

# Introduction Web Sémantique

Ghislain Atemezing, PhD

# Agenda

- Histoire du Web Sémantique
- RDF - cadre descriptif de ressources
- Linked Data - Données liées en RDF
- Formalisation des connaissances

# Histoire du Web sémantique

**...Une histoire du Web il y a 36 ans**

# What is the evolution of the internet to 2020?

INCREASING KNOWLEDGE CONNECTIVITY & REASONING



SOURCE: NOVA SPIVAK, RADAR NETWORKS; JOHN BRESLIN, DERZ; & MILL & DAVIS, PROJECTION

# Historique de la mise en réseau

## ~1970 : INTERNET

- **Machines** identifiées, connectées et adressables
- Réseau matériel, logistique de l'adressage

## ~1990 : WEB 1.0

- **Documents** identifiés, connectés et adressables
- Réseau logiciel, protocoles d'échanges
- Le Web est né au CERN (1989) et le W3C (1991)

## ~ 2001 : WEB sémantique

- **Choses** (entités , concepts) identifiées, connectées et adressables
- Réseau conceptuel, architecture logique
- Autres noms : Web 3.0, Web of Data, Knowledge Graph, Web of Things ...

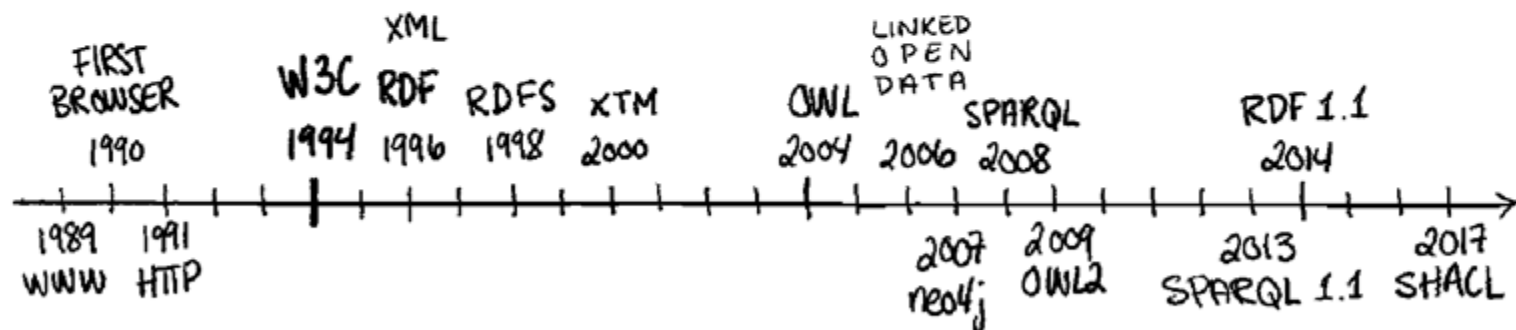


Tim Berners-Lee

Vint Cerf

Once upon a time...

## TIMELINE OF GRAPH ON THE WEB



# Le Web devient “chose centrique”

- **1996** : Le Wiki, une technologie sémantique avant la lettre
  - Un nom = une URI = un sujet de conversation
  - Technologie simple et robuste
- **2001** : Wikipedia popularise la technologie Wiki
  - Nombreuses améliorations d'interface, mais en gardant les principes
  - Les URI Wikipedia deviennent des « références de sujet »
  - Liens interwiki = aspects du sujet dans différentes langues
- **2006** : DBpedia devient le nœud central du Linked Open Data
  - Reprise du contenu structuré de Wikipedia en RDF
  - Aujourd'hui nombreuses versions multilingues interconnectées
- **2011** : Google lance le Knowledge Graph
  - Sources principales : Freebase ... et Wikipedia ...
- **2012**: Lancement de Wikidata
  - Données structurées sur du Wiki

## VISION INITIALE, UN PEU SIMPLISTE

- Les machines du réseau sont géo-localisées
  - Inutile de savoir où, l'adresse IP suffit
- Les documents sont stockés sur des machines
  - Inutile de savoir sur lesquelles, l'URL suffit
- Les entités sont décrites dans des documents
  - Inutile de savoir dans lesquels, l'URI suffit
- Un URI identifie une **ressource** ...
  - ... décrite quelque part dans un « document »
  - ... stocké quelque part sur une machine.



# Vers un Web plus intelligent, et ouvert

- 2001 : Article fondateur du Web Sémantique
  - Tim Berners-Lee et Cie proposent une vision des agent Web intelligents
- En ajoutant des annotations sur les pages existantes
  - Afin que les machines puissent les interpréter
  - Afin que nous puissions mieux interagir avec les machines
- Un réseau de connaissance est ainsi crée en liant différentes parties
  - Suivre les liens .. Tout simplement
- Hendler Hypothesis: « **A Little Semantics Goes a Long Way** »
- Le CERN rend le Web accessible à tous en **1993**

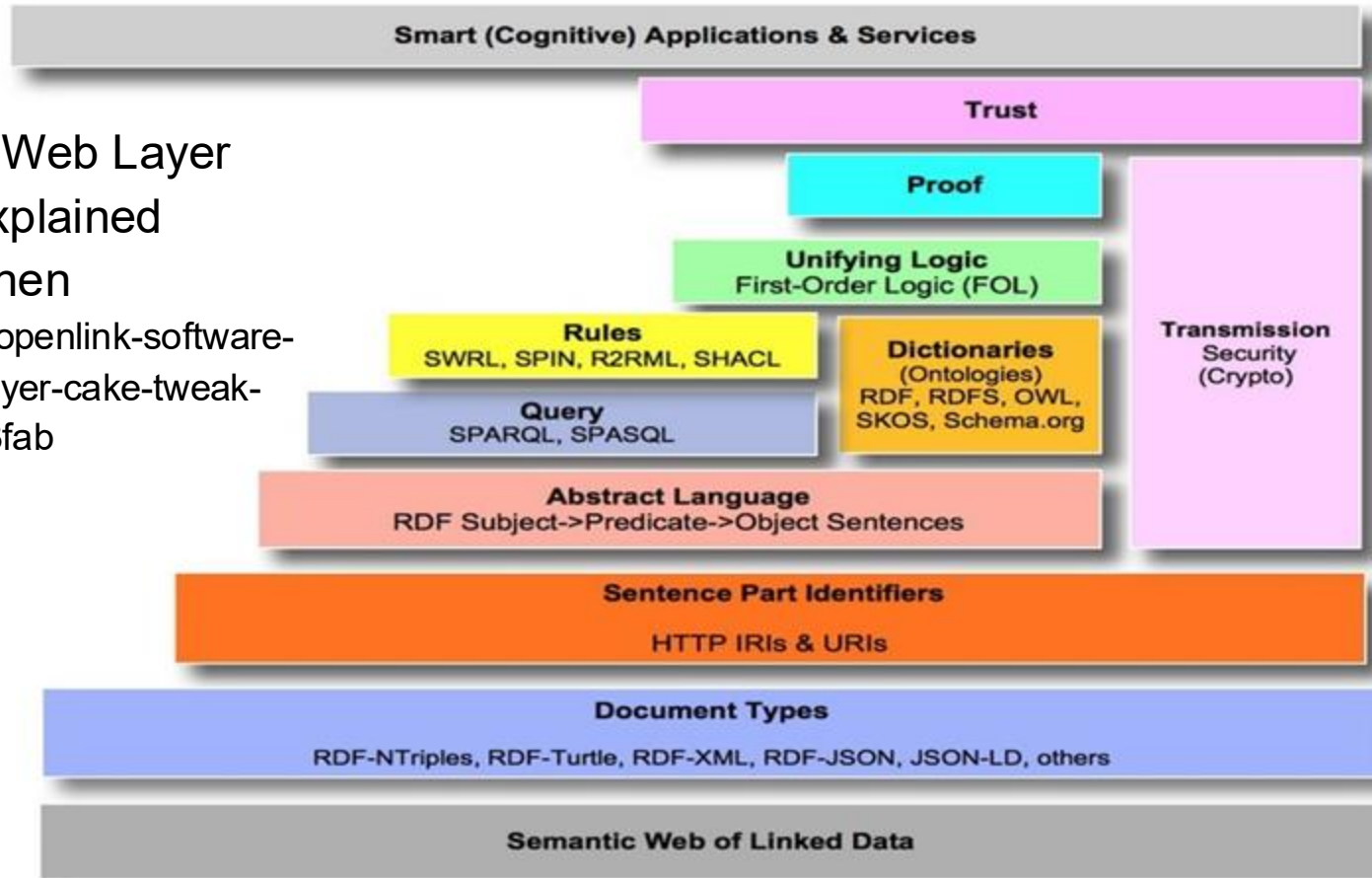


# Sur le Web...

- Une ressource est ce qui est identifié ou **signifié** par un URI
- Un URI est un référent indépendant du contexte d'utilisation
  - En tout cas c'est l'accord implicite entre tous les utilisateurs de l'URI
- Les « schémas d'URI » peuvent être associés à des protocoles
  - Un protocole est un contexte d'utilisation par défaut (http: mailto: ...)
  - Un protocole décrit les fonctionnalités associées aux URI
- En particulier : le protocole HTTP
  - Permet d'échanger des représentations de ressources sur le Web
  - Ces représentations dépendent du contexte client-serveur
- Un URI est un signe (indépendant du temps)
- La signification est un processus, dont le résultat dépend du contexte

# Technologies Web Sémantique

Credit Semantic Web Layer  
Cake Tweak, Explained  
Kingsley Uyi Idehen  
<https://medium.com/openlink-software-blog/semantic-web-layer-cake-tweak-explained-6ba5c6ac3fab>



# URI - une syntaxe générique

scheme ":" [ "//" authority "/" ] [ path ] [ "?" query ] [ "#" fragment]

- Le **nom du schéma** est une lettre suivie de n'importe quelle combinaison de lettres, de chiffres, du signe plus (+), du point (.) ou d'un tiret (-) et se termine par deux points (:).
- La **partie hiérarchique** de l'URI est prévue pour contenir les informations d'identification de la ressource, hiérarchique par nature. En général, cette partie commence par un double slash (//), suivi par le *domaine* puis un *chemin* optionnel.
- Le **domaine** est principalement constitué du nom d'hôte ou de l'adresse IP, cette information peut être encadrée par deux autres champs optionnels : avant, les informations concernant l'utilisateur délimitées par un @ ; après, un numéro de port, précédé par un deux-points (:).
- Le **chemin** est une séquence de segments (similaires à des répertoires), séparés par des slash (/). Chaque segment peut contenir des paramètres séparés par des points virgules (;), bien que cela soit rarement utilisé.
- La **requête** est une partie optionnelle séparée par un point d'interrogation qui contient des informations complémentaires qui ne sont pas de nature hiérarchique. Le format de la requête n'est pas défini de manière générale, mais est souvent formée d'une suite de paires <def>=<valeur> séparées par des points virgules ou par des esperluettes.
- Le **fragment** est une partie optionnelle séparée par un dièse. Il fournit des informations supplémentaires permettant d'accéder à une ressource secondaire. Lorsque la ressource est un document HTML, le **fragment** est souvent l'id d'un élément et le navigateur fera le nécessaire pour que cet élément soit visible.



Internationalized Resource Identifier (RFC 3987): une extension pour supporter l'encodage Unicode

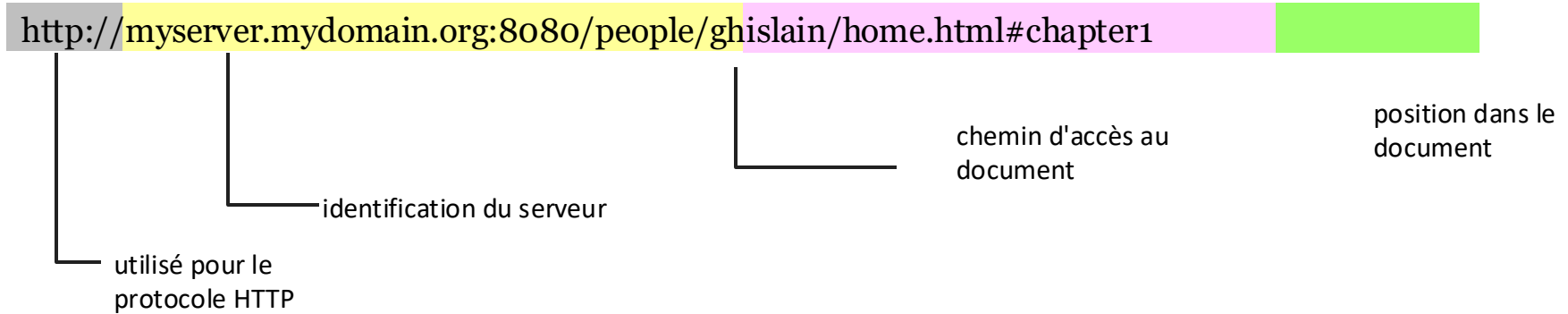
<https://hy.wikipedia.org/wiki/qnwpwpn>



## En détail

scheme ":" [ "//" authority "/" ] [ path ] [ "?" query ] [ "#" fragment ]

### Décortiquons un exemple





Vous reconnaissez ...





# Comprendre la négociation de contenu

- Par exemple, l'utilisateur essaie d'accéder à un URI qui fait référence à la Tour Eiffel elle-même avec <http://dbpedia.org/resource/Abidjan> .
- S'il le fait par une requête HTTP GET, comme la Tour Eiffel elle-même n'est pas une ressource d'information, aucune représentation Web n'est directement disponible.
- Au lieu de cela, l'utilisateur reçoit un message 303 "See Other" qui le redirige vers une description associée qui héberge des représentations Web sur la Tour Eiffel, comme notamment la ressource d'information <http://dbpedia.org/page/Abidjan> .
- Lorsque cet URI renvoie le code d'état 200 en réponse à une demande HTTP GET, l'agent en déduit que <http://dbpedia.org/page/Abidjan/> est une ressource d'information.

# Qu'est-ce qu'un bon URI?

- « Cool URIs don't change ! » - Cf. Sir TimBL
  - Persistance et stabilité de l'URI
  - Persistance de la réponse dans le protocole associée
  - Persistance du référent de la ressource
- Structure cohérente d'espaces de noms
  - Respecter les usages et bonnes pratiques
  - Eviter les URI trop compliqués!
- Les URI peuvent utiliser des fragments en langage naturel
  - C'est pratique pour les humains qui les utilisent aussi
  - Mais les humains ont beaucoup de langages naturels ...
  - Et on risque d'interpréter le nom au lieu de la description ...
  - Les « identifiants opaques » sont une solution plus simple
  - Exemple : <http://sws.geonames.org/3014258/>
- Référence : Datalift Livrable 3.1 : Politique d'identification des ressources  
<https://gforge.inria.fr/docman/view.php/2935/8286/D3.1+Politique+d%27identification+des+ressources+v1.0.pdf>





# 10 règles pour des URIs pérennes



## Follow the pattern

e.g. `http://[domain]/[type]/[concept]/[reference]`

## Re-use existing identifiers

e.g. `http://education.data.gov.uk/id/school/123456`

## Link multiple representations

e.g. `http://data.example.org/doc/foo/bar.html`

e.g. `http://data.example.org/doc/foo/bar.rdf`

## Implement 303 redirects for real-world objects

e.g. `http://www.example.com/id/alice_brown`

## Use a dedicated service

i.e. independent of the data originator

# 10 rules for persistent URIs



## Avoid stating ownership

e.g. `http://education.data.gov.uk/ministryeducation/id/school/123456`

## Avoid version numbers

e.g. `http://education.data.gov.uk/doc/school1/123456`

## Avoid using auto-increment

e.g. `http://education.data.gov.uk/id/school1/123456`

e.g. `http://education.data.gov.uk/id/school1/123457`

## Avoid query strings

e.g. `http://education.data.gov.uk/doc/school1?id=123456`

## Avoid file extensions

`http://education.data.gov.uk/doc/schools/123456.cx`

See also: 10 Rules for Persistent URIs. <https://joinup.ec.europa.eu/node/53858>



## Des URIs Persistants - 5 DO

- Suivez le modèle  
<http://{domain}/{type}/{concept}/{reference}>
- Réutilisez les identifiants existants  
<http://education.data.gov.uk/id/school/123456>
- Liez plusieurs représentations  
<http://data.example.org/doc/foo/bar.html>  
<http://data.example.org/doc/foo/bar.rdf>
- Mettez en place des redirections pour les objets du monde réel  
<http://www.example.com/id/JeanDupont>
- Utilisez un service dédié  
indépendant de l'origine des données

## Des URIs Persistents - 5 DON'T

- Évitez de déclarer la propriété  
<http://edu.gov.uk/ministryofeducation/id/school/12345>
- Évitez les numéros de version  
<http://edu.gov.uk/doc/school/v01/12345>
- Évitez d'utiliser l'auto-incrément  
<http://edu.gov.uk/id/school1/12345>  
<http://edu.gov.uk/id/school1/12346>
- Évitez les chaînes de requête  
<http://edu.gov.uk/doc/school?id=12345>
- Évitez les extensions de fichiers  
<http://edu.gov.uk/doc/schools/12345.csv>

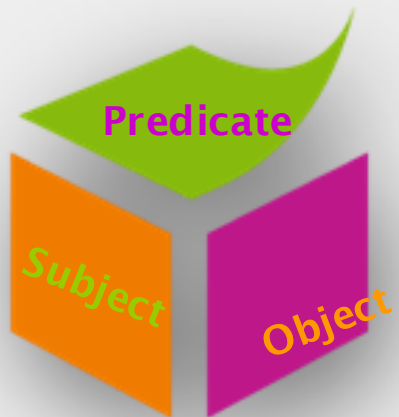
# **RDF (Resource Description Framework)**

**Le cadre formel pour  
générer les graphes de  
connaissance**



La molécule de base : le triplet

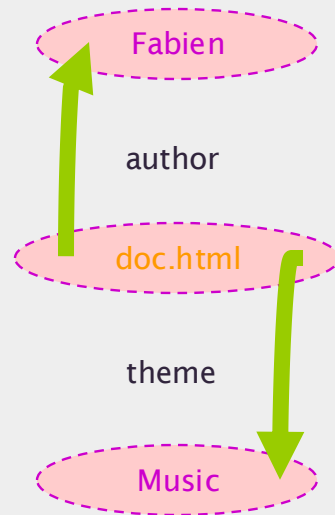
**RDF** Toute connaissance peut être ramenée à un ensemble équivalent de triplets (**sujet**, **prédictat**, **objet**).



doc.html was written by  
Fabien and has the  
theme Music



doc.html was written by Fabien  
doc.html has the theme Music

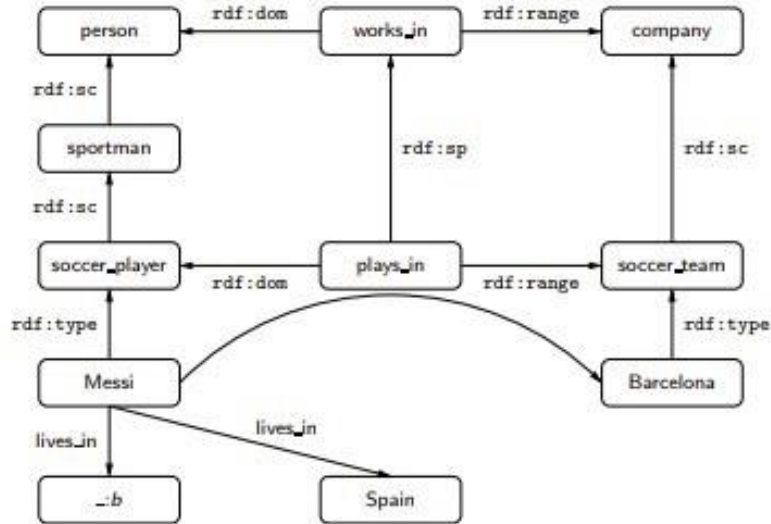
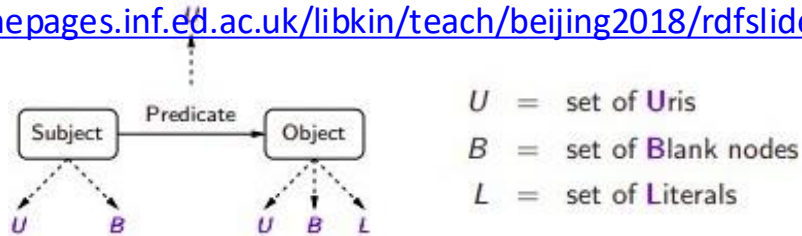


# Resource Description Framework RDF

Credit Semantic Web Data/RDF/SPARQL

<http://homepages.inf.ed.ac.uk/libkin/teach/beijing2018/rdfslides.pdf>

Triplet RDF

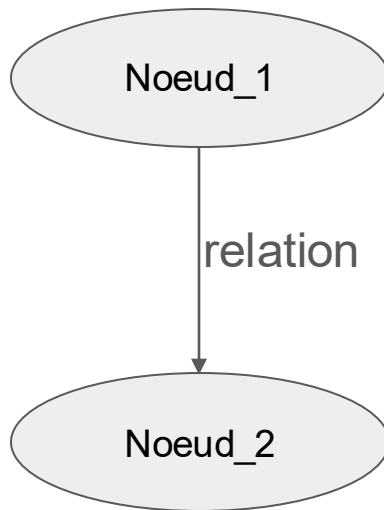




## Faisons quelques exemples

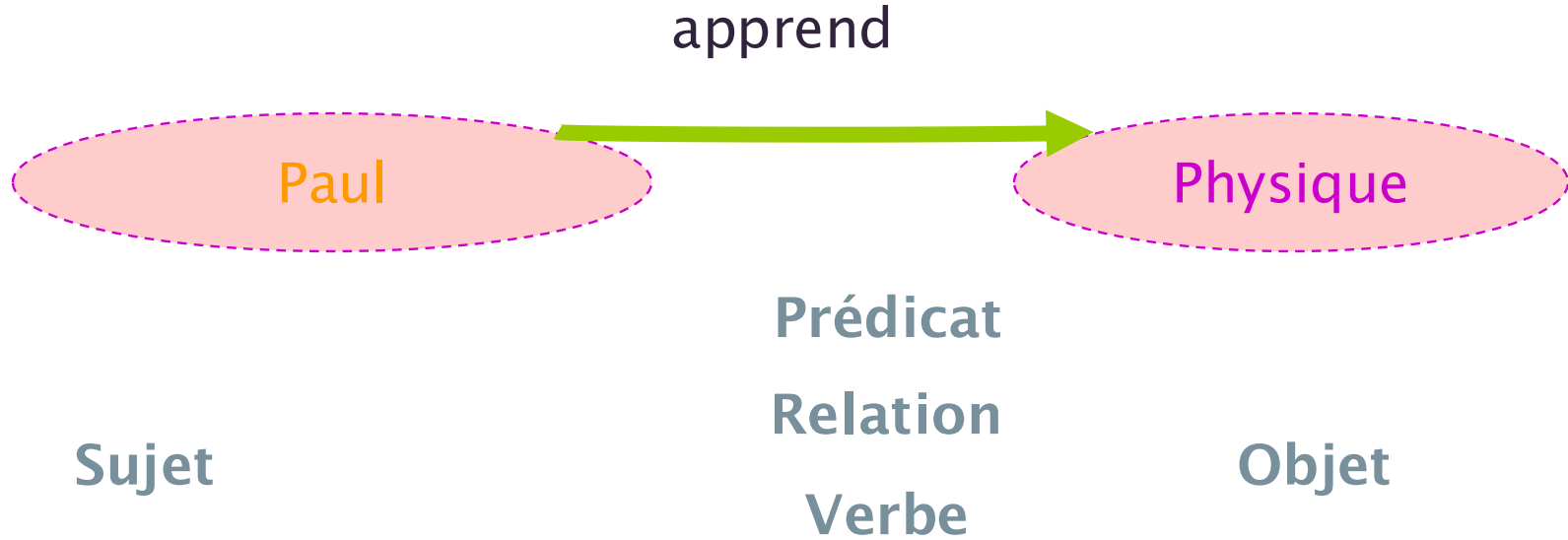
→ Formez des triplets RDF pour les phrases suivantes :

- ◆ Paul étudie économie appliquée
- ◆ Paul aime Berniss
- ◆ Paul and Berniss vivent en Côte d'ivoire
- ◆ Edouard est le fils de Katharina et Paul
- ◆ Marie est la soeur d'une amie de Paul
- ◆ Paul est né à Abidjan le 22 avril 2001



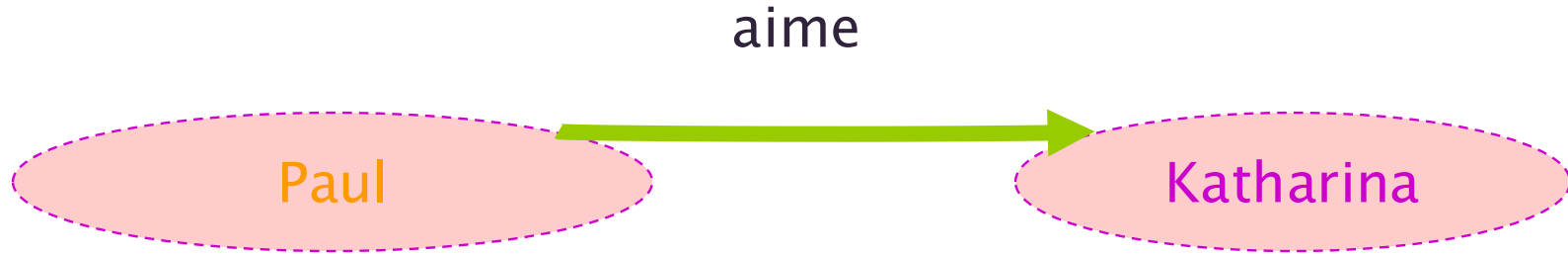


“Paul apprend la physique”





Dessinez : Paul aime Katharina

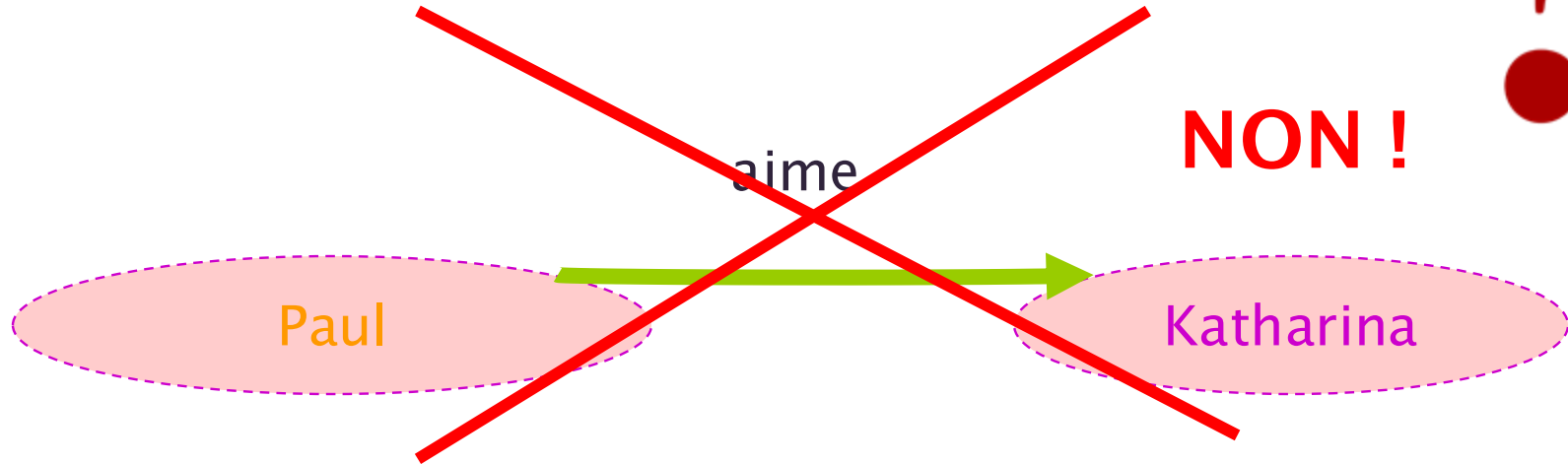


( Paul - aime - Katharina )





Dessinez : Paul aime Katharina



(Katharina - aime - Paul)

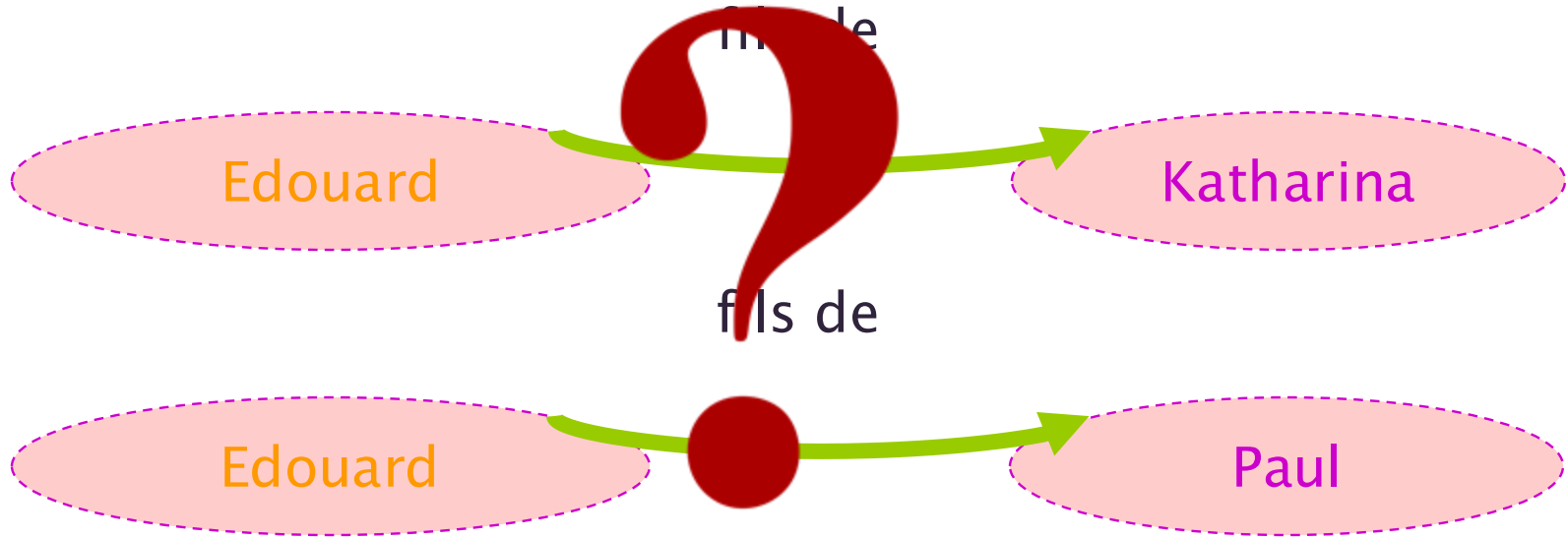


## L'hypothèse du monde ouvert

- L'hypothèse du monde fermé (CWA – Closed World Assumption) est l'hypothèse selon **laquelle ce qui n'est pas connu comme vrai doit être faux.**
- L'hypothèse du monde ouvert (OWA – Open World Assumption) est l'inverse. En d'autres termes, elle part du principe que **ce qui n'est pas connu comme étant vrai est tout simplement inconnu.**



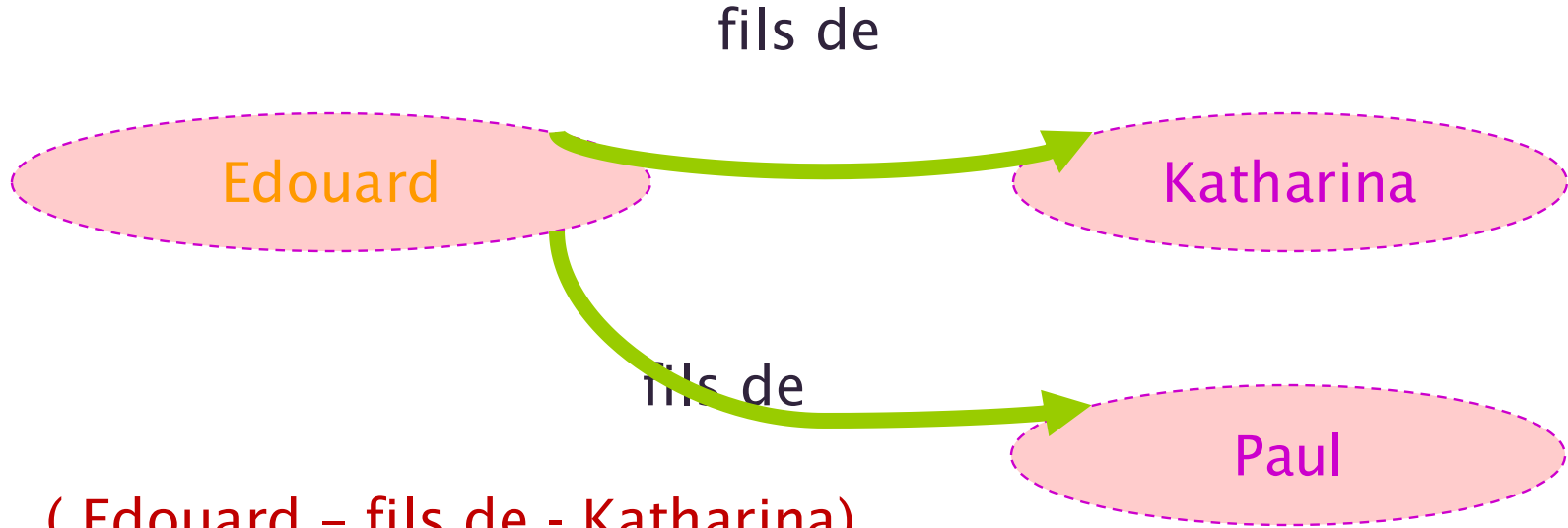
Edouard est le fils de Katharina et Paul



( Edouard – fils de - Katharina)  
( Edouard – fils de - Paul)



Edouard est le fils de Katharina et Paul



( Edouard – fils de - Katharina)  
( Edouard – fils de - Paul)  
( Edouard – fils de – Katharina, Paul)



Edouard est le fils de Katharina et Paul



( Edouard – fils de – Katharina et Paul)

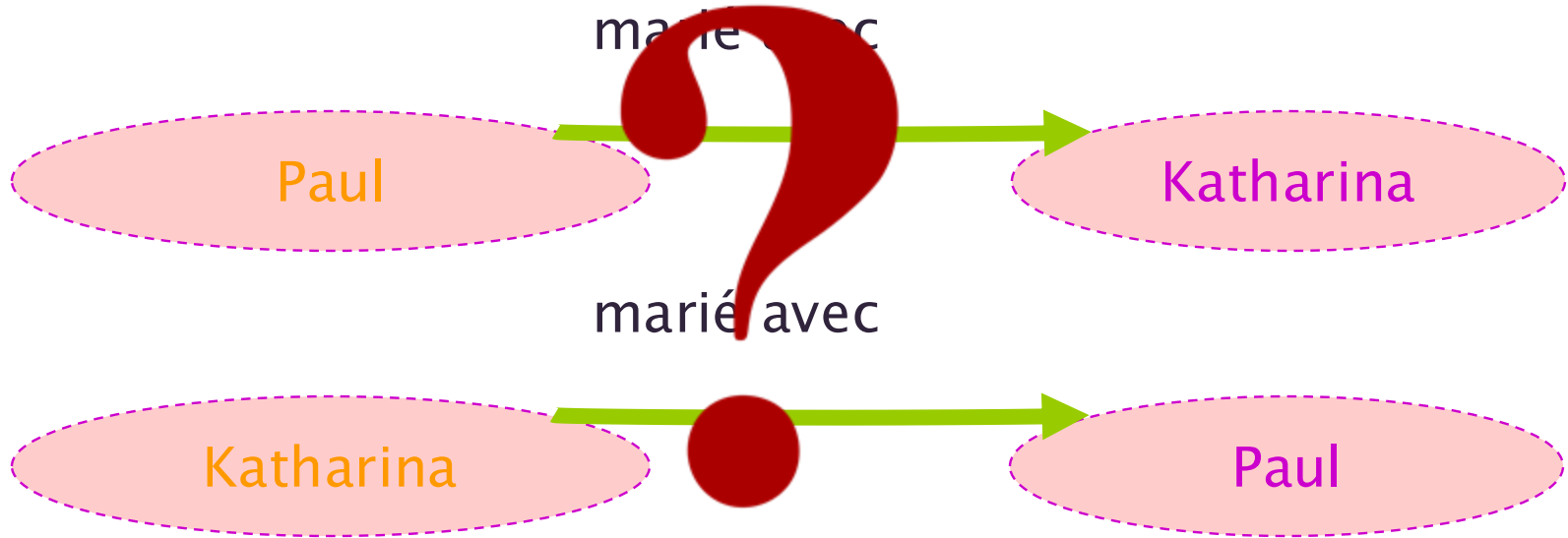


## Edouard est le fils de Katharina et Paul

- Nous ne savons pas ce que "mère de" et "père de" signifient.
- La seule chose connue est "Edouard est le fils de Katharina et Paul"
  - Sujet : Edouard
  - Verbe/Relation/Prédicat : fils de
  - Objet(s) : Katharina et Paul
- Donc, les deux prédicats "mère de" et "père de" ne sont pas définis : ils sont indéfinis, inconnus.

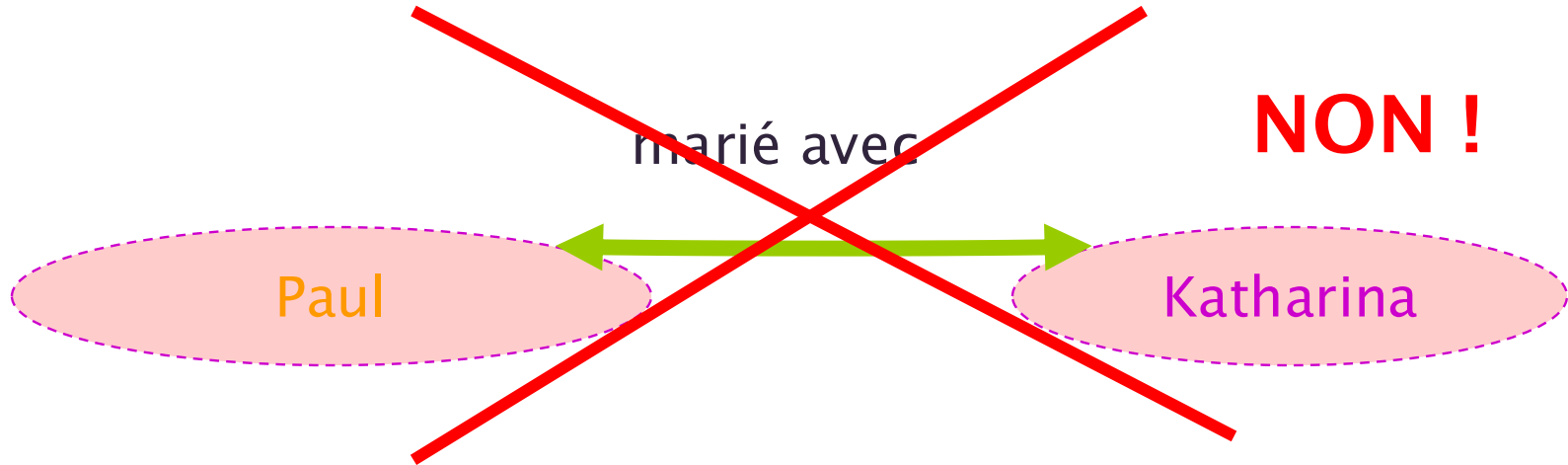


Dessinez : Paul et Katharina sont mariés





Dessinez : Paul et Katharina sont mariés



Parce que le graphe doit être orienté  
avec une source et une cible. Autrement dit,  
avec un sujet, un verbe et un objet.





Dessinez : Paul et Katharina sont mariés



( Paul – marié avec - Katharina)  
(Katharina – marié avec - Paul)



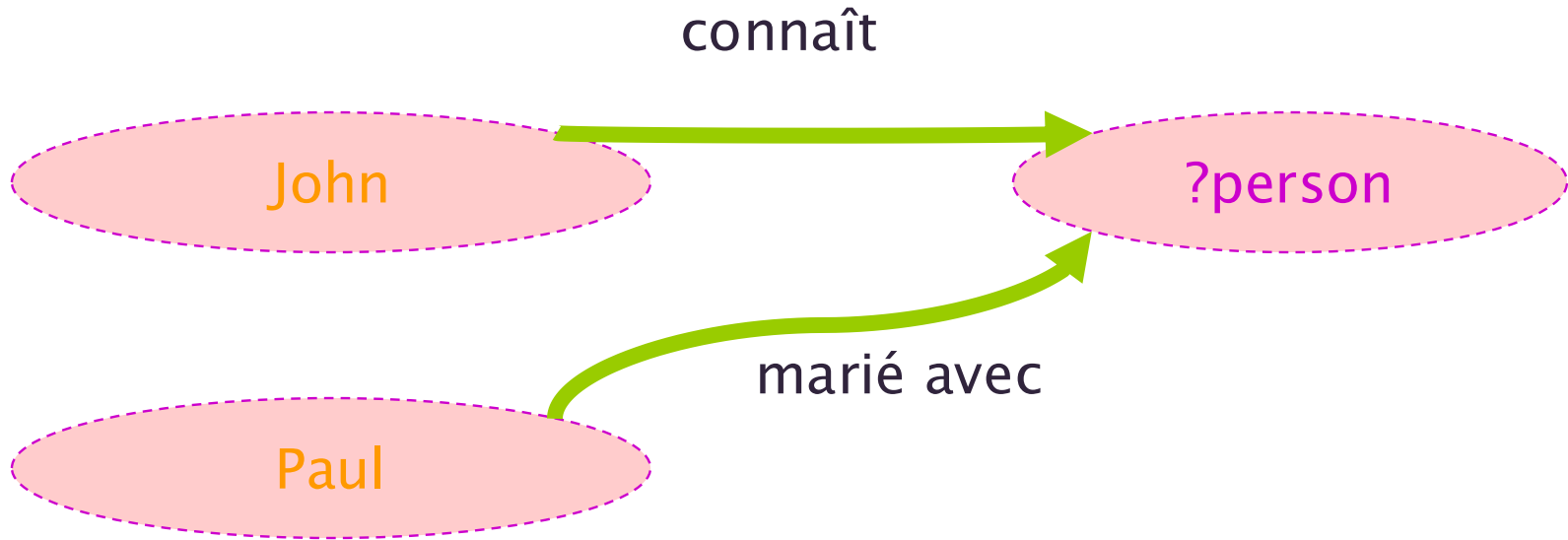
Dessinez : Paul et Katharina sont mariés



Ce sont deux relations distinctes  
puisque le sujet change

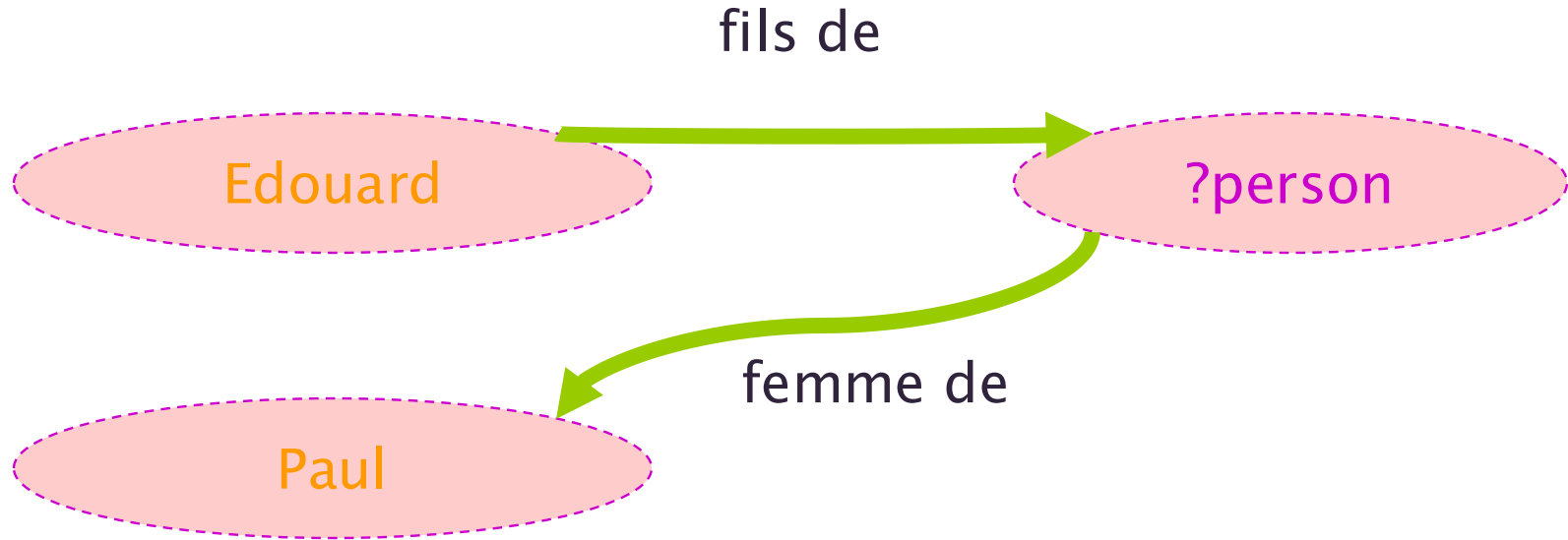


Lisez le graphe



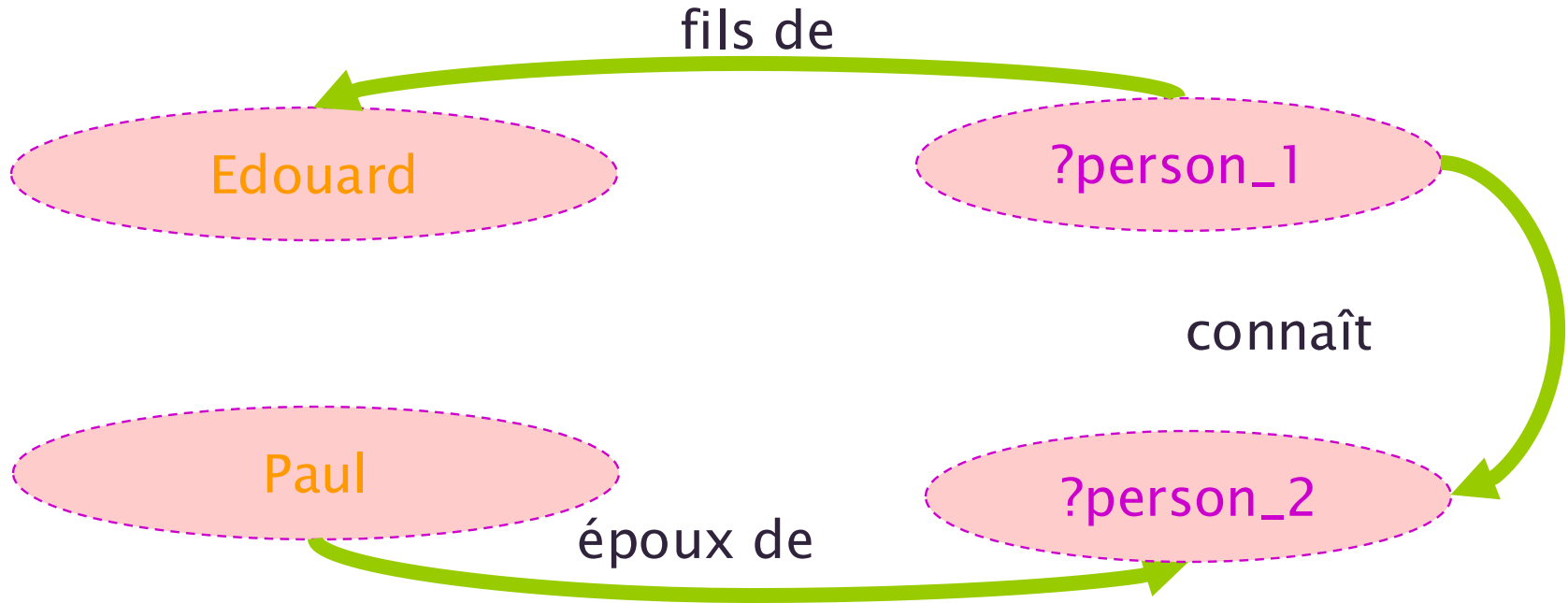


Lisez le graphe



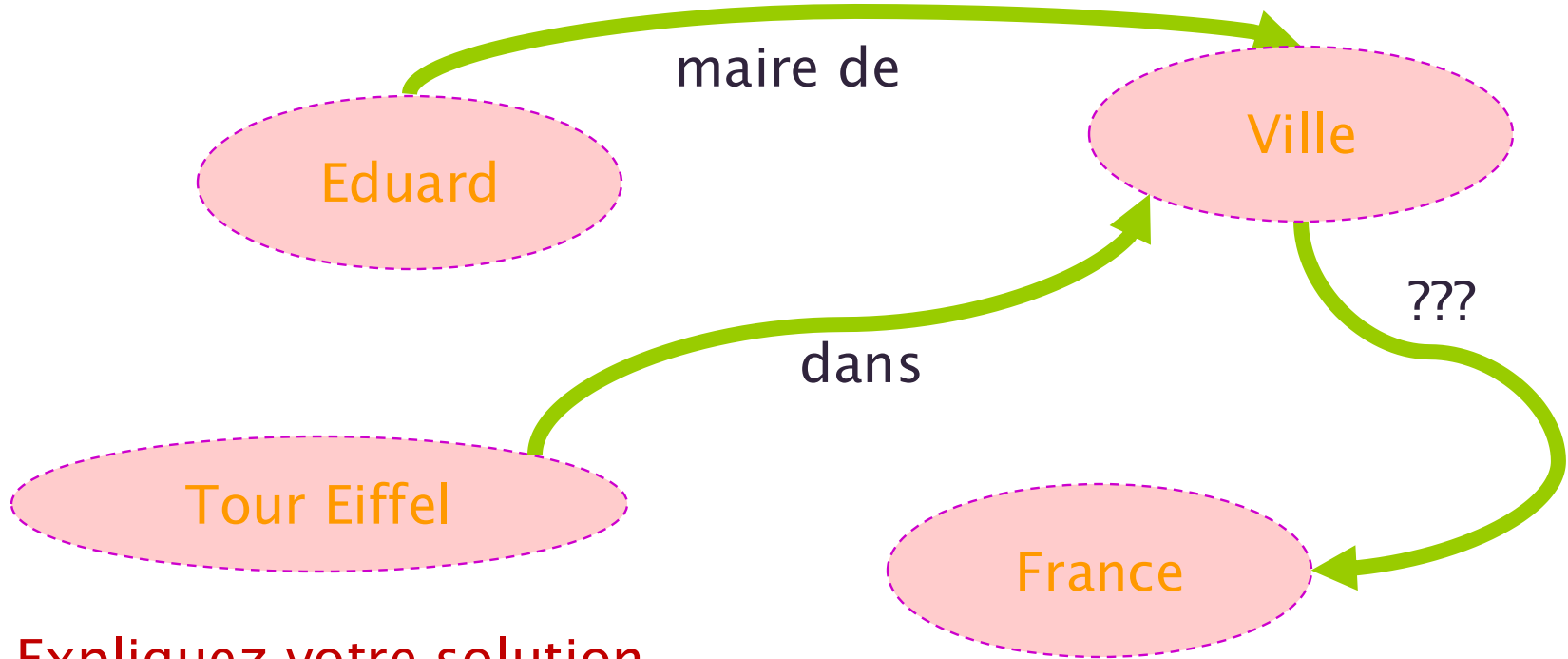


Lisez le graphe





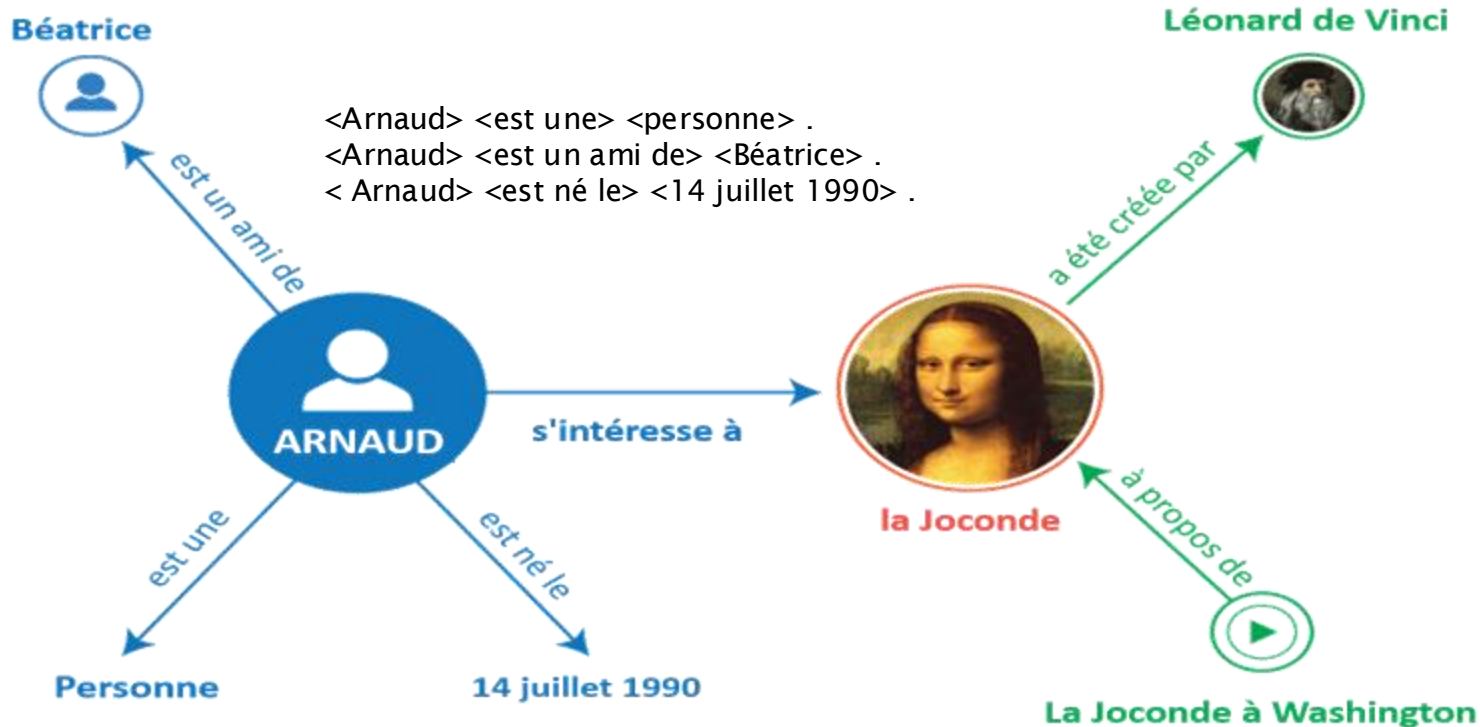
Lisez et complétez le graphe



Expliquez votre solution



# La même ressource est souvent référencée dans plusieurs triplets



<Arnaud> <s'intéresse à> <la Joconde> .  
<la Joconde> <a été créée par> <Léonard de Vinci> .  
<La vidéo 'La Joconde à Washington'> <est à propos de> <la Joconde> .

# Données FAIR



Credit FAIR in a nutshell. Image: ARDC 2018 - CC-BY 4.0.

<https://librarycarpentry.org/Top-10-FAIR/2018/12/01/aus-government-data/>





## Des formats de sérialisation pour RDF

- N-Triples [[N-TRIPLES](#)] et N-Quads [[N-QUADS](#)] (formats d'échange basés sur les lignes)
- Turtle [[TURTLE](#)] et TriG [[TRIG](#)]
- JSON-LD [[JSON-LD](#)] (basé sur JSON)
- RDFa [[RDFa-PRIMER](#)] (pour une encapsulation dans HTML)
- RDF/XML [[RDF11-XML](#)] (la syntaxe originale et historique (2004), mise à jour avec RDF 1.1)

# **Linked (Open) Data - réseau gigantesque de données en RDF**

# LINKED DATA PRINCIPLES

- **Tim Berners Lee [2006] ([Design Issues](#))**
    1. Use URIs to identify things (anything, not just documents);
    2. Use HTTP URIs – globally unique names, distributed ownership – so that people can look up those names;
    3. Provide useful information in RDF – when someone looks up a URI;
    4. Include RDF links to other URIs – to enable discovery of related information
-

# SCRAPING INFOBOX DATA

## About: [Bogotá](#)

An Entity in Data Space: [dbpedia.org](#)

Bogotá (anciennement Santa fé de Bogotá Distrito Capital) est la capitale de la Colombie et du dép

Property	Value
<a href="#">dbpedia-owl:areaLand</a>	■ 668.7
<a href="#">dbpedia-owl:areaTotal</a>	■ 1587
<a href="#">dbpedia-owl:elevation</a>	■ 2640
<a href="#">dbpedia-owl:establishedTitle</a>	■ Foundation
<a href="#">dbpedia-owl:foundingDate</a>	■ 1538-08-06 (xsd:date)
<a href="#">dbpedia-owl:foundingPerson</a>	■ <a href="#">dbpedia:Gonzalo_Jim%C3%A9nez_de_Quesada</a>
<a href="#">dbpedia-owl:leaderName</a>	■ <a href="#">dbpedia:Alternative_Democratic_Pole</a> ■ <a href="#">dbpedia:Samuel_Moreno_Rojas</a>
<a href="#">dbpedia-owl:leaderTitle</a>	■ Mayor
<a href="#">dbpedia-owl:motto</a>	■ Bogotá, 2600 meters closer to the stars ■ Bogotá, 2600 metros más cerca de las estrellas
<a href="#">dbpedia-owl:populationAsOf</a>	■ 2005
<a href="#">dbpedia-owl:populationDensity</a>	■ 22593 ■ 11071
<a href="#">dbpedia-owl:populationMetro</a>	■ 7881156
<a href="#">dbpedia-owl:populationTotal</a>	■ 6776009

<http://dbpedia.org/resource/Bogotá>

Localities (*localidades*) of Bogotá  
Coordinates:  4°35'53"N 74°4'33"W

<b>Country</b>	Colombia
<b>Department</b>	Cundinamarca
<b>Foundation</b>	August 6, 1538
<b>Founder</b>	Gonzalo Jiménez de Quesada
<b>Government</b>	
<span> </span> - <b>Mayor</b>	Samuel Moreno Rojas (PDA)
<b>Area</b>	
<span> </span> - <b>City</b>	1,587 km <sup>2</sup> (612.7 sq mi)
<span> </span> - <b>Land</b>	1,731.9 km <sup>2</sup> (668.7 sq mi)
<b>Elevation</b> <sup>[1]</sup>	2,640 m (8,661 ft)
<b>Population</b> (2005 census) <sup>[2]</sup>	
<span> </span> - <b>City</b>	6,776,009 Ranked 1 <sup>st</sup>
<span> </span> - <b>Density</b>	22,593/km <sup>2</sup> (11,071/sq mi)
<span> </span> - <b>Metro</b>	7,881,156
<b>HDI</b> (2006)	0.880 — high <sup>[3]</sup>
<b>Website</b>	<a href="#">City Official Site</a>  <a href="#">Mayor Official Site</a>  <a href="#">Bogotá Tourism</a> 

# THING VS DESCRIPTION

- **Eiffel Tower described in DBpedia:**

- [http://dbpedia.org/resource/Eiffel\\_Tower](http://dbpedia.org/resource/Eiffel_Tower) (*thing*)
- [http://dbpedia.org/page/Eiffel\\_Tower](http://dbpedia.org/page/Eiffel_Tower) (*HTML page*)
- [http://dbpedia.org/data/Eiffel\\_Tower](http://dbpedia.org/data/Eiffel_Tower) (*RDF description*)

- **Handle by content negotiation (defined in HTTP)**

GET /resource/Eiffel\_Tower HTTP/1.1 Host:

dbpedia.org

Accept: text/html;q=0.5, application/rdf+xml

HTTP/1.1 303 See Other

Location: [http://dbpedia.org/data/Eiffel\\_Tower](http://dbpedia.org/data/Eiffel_Tower) Vary: Accept

---

# AUTOMATIC LINKS AMONG OPEN DATASETS

<http://dbpedia.org/resource/Bogotá>

owl:sameAs <http://sws.geonames.org/3688689/> owl:sameAs

<http://rdf.freebase.com/ns/guid.9202a8c04000641f8000000000167bab>

dbpedia:population "6776009"

...

DBpedia



<http://sws.geonames.org/3688689/>

owl:sameAs <http://dbpedia.org/resource/Bogotá>

wgs84\_pos:lat "4.6"

wgs84\_pos:long "-74.0833333"

geo:population "7102602"

...

Geonames



# WIKIDATA – DONNÉES STRUCTURÉES

- Base de connaissance libre maintenue par Wikimedia et éditée par des bénévoles (depuis 2012). Lien: <https://www.wikidata.org>
- Taille des triplets: > 5,8 Milliards (Juin 2018)
- Liens vers plus de 30 bases RDF dans le LOD Cloud.

The image shows a Wikidata profile for Douglas Adams (Q42) with various annotations explaining the structure of the data. The annotations are as follows:

- libellé**: Points to the name "Douglas Adams (Q42)".
- description**: Points to the description "écrivain anglais de science-fiction" and the aliases "Douglas Noel Adams | Douglas Noel Adams".
- propriété**: Points to the "scolarité" property.
- rang**: Points to the "Déclarations" section.
- groupe de déclarations**: Points to the "scolarité" property group.
- identifiant de l'élément**: Points to the "(Q42)" part of the name.
- alias**: Points to the aliases "Douglas Noel Adams | Douglas Noel Adams".
- valeur**: Points to the "St John's College" value.
- qualificatifs**: Points to the "date de fin", "principale spécialité d'étude", "diplôme", and "date de début" qualifiers.
- références dépliées**: Points to the "2 références" section.
- références repliées**: Points to the "0 références" section.

The profile itself contains the following information:

**Douglas Adams (Q42)**

écrivain anglais de science-fiction  
Douglas Noel Adams | Douglas Noel Adams  
► Plus de langues

**Déclarations**

**scolarité**

**St John's College**

date de fin: 1974  
principale spécialité d'étude: littérature anglaise  
diplôme: baccalauréat en arts  
date de début: 1971

▼ 2 références

affirmé dans: Encyclopedia Britannica Online  
URL de la référence: <http://www.oxford.com/people/731/000023662/>  
langue originale: anglais  
date de consultation: 7 décembre 2013  
éditeur: NND®  
titre: Douglas Adams (anglais)

► ajouter une référence

**Brentwood School**

date de fin: 1970  
date de début: 1959

► 0 références

► ajouter

# SAMEAS.ORG – LIENS ENTRE URI

## <sameAs> interlinking the Web of Data

The Web of Data has many equivalent URIs.  
This service helps you to find co-references  
between different data sets.

Enter a known URI, or use Sindice to search first.

<sameAs>



Equivalent URIs for <http://dbpedia.org/resource/Bogotá> –

<sameAs>



1. <http://dbpedia.org/resource/Bogotá>
2. <http://mpii.de/yago/resource/Bogotá>
3. <http://rdf.freebase.com/ns/guid.9202a8c04000641f8000000000167bab>
4. <http://sws.geonames.org/3688689/>
5. <http://telegraphis.net/data/capitals/CO/Bogotá#Bogotá>
6. <http://umbel.org/umbel/ne/wikipedia/Bogotá>

[rdf+xml](#) · [n3](#) · [json](#) · [text](#) · [show fewer items](#)



# Linked Open Data Cloud

- ..Depuis Mai 2007 - <http://lod-cloud.net>
  - Jeux de données sur le Web- Lien: <https://lod-cloud.net/datasets>
  - Graphes de données interconnectés
  - Rendu possible par la notion de **Linked Open Data** (données ouvertes et liées)
- Tout jeu de données possède
  - Plus de **1000 triplets**
  - Plus de 50 liens vers d'autres jeux du graphe
  - Toutes les données sont accessibles par des URIs
  - L'accès aux données peuvent se faire via un crawling RDF, un dump RDF ou un endpoint SPARQL
- Données en mars 2019 vs Mai 2020:
  - 1239 jeux de données / 1260 jeux de données
  - 16 147 liens / 16 187 liens
  - 9 sous-domaines (Géographie, Media, Science, publications, réseaux sociaux, etc.)

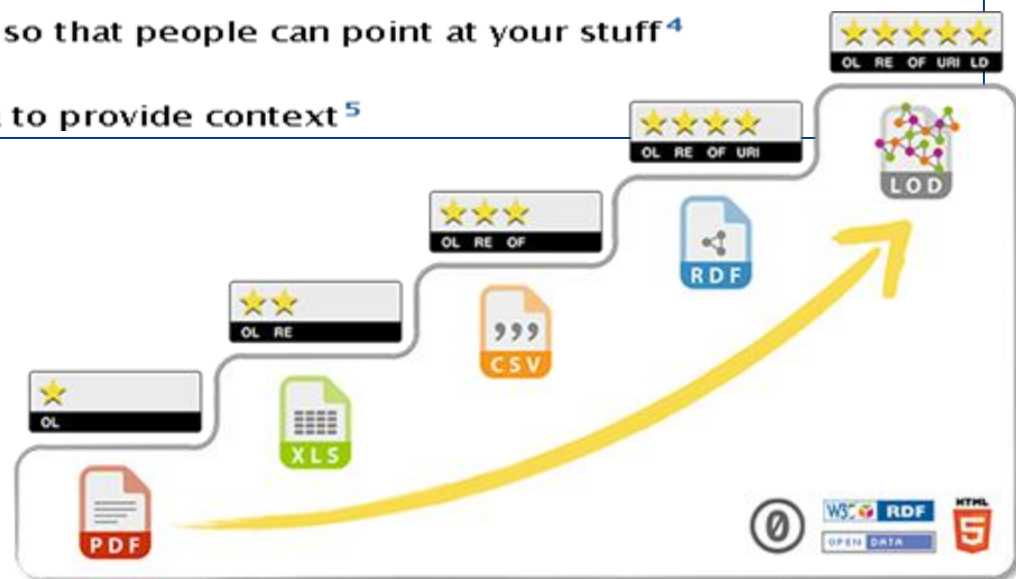


# Modèle 5 étoiles

Credit Tim Berners-Lee,

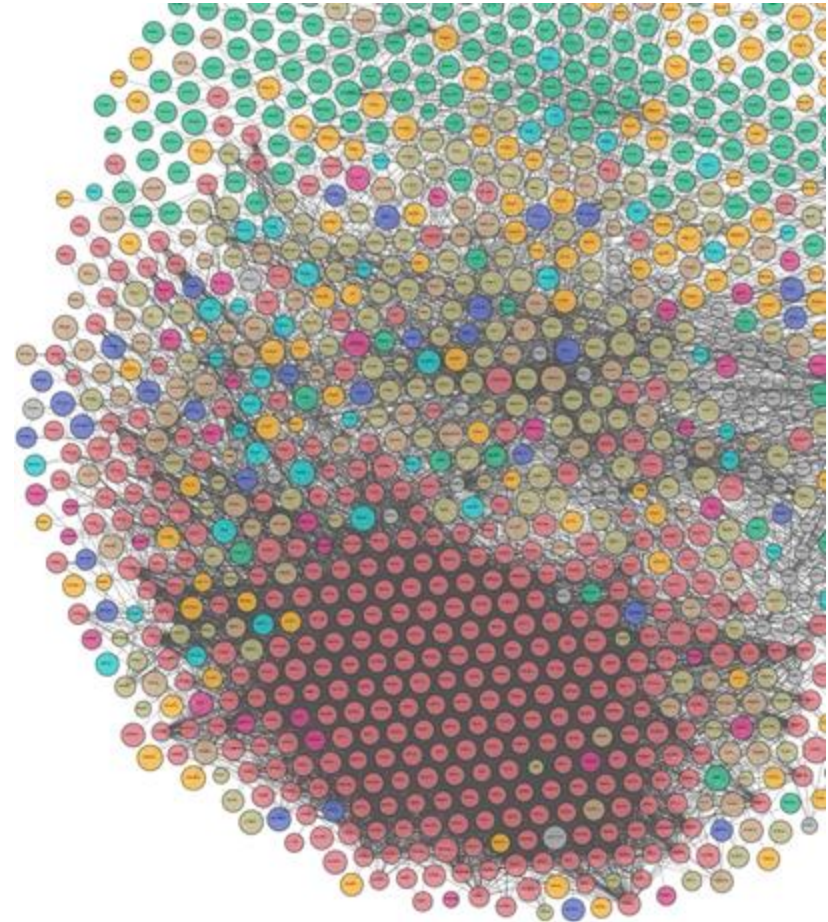
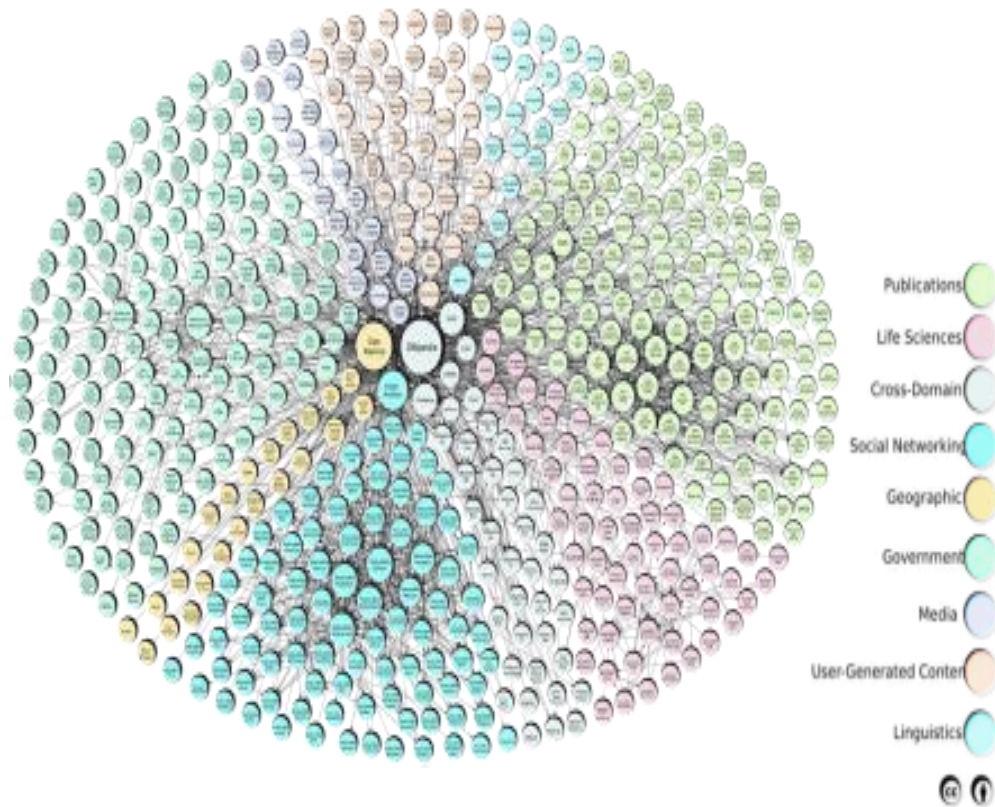
~~https://5stardata.info/en/~~

- ★ make your stuff available on the Web (whatever format) under an open license <sup>1</sup>
- ★★ make it available as structured data (e.g., Excel instead of image scan of a table) <sup>2</sup>
- ★★★ use non-proprietary formats (e.g., CSV instead of Excel) <sup>3</sup>
- ★★★★ use URIs to identify things, so that people can point at your stuff <sup>4</sup>
- ★★★★★ link your data to other data to provide context <sup>5</sup>



<https://5stardata.info/en/>

# LINKED DATA CLOUD



Linking Open Data cloud diagram : <http://lod-cloud.net>

# Outils RDF en ligne et gratuits



# Convertisseur RDF

**EASYRDF** Documentation Examples Converter Support Downloads

## Converter

Input Data:

or Uri:

http://njh.me/

Input Format:

Guess

Output Format:

Turtle Terse RDF Triple Language

☐ Raw output

Clear Submit

This converter is running version **0.9.0** of EasyRdf.

### Format en entrée

RDF/PHP  
RDF/JSON Resource-Centric  
JSON-LD  
N-Triples  
Turtle Terse RDF Triple Language  
RDF/XML  
RDFa

### Format en sortie

RDF/PHP  
RDF/JSON Resource-Centric  
JSON-LD  
N-Triples  
Turtle Terse RDF Triple Language  
RDF/XML  
Graphviz  
Notation3 (n3)  
Portable Network Graphics (PNG)  
Graphics Interchange Format (GIF)  
Scalable Vector Graphics (SVG)



# Outil de visualisation Linked Data

lodlive IT / EN / FR / GL

simple recherche \*

CHOISIR UN DATASET

choisir...

INTRODUIRE UN MOT

départ >>

insérer une URI \*

COLLER UNE ADRESSE

départ >>

parcourir vos données \*

INTRODUIRE RDP

INTRODUIRE

coming soon

départ >>

Live  
on  
LodLive

CHOISIR UN ENDPOINT

choisir...

fr.dbpedia.org \*

RESOURCE

choisir...

départ >>

linked.opendata.cz \*

RESOURCE

choisir...

départ >>

>



# RDF Grapher



## RDF Grapher

RDF grapher is a web service for parsing RDF data and visualizing it as a graph.

The service is based on [Redland Raptor](#) and [Graphviz](#).

Supported RDF serialization formats: Turtle, RDF/XML, RDF/JSON, N-Triples, TriG, and N-Quads.

Supported image formats: PNG, SVG, PDF, PS, EPS, GIF, and JPG.

Usage:

`http://www.lda.fi/service/rdf-grapher?rdf=DATA\_OR\_URI&from=FORMAT&to=FORMAT`

GET/POST parameters:

<code>rdf</code>	RDF data or URI
<code>from</code>	input serialization format (ttl, xml, json, nt, trig, nq), default: ttl
<code>to</code>	output image format (png, svg, pdf, ps, eps, gif, jpg), default: png

Examples:

`http://www.lda.fi/service/rdf-grapher?rdf=<http://example.com/s>+<http://example.com/p>+<http://example.com/o>+.&from=ttl&to=png`

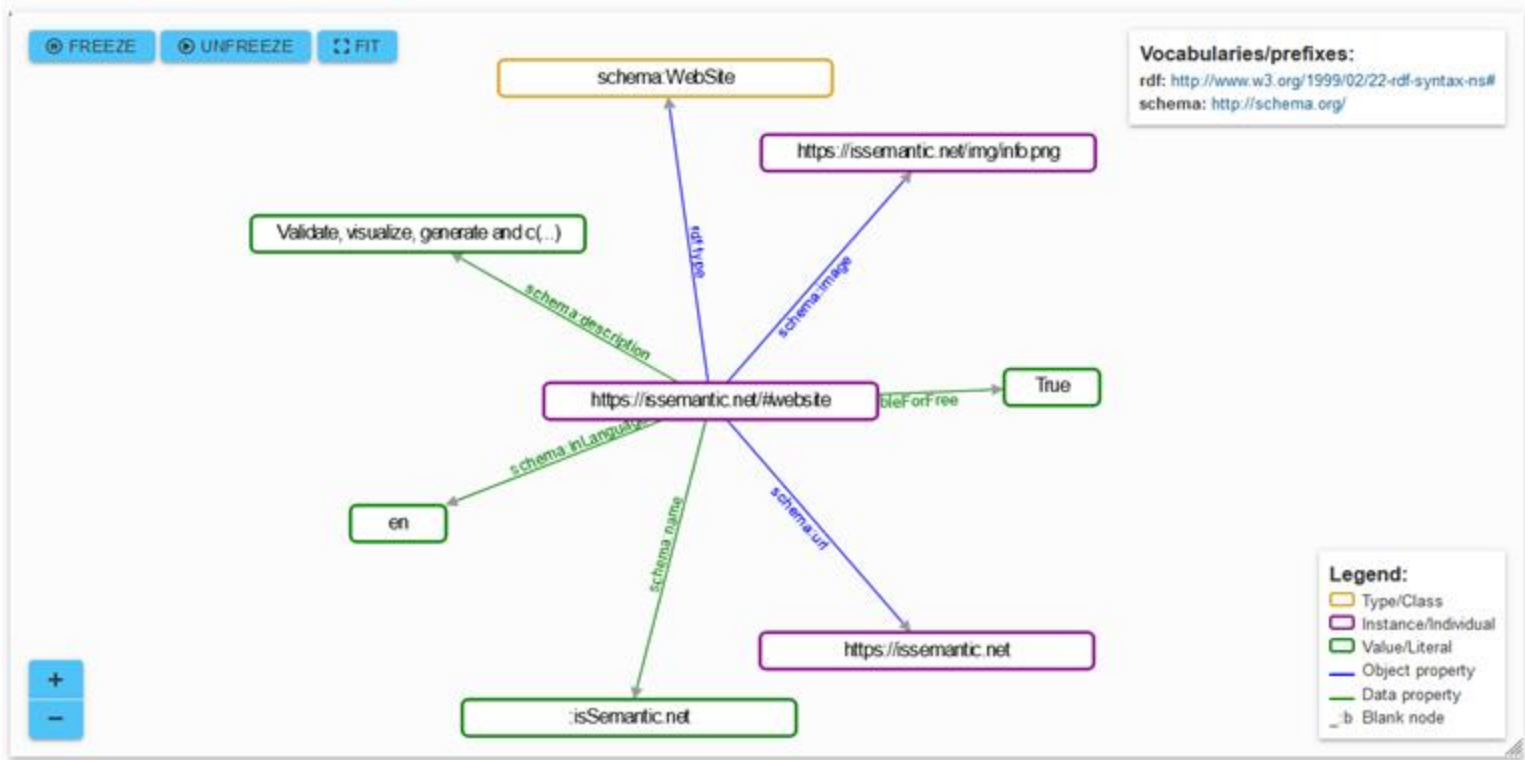
`http://www.lda.fi/service/rdf-grapher?rdf=http://dbpedia.org/resource/Helsinki&from=xml&to=png`

<http://www.lda.fi/service/rdf-grapher>





# isSemantic



<https://issemantic.net/rdf-visualizer>



# **Introduction à la formalisation des connaissances sous forme d'ontologies**

# Définitions

*“Une donnée est une information qui peut être enregistrée, traitée, analysée ou communiquée, quelle que soit sa nature. Lorsque la donnée n’a pas encore été traitée ou contextualisée, on parle de donnée brute”. (Guide pratique pour la gestion des données de la recherche irstea)*

34

**Donnée:** un élément

d’information, percevable  
par l’humain, manipulable  
par la machine



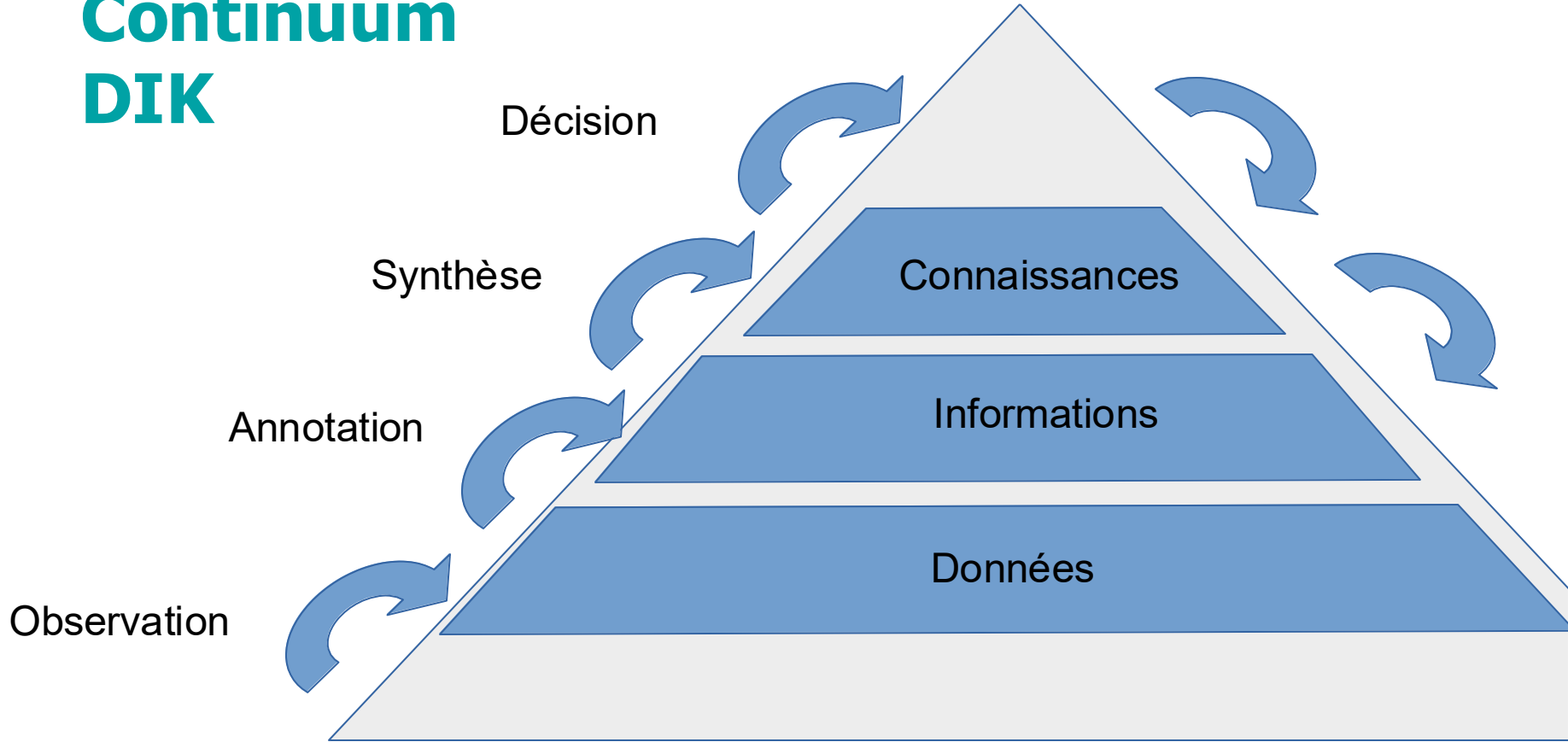
**Information:** donnée + sens

Sens : il faut fournir à l’humain du contexte

Type → structure de données



# Continuum DIK



# Ontologie: un peu d'histoire

Il était une fois, les systèmes experts une branche de l'Intelligence Artificielle

dans les années 90, évolution vers les systèmes à base de connaissances

Séparer les connaissances de résolution de problème des connaissances du domaine

**Ontologie**: formalisation des connaissances du domaine consensuelles et stables

**Ontologie**: *“spécification explicite d’une conceptualisation”* (Gruber, 1993)

**Problème**: spécification + conceptualisation = diagramme UML

**Ontologie** *“une spécification explicite et formelle d’une conceptualisation partagée”* (Studer et al., 1998)

**Ontologie du web**: les schémas de données formalisés dans un des langages des technologies du Web sémantique

# Le spectre sémantique - que devez-vous gérer?

Simple

Complexe

Vocabulaire contrôlé	Taxonomie	Thésaurus	Ontologie
<ul style="list-style-type: none"><li>• Une liste finie ou non</li><li>• Un unique niveau</li></ul>	<ul style="list-style-type: none"><li>• Un arbre</li><li>• Une hiérarchie</li><li>• Plusieurs niveaux imbriqués</li></ul>	<ul style="list-style-type: none"><li>• Un treillis</li><li>• Un dictionnaire</li><li>• Des classes d'équivalence</li><li>• Une hiérarchie avec des relations d'ordre partiel (généralisation / spécialisation)</li></ul>	<ul style="list-style-type: none"><li>• Un graphe orienté</li><li>• Une hiérarchie de classes et d'autres liens orientés spécifiques</li><li>• Des fonctions</li><li>• Des contraintes</li></ul>
<ul style="list-style-type: none"><li>• Appartenir ou ne pas appartenir à la liste</li></ul>	<ul style="list-style-type: none"><li>• Avoir pour père</li><li>• Avoir pour enfant</li><li>• Avoir un même père</li><li>• Etc.</li></ul>	<ul style="list-style-type: none"><li>• Avoir pour pères</li><li>• Avoir pour enfant préféré</li><li>• Avoir les mêmes pères</li><li>• Etc.</li></ul>	<ul style="list-style-type: none"><li>• Avoir une relation avec</li><li>• Avoir un chemin de relations qui conduit à</li><li>• Etc.</li></ul>

# Ontologies, pourquoi?

## 3 bonnes raisons





# 1- Organisation des données

- Il est très facile de se perdre dans de grandes collections de données ou de documents.
- Une ontologie est un bon moyen d'organiser (ou de structurer) cette collection.
- Elle facilite la navigation afin de trouver la bonne information. Cette organisation est flexible et évolutive.
- Elle structure les informations de manière multidimensionnelle.



## 2- Amélioration de la recherche

- Les ontologies sont utiles pour améliorer la précision de la recherche sur le Web.
- Sans cela, le site peut être un ensemble de pages dans lesquelles des termes ambigus se produisent.



- L'ontologie pour les moteurs de recherche (Google, Yahoo, Yandex, Microsoft)
- Meilleure recherche par types.
- Développement des services au-dessus.
- Exemple: Datasearch de Google  
<https://datasetsearch.research.google.com/>
- Plus de 10 millions de sites utilise ce vocabulaire pour structurer l'information

## Nelson Mandela

Ancien président de la république d'Afrique du Sud



Nelson Rolihlahla Mandela, dont le nom du clan tribal est « Madiba », né le 18 juillet 1918 à Mvezo et mort le 5 décembre 2013 à Johannesburg, est un homme d'État sud-africain. [Wikipédia](#)

**Date/Lieu de naissance :** 18 juillet 1918, [Mvezo](#), [Afrique du Sud](#)

**Date de décès :** 5 décembre 2013, [Houghton Estate](#), [Johannesburg](#), [Afrique du Sud](#)

**Parti :** [Parti communiste sud-africain](#)

**Épouse :** [Graça Machel](#) (m. 1998–2013), [Winnie Mandela](#) (m. 1958–1996), [Evelyn Ntoko](#) (m. 1944–1958)



## 3- Intégration de données hétérogènes

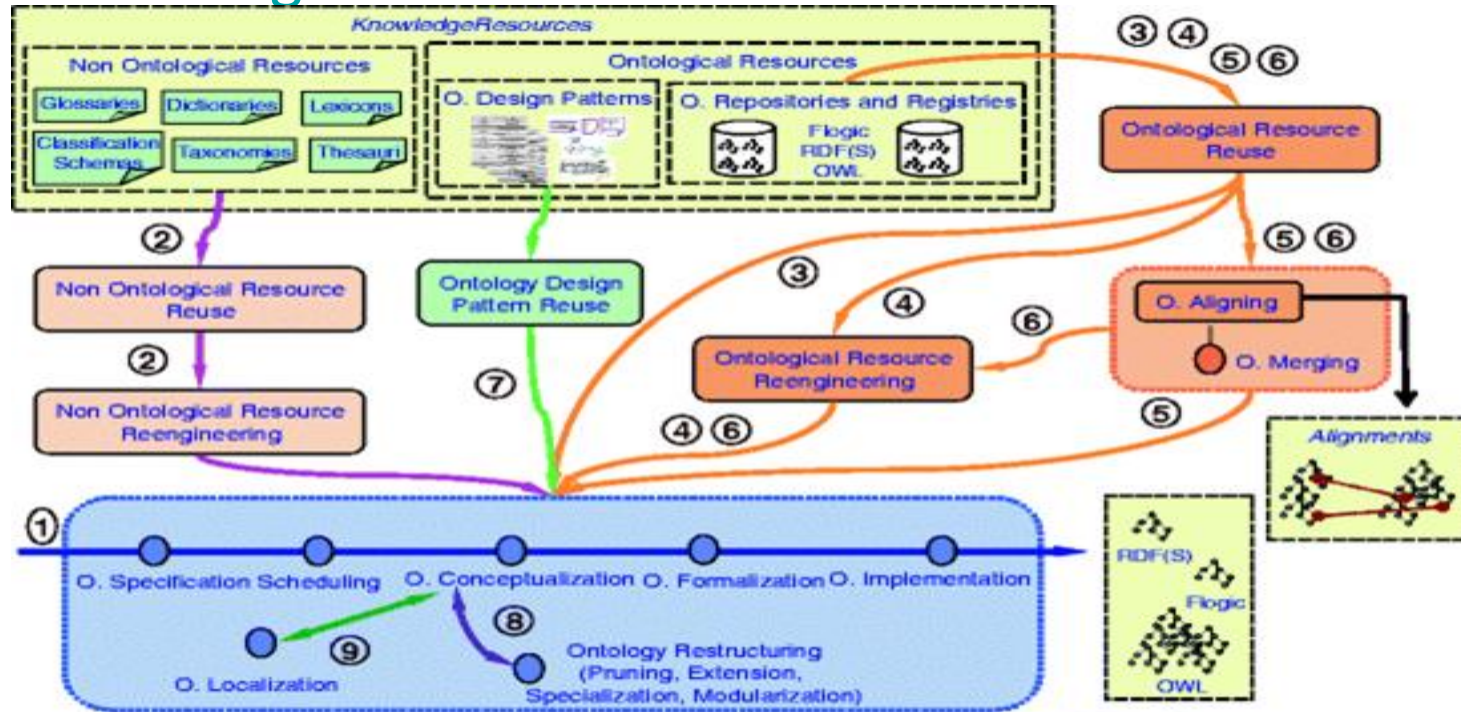
- Les ontologies servent d'interconnexion sémantique entre des sources de données hétérogènes.
- La possibilité de faire des inférences sur les données facilitent aussi la recherche d'information.
- En alignant/(ré)utilisant les ontologies, on peut intégrer de nombreuses sources et offrir un point d'entrée unique aux informations qu'elles fournissent.



# **Ingénierie des connaissances**

# Ingénierie des ontologies

la méthodologie NeOn = un ensemble de méthodes



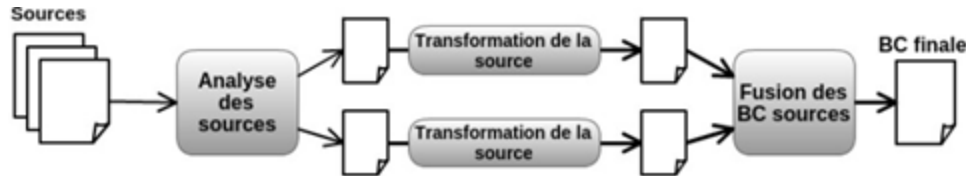
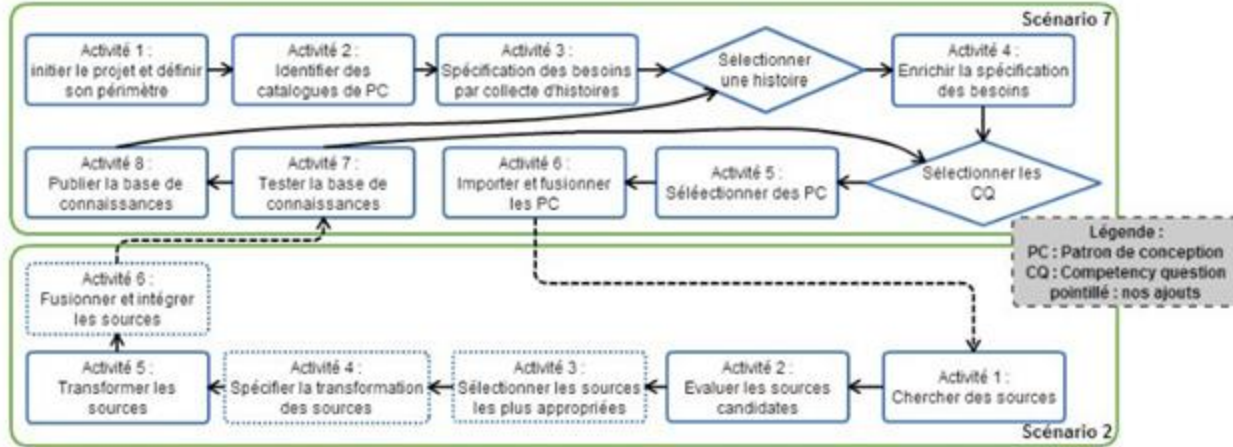
Crédit Suárez-Figueroa M.C., Gómez-Pérez A., Fernández-López M. (2012) The NeOn Methodology for Ontology Engineering.

In: Suárez-Figueroa M., Gómez-Pérez A., Motta E., Gangemi A. (eds)

[https://doi.org/10.1007/978-3-642-24794-1\\_2](https://doi.org/10.1007/978-3-642-24794-1_2)

# Ingénierie des

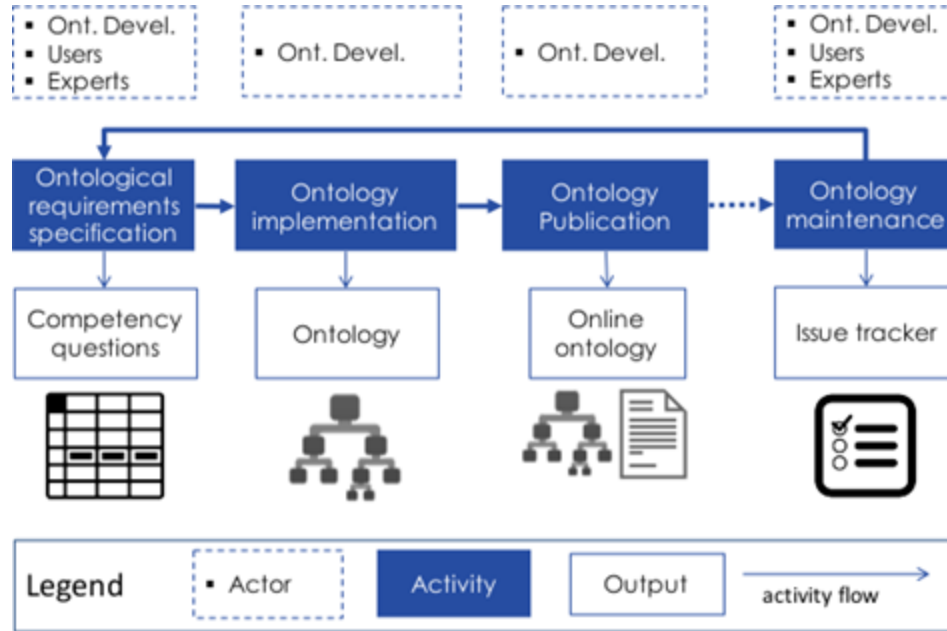
Non automatisés : plusieurs scénarios que l'on peut



Crédit Fabien AMARGER, “Vers un système intelligent de capitalisation de connaissances pour l'agriculture durable : construction d'ontologies agricoles par transformation de sources existantes” - Thèse Toulouse 2

# Ingénierie des

## ontologies (LOT)



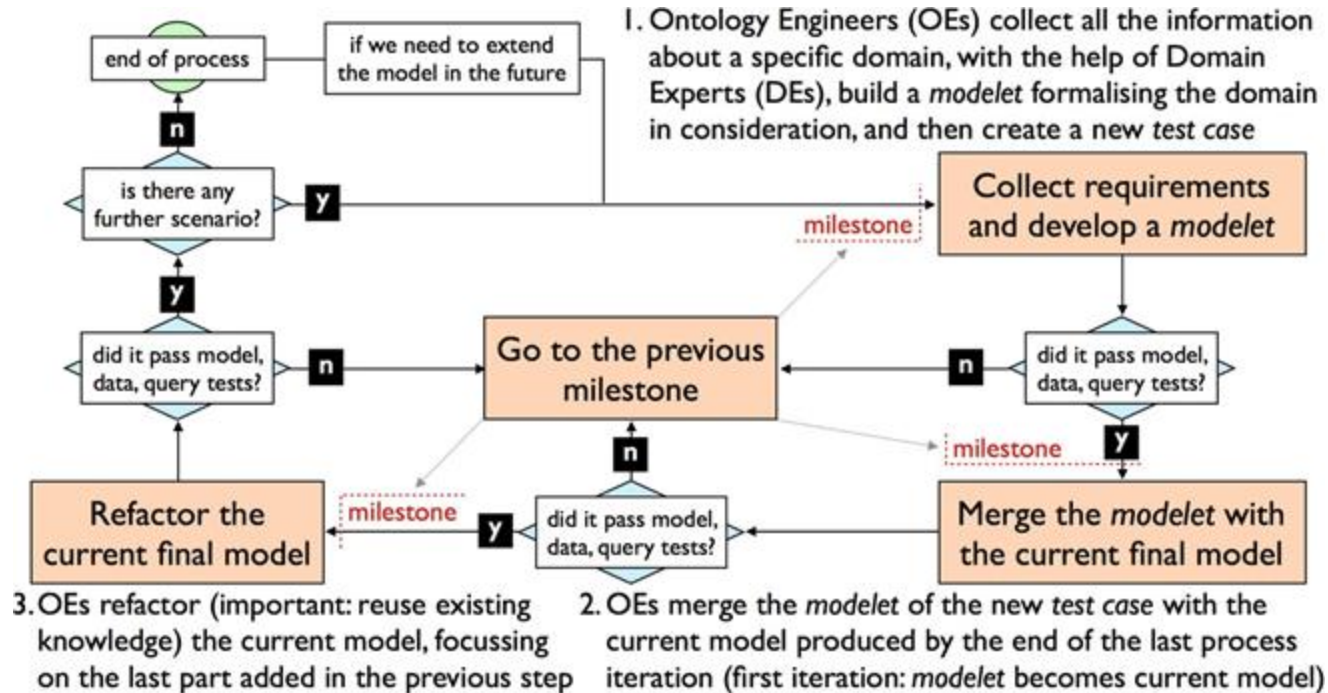
Crédit Espinoza-Arias, P., Poveda-Villalón, M. & Corcho, O.

Using LOT Methodology to develop a noise pollution ontology: a Spanish use case.

<https://doi.org/10.1007/s12652-019-01561-2>

# Ingénierie des ontologies

## La méthode SAMOD



Crédit Silvio Peroni,

SAMOD: an agile methodology for the development of ontologies,

# Ontology Design Pattern

Design Pattern : une modélisation de petite taille répondant à un problème précis

ODP sont présents :

- dans les ontologies fondationnelles BFO ou DoIcE
- Dans les ontologies coeur de domaine (core domain) comme Semantic Sensor Network

Ils peuvent prendre différents noms et forme:

- Le Module AgronomicTaxon pour la Thèse de Fabien
- Amarger Le modelet dans la méthode SAMOD



# Consensus : Comment l'atteindre ?

- Thèse de Abdel Kader Keita

Aider les experts du domaine à se mettre d'accord sur les définitions des concepts (méthode Delphi)

Illustrer les liens entre concepts par des images (photo, carte, ...)

- Thèse de Fabien Amarger

Extraire des sources des informations communes et les qualifier :

- Détecter les candidats
- Évaluer leur degré de confiance Gérer les incohérences

# Où trouver des ontologies FAIR ?

## Organismes de

### normalisation

- World Wide Web Consortium (W3C) <https://www.w3.org/>
- European Telecom Standard Institut (ETSI) <https://www.etsi.org/>

## Répertoires

- Linked Open Vocabulary (LOV) <https://lov.linkeddata.es/dataset/lov/>
- EMBL-EBI Ontology Lookup Service (OLS) → biomedical <https://www.ebi.ac.uk/ols/index>
- Agroportal → agriculture, environnement, INRAE , ... <http://agroportal.lirmm.fr/>

## Communauté de

### pratique

- Community group CROP ontology <https://www.croponontology.org/>
- Basic Formal Ontology <https://basic-formal-ontology.org/>

# Conclusion

- Les ontologies du web sont des schémas de données FAIR qui contiennent des patrons de conception (des bonnes pratiques de modélisation)

Les graphes de connaissance permettent aujourd'hui d'apporter une valeur ajoutée dans l'intégration des sources hétérogènes

- Les technologies du Web Sémantique changent les paradigmes de modélisation: base de graphes, Property First Class Citizen, Open World Assumption, ...

# Bienvenue dans le monde sémantique!!



Image générée par Gemini 3