



**TIMCO**



# Le Linked Open Data et la plateforme DaaS de IO.CDS <https://io.datascience-paris-saclay.fr>

Plateforme "Data as a Service" du  
Center for (Input/Output) Data Science

**Karima Rafes, 25 novembre 2014**

EQUIPE PROJET  
**TAO**  
Paris-Saclay

# SOMMAIRE

1. Les mots clés de l'Open Data
2. Plateforme CDS de Paris-Saclay
3. Les autres approches

# 1

## Les mots-clés de l'Open Data

Open Data

Datahub

CSV

Excel

JSON

XML

Linked data

SPARQL

No-Sql

BigData

Dbpedia

Wikidata

Wikipedia

Intéropérabilité

Triplet

Endpoint

Triplestore

FOAF

LOD

Ontologie

Implication

IRI

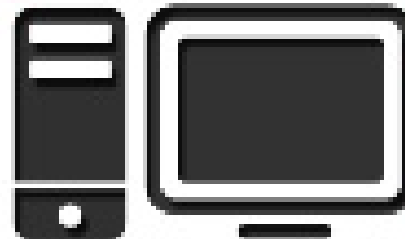
Propriété

Hackathon

# Open Data

- Informations au sein de documents sur le Web
- Lisible par un être humain
- Transformable gratuitement

Les  
administrations  
publiques  
partagent des  
données en ligne



Source Open Data Support EU

<http://fr.slideshare.net/OpenDataSupport/d211-training-module-15-promoting-the-reuse-of-open-data-through-odip-v012-fr>

# Open Data

→ Création de valeurs pour la société

## Publier des données



## Réutilisation des données

# Wikipedia, Wikimedia, MediaWiki, Wikiversité, Commons, Wikidata, Wikibase...

La première source d'Open Data sur le Web est **Wikipedia**.

La **fondation Wikimedia** se charge d'héberger les projets comme **Wikipedia, Commons, Wikiversité, Wikidata** etc. Elle récolte les fonds et les redistribuent.

Les entités comme **Wikimedia France** reversent une partie de ses fonds à la **fondation Wikimedia**.

La **fondation Wikimedia** et les associations n'ont aucun pouvoir éditorial direct sur les projets.

Tous les projets fonctionnent à l'aide d'un logiciel : **Mediawiki** avec des extensions comme **WikiBase**.

# Datahub.io

CSV (916)

example/rdf+xml (731)

api/sparql (560)

HTML (397)

application/rdf+xml (307)

meta/void (285)

XLS (284)

PDF (251)

aspx (211)

JSON (193)

spreadsheet (184)

text/turtle (153)

XML (145)

- Le catalogue officiel de l'Open Data dans le monde.
  - Gestion par la communauté Open Knowledge Foundation (OKF créé en 2004 à Cambridge, UK.)
  - Peut stocker une copie des données ouvertes et fournir des outils de base de visualisation.
  - La majorité des portails nationaux sur l'Open Data sont des copies de ce CMS.
  - La majorité des agrégateurs d'Open Data utilise l'API de Datahub.io.
- ➔ Plus de la moitié de l'Open Data sont des fichiers pour tableurs.

Source : Datahub.io

[http://datahub.io/dataset?\\_res\\_format\\_limit=0](http://datahub.io/dataset?_res_format_limit=0)

# CSV ou Excel (format pour tableurs)

● CSV (916)

example/rdf+xml (731)

api/sparql (560)

HTML (397)

application/rdf+xml (307)

meta/void (285)

● XLS (284)

PDF (251)

aspx (211)

JSON (193)

● spreadsheet (184)

text/turtle (153)

XML (145)

● application/vnd.ms-... (134)

meta/rdf-schema (127)

- Format de fichier pour pouvoir être ouvert dans Excel
- Le contenant respectent rarement un standard
- Plus de la moitié de l'Open Data son de ce type en 2014 plus de 95%, il y 4 ans.
- Données souvent enregistrées manuellement avec toutes les erreurs que cela implique.
- Données dont les identifiants ont disparu... difficile à mettre à jour automatiquement ou même à vérifier.

➔ De nombreux outils et de nouveaux métiers pour corriger des erreurs humaines... processus rarement automatique ou reproductible



# L'interopérabilité un luxe ou une nécessité ?

L'objectif de création de valeur via l'Open Data des fichiers CSV à de nombreux travers... dont techniques qui empêchent la réutilisation réelle de la majorité de ces données.



# L'interopérabilité une nécessité...

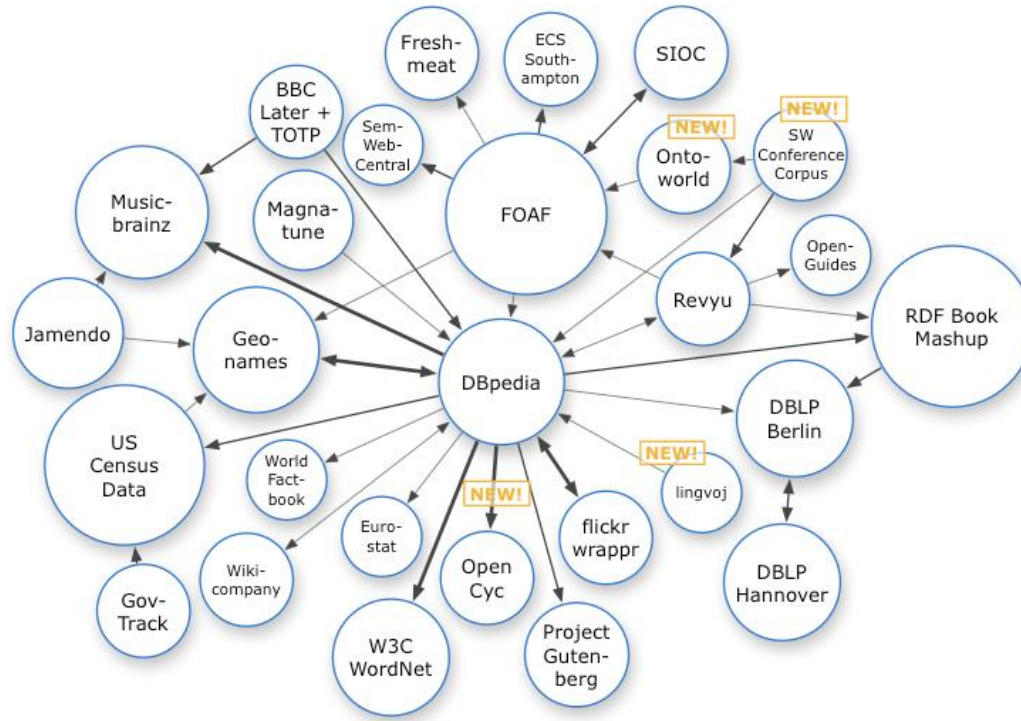
- Données accessibles de manière atomique
- Lisible par une machine et un humain
- Interrogeable gratuitement directement à la source



# Le LOD (Linked Open Data)...

## Le GGG (Geant Global Graph)

Et si le Web devenait une base de données ?  
2005 à 2007 explosions des démos...





<http://lod-cloud.net/versions/2007-11-10/lod-cloud.png>

# Linked Open Data + tremblements de terre



# Linked Open Data + TV (BBC Backstage)

	Series 2	5	Rise of the Cybermen	The TARDIS is trapped on a parallel Earth, where Rose discovers her father is still alive.
	Series 2	6	The Age of Steel	The Cybermen take control of London and start converting the populace.

# Le Web de données ou le Linked Data

Initiative visant à favoriser la publication de données :

- structurées
- accessibles à travers le Web
- reliées entre elles pour constituer un réseau global d'informations

Le lancement de l'initiative, en 2006, avait pour objectifs de :

- promouvoir une vision du Web comme une base de données globale
- naviguer à travers les données sur le Web de la même façon que l'hypertexte permet de relier des pages Web

Le Consortium World Wide Web (W3C) à travers cette initiative, poursuit sa mission qui est d'assurer au Web une croissance à long terme

# IRI (Internationalized Resource Identifier)

Chaque chose doit être représentée par un IRI.

Un IRI doit être :

- Unique sur le Web
- Valide : un IRI dans un navigateur doit afficher un document
- Lisible : (si possible) dans la langue de l'être humain qui l'a créé
- Lié : si vous décrivez une chose qui est déjà décrite sur le web avec un IRI, il faut lier les IRIs.

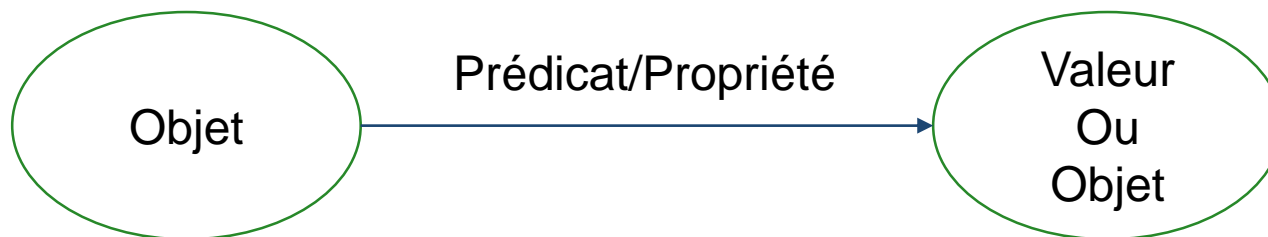
Une IRI pointe sur un document qui doit également être lisible pour les humains et les machines. Un IRI affiche

- une page HTML pour un être humain,
- un document RDF pour une machine.

# RDF (Resource Description Framework) Triplets

Modèle de graphe destiné à **décrire de façon formelle les ressources Web** et leurs métadonnées, de façon à permettre le traitement automatique (par des machines) de telles descriptions.

Un document structuré en RDF est constitué d'un ensemble de triplets.



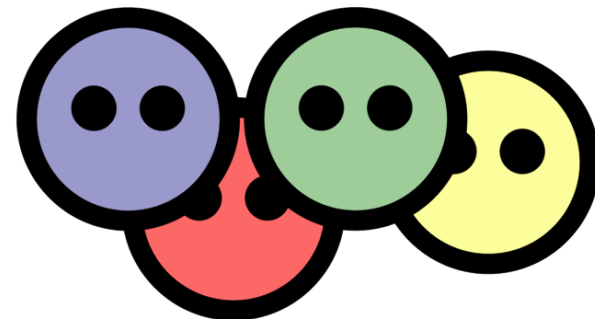
[https://en.wikipedia.org/wiki/Voyager\\_1](https://en.wikipedia.org/wiki/Voyager_1)

<https://www.wikidata.org/wiki/Property:P18> (Image)

<https://commons.wikimedia.org/wiki/File:Voyager.jpg>



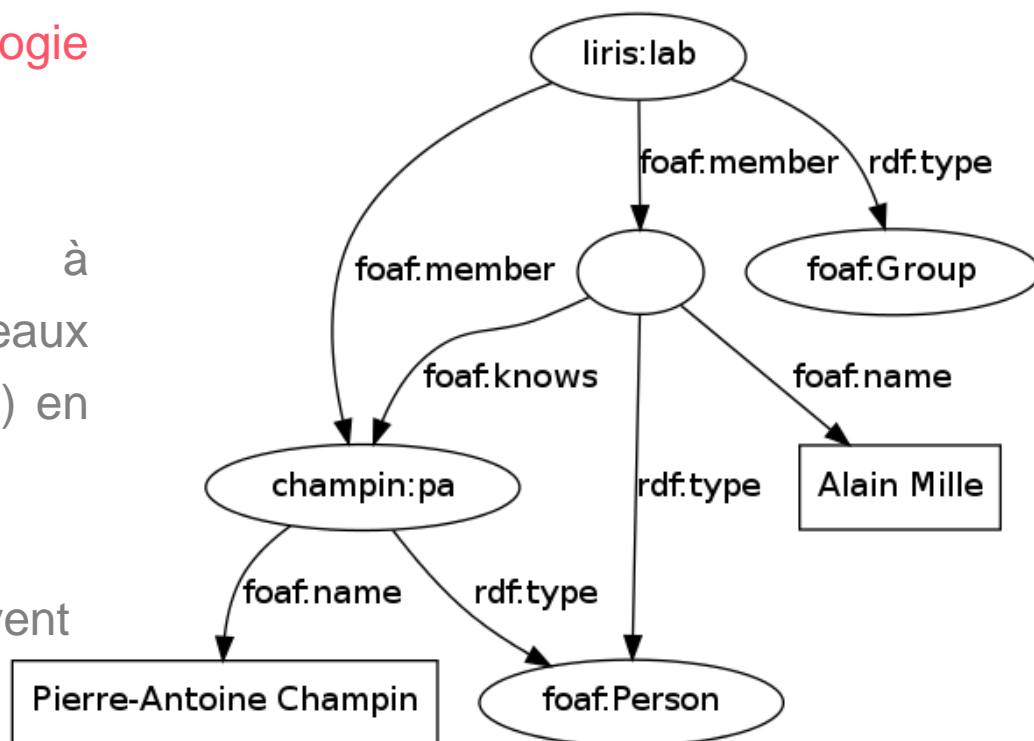
# Ontologie... FOAF (Friend of a friend)



Entre 2001 et 2007, les experts du domaine partagent en RDF leurs contacts en respectant une **ontologie partagée**.

Des services commencent à reproduire les usages des réseaux sociaux (le buzzword du moment) en agrégeant les fichiers RDF.

Les ontologies comme FOAF servent aux applications métiers pour exploiter ces données.



<http://liris.cnrs.fr/~pchampin/2012/lod/index.html#>

# Triplestore, Quadstore... BDD RDF

En 2006, avec ces ressources structurées et les buzzwords du moment, les bases de données commencent à agréger les données RDF pour offrir des services en ligne.

Une BDD RDF :

- contient des triplets RDF
- Est agnostique à priori au contexte métier. On peut enregistrer des données respectant n'importe quelles ontologies.

Premières implémentations de BDD qui débouchent sur la première recommandation au W3C en 2008 pour définir un protocole d'accès aux données et un langage pour effectuer des requêtes à travers le Web. Il donne naissance au SPARQL.

# SPARQL

## prononcer « sparkle »

SPARQL doit devenir au Web ce que le SQL est à une base de données.

```
SELECT ?image
WHERE {
  <https://en.wikipedia.org/wiki/Voyager\_1>
    <https://www.wikidata.org/wiki/Property:P18>
      ?image .}
```

image

<https://commons.wikimedia.org/wiki/File:Voyager.jpg>



# Endpoint Sparql

Une BDD RDF sur le Web fournit un endpoint Sparql :

- Une adresse HTTP : <http://www.example.org/sparql/>
- Paramètres Query GET ou POST pour passer une requête

Exemple de requête :

[http://www.example.org/sparql/?query= SELECT ?image WHERE..  
...&output=json](http://www.example.org/sparql/?query=SELECT%20?image%20WHERE...&output=json)

# JSON, CSV, XML, Turtle, n3, etc...

## Formats en sortie après une requête SPARQL

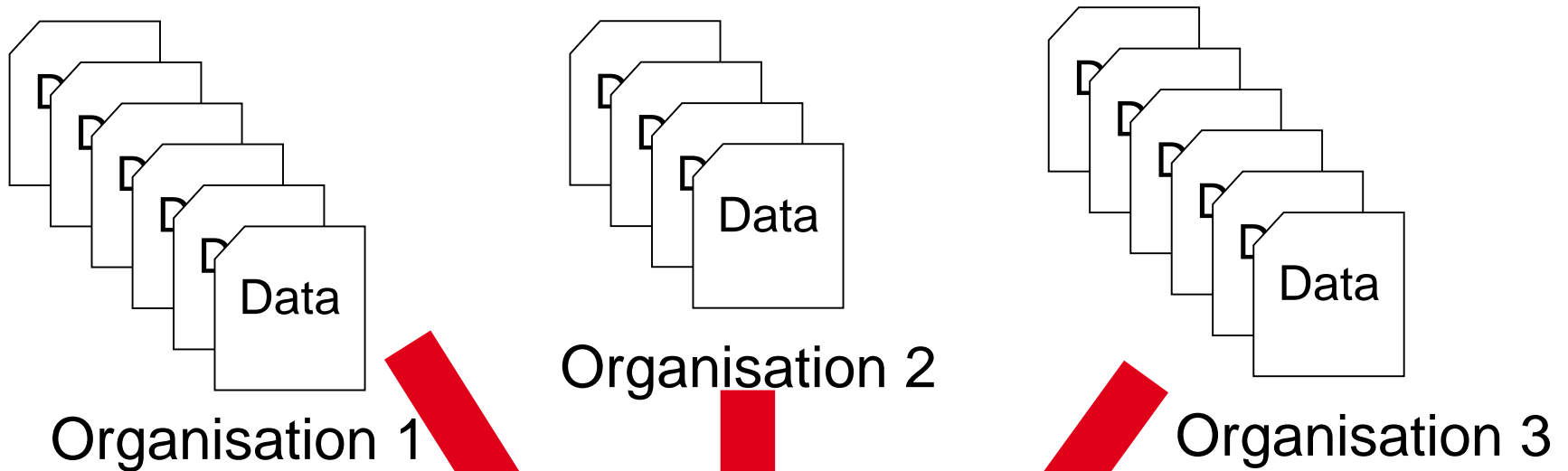
Le format de sortie d'une requête SPARQL en lecture est un tableau comme en SQL.

Le développeur peut choisir le format en fonction du contexte de son application. Par exemple, pour afficher des données :

- dans une page Web en Javascript : JSON
- dans une application : XML
- dans son tableur : CSV
- etc.

C'est en fonction de la base de données et des préférences de la personne qui développe l'application.

# SPARQL n'est pas du NoSQL (BigData)

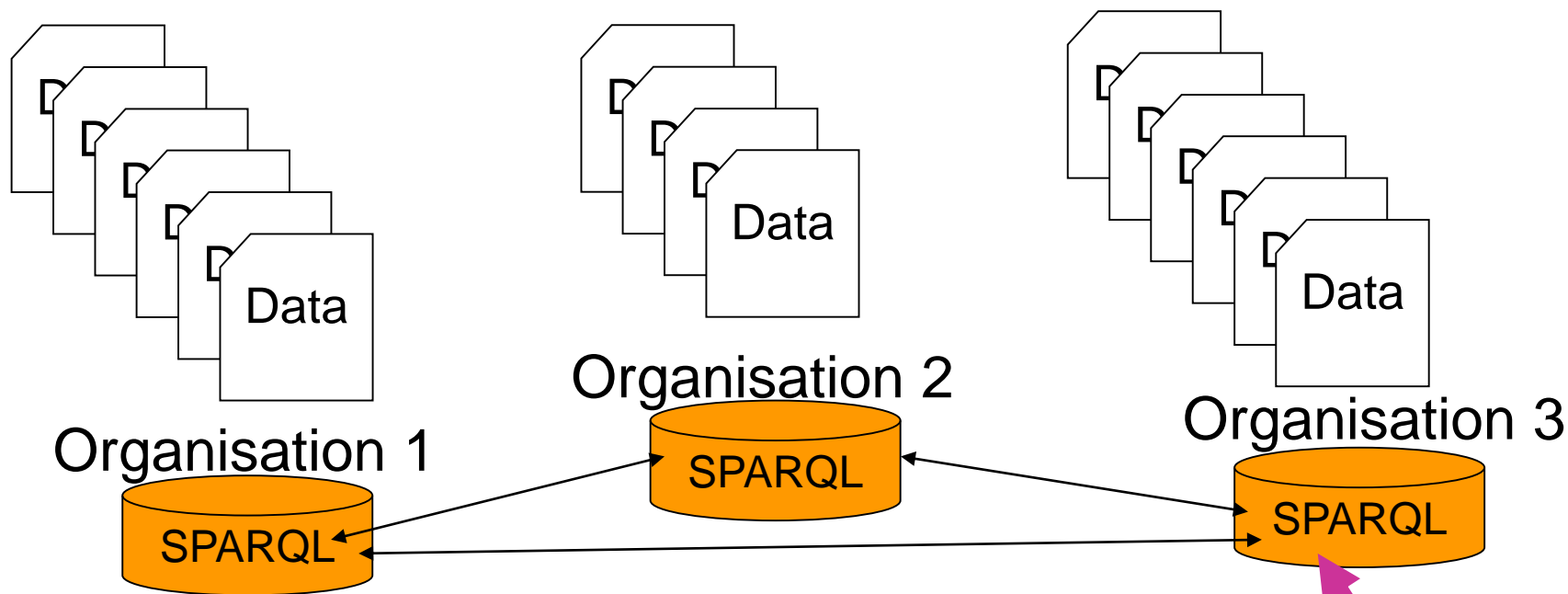


NoSQL = une BDD pour tous  
(exemple: Google, Facebook, etc...)



# SPARQL est aussi un protocole Web

## Les requêtes fédérées



Décentralisation et mise en relation des données à travers le Web par les créateurs/propriétaires des données eux-mêmes.

Le croisement des données peut se faire du côté du client (respect des données privées) ou des serveurs

# Exemple d'éditeurs SPARQL

<http://ldlab.ifs.tuwien.ac.at/node/7>

The screenshot displays a web-based Mashup Platform interface. At the top, there are navigation tabs: "Mashup Platform", "Widget Collection", "Widget Annotator", and "Home page". Below these are buttons for "Save", "Clear", "Load", and "Run".

The interface is divided into several sections:

- DBpedia Basics:** Includes "Json Viewer", "Item Extractor", and "Instance Viewer".
- Maps:** Includes "Google Map", "Location", "Geo Merger", and "Citybike".
- Terminal "output" can be connected with:** A section for connecting the mashup's output to other widgets.
- Definite results:** Lists the inputs used in the mashup: "input1 Geo Merger", "input Google Map", "input2 Geo Merger", and "input Instance Viewer".

The central workspace contains three widget editors connected by blue lines:

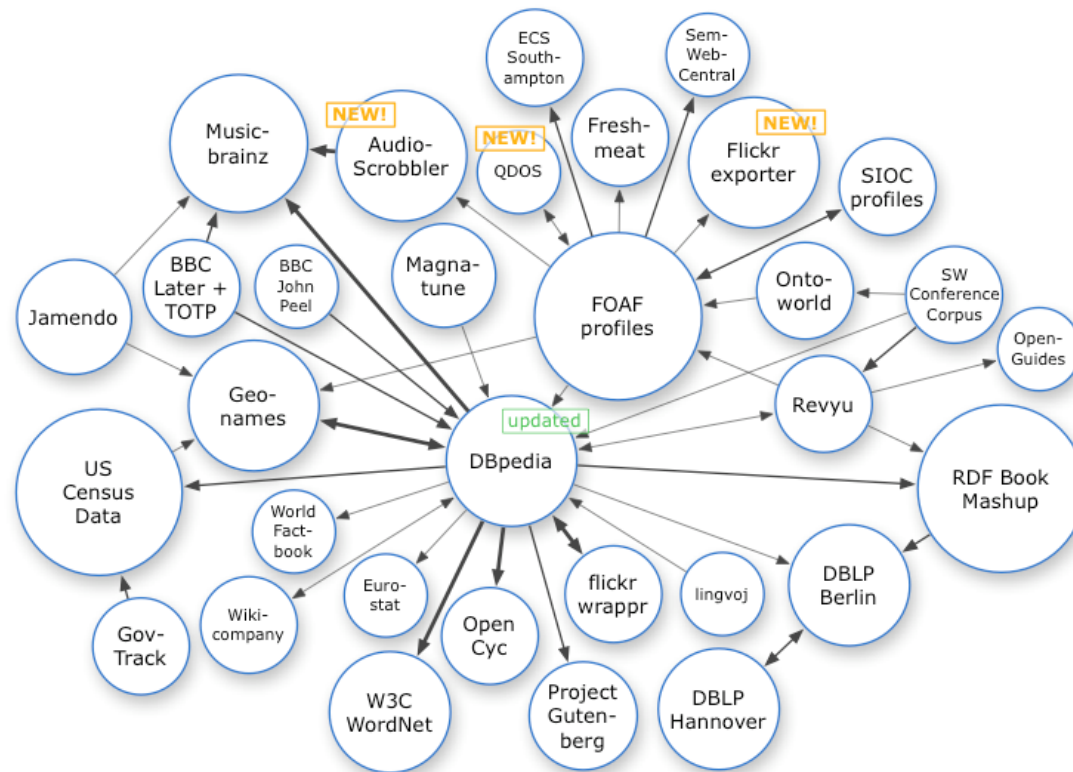
- Citybikes:** A widget with a "Run" button, a "Street:" input field, and a "Looking for:" dropdown menu set to "Free box".
- Location:** A widget with a "Run" button and a "Choose location type:" dropdown menu set to "Parks".
- Geo Merger:** A widget with a "Run" button, "Maximum distances:" set to "100 meters", and "Minimum distances:" set to "None".

On the right side, a "Google Map" widget displays a map of Vienna, Austria. It includes a "Run" button, "Previous" and "Next" navigation buttons, and a status bar showing "Found: 43. Current: 1". A blue pin is placed on the map near the "Rathaus" (City Hall) area.



# SPARQL, l'étincelle du LOD (Linked Open Data)

Première recommandation du W3C en 2008 mais des BDD seulement en lecture  
« sparkle », en anglais « étincelle »...






<http://lod-cloud.net/versions/2008-02-28/lod-cloud.png>

# DBPedia, Yago, etc...

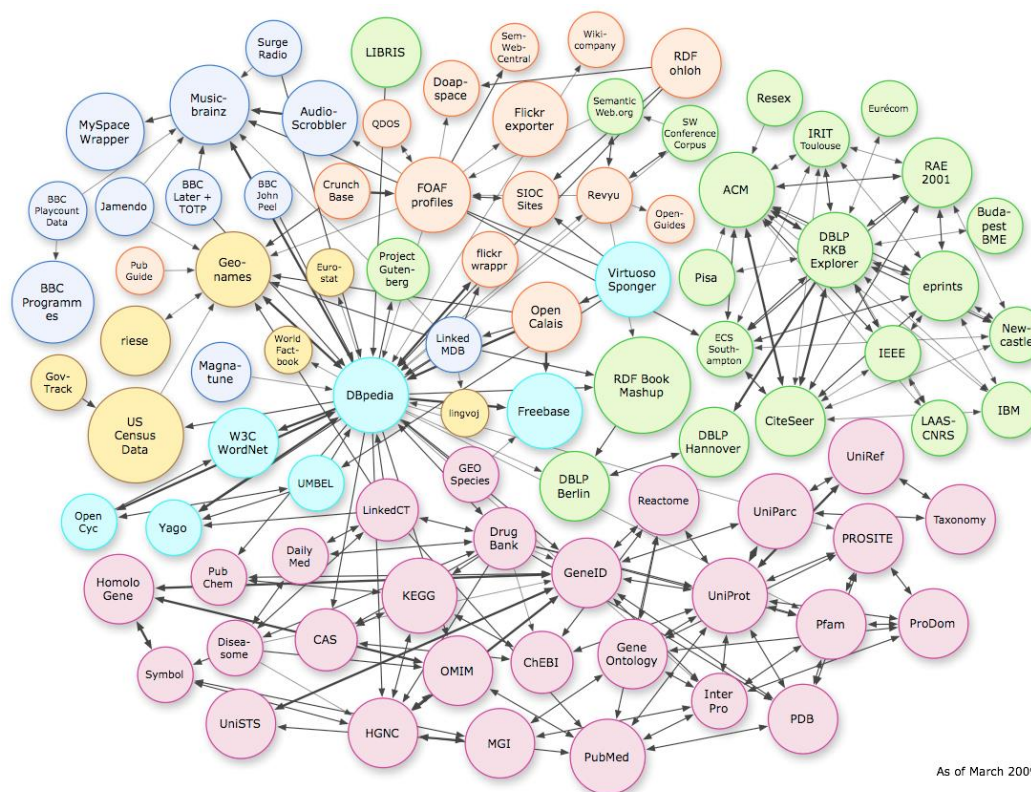
Extraction automatique des pages de Wikipédia.

En 2008, seulement des démos... mais surtout des IRI uniques.

image 	artist 	comment 
	Agnolo Bronzino	La Déploration sur le Christ mort est une peinture de l'artiste italien maniériste Angelo di Cosimo di Mariano, exécutée entre 1540 et 1545. L'œuvre est conservée au Musée des Beaux-Arts et d'archéologie de Besançon.

# 2009, les sciences s'emparent du LOD

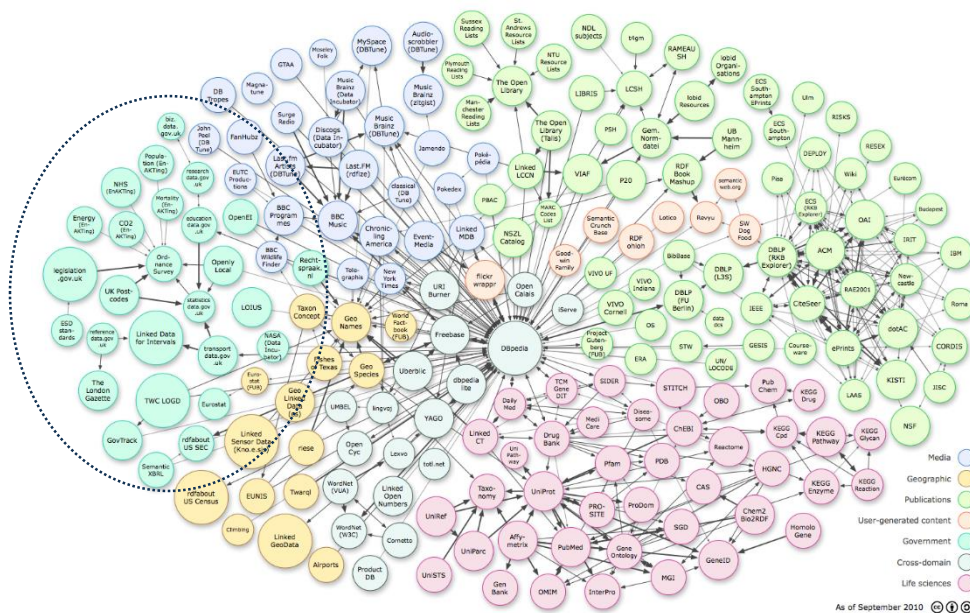
Dbpedia permet d'offrir des identifiants uniques de manière transversale aux différents domaines scientifiques.



[http://lod-cloud.net/versions/2009-03-05/lod-cloud\\_colored.png](http://lod-cloud.net/versions/2009-03-05/lod-cloud_colored.png)

# 2010, L'Open Data bascule dans le LOD

L'Open Data au UK bascule à son tour. On commence à parler de Web Science. En France, le buzzword de l'Open Data commence mais la question du format des données est clairement éludée.



[http://lod-cloud.net/versions/2010-09-22/lod-cloud\\_colored.png](http://lod-cloud.net/versions/2010-09-22/lod-cloud_colored.png)

## Exemple : Data.gov.uk Education

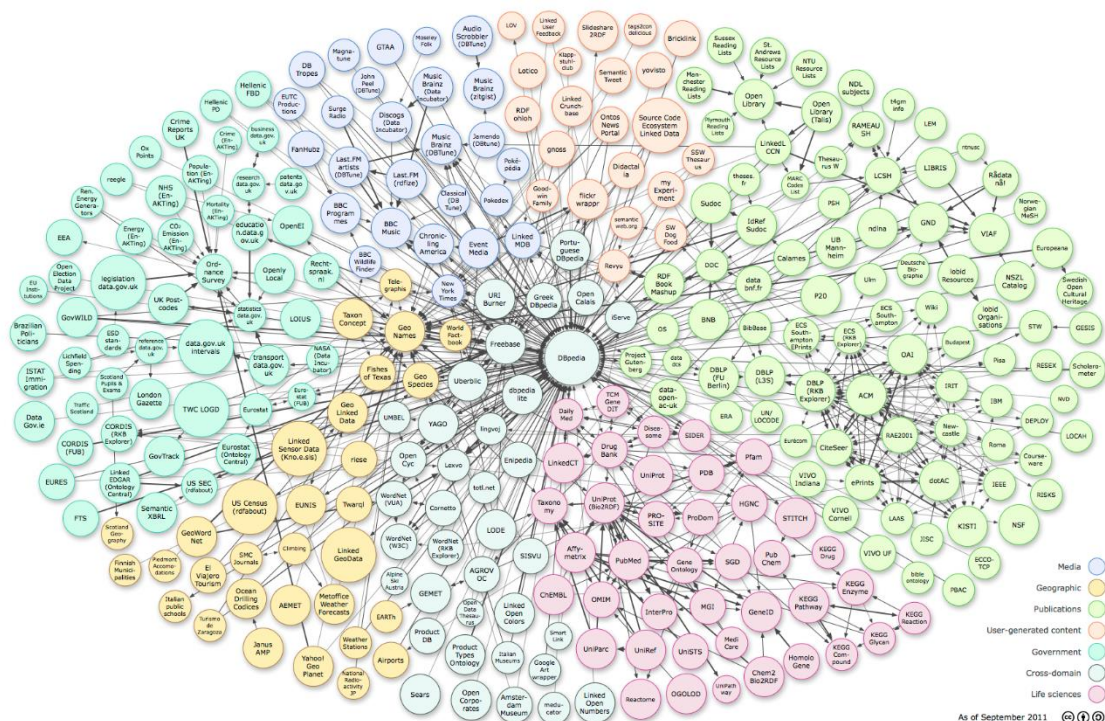


# 2011, Draft en urgence de SPARQL 1.1

Dbpedia au cœur du Linked Open Data...

Les BDD RDF se multiplient et étendent SPARQL pour répondre aux besoins de leurs clients.

Les BDD RDF divergent et une nouvelle recommandation est proposée pour faire de SPARQL un langage à la hauteur du SQL.



[http://lod-cloud.net/versions/2011-09-19/lod-cloud\\_colored.png](http://lod-cloud.net/versions/2011-09-19/lod-cloud_colored.png)

# Hackathon

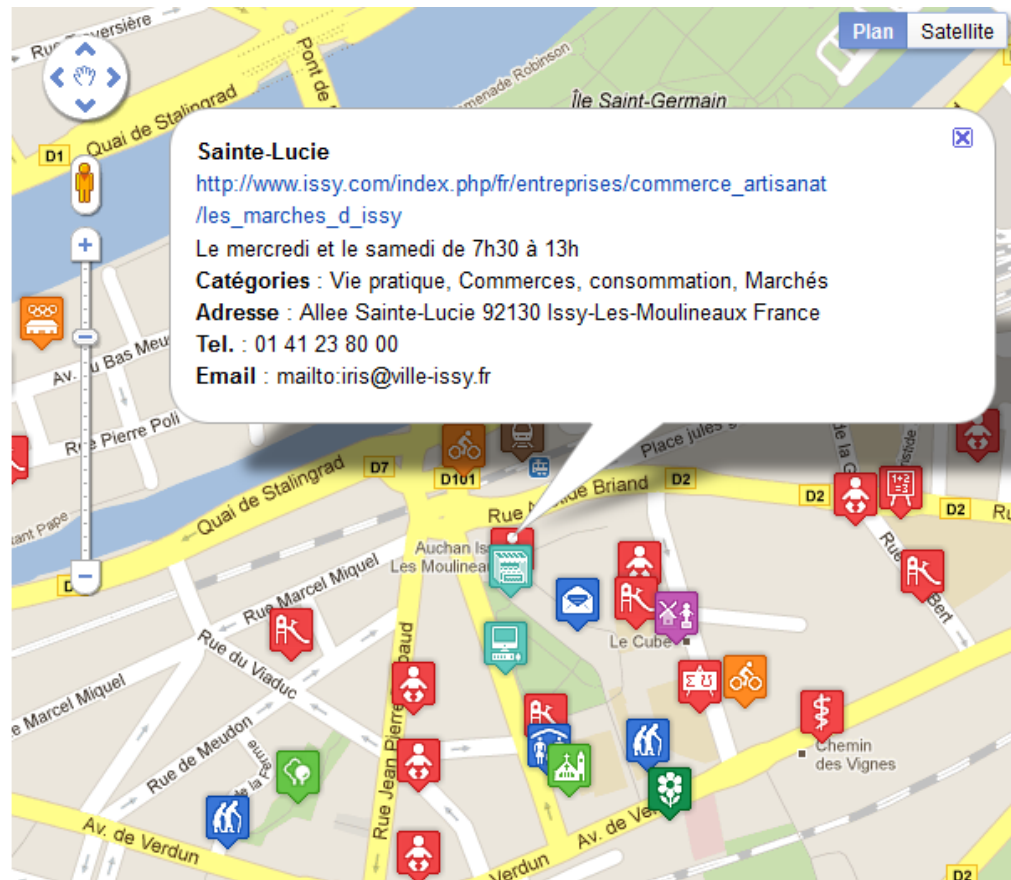
Stabilité des technos + des données = développeurs

Les widgets et applications pour mobile se multiplient.

Le SPARQL occulte complètement le RDF pour les consommateurs de données. RDF devient un format "Dump" de BDD.

Buzzword des Hackathons dans le monde.

En France, on parle licences de données...



# Linked Data = Interopérabilité des données

- Lisible par une machine = interopérable
- Interopérable = réutilisable par un développeur
- Interopérable  $\neq$  réutilisable par une machine

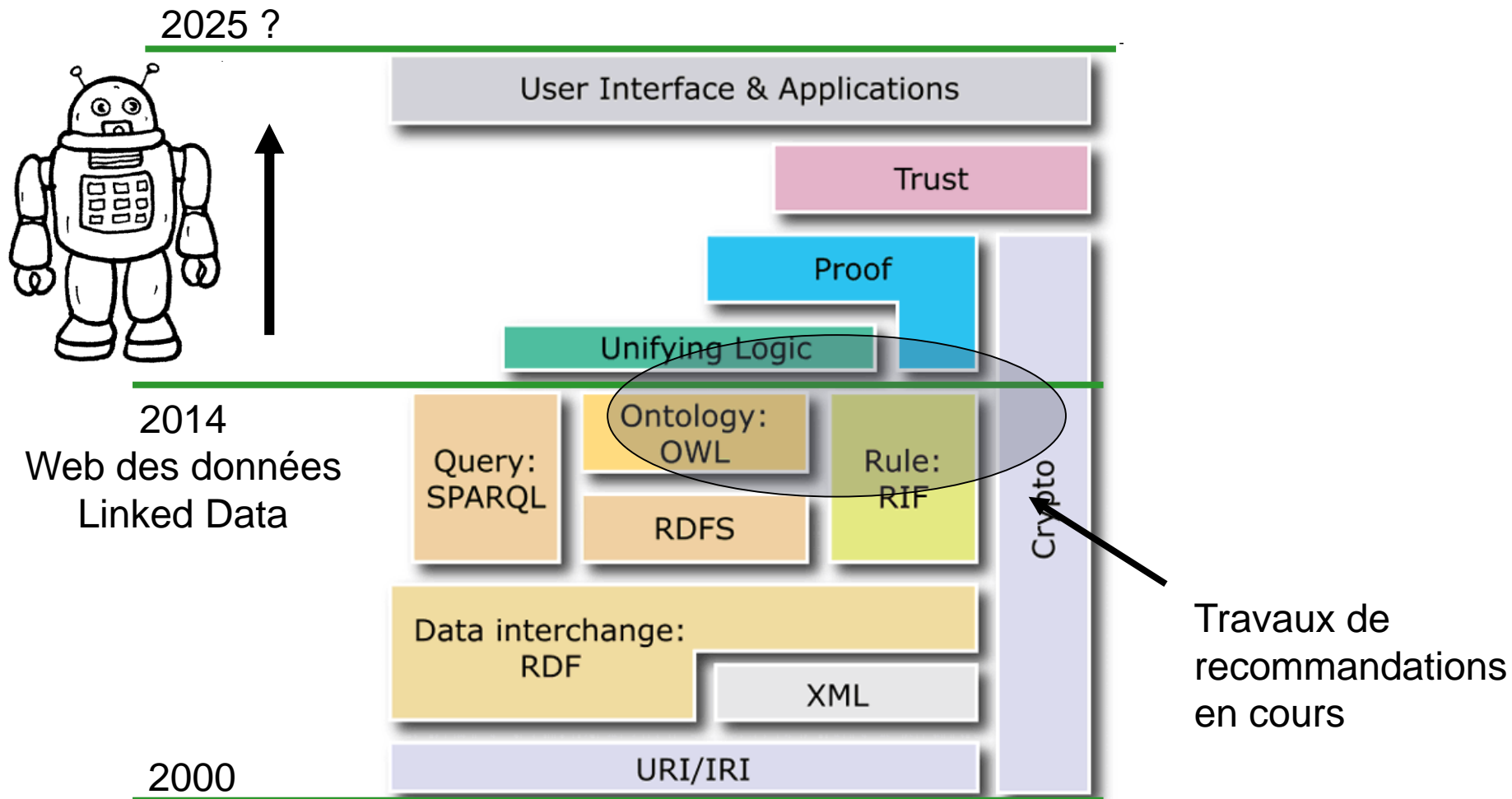




# Web sémantique

- Mouvement/Consensus/Roadmap technologique
- Rendre réutilisable les données/informations via Internet pour les êtres humains et les machines
- ➔ Rendre lisibles les connaissances aux machines pour faciliter l'émergence de nouvelles connaissances par les êtres humains
- ➔ Linked Data est une des étapes intermédiaire de ce mouvement.

# Point de vue Recherche



Applications et interfaces				
Semantic Web	<b>Confiance</b>	Recherche en cours		
	<b>Preuve</b>	Recherche en cours		
	<b>Logique</b>	Recherche en cours		
	<b>Stockage : Triplestore</b>	<b>Règles</b>	RIF · Semantic Web Rule Language	
		<b>Structure</b>	<b>Ontologies</b>	OWL · SKOS
			<b>Taxonomies</b>	RDFS
<b>Requête</b>	SPARQL			
<b>Modèle de données</b>	RDF ( <a href="#">RDF/XML (en)</a> ) · <a href="#">Notation3</a> · <a href="#">Turtle</a> · <a href="#">N-Triplets (en)</a> )			
Web	<b>Document</b>	HTML5 ( <a href="#">Microdonnées</a> ) · CSS3 · HTML ( <a href="#">Avec RDFa</a> · <a href="#">GRDDL</a> )		
	<b>Syntaxe</b>	XML · XSD		
	<b>Protocole</b>	DNS · HTTP · IRI ( <a href="#">URI</a> · <a href="#">URL</a> )		
	<b>Identifiant</b>	IRI · ( <a href="#">URI</a> · <a href="#">URL</a> )		
Internet (Modèle OSI)	<b>6. Présentation</b>	Unicode		
	<b>5. Session</b>	TLS		
	<b>4. Transport</b>	TCP · UDP		
	<b>3. Réseau</b>	Ex. IPv4 · IPv6 · ARP · IPX · ICMP		
	<b>2. Liaison</b>	Ex. Ethernet · Wi-Fi · PPP · ATM		
	<b>1. Physique</b>	Ex. Ligne de transmission · Fibre optique · Radio		

# Système d'implication (entailment regime) Inférence

Les BDD RDF commencent à **implémenter** les implications des ontologies recommandées par le W3C comme le vocabulaire **RDF et RDFS**.

**La performances** des bases de données est cependant **la priorité** pour surfer sur la vague du Big Data.

L'**intégration** de OWL (Web Ontology Language) dans les BDD RDF est... **en cours**... Et les **inférences** suivront (<http://www.w3.org/standards/semanticweb/inference>)

Les **systèmes d'implication font s'écrouler les performances** des BDD et leurs implémentations suivront la même courbe que l'évolution de la puissance des machines.

# Prise de conscience des Wikipédiens...



En 2010, Dbpedia commence à exporter de plus en plus de données de Wikipedia et fusionne ses données avec des BDD commerciales.

Les Wikipédiens commencent à s'en **exaspérer** :

- Les développeurs commencent à **impacter le contenu** de Wikipedia et la structure des pages se complexifie **pour en faciliter l'extraction**.
- Les **données** sont réutilisées dans des applications où les données sont **sorties du contexte de la page** et les pages d'origines devient du contenu secondaire.
- **L'objectif** n'est plus d'obtenir des données de **qualité** mais d'être **exhaustif**.

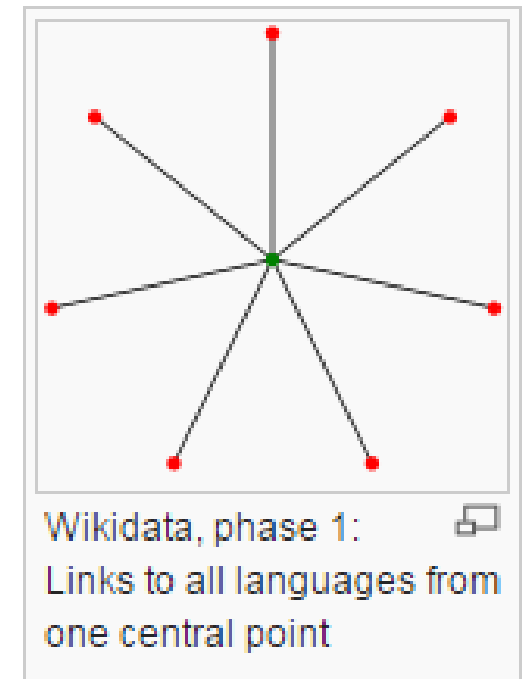
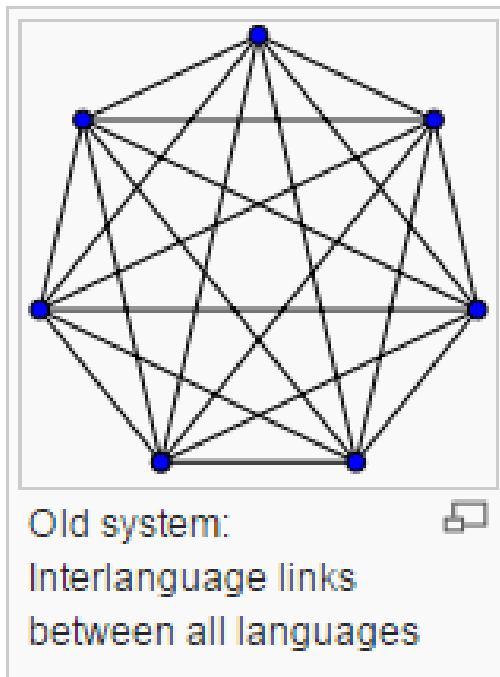
**Crise d'identité** : qu'est-ce que Wikipédia ? Quelle doit être son rôle ?

## La fronde des Wikipédiens...



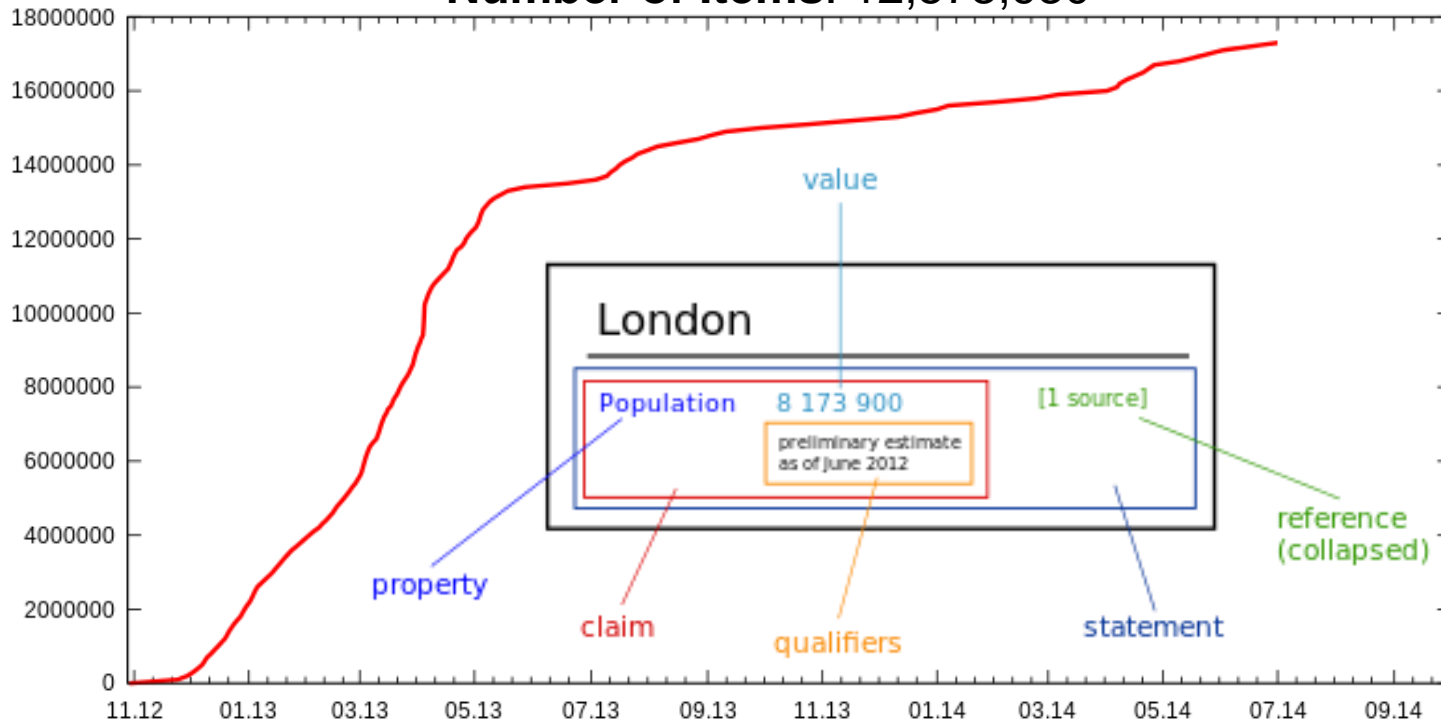
Philip Kaufman's 1978 Invasion Of The Body Snatchers

# Contre attaque des Wikipédiens : Wikidata Phase 1 : IRI unique (2012)



# Contre attaque des Wikipédiens : Wikidata Phase 2 : centraliser le contenu structuré (2013)

Number of Items: 12,573,930



## Properties:

**Number of properties:** 436

**Number of labels:** 13,892 (avg. 32 per Property )

**Number of descriptions:** 6,400 (avg. 15 per Property )



# Contre attaque des Wikipédiens : Wikidata

## Phase 3 : Réinjection du contenu dans les infoboxes (2014)

**Voyager 1** (Q48469) [edit]

space probe and the farthest man-made object from Earth [edit]

Also known as: [edit]

In other languages | Statements | Wikipedia pages linked to this item | Wikinews pages linked to this item | Wikiquote pages linked to this item | Wikisource pages linked to this item | Wikivoyage pages linked to this item | Pages on other sites linked to this item

**In other languages** [edit]

	Voyager 1
français	No description defined yet
	Also known as:
	Voyager 1
occitan	No description defined yet
	Also known as:
	Voyager 1
italiano	sonda interplanetaria
	Also known as:

**Statements**

<b>COSPAR ID</b>	1977-084A#	<span>[edit]</span>
	↑ 1 reference	
		<span>[add]</span>
<b>Instance of</b>	robotic spacecraft	<span>[edit]</span>
	↳ 0 references	
		<span>[add reference]</span>
		<span>[add]</span>



**Voyager 1**

Voyager 1, its extended booms deploying instruments (right) and RTGs (left). The golden record, included to communicate with extraterrestrials, is below the high-gain antenna.

**Mission type** Outer planetary, heliosphere, and interstellar medium exploration

**Operator** NASA / JPL

**COSPAR ID** 1977-084A<sup>[1]</sup>

**SATCAT №** 10321<sup>[2]</sup>

**Website** [voyager.jpl.nasa.gov](http://voyager.jpl.nasa.gov) [edit]

**Mission duration** 37 years, 2 months and 17 days elapsed

Planetary mission: 3 years, 3 months, 9 days  
Interstellar mission: 33 years, 11 months and 8 days elapsed (continuing)

**Spacecraft properties**

**Manufacturer** Jet Propulsion Laboratory


**Launch mass** 721.9 kilograms (1,592 lb)

# Contre attaque des Wikipédiens : Wikidata

## Phase 3 bis : Mise à disposition des données RDF (2014)

### Index of /wikidata-exports/rdf/exports/

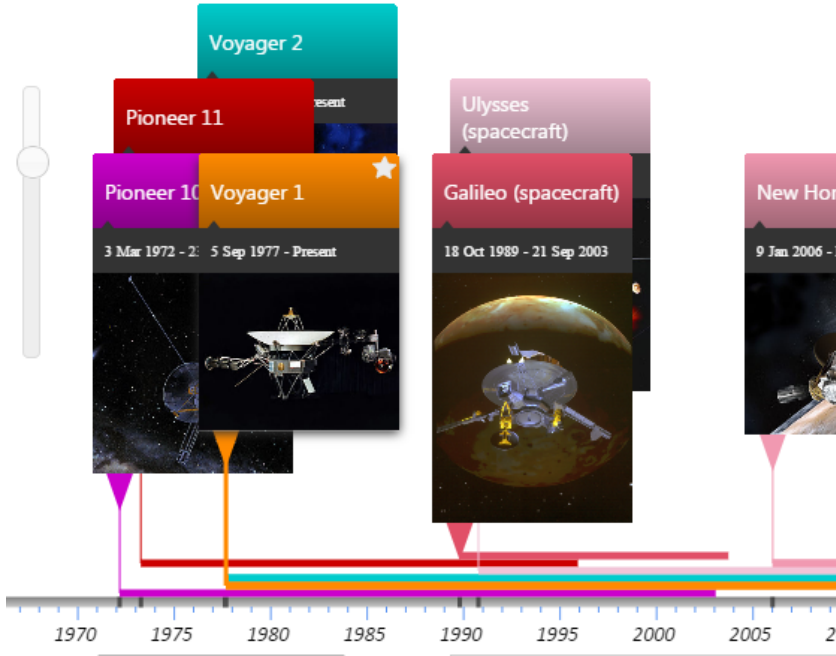
Name	Last Modified	Size	Type
Parent Directory/			
<a href="#">20140420/</a>			
<a href="#">20140526/</a>			
<a href="#">20140804/</a>			
<a href="#">20141013/</a>			



Start searching for your topic...

Medium ▾
Save
Sign In



#### Voyager 1

**Voyager 1** is a 722-kilogram (1,592 lb) [space probe](#) launched by [NASA](#) on September 5, 1977, to study the outer [Solar System](#). Operating for 37 years, 2 months and 16 days as of November 21, 2014, the spacecraft communicates with the [Deep Space Network](#) to receive routine commands and return data. At a distance of about 130.29 AU ( $1.949 \times 10^{10}$  km) (approximately 12 billion miles) from [Earth](#) as of November 11, 2014,<sup>[3][4]</sup> it is the farthest spacecraft from Earth.<sup>[5][6]</sup>

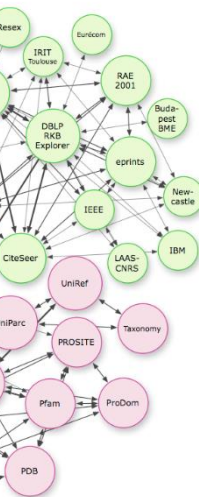
The primary mission ended on November 20, 1980, after encounters with the [Jovian](#)

# Wikidata et Dbpedia

Probablement une fusion des projets à terme.

Nombreux outils commencent à apparaître

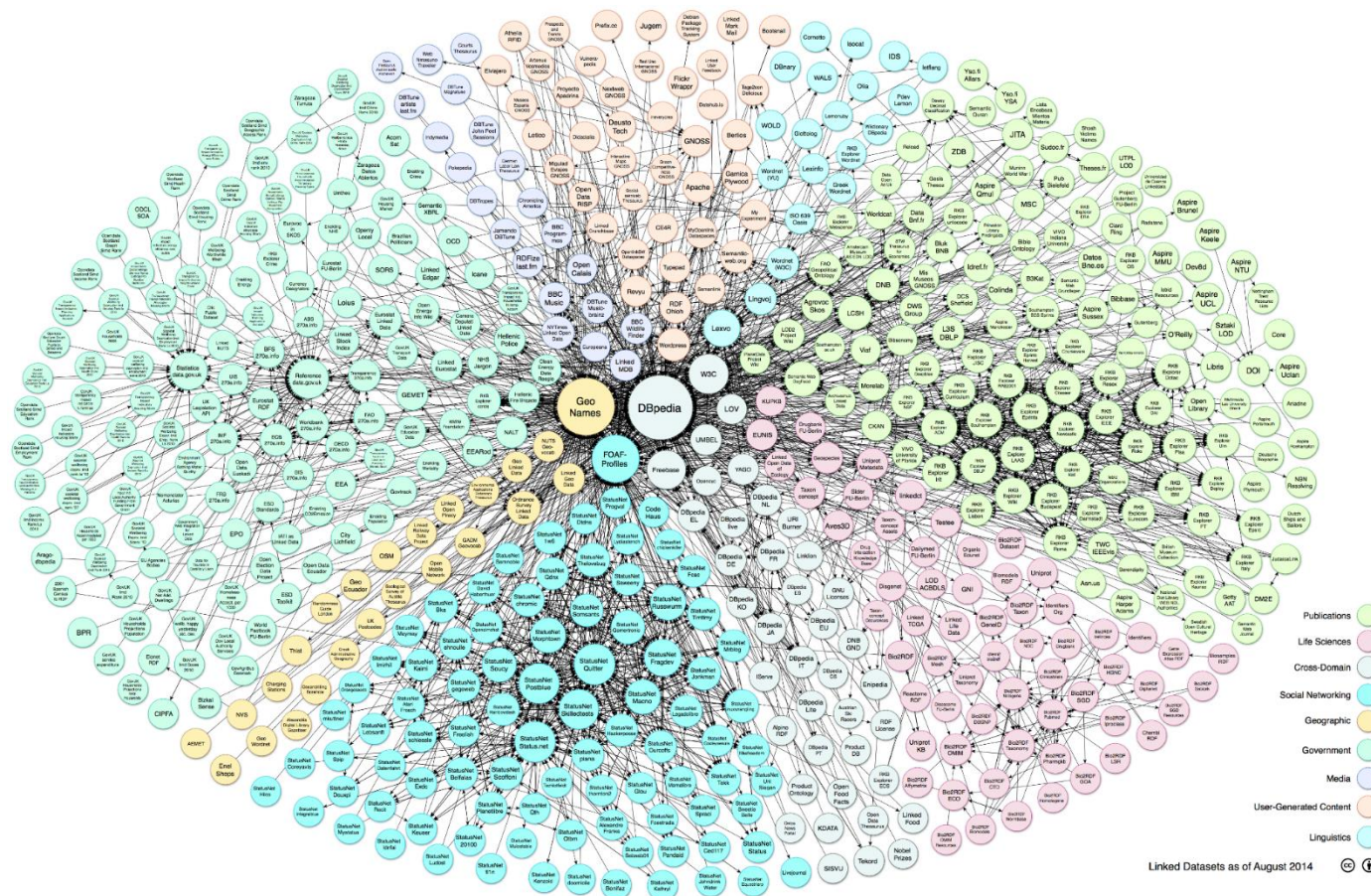
Wikidata offrira des IRI unique et une ontologie globale partagée nait d'un consensus mondial.



As of March 2009



# 2014, 570 Endpoints SPARQL dans le LOD



[http://lod-cloud.net/versions/2014-08-30/lod-cloud\\_colored.png](http://lod-cloud.net/versions/2014-08-30/lod-cloud_colored.png)

# Conclusion personnelle

53.65% (279/520) ne fonctionne pas correctement en réalité (<http://sparqls.okfn.org/availability>)



La QoS est indissociable de la mise à disposition des données sinon autant mettre à disposition des fichiers. De nombreux projets H2020 de l'UE sont créés pour résoudre ce problème afin d'industrialiser ces technologies.

L'ontologie globale de Wikidata est en concurrence avec Schema.org de Google. Qui va l'emporter ?

On ne sait qu'une chose... Quelque soit l'ontologie que vous utiliserez dans l'avenir, elle ne cessera pas d'évoluer et il faut le prendre en compte dans vos systèmes.

L'architecture d'hébergement et de manipulation des données évoluent très très rapidement. Il faut intégrer les technologies du Cloud pour en faciliter la gestion pour mettre à disposition ses technologies plus simplement.

# Buzzwords du moment... SmartCities, IOT, etc...

Ce qui restera : le Web des objets alimentent le LOD



Prototypage rapide d'un client qui diffusera des données en temps réel via Wifi sans aucun intermédiaire.

# 2

**Plateforme CDS de Paris-Saclay**  
**<https://io.datascience-paris-saclay.fr>**

Plateforme "Data as a Service"  
du Center for Data Science

## Objectifs de l'action Data de CDS

**Phase 1 : Référencer les données** sur le plateau de Paris-Saclay :

- Données ouvertes ou privées
- Données partagées entre chercheurs du plateau

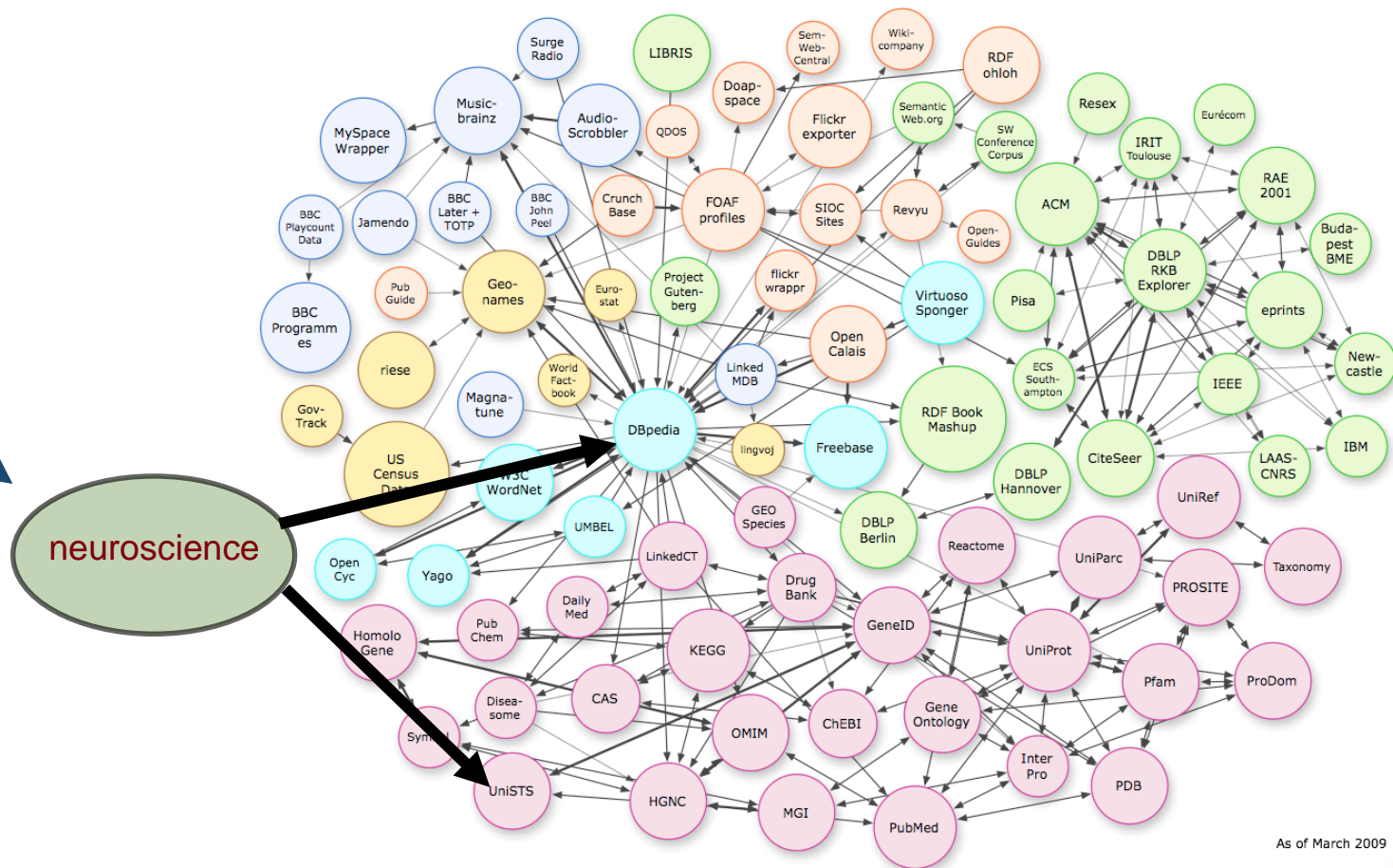
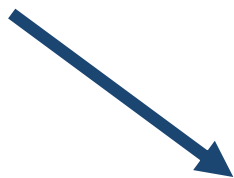
**Phase 2 : Mettre à disposition** des laboratoires sur le plateau les moyens de déployer eux-mêmes des données.

- Dans leur infrastructure
- Ou dans une infrastructure mutualisée Cloud pour les expérimentations (de technologies, d'interdisciplinarité...)

**Phase 3 : Contribuer à la réflexion et l'expérimentation** de mise à dispositions des technologies du Web afin d'accélérer la recherche.



# Objectifs en 1 image : mieux diffuser



As of March 2009

# Phase 1 : Référencer les données

Atelier en janvier pour enregistrer vos données  
et manipuler les données déjà disponible sur le plateau de Saclay.

<https://io.datascience-paris-saclay.fr>

ALPHA Version 0.1

Paris-Saclay  
Center for Data Science

DATA ▾ DOCS ▾ APP ▾

Log in Register

Search a dataset... very soon

Search Advanced

## Search an Open Dataset at Paris-Saclay

Locate on the map the actual open datasets.

Map Graph

test

computar science

Linked data

test test

Download ▾ Endpoint ▾ Examples ▾

Leaflet | © Saclay Carte Ouverte © OpenStreetMap contributors

- Dashboard
- Users **1**
- Docs
- Settings
- Dataset
  - Add
  - List

# Add a new dataset

Dataset / Add

Access level  Open  Private  Share only with users of CDS.

Title

Domain

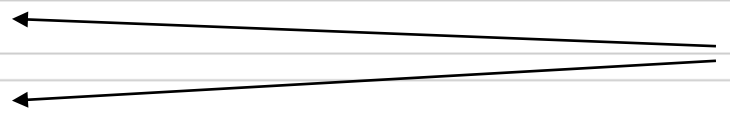
Tag

Description

Website

Email

Location Latitude  Longitude



# Phase 1 : Référencer les données

## En janvier, déclaration des API

4 ★ Title

Physics

Keyword1 Keyword2 Keyword3

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt

Home Envelope Download Endpoint Examples Refresh


Last check : 2014-08-21 04:46:05 sparql-results+xml

Digit Lab ✘ application/sparql-results+json

# Phase 2: déployer vos données

## Choisir vos logiciels

### 1 Contrôle qualité des logiciels de stockage RDF



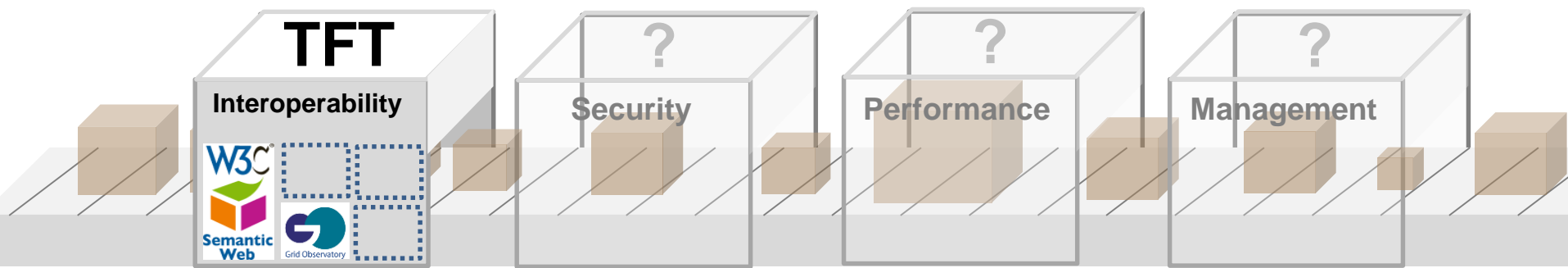
**QUALITY CONTROL**  
**DATABASE SYSTEM**

<http://SPARQLSCORE.com>

THIS SOFTWARE SCORES **439** OUT OF 459 POINTS

Grid Observatory 3.0	5/6	SPARQL L1 Entailment Regimes	11/13
10 Subtract date	Partial	Literal with language tag test	Pass

Each day, we test the last build of Fuseki, Marmotta, Virtuoso, 4store...



# Phase 2: déployer vos données

## Automatiser le déploiement de vos données

Sélection des données qui ont vocation à être partagées via le Web

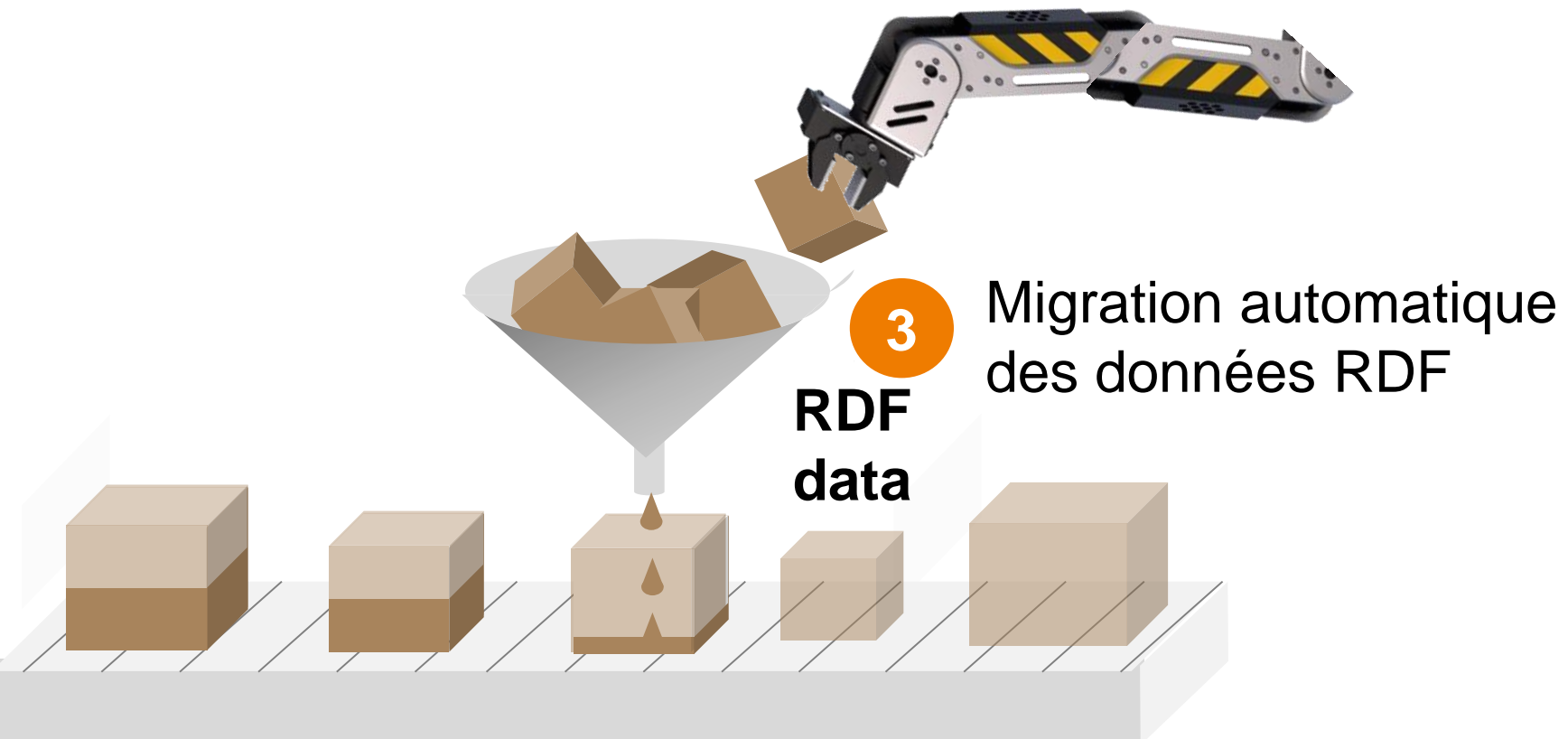


2

Sélection de la BDD RDF en fonction de vos critères

# Phase 2: déployer vos données

Automatiser la migration des données RDF



# Phase 2: déployer vos données

## Automatiser le déploiement des données RDF



5

4

Simplification de travailler avec les données d'autres laboratoires

Déploiement des données dans le Linked Data Science



# Phase 2: déployer vos données

## Enregistrer les données

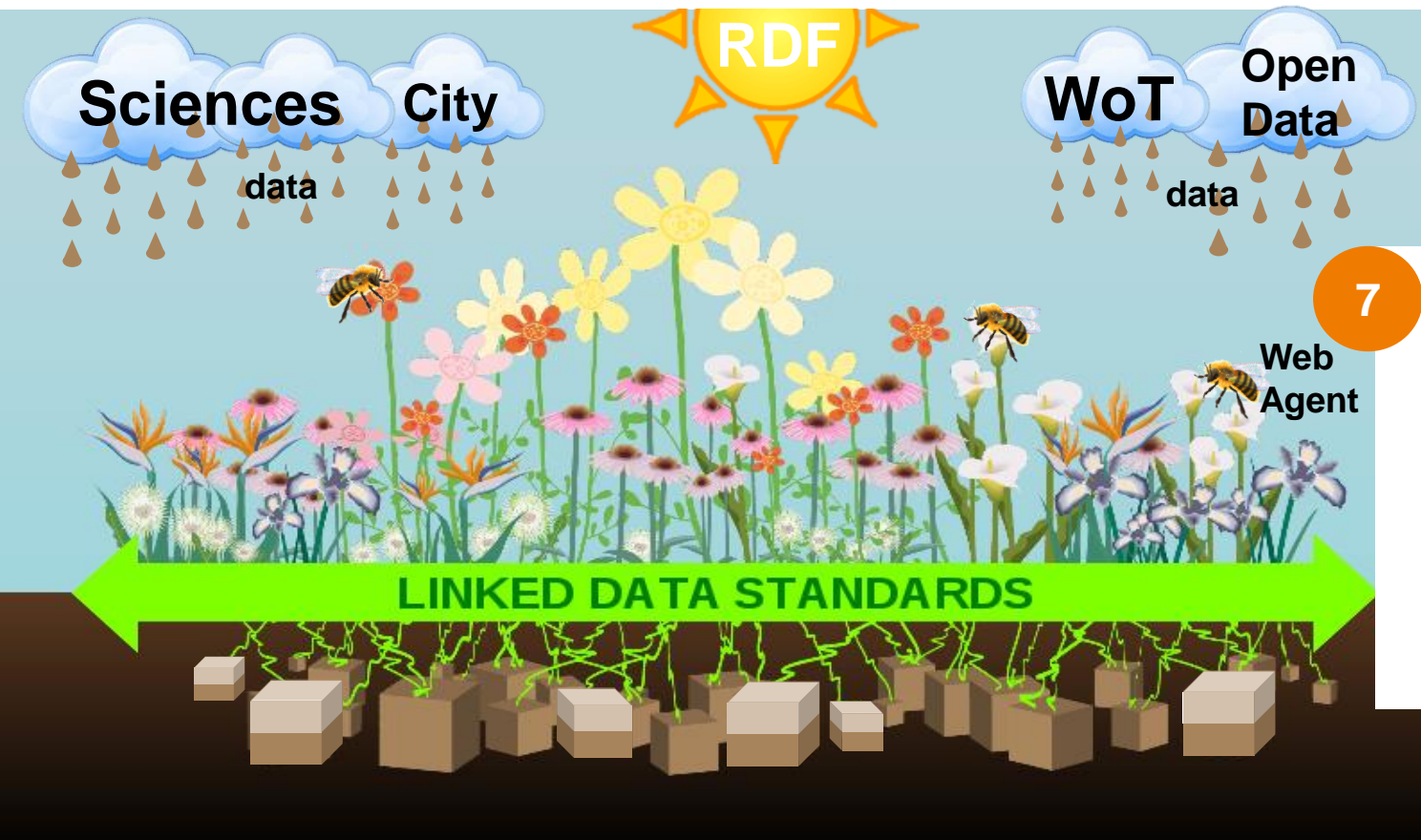
Permettre l'enregistrement de données en temps réel.



08/10/2014

# Phase 3: déployer vos données dans le Cloud

## Réutiliser les outils du Web



Faciliter l'intégration de Web Agent pour aider les chercheurs

08/10/2014

# 3

## Les autres approches

Sous-titre facultatif

# Open Data Interoperability Plateform (ODIP) de l'Union Européenne

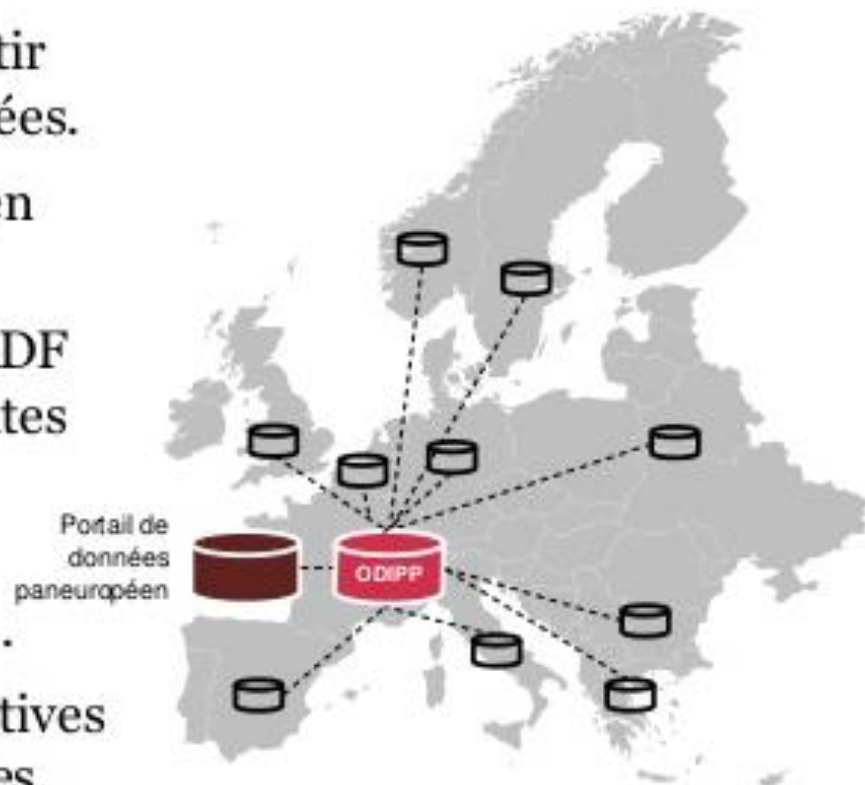
Afin d'améliorer la *visibilité* et de faciliter l'*accès* vers les données publiées sur des portails locaux et nationaux de données ouvertes afin d'accroître leur *réutilisation* à l'intérieur et au delà des frontières.

Fournissant un *accès homogénéisé* à des descriptions de métadonnées d'ensembles de données ouverts via un *point d'accès unique*



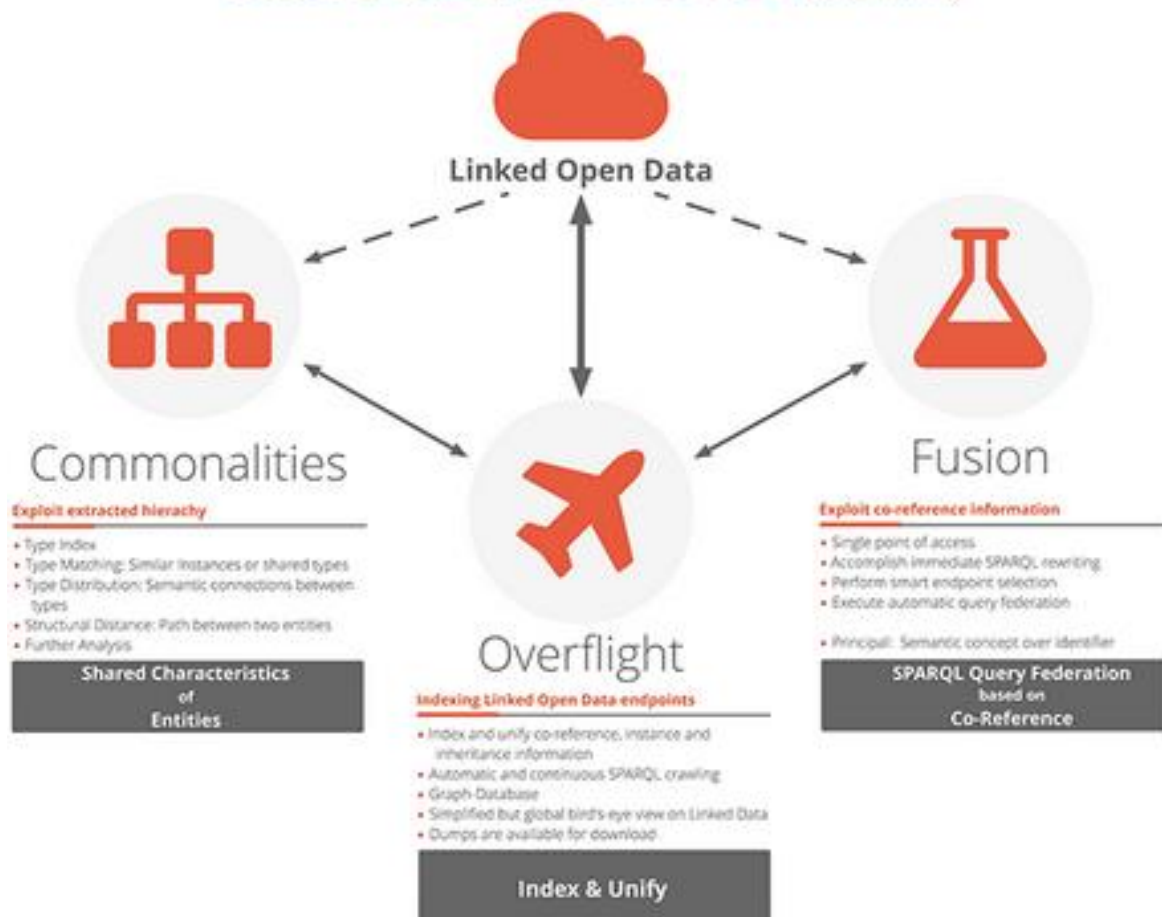
# Open Data Interoperability Platform (ODIP) de l'Union Européenne

- **Récolter** des métadonnées à partir d'un portail d'ouverture des données.
- **Transformer** les métadonnées en RDF.
- **Harmoniser** les métadonnées RDF produite dans les étapes précédentes avec « DCAT-AP ».
- **Valider** les métadonnées harmonisées contre « DCAT-AP ».
- **Publier** les métadonnées descriptives comme métadonnées ouvertes liées.



# Balloon : Big Data du Linked Data...

## Linked Data as a Service (LDaaS)



<http://schlegel.github.io/balloon/about.html>

# IO.CDS s'inscrit dans le mouvement du Web sémantique

IO.CDS veut :

- Mettre en œuvre une **plateforme interopérable** des données entre les chercheurs afin d'accélérer la recherche
- **Simplifier l'hébergement** des données en facilitant l'accès à des données dispersées
- Permettre de **choisir le niveau d'accès** à ses données
- **Réfléchir** à l'Open Data mais aussi aux nouveaux business modèles qu'offriront une **consommation des données** qui va se généraliser dans tous les secteurs d'activités.

## En bref.

Action Data CDS	Les autres
Collaborative	Centralisatrice
Respectueuse	Intrusive
Volontaire	Automatisée (crawling)
Niveaux d'accès	Tout gratuit/payant



**Merci**  
**Des questions ?**

*Inria*  
INVENTEURS DU MONDE NUMÉRIQUE

# Sources

Promouvoir la réutilisation des données publiques ouvertes à travers l'Open Data Interoperability Platform (ODIP)

<http://fr.slideshare.net/OpenDataSupport/d211-training-module-15-promoting-the-reuse-of-open-data-through-odip-v012-fr>

Histropedia: The Timeline of Everything

<http://www.histropedia.com/works.html>

SPARQL Protocol and RDF Query Language

[https://fr.wikiversity.org/wiki/SPARQL\\_Protocol\\_and\\_RDF\\_Query\\_Language](https://fr.wikiversity.org/wiki/SPARQL_Protocol_and_RDF_Query_Language)