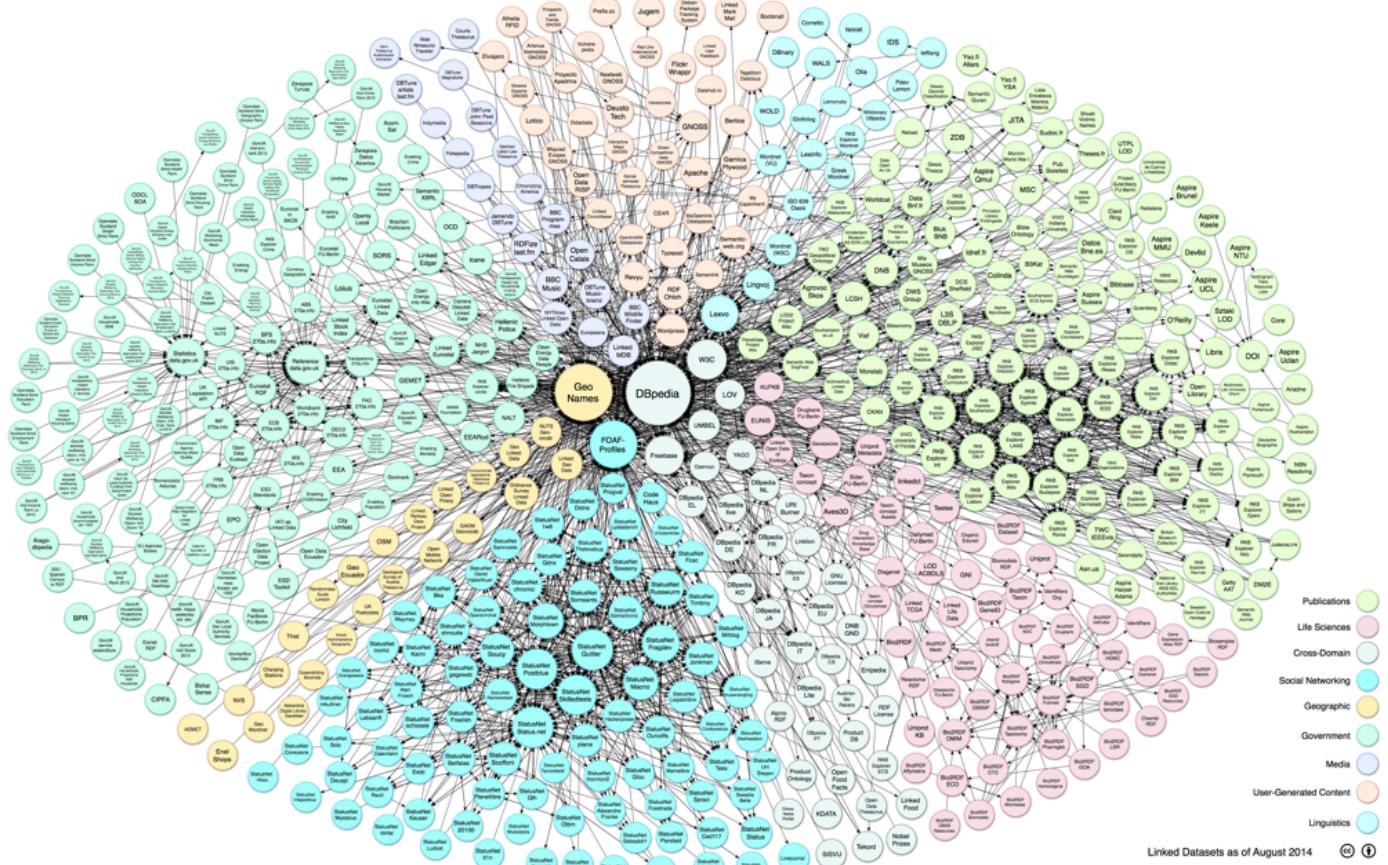
- soutenir la R&D
- soutenir l'innovation
- ...

en 2007 ...



# ‘Linking Open Data’

- En juillet 2009 : 4,7 milliards de triplets RDF - <http://linkeddata.org>



# Epilogue

# Faisons le point

- **Slogan AAA** : Anyone can say anything about anything
- **RDF** : un cadre de modélisation de données
- **Triplets/Graphes** : modèle simple
- **Fusion des données aisée** : agrégation de triplets
- **Infrastructure du Web et URI**, espace de nommage
- **Syntaxes concrètes de RDF** indépendantes du modèle abstrait
- **Conteneurs et collections**
- **Monde ouvert**



*Donner du sens à l'information que l'on manipule afin de pouvoir la réutiliser et raisonner : de la logique !*

# ce n'est pas fini...

- Notions de graphes nommés pour identifier des graphes/sous-graphes RDF avec des URI, afin de les annoter, par leur provenance, leur date de création/péremption
- Provenance Interchangeable Language : vocabulaire pour la publication de provenances
- RDFA1.1 pour une intégration avec HTML 5
- ...

RDF 1.1 Concepts and Abstract Syntax

W3C Recommendation 25 February 2014

This version:

<http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>

Latest published version:

<http://www.w3.org/TR/rdf11-concepts/>

Previous version:

<http://www.w3.org/TR/2014/PR-rdf11-concepts-20140109/>

Previous Recommendation:

<http://www.w3.org/TR/rdf-concepts>

Editors:

[Richard Cyganiak](#), DERI, NUI Galway

[David Wood](#), 3 Round Stones

[Markus Lanthaler](#), Graz University of Technology

Previous Editors:

Graham Klyne

Jeremy J. Carroll

Brian McBride

RDFA 1.1 Primer - Second Edition

Rich Structured Data Markup for Web Documents

W3C Working Group Note 22 August 2013

This version:

<http://www.w3.org/TR/2013/NOTE-rdfa-primer-20130822/>

Latest published version:

<http://www.w3.org/TR/rdfa-primer/>

Previous version:

<http://www.w3.org/TR/2012/NOTE-rdfa-primer-20120607/>

Editors:

[Ivan Herman](#), W3C, [ivan@w3.org](mailto:ivan@w3.org)

[Ben Adida](#), Creative Commons, [ben@adida.net](mailto:ben@adida.net)

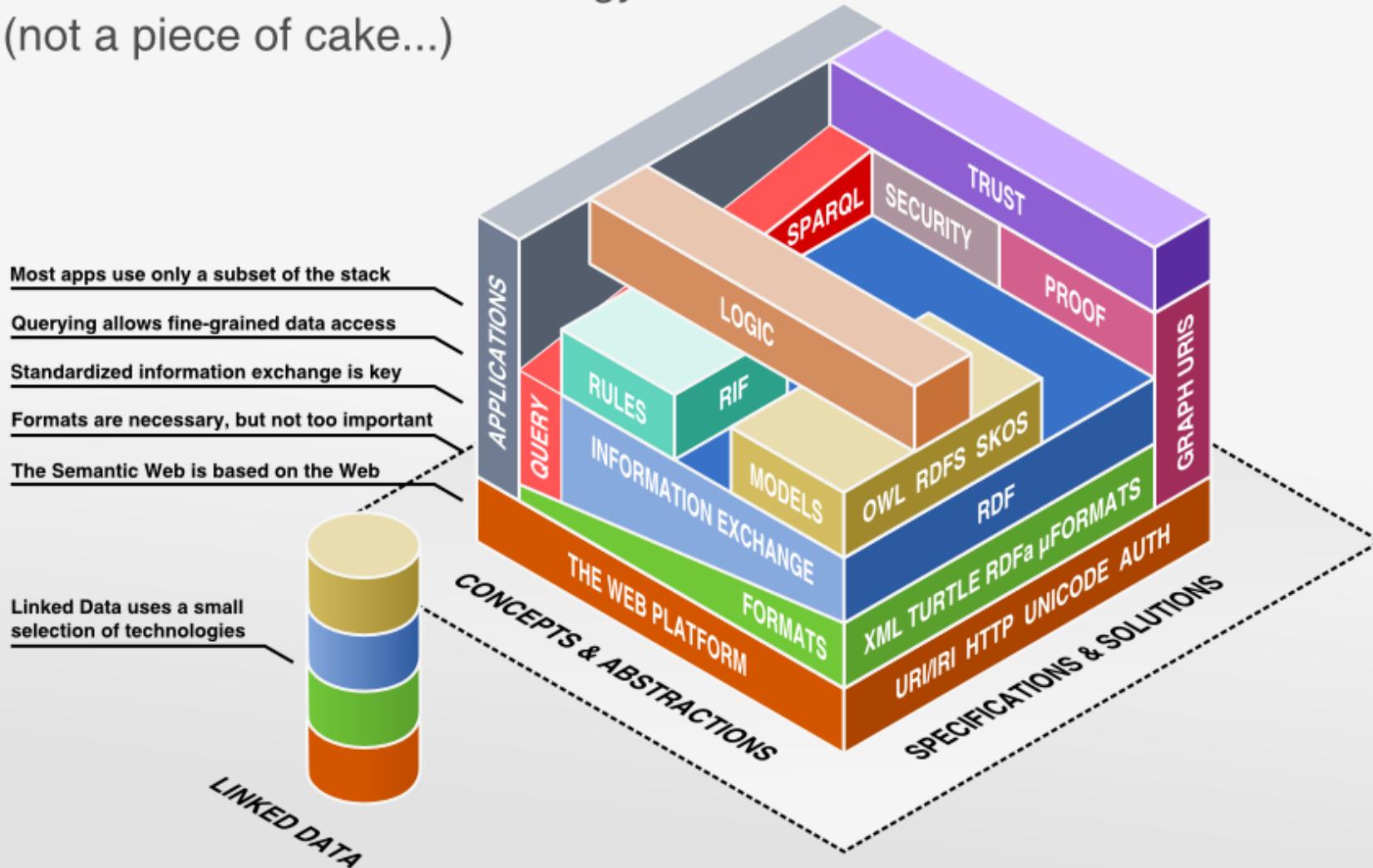
[Manu Sporny](#), Digital Bazaar, [msporny@digitalbazaar.com](mailto:msporny@digitalbazaar.com)

Mark Birbeck, webBackPlane.com, [mark.birbeck@webBackplane.com](mailto:mark.birbeck@webBackplane.com)

# Cours suivant

## SPARQL

The Semantic Web Technology Stack  
(not a piece of cake...)



# Courte biblio-webographie

- Harald Sack. **Semantic Web technologies**. Open course Hasso Plattner Institut (2013) : <https://open.hpi.de/course/semanticweb>
- Gandon F., Faron-Zucker C., Corby O. **Le Web sémantique. Comment lier les données et les schémas sur le Web ?** Dunod eds. 2012. (et MOOC INRIA correspondant).
- Allemang D., Hendler J. **Semantic Web for the working ontologist : Effective modelling in RDFS and OWL**. Morgan Kaufmann eds. 1st edition (2008)
- Cours de **F. Badra** ([fadi.lautre.net](http://fadi.lautre.net)) et références y figurant
- Actes annuels des conférences **WWW, ISWC, ESWC** (voir [www.informatik.uni-trier.de/~ley/db](http://www.informatik.uni-trier.de/~ley/db))



*Les recommandations et RFC sont d'excellentes sources !*