



HAL
open science

Découverte des outils du web sémantique

Magalie Weber

► **To cite this version:**

Magalie Weber. Découverte des outils du web sémantique. Journée des Techniques de l'Unité BIA, Mar 2018, Nantes, France. hal-02786661

HAL Id: hal-02786661

<https://hal.inrae.fr/hal-02786661v1>

Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Journée des Techniques BIA

Découverte des outils du web sémantique :
exemple d'application sur une terminologie
dédiée au pain.

Magalie Weber – Cellule logistique



29 mars 2018



Découverte du web sémantique : origine et principes



Succès du web

- Documents
- Images
- Vidéos
- Applications
- Services
- Vie Sociale
- Business

Cependant



Problème du web

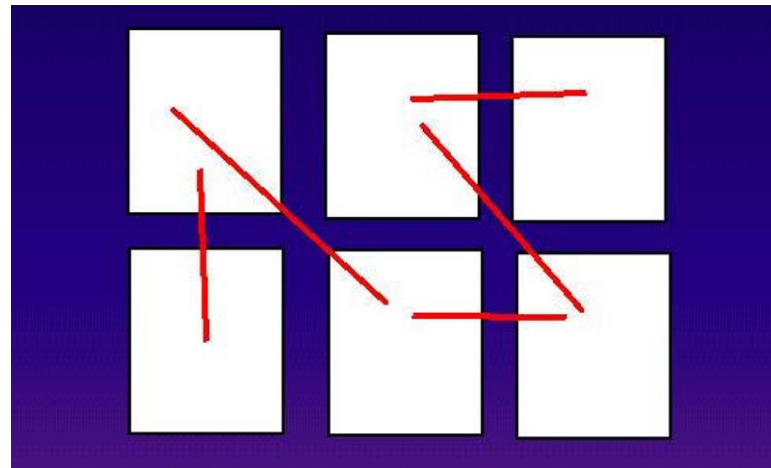
Qui peut exploiter le Web actuel ?

- Les humains ?
- Les moteurs de recherche ?

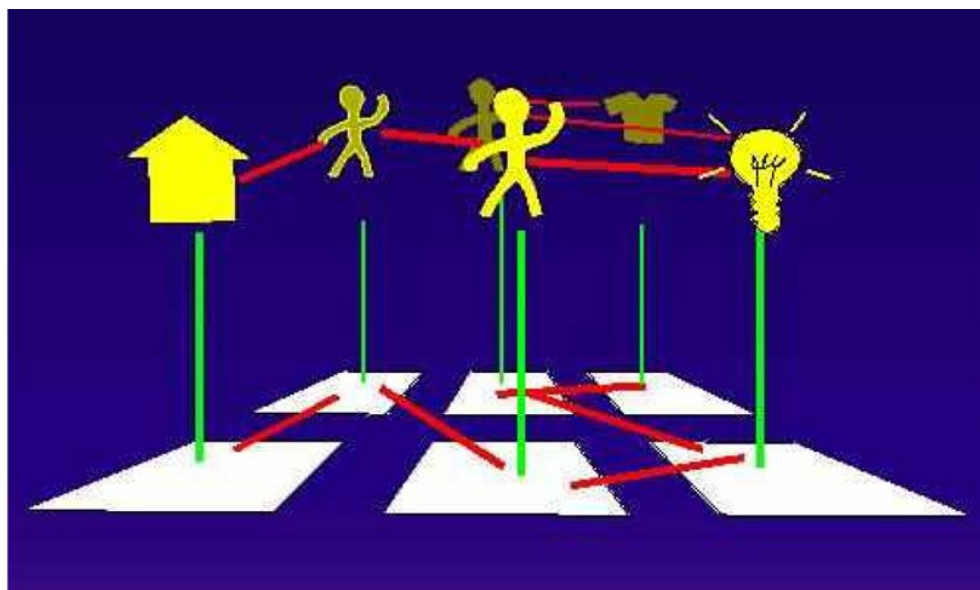




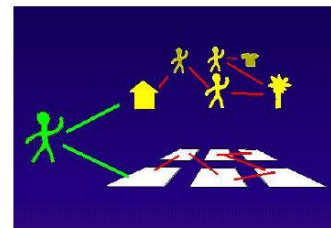
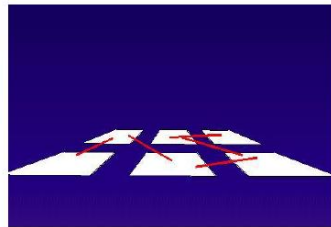
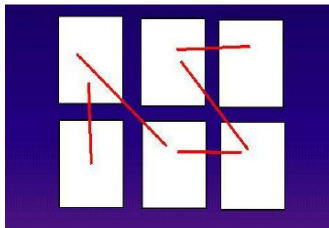
Pour un ordinateur, le web c'est ça :



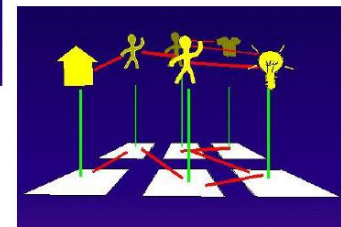
Le web sémantique, c'est ça :



Le Web Sémantique mentionné par Tim Berners-Lee en 1994 à WWW



... à un web de « choses » reliées à la réalité et compréhensibles par les machines



Passer du web des liens entre les pages peu compréhensibles par les machines.....

**Web sémantique, Web 3.0, Web de données :
3 appellations pour une même notion**

Ce qu'il faut faire :

Identifier

URI : Uniforme Resource Identifier

URL

Identifier ce qui existe sur le web.



<http://mon-site.dz>

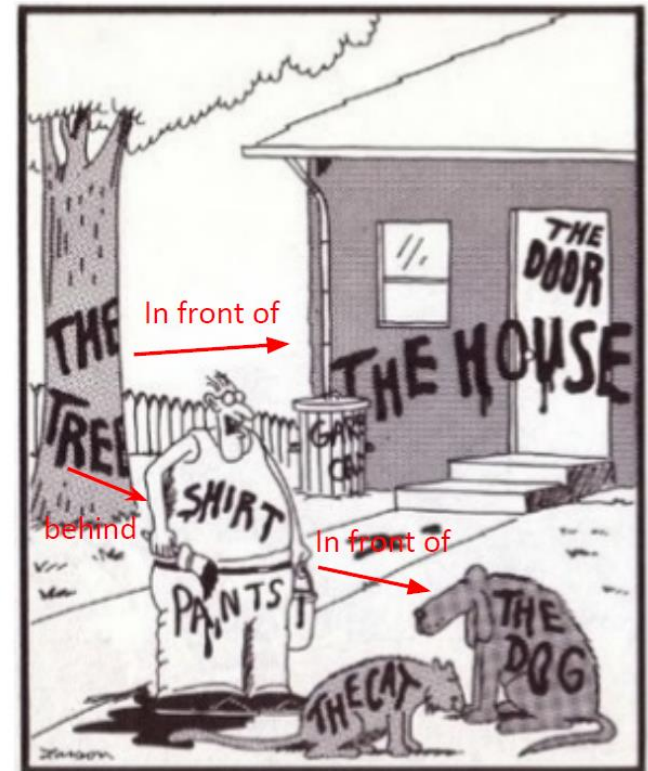
URI

Identifier, sur le web, ce qui existe.



<http://animaux.dz/ce-zebre>

Expliciter



"Now! ... That should clear up a few things around here!"

Comment faire ?

Représenter les informations ou « données » sous forme de phrase grâce au format RDF



RDF : un modèle de triplets

(**sujet** , **predicat** , **objet**)

Structure d'une phrase simple

(**sujet** , **verbe** , **complément**)

Le format RDF

Resource: entreprises, livres, personnes, perceuses, média, services, idées...

Description: attributs, qualités, et des relations entre les ressources

Framework: modèle, langages et syntaxes pour ces descriptions





Exemple de texte en langage « naturel »

Contenu non interprétable par la machine

Timothy John Berners-Lee,
né le 8 juin 1955 à Londres.

Il est le principal inventeur du World Wide Web.
Il préside le World Wide Web Consortium (W3C),
organisme qu'il a fondé.

Auteur du livre *A Framework for Web Science*.

Exemple « explicité »

Identification : Ressources & Relations

Timothy John Berners-Lee,

né le 8 juin 1955 à Londres.

Il est le principal inventeur du World Wide Web.

Il préside le World Wide Web Consortium (W3C),
organisme qu'il a fondé.

Auteur du livre *A Framework for Web Science*.

Traduction dans une forme « standard »

Phrases simples

TimBL has date of birth 1955-06-08.

TimBL has place of birth London.

TimBL is inventor of WWW.

TimBL is director of W3C.

TimBL is creator of W3C.

TimBL is author of Frame work for Web Science.

sujet verbe complément

Transcription au format RDF

Triplets **RDF** : interprétables par la machine

(TimBL, dateOfBirth, 1955-06-08)

(TimBL, placeOfBirth, London)

(TimBL, isInventorOf, WWW)

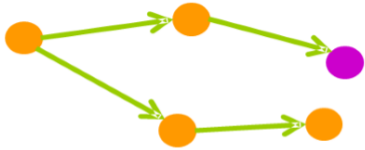
(TimBL, isDirectorOf, W3C)

(TimBL, isCreatorOf, W3C)

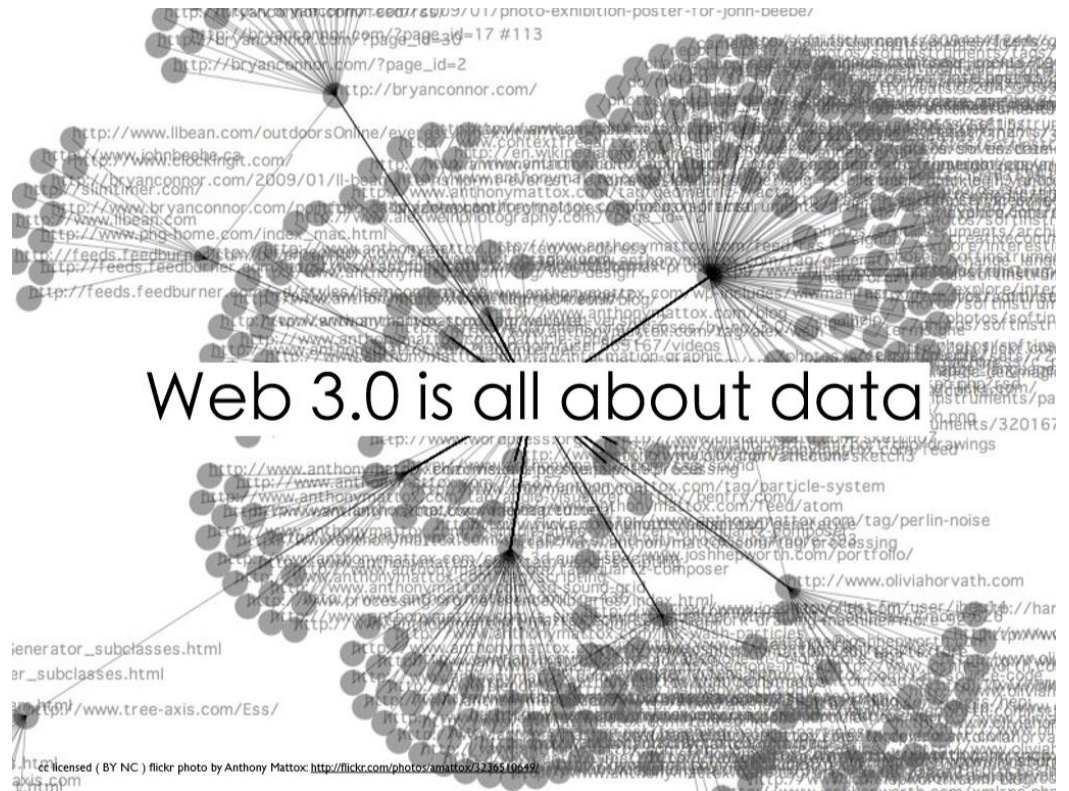
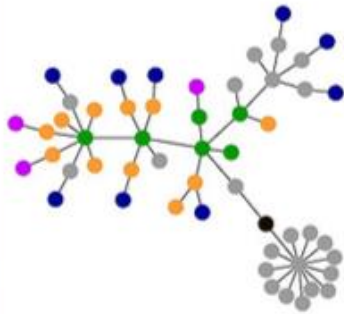
(TimBL, isAuthorOf, Web_Science)

(Sujet, predicat, Objet)

RDF est aussi un modèle de graphe orienté



(sommets, arcs, sommets)



Web 3.0 is all about data

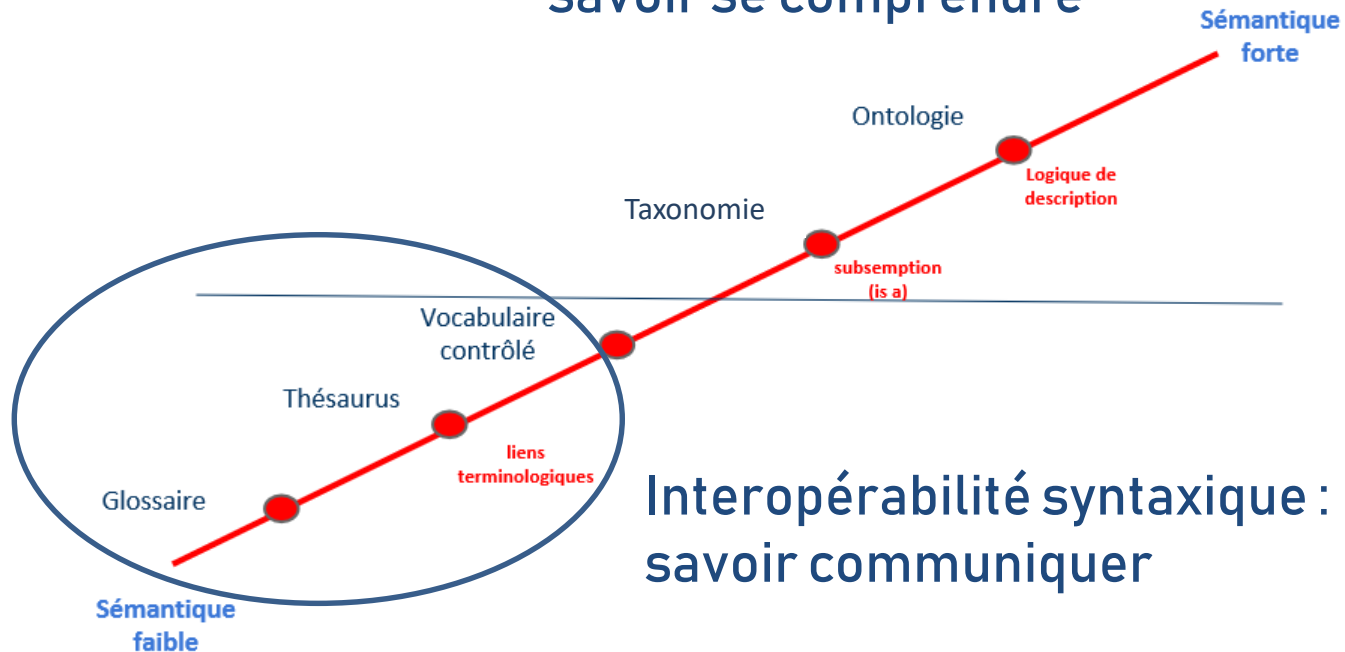


Comment donner du sens aux informations de type texte et les rendre compréhensibles par les machines ?



Niveaux d'interopérabilité

Interopérabilité sémantique :
savoir se comprendre



Interopérabilité syntaxique :
savoir communiquer



**Savoir communiquer
(interopérabilité syntaxique)**



Vocabulaire contrôlé + Données structurées selon un modèle



Liste de termes ou de concepts

Glossaire, lexique, terminologie...



Syntaxe = structure

**Format
RDF/SKOS**

Format RDF/SKOS

SKOS ou **Simple Knowledge Organization System** (« *système simple d'organisation des connaissances* ») est une recommandation du **W3C** publiée le 18 août 2009 pour représenter des thésaurus, classifications ou d'autres types de vocabulaires contrôlés ou de langages documentaires en s'appuyant sur le modèle de données RDF.

Chaque « concept SKOS » est défini comme une ressource RDF.

https://fr.wikipedia.org/wiki/Simple_Knowledge_Organization_System

Format RDF/SKOS

À chaque concept peuvent être attachées des propriétés RDF :

- au maximum un *terme préférentiel* par langue ;
- *des synonymes*, avec spécification possible de la langue ;
- *des définitions et notes*, avec spécification possible de la langue

Les concepts peuvent être liés entre eux par des relations hiérarchiques ou associatives.

Concept en RDF/SKOS : sujet du triple RDF

rdf:type : skos:Concept

Attributs :

skos:prefLabel, skos:altLabel et skos:hiddenLabel

Relations :

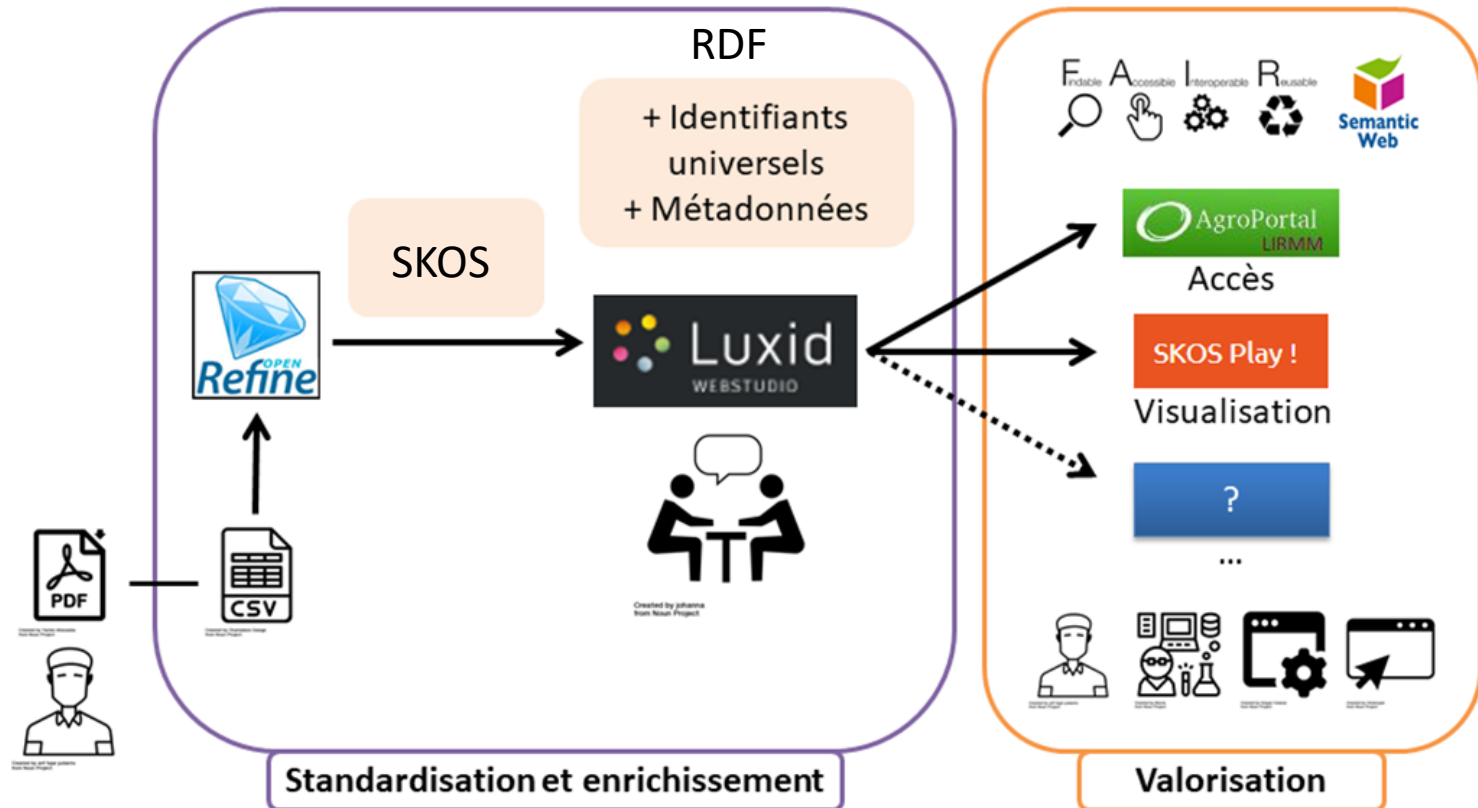
skos:broader (terme générique)
skos:narrower (terme spécifique)
skos:related (terme lié)



Application au glossaire du pain

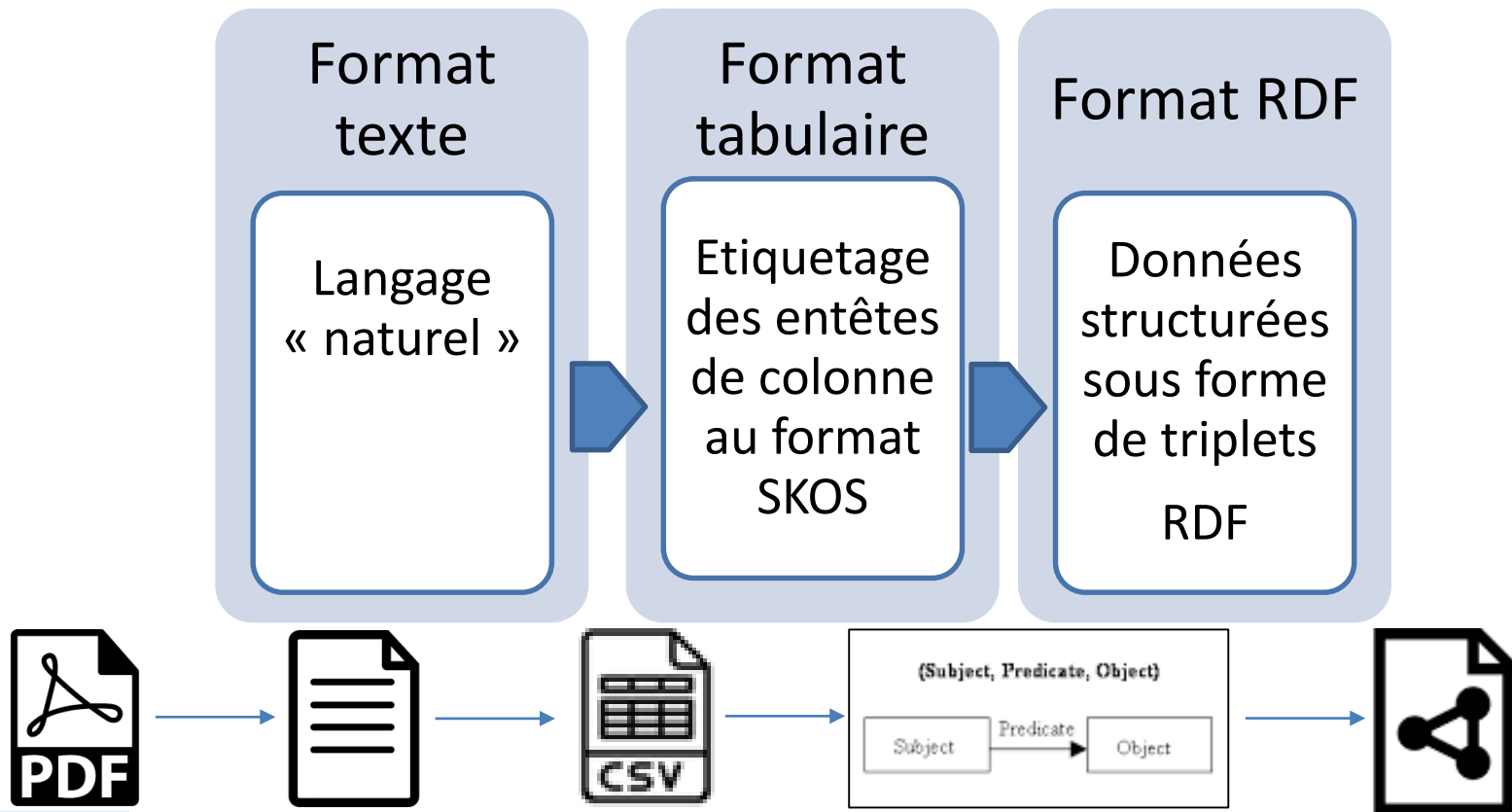


La démarche de transformation des données

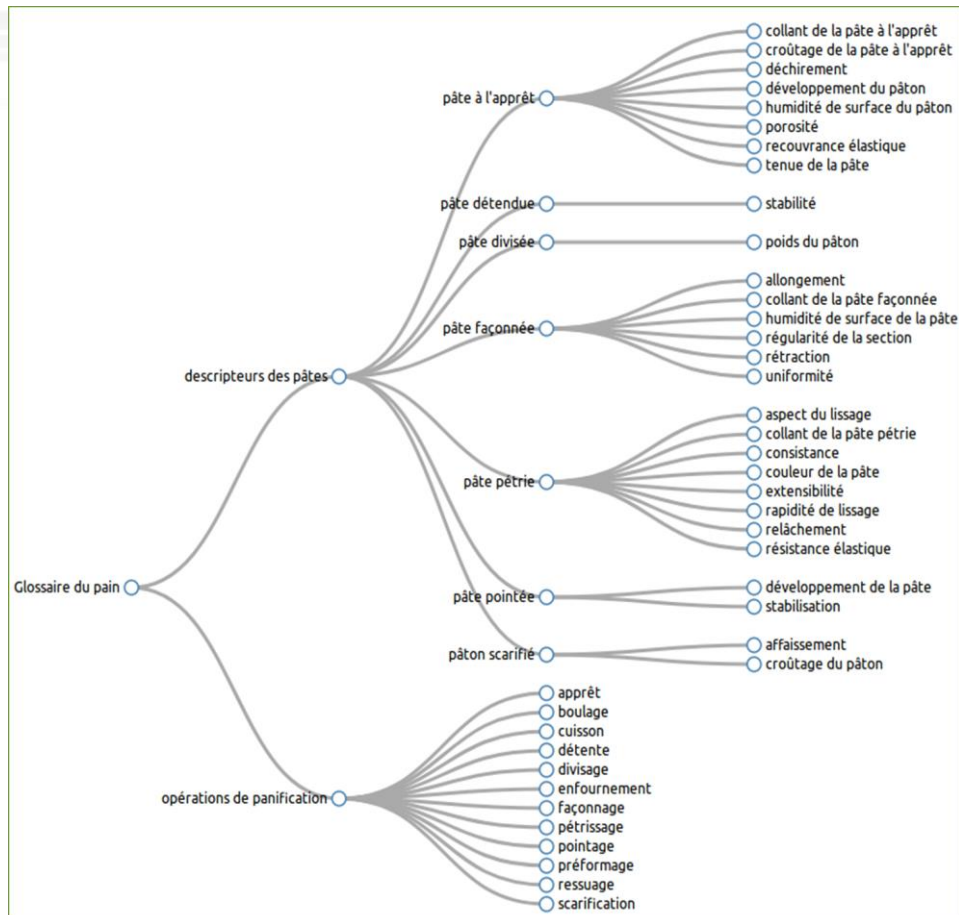


La modélisation des données :

Comment passer du format texte au standard RDF/SKOS ?

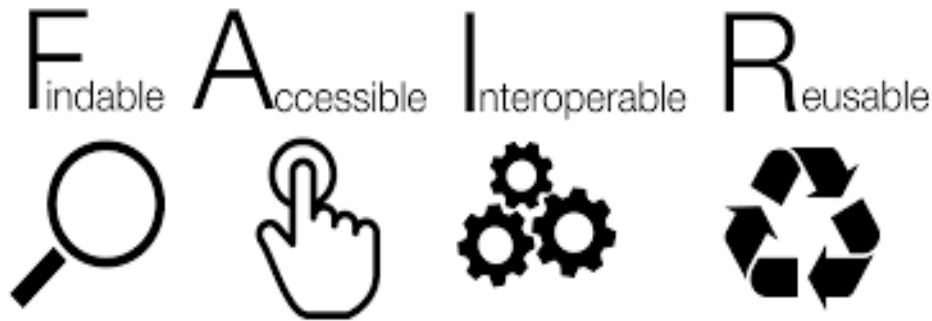


Exemple de sortie avec SKOS play

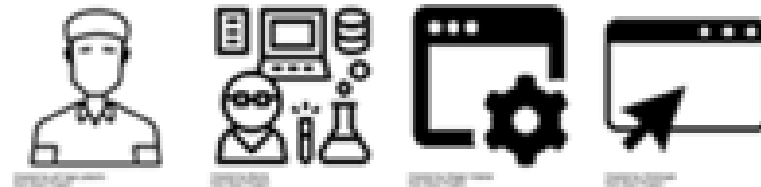


Visualisation arborescente

Conclusion et perspectives

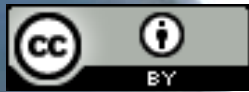


De nouvelles applications possibles pour de nouveaux utilisateurs et de nouveaux usages



Merci pour votre attention

link
to the world



Cette présentation a été réalisée à partir d'un ensemble de présentations de Fabien Gandon, chercheur à l'INRIA et membre du W3C. Merci à lui.