



HAL
open science

Enrichir mon corpus massivement de manière automatique : le web sémantique et SPARQL

Victoria Le Fournier, Florence Perret

► To cite this version:

Victoria Le Fournier, Florence Perret. Enrichir mon corpus massivement de manière automatique : le web sémantique et SPARQL. École thématique. France. 2021, pp.62. hal-03674634

HAL Id: hal-03674634

<https://hal.science/hal-03674634>

Submitted on 20 May 2022

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

ENRICHIR MON CORPUS MASSIVEMENT DE
MANIÈRE AUTOMATIQUE : LE WEB
SÉMANTIQUE ET SPARQL

NUMÉRIQUE



Victoria Le Fournier
Florence Perret

EXEMPLES DE RÉSULTATS



EXEMPLES DE VISUALISATIONS DE DONNÉES OBTENUES GRÂCE À DES REQUÊTES SPARQL SUR WIKIDATA :

The screenshot shows the Wikidata Query Service interface. The SPARQL query is as follows:

```
1 #Éducation
2 #defaultView:Map
3 SELECT ?personne ?personneLabel ?institutionLabel ?coord WHERE {
4   ?personne wdt:P31 wd:Q5 ; ?personne est un humain
5   wdt:P166 wd:Q3104 ; ?personne a reçu le prix nobel
6   p:P69 ?scolarité . ?personne a une "scolarité"
7   ?scolarité ps:P69 ?institution . ?scolarité a une valeur "institution"
8   ?institution wdt:P625 ?coord . ?institution a des coordonnées
9
10
11
12
13 SERVICE wikibase:label { bd:serviceParam wikibase:language "(AUTO_LANGUAGE),en" . }
14 }
```

The map below the query shows red dots representing the locations of laureates, primarily concentrated in Europe and North America.

Exemple 1 : une requête pour obtenir une carte des lieux où les lauréats du prix Nobel de physique ont fait leurs études

The screenshot shows the Wikidata Query Service interface. The SPARQL query is as follows:

```
1 #defaultView:ImageGrid
2 SELECT ?eglise ?egliseLabel (YEAR(?date) AS ?finConstruction)?image ?coord
3 WHERE {
4   ?eglise wdt:P31 wd:Q16970.
5   (?eglise wdt:P131 wd:Q220 .)
6   UNION
7   (?eglise wdt:P131 ?arr .
8    ?arr wdt:P131 wd:Q220 .)
9
10 OPTIONAL (?eglise wdt:P571 ?date .)
11 OPTIONAL (?eglise wdt:P18 ?image .)
12 OPTIONAL (?eglise wdt:P625 ?coord .)
13
14 SERVICE wikibase:label { bd:serviceParam wikibase:language "(AUTO_LANGUAGE),fr,en" . }
15 }
```

The gallery below the query displays four images of churches in Rome.

Exemple 2 : une requête pour obtenir une galerie d'image des églises de Rome

VUE DU TABLEAU OBTENU À PARTIR DE LA REQUÊTE DE L'EXEMPLE 1

personne	personneLabel	institutionLabel	coord
Q wd:Q937	Albert Einstein	Université de Zurich	Point(8.548333333 47.374722222)
Q wd:Q937	Albert Einstein	ancienne école cantonale d'Aarau	Point(8.05321372 47.39333869)
Q wd:Q937	Albert Einstein	Luitpold-Gymnasium	Point(11.5733 48.132)
Q wd:Q949	John Bardeen	université de Princeton	Point(-74.656111111 40.345277777)
Q wd:Q949	John Bardeen	université du Wisconsin à Madison	Point(-89.40972222 43.07527778)
Q wd:Q7085	Niels Bohr	université de Copenhague	Point(12.5725 55.679722222)
Q wd:Q7085	Niels Bohr	Trinity College	Point(0.116944444 52.206944444)
Q wd:Q7186	Marie Curie	université de Paris	Point(2.34333 48.8486)
Q wd:Q7186	Marie Curie	université de Paris	Point(2.34333 48.8486)
Q wd:Q7186	Marie Curie	université de Paris	Point(2.34333 48.8486)
Q wd:Q7186	Marie Curie	Sorbonne	Point(2.34356 48.84864)
Q wd:Q8753	Enrico Fermi	université de Göttingen	Point(9.93792 51.533979833)
Q wd:Q8753	Enrico Fermi	université de Leyde	Point(4.485404 52.157073)
Q wd:Q8753	Enrico Fermi	École normale supérieure de Pise	Point(10.400225 43.719611)
Q wd:Q9021	Max Planck	université Louis-et-Maximilien de Munich	Point(11.580277777 48.150833333)
Q wd:Q9021	Max Planck	université Humboldt de Berlin	Point(13.393333333 52.518055555)
Q wd:Q9021	Max Planck	Maximiliansgymnasium München	Point(11.5847 48.1646)
Q wd:Q9021	Max Planck	université Frédéric-Guillaume	Point(13.393333 52.518056)

COMMENT EN ARRIVER LÀ?

```
1 #éducation
2 #defaultView:Map
3 SELECT ?personne ?personneLabel ?institutionLabel ?coord WHERE {
4   ?personne wdt:P31 wd:Q5 ; #personne est un humain
5             wdt:P166 wd:Q38104 ; #personne a reçu le prix nobel
6             p:P69 ?scolarite . #personne a une "scolarité"
7   ?scolarite ps:P69 ?institution . #scolarité a une valeur "institution"
8   ?institution wdt:P625 ?coord . #institution a des coordonnées
9
10
11
12
13 SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
14 }
```

Focus sur la requête permettant d'obtenir la carte des des lieux où les lauréats du prix Nobel de physique ont fait leurs études

REPARTIR D'UN EXEMPLE CONCRET



La représentation des sculpteurs parisiens (1500-1800) dans les musées internationaux

LA REPRÉSENTATION DES SCULPTEURS PARISIENS (1500-1800) DANS LES MUSÉES INTERNATIONAUX

Comment faire ?

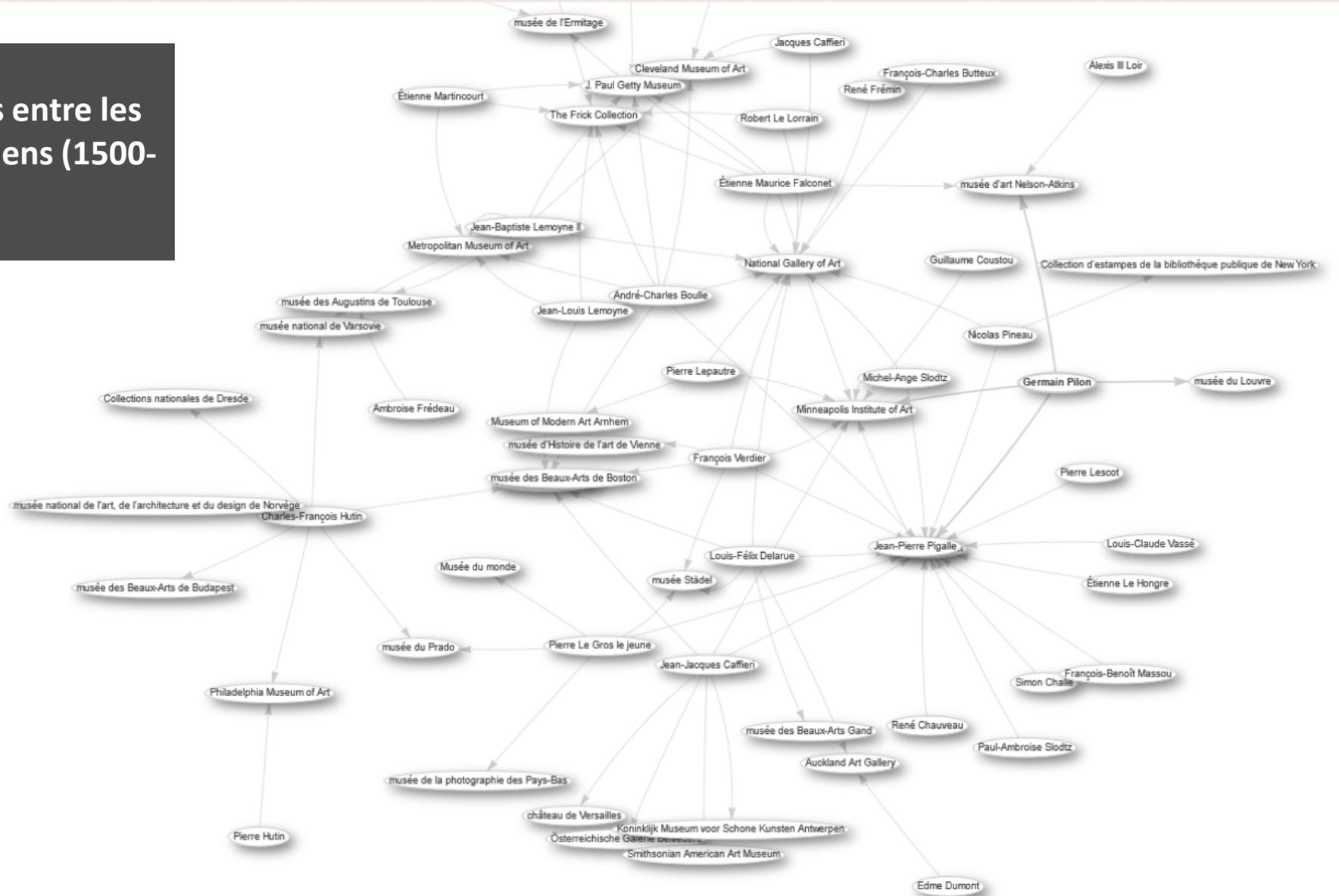
Les étapes :

- Lister tous les sculpteurs parisiens
- Trouver leurs dates de naissance
- Trouver tous les catalogues des musées pour voir s'ils sont dans le département de sculpture
- Créer un tableau pour reporter toutes les informations
- Générer un graphique

Un travail colossal qui prendrait plusieurs jours...

Et si on le faisait plutôt en 10 minutes ?

Graph des liens entre les sculpteurs parisiens (1500-1800)



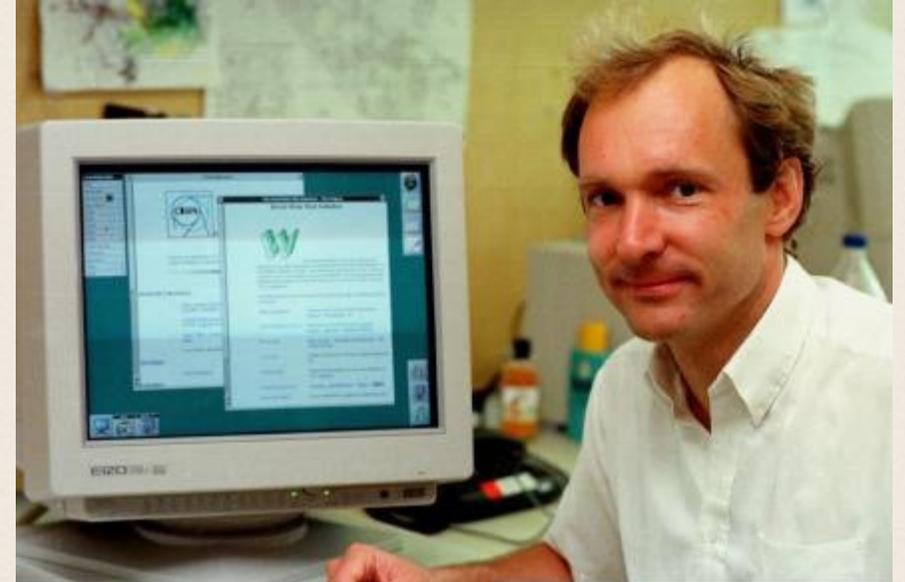
BESOIN DE DONNÉES STRUCTURÉES

- Il faut que la description de l'information soit **identique partout**
- Que je puisse la récupérer facilement
- Que je puisse l'exploiter facilement

Il faut donc une structure commune

LA RÉPONSE DU WEB

Permettre l'échange de données entre des agents (machines/humains)

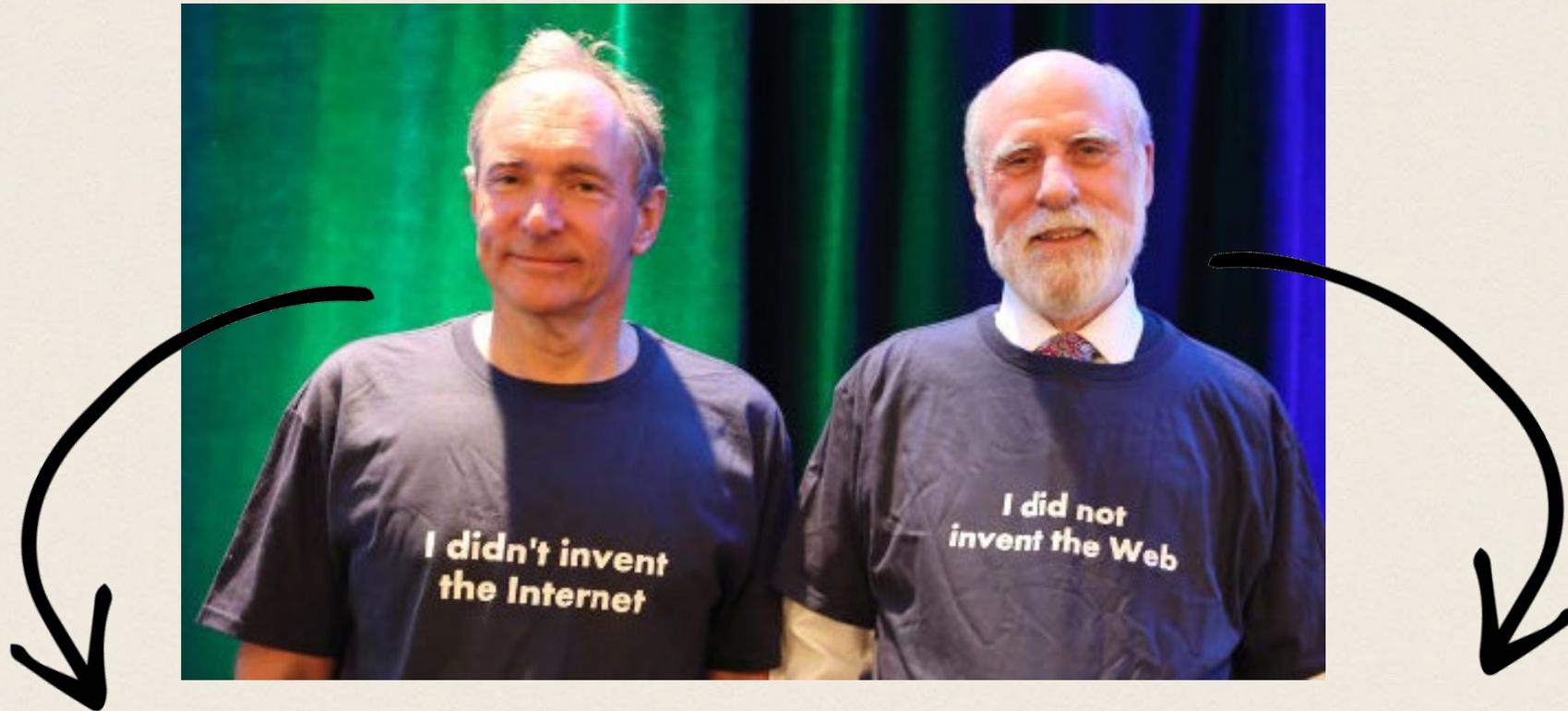


REPARTONS DU DÉBUT



Qu'est ce que le web sémantique en fait ?

LE WEB CE N'EST PAS INTERNET !



Tim Berners-Lee

Vinton Cerf (TCP/IP)

LE WEB SÉMANTIQUE

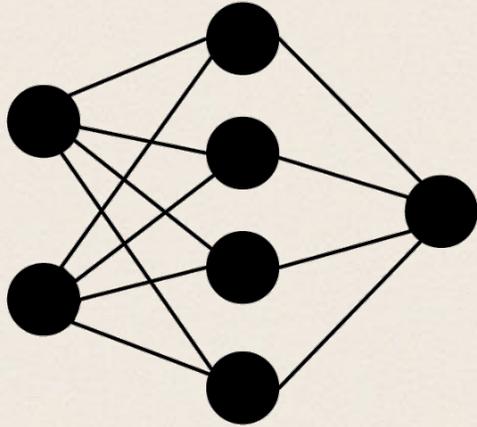
« The Semantic Web is a **web of data**, in some ways like a **global database** »

« The Semantic Web is not a separate Web but an **extension of the current one**, in which information is given **well-defined meaning**, better enabling computers and people to work in cooperation. »

LE WEB SÉMANTIQUE, POUR QUOI FAIRE ?

- Utiliser les agents du web pour traiter des données
- Communiquer/utiliser des données du web pour enrichir des bases de données internes
- Communiquer/utiliser les bases de données internes pour enrichir la base de connaissance du web.

FONCTIONNEMENT

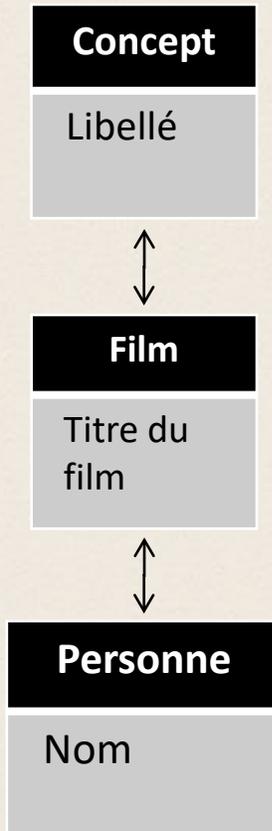


Le graphe pour modéliser
Décrire le monde

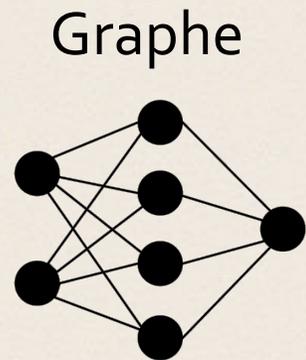
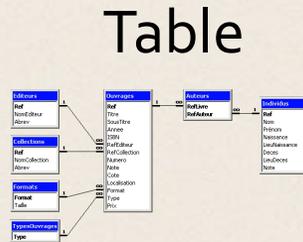
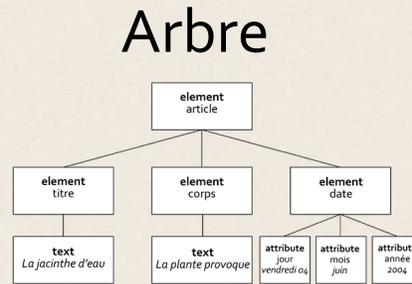


HTTP, URI et RDF pour placer dans le Web
Webifier le monde

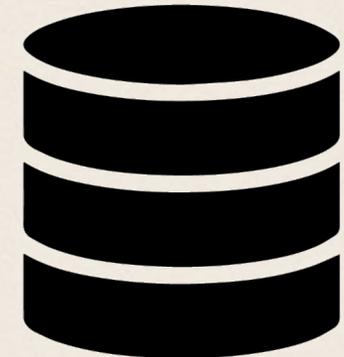
MODÈLE CONCEPTUEL



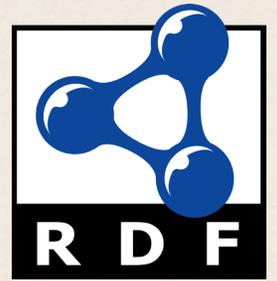
MODÈLE LOGIQUE



MODÈLE PHYSIQUE



RDF – RESSOURCE DESCRIPTION FRAMEWORK



- Modèle de graphe destiné à décrire formellement les ressources Web et leurs métadonnées, afin de permettre le traitement automatique de telles descriptions.
- Développé par le W3C, RDF est le langage de base du Web sémantique.
- Le principe de RDF est d'exprimer l'information sous forme de phrases simples - **triplets**

STRUCTURE D'UNE PHRASE SIMPLE :

Sujet verbe complément

Eugène Delacroix est l'auteur de La liberté guidant le peuple

sujet



prédicat



objet



Eugène Delacroix

est l'auteur de

La liberté guidant le peuple

objet



prédicat



http://www.mied.org/personne/Eugene_Delacroix

<http://www.mied.org/ontologie/auteur>

http://www.mied.org/oeuvre/Liberte_guidant_le_peuple



est mort à

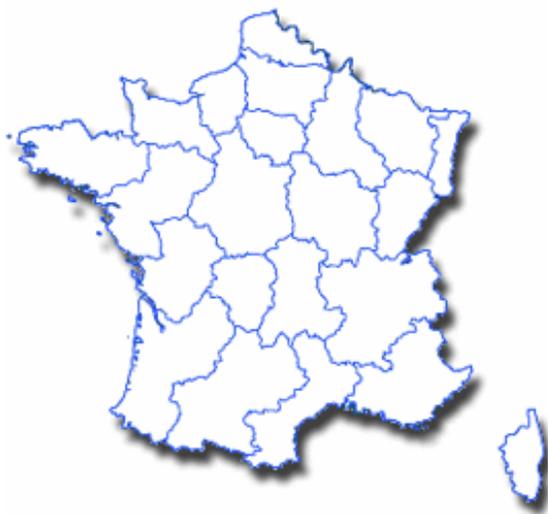
<http://www.mied.org/lieu/Paris>



est un



fait partie de

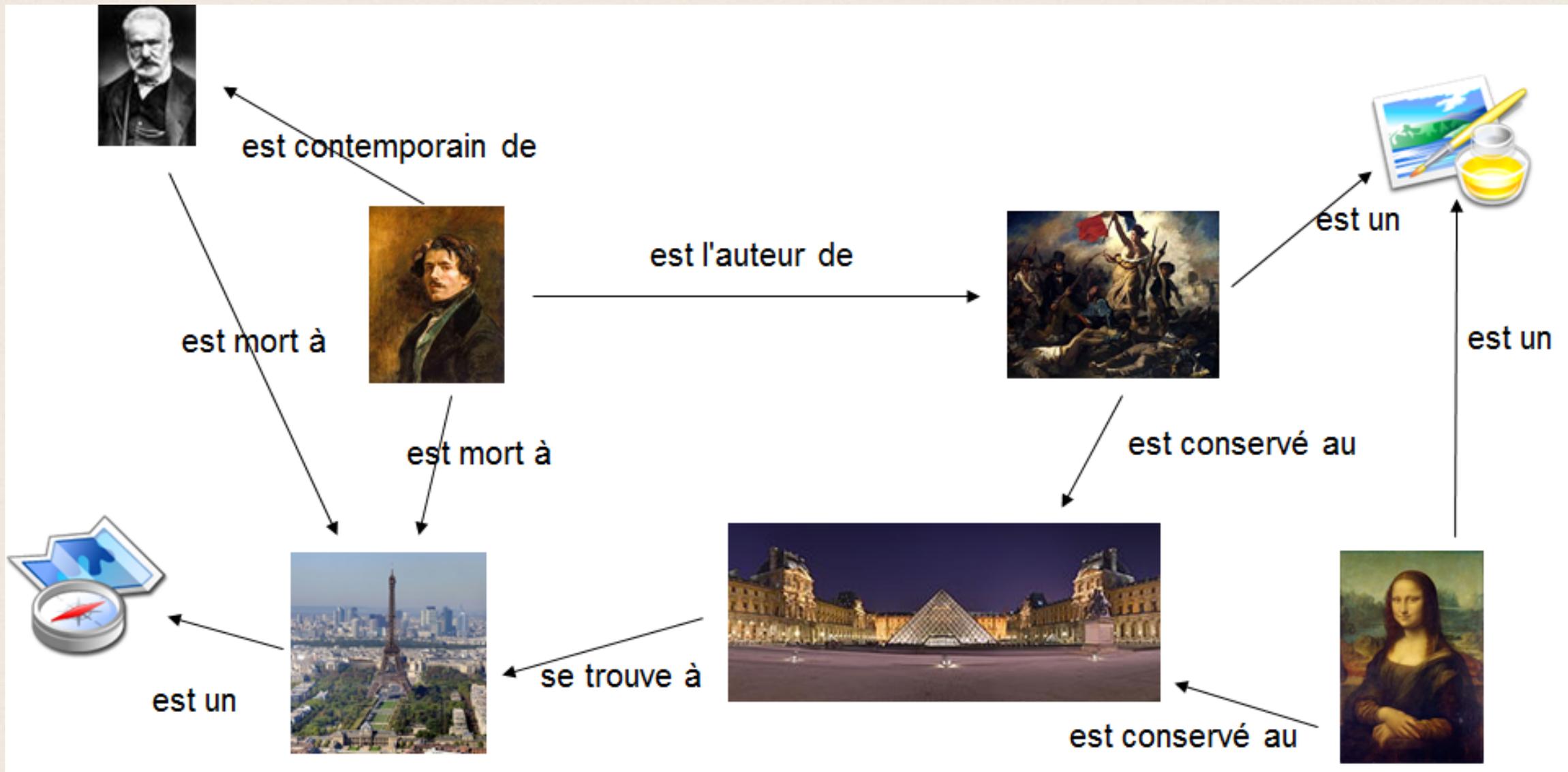


label

“Paris”@fr

label

“Parigi”@it



STRUCTURE D'UN TRIplet RDF :

Sujet prédicat objet

<<https://data.bnf.fr/ark:/12148/cb119298072#about>>
<<http://rdaregistry.info/Elements/a/P50121>>
<<https://data.bnf.fr/date/1873/>> .

Colette est née en 1873

SÉRIALISATION RDF

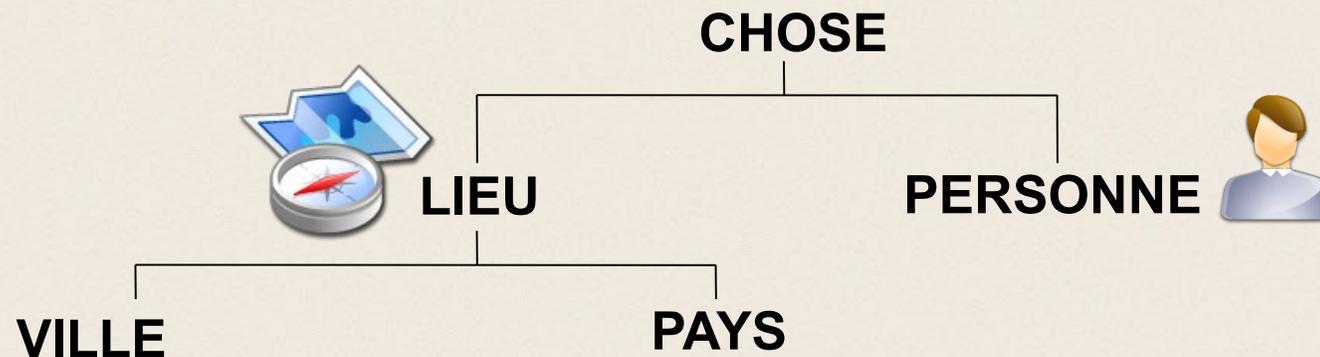
RDF est un modèle de données et non un format.

- RDF/XML ;
- Turtle sous-ensemble de Notation3 ;
- N-Triples sous-ensemble de Turtle ;
- JSON-LD ;
- RDFa;
- Web Ontology Language (OWL)

CLASSES ET HIÉRARCHIES DE CLASSES

En RDF, toute "chose" a un type (ou plusieurs). On appelle cela une "classe".

Les classes sont organisées en hiérarchies (classes, sous-classes)



RELIER L'INSTANCE À LA CLASSE



rdf:type



wdt:P31



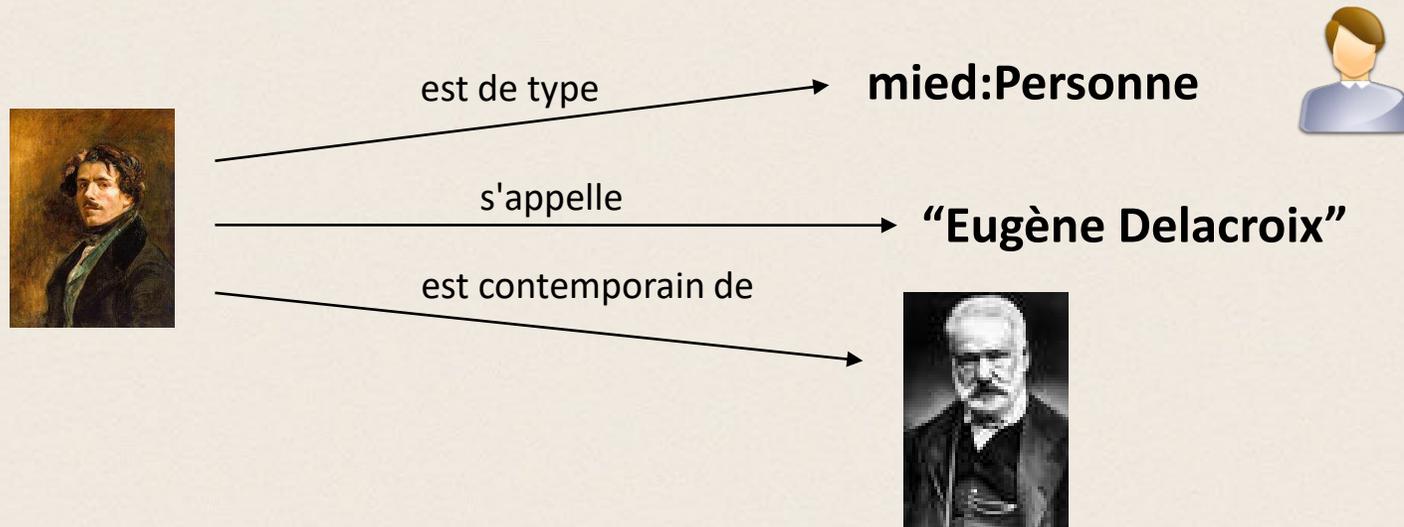
<http://www.mied.org/ontologie/Personne>



On désigne la classe par son URI

PROPRIÉTÉS

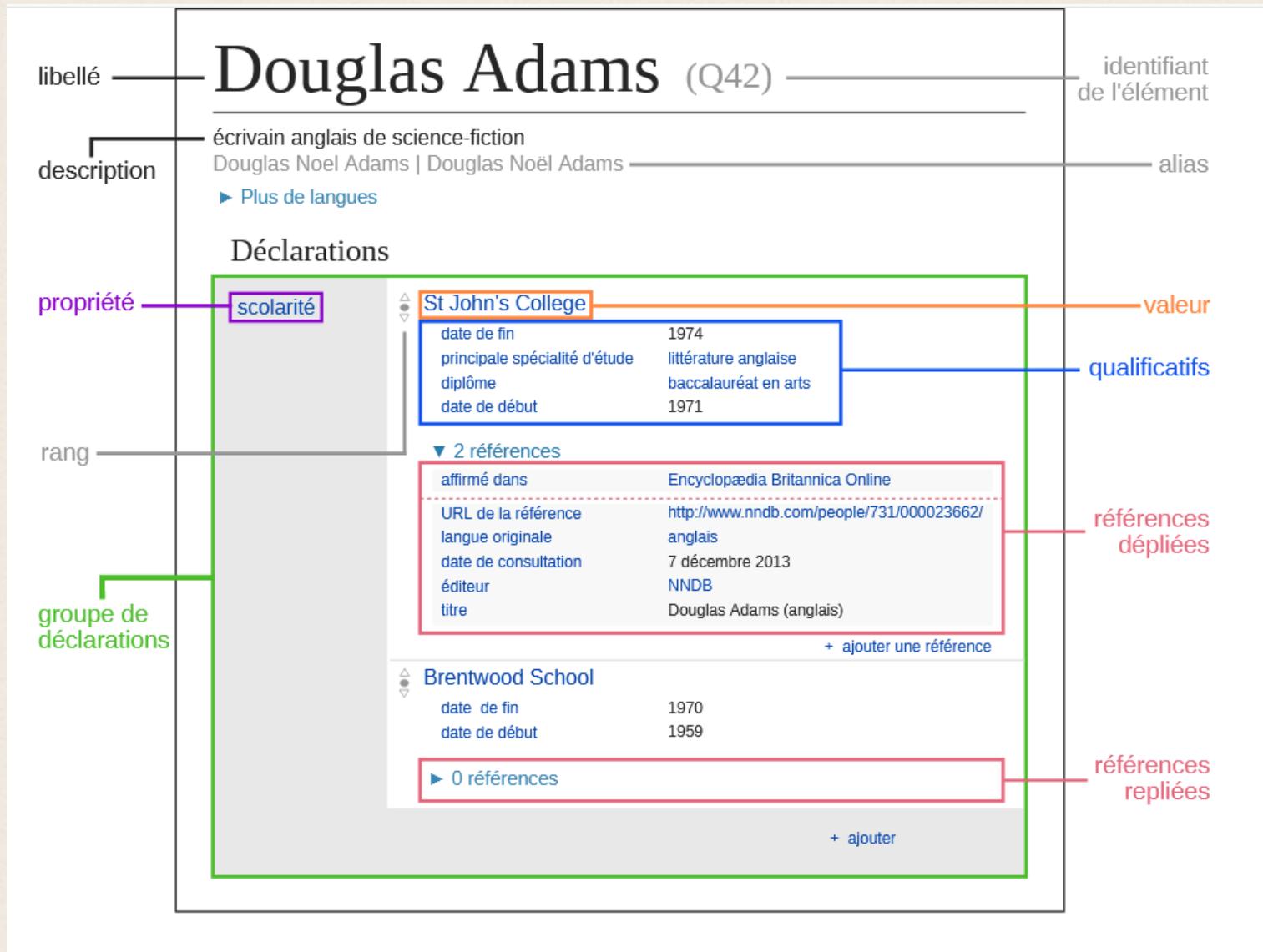
Les propriétés correspondent au “verbe” dans le triplet.



WIKIDATA

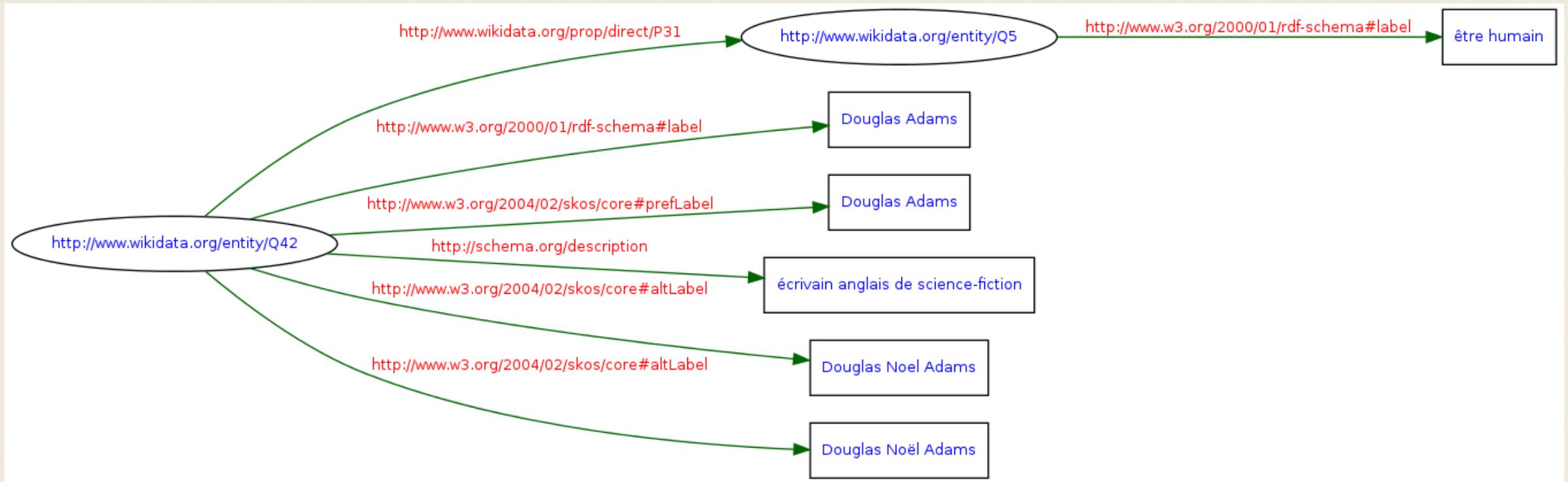


- Projet de la Wikimedia Foundation développé à partir de 2012 à l'initiative de l'association Wikimedia Deutschland
- **centralisateur d'informations** au service des Wikipedias
- des **données structurées** sous la forme de **déclarations**

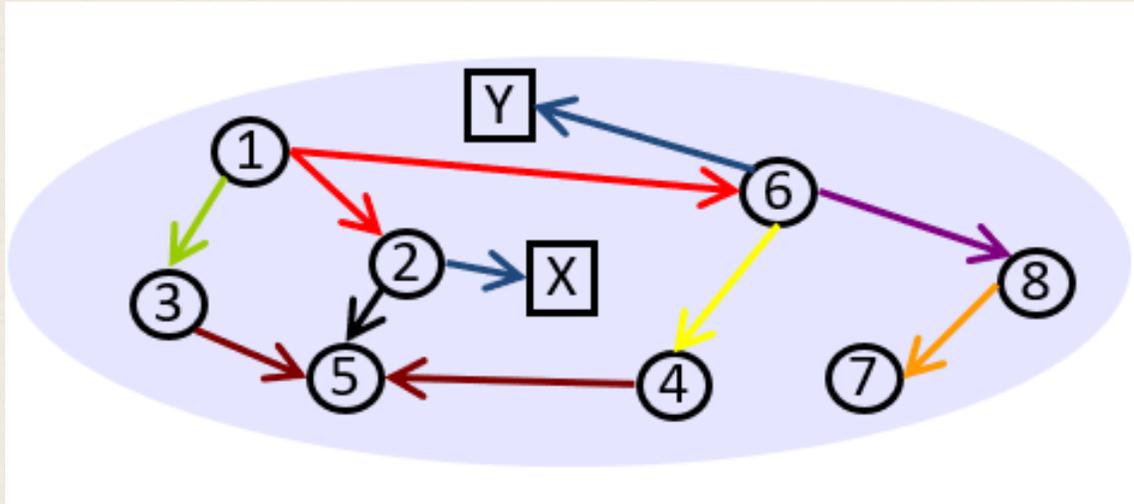


Organisation d'une page wikidata

STRUCTURE DES DONNÉES DANS WIKIDATA



EXERCICE – PARCOURIR LE GRAPHE



<1> <rouge> <2>.

<1> <rouge> <6>.

<1> <vert> <3>.

<2> <bleu> "X".

<2> <noir> <5>.

<6> <bleu> "Y".

<6> <violet> <8>.

<8> <orange> <7>.

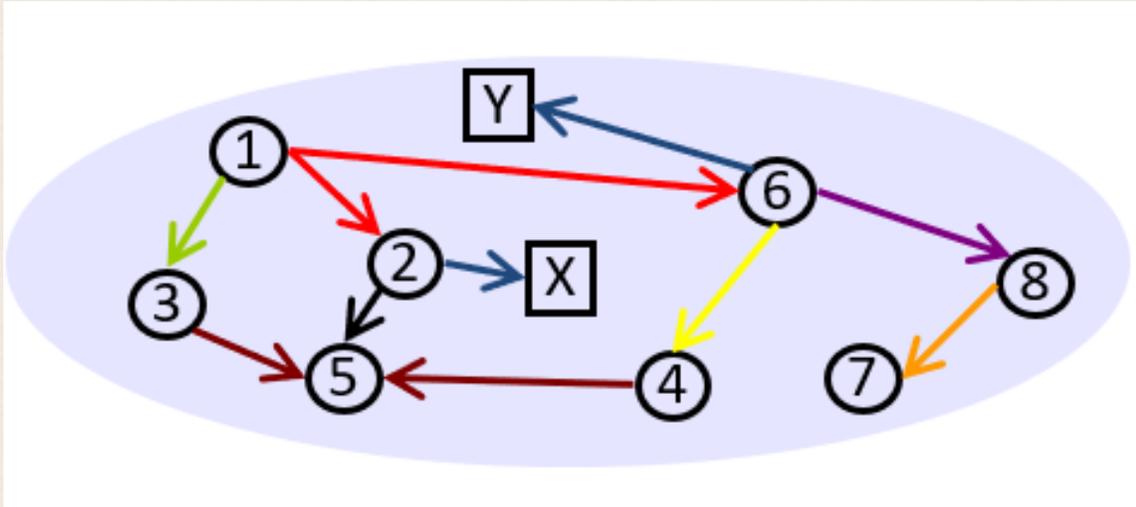
<6> <j jaune> <4>.

<4> <marron> <5>

<3> <marron> <5>.

→ Comment atteindre les ressources <2> et <6> à partir de <1> ?

EXERCICE – PARCOURIR LE GRAPHE



<1> <rouge> <2>.
<1> <rouge> <6>.
<1> <vert> <3>.
<2> <bleu> "X".
<2> <noir> <5>.
<6> <bleu> "Y".

<6> <violet> <8>.
<8> <orange> <7>.
<6> <j jaune> <4>.
<4> <marron> <5>
<3> <marron> <5>.

Comment atteindre les ressources <2> et <6> à partir de <1> ?

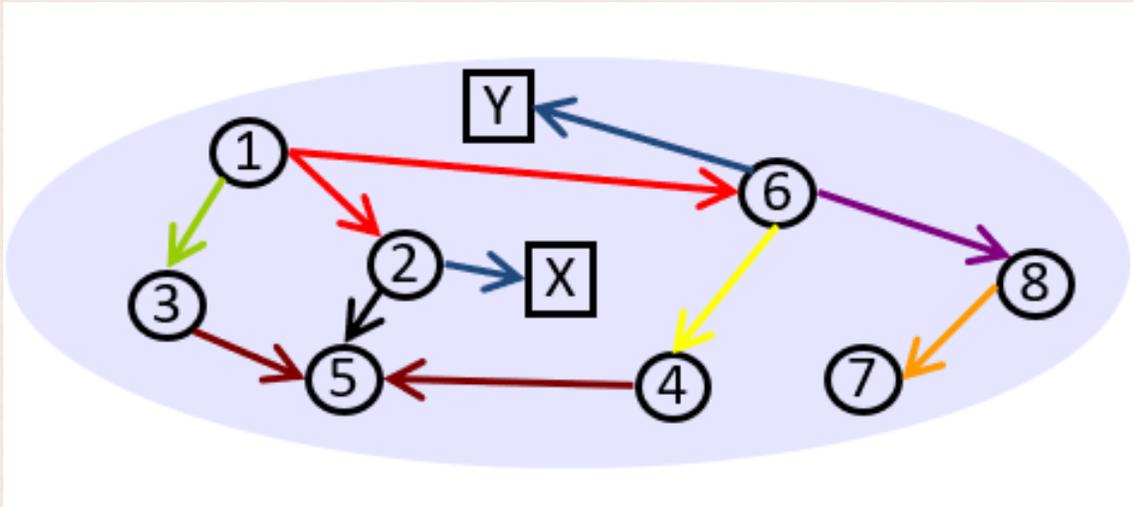
Reformuler la question

Sujet = <1>

Prédicat = ?

Objet = <2> et <6>

Exercice – Parcourir le graphe - Prédicat



<p><1> <rouge> <2>. <1> <rouge> <6>. <1> <vert> <3>. <2> <bleu> "X". <2> <noir> <5>. <6> <bleu> "Y".</p>	<p><6> <violet> <8>. <8> <orange> <7>. <6> <jaune> <4>. <4> <marron> <5>. <3> <marron> <5>.</p>
---	---

Comment atteindre les ressources <2> et <6> à partir de <1> ?

Reformuler la question

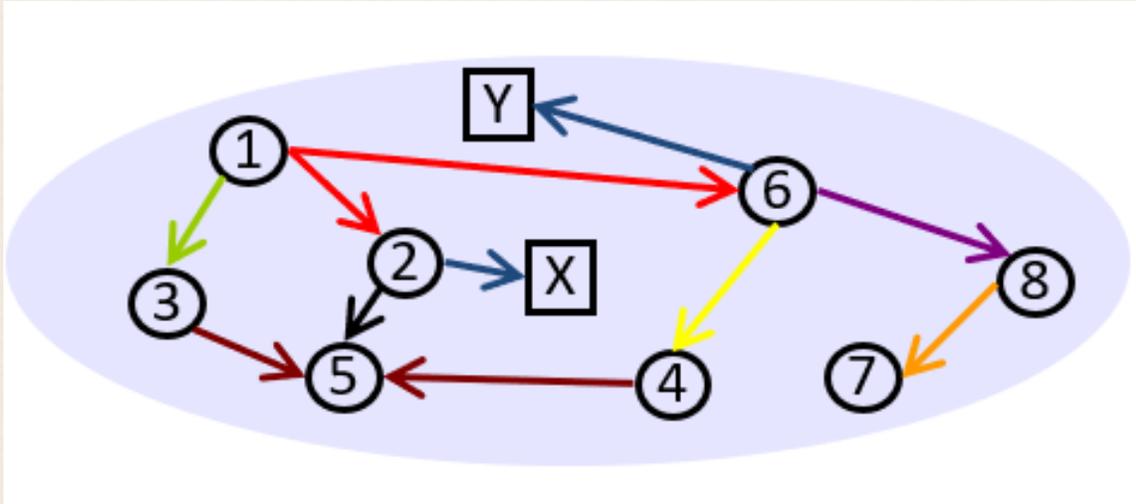
Sujet = <1>
 Prédicat = ?
 Objet = <2> et <6>



Sous forme de triplets

<1> ? <2>.
 <1> ? <6>.

EXERCICE – PARCOURIR LE GRAPHE



<1> <rouge> <2>.

<1> <rouge> <6>.

<1> <vert> <3>.

<2> <bleu> "X".

<2> <noir> <5>.

<6> <bleu> "Y".

<6> <violet> <8>.

<8> <orange> <7>.

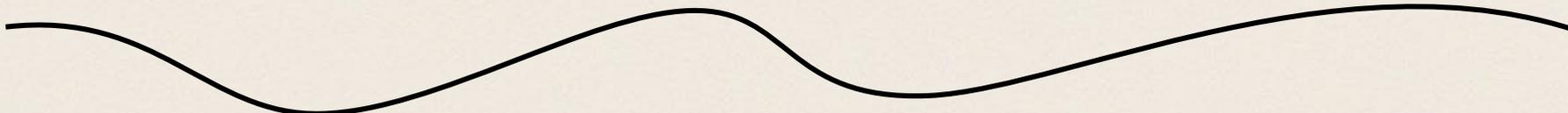
<6> <j jaune> <4>.

<4> <marron> <5>

<3> <marron> <5>.

Quelle est la ressources liée en tant qu'objet à la ressource <6> par le prédicat <violet> ?

CONSTRUIRE UNE REQUÊTE SPARQL AVEC WIKIDATA



<https://query.wikidata.org/>

 Wikidata Query Service

[Exemples](#) [Aide](#) [Davantage d'outils](#)

[français](#)

1 | (Entrez une requête SPARQL ou choisissez un exemple de requête)

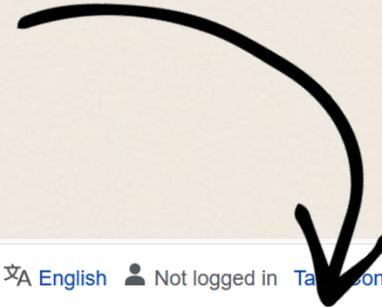




```
1 SELECT *
2
3
4 WHERE
5 {
6
7
8   SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
9
10 }
11 LIMIT 100
12
13
14
15
```

Préparer ma requête avec
l'assistant

SUR WIKIDATA, CHERCHER UN SCULPTEUR



The screenshot shows the Wikidata interface. At the top right, there are links for 'English', 'Not logged in', 'Talk', 'Contributions', 'Create account', and 'Log in'. Below this is a search bar containing the text 'rodin'. To the left of the search bar are tabs for 'Main Page' and 'Discussion', and buttons for 'Read', 'View source', and 'View history'. The search results are listed on the right side of the page:

- Auguste Rodin (*Rodin*)**
French sculptor
- Raduń (*Rodin*)**
urban-type settlement in Belarus
- 6258 Rodin (*Rodin*)**
asteroid
- Rodin**
family name
- Rodin**
2017 film by Jacques Doillon
- Rodin**
crater on Mercury
- Rodin**
Wikimedia disambiguation page

Below the search bar, there is a 'more' link and a search box for pages containing 'rodin'. In the center of the page, there is a large grey box with the text 'Welcome to Wikidata' and 'the free knowledge base with 93,808,630 data items that anyone can edit'. Below this, there are links for 'Introduction', 'Project Chat', 'Community Portal', and 'Help'. At the bottom of the grey box, there is a link that says 'Want to help translate? Translate the missing messages.' The background of the page features a network graph with red and green lines connecting various nodes.

FICHE DE L'ITEM AUGUSTE RODIN :

Auguste Rodin (Q30755)

French sculptor

Rodin | François-Auguste-René Rodin | François Auguste Rodin | Francois A. Rene Rodin | Rodan | Ogiust Roden | François Auguste René Rodin | René François Auguste Rodin | august rodin | rodin | a. rodin | rodin auguste | rodin a. | aug. rodin | e. rodin | Lo-tan

▼ In more languages

Configure

Language	Label	Description	Also known as
English	Auguste Rodin	French sculptor	Rodin François-Auguste-René Rodin François Auguste Rodin Francois A. Rene Rodin Rodan Ogiust Roden François Auguste René Rodin René François Auguste Rodin august rodin rodin a. rodin rodin auguste rodin a.

writing language French
 ▾ 0 references

occupation

- sculptor ▸ 1 reference
- graphic artist ▸ 1 reference
- drawer ▸ 2 references
- photographer ▾ 0 references



sculptor (Q1281618)

artist specializing in sculpture
 sculptress | sculptors

[In more languages](#)
[Configure](#)

Language	Label	Description	Also known as
English	sculptor	artist specializing in sculpture	sculptress sculptors
French	sculpteur	artiste pratiquant la sculpture	sculptrice
Spanish	escultor	artista que practica la escultura	escultora
German	Bildhauer	künstlerischer beziehungsweise handwerklicher Beruf	Bildhauerin Plastiker Plastikerin

[All entered languages](#)

CONSTRUCTION DE MA PHRASE :

Auguste Rodin exerce comme sculpteur

Q30755

P106

Q1281618

writing language	 French
	▼ 0 references
<u>occupation</u>	 sculptor
<small>Property:P106</small>	▶ 1 reference
	 graphic artist
	▶ 1 reference
	 drawer
	▶ 2 references
	 photographer
	▼ 0 references

Property [Discussion](#) [Read](#) [View history](#) Q

instance of (P31)

that class of which this subject is a particular example and member

is a | is an | has class | has type | is a particular | is a specific | is an individual | is a unique | is an example of | member of | unique individual of | distinct member of | unitary element of class | distinct element of | distinct individual member of | rdf:type | type | main type | is a(n) | type of | is a type of

[In more languages](#)
Configure

Language	Label	Description	Also known as
English	instance of	that class of which this subject is a particular example and member	is a is an has class has type is a particular is a specific is an individual is a unique is an example of member of unique individual of distinct member of

QUELLE EST LA PHRASE?

Item [Discussion](#) [Read](#) [View history](#) Q

human (Q5)

common name of Homo sapiens, unique extant species of the genus Homo

human being | humankind | people | homosapiens | person | mankind | peoplekind | personkind | persons | humans

[In more languages](#)
Configure

Language	Label	Description	Also known as
English	human	common name of Homo sapiens, unique extant species of the genus Homo	human being humankind people homosapiens person mankind peoplekind personkind persons humans
French	être humain	individu appartenant à l'espèce Homo sapiens, la seule espèce restante du genre Homo – distinct de « humain fictif » et de « humain possiblement fictif »	etre humain Homme Homo sapiens sapiens Homo sapiens

Auguste Rodin est un humain

RECONSTRUIRE LE GRAPHE POUR TROUVER LES SCULPTEURS

?nom wdt:P31 wd:Q5. #chose1 qui est un humain

?nom wdt:P569 ?date_naissance. # chose1 est né le chose2

?nom wdt:P19 wd:Q90. # chose1 est né à Paris

?nom wdt:P106 wd:Q1281618. # chose1 exerce comme sculpteur

Une phrase se termine toujours par un point.

```

1 #La représentation des sculpteurs parisiens (1500-1800) dans les musées internationaux
2
3 SELECT ?nom ?nomLabel ?date_naissance ?collection ?collectionLabel
4
5
6 WHERE
7 {
8 #Recherche sur le sculpteur
9   ?nom wdt:P31 wd:Q5. #choses1 qui est un humain
10  ?nom wdt:P569 ?date_naissance. # choses1 est né le choses2
11  ?nom wdt:P19 wd:Q90. # choses1 est né à Paris
12  ?nom wdt:P106 wd:Q1281618. # choses1 exerce comme sculpteur
13
14
15
16 #Recherche sur les collections
17   ?nom wdt:P6379 ?collection. # choses1 a son travaille dans la collection de choses3
18
19 #Filtrage sur les dates de naissances des sculpteurs pour arriver ) l'époque qui nous intéresse
20 filter (?date_naissance > "1500-01-01"^^xsd:dateTime && ?date_naissance < "1800-01-01"^^xsd:dateTime)
21
22   SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en,fr". }
23
24 }
25

```

LES RÉSULTATS

nom	nomLabel	date_naissance	collection	collectionLabel
Q wd:Q289246	Étienne Maurice Falconet	1 décembre 1716	Q wd:Q132783	musée de l'Ermitage
Q wd:Q289246	Étienne Maurice Falconet	1 décembre 1716	Q wd:Q214867	National Gallery of Art
Q wd:Q289246	Étienne Maurice Falconet	1 décembre 1716	Q wd:Q682827	The Frick Collection
Q wd:Q289246	Étienne Maurice Falconet	1 décembre 1716	Q wd:Q731126	J. Paul Getty Museum
Q wd:Q289246	Étienne Maurice Falconet	1 décembre 1716	Q wd:Q1700481	Minneapolis Institute of Art
Q wd:Q289246	Étienne Maurice Falconet	1 décembre 1716	Q wd:Q1976985	musée d'art Nelson-Atkins
Q wd:Q322582	Paul Delaroche	17 juillet 1797	Q wd:Q49133	musée des Beaux-Arts de Boston
Q wd:Q322582	Paul Delaroche	17 juillet 1797	Q wd:Q132783	musée de l'Ermitage
Q wd:Q322582	Paul Delaroche	17 juillet 1797	Q wd:Q153306	musée national de Varsovie

DIFFÉRENTS TYPES DE VISUALISATION DU RÉSULTAT

The screenshot shows a web application interface for data visualization. At the top right, there is a refresh icon and a status bar indicating "260 résultats en 1258 ms". Below this, there are buttons for "Code", "Télécharger", and "Lien". A search bar is also present. The main content is a table with the following columns: "nomLabel", "date_naissance", "collection", and "collectionLabel". The table contains 10 rows of data, including names like "Etienne Maurice Falconet" and "Paul Delaroche", along with their birth dates and the museums where their works are held. On the left side, a menu is open, listing various visualization types: Table, Image grid, Graph builder, Map, Line chart, Bar chart, Scatter chart, Area chart, Bubble chart, Tree map, Tree, Timeline, Dimensions, and Graph. The "Table" option is currently selected.

nomLabel	date_naissance	collection	collectionLabel
Etienne Maurice Falconet	1 décembre 1716	wd:Q132783	musée de l'Ermitage
Etienne Maurice Falconet	1 décembre 1716	wd:Q214867	National Gallery of Art
Etienne Maurice Falconet	1 décembre 1716	wd:Q682827	The Frick Collection
Etienne Maurice Falconet	1 décembre 1716	wd:Q731126	J. Paul Getty Museum
Etienne Maurice Falconet	1 décembre 1716	wd:Q1700481	Minneapolis Institute of Art
Etienne Maurice Falconet	1 décembre 1716	wd:Q1976985	musée d'art Nelson-Atkins
Paul Delaroche	17 juillet 1797	wd:Q49133	musée des Beaux-Arts de Boston
Paul Delaroche	17 juillet 1797	wd:Q132783	musée de l'Ermitage
Paul Delaroche	17 juillet 1797	wd:Q153306	musée national de Varsovie
Paul Delaroche	17 juillet 1797	wd:Q160112	musée du Prado

CRÉATION DU GRAPHIQUE

Graph builder ? 260 résultats en 12 ms </> Code ↓ Télécharger 🔗 Lien

polesstar 🔖 Bookmarks (0) ↶ Undo ↷ Redo 📄 Export 🔍 SWAP X/Y 🔖 BOOKMA

Encoding 🗑️ Clear

Positional

x ▼ **A** collectionLabel ✕

y ▼ **A** nomLabel ✕

column

row

Marks point ▼

size

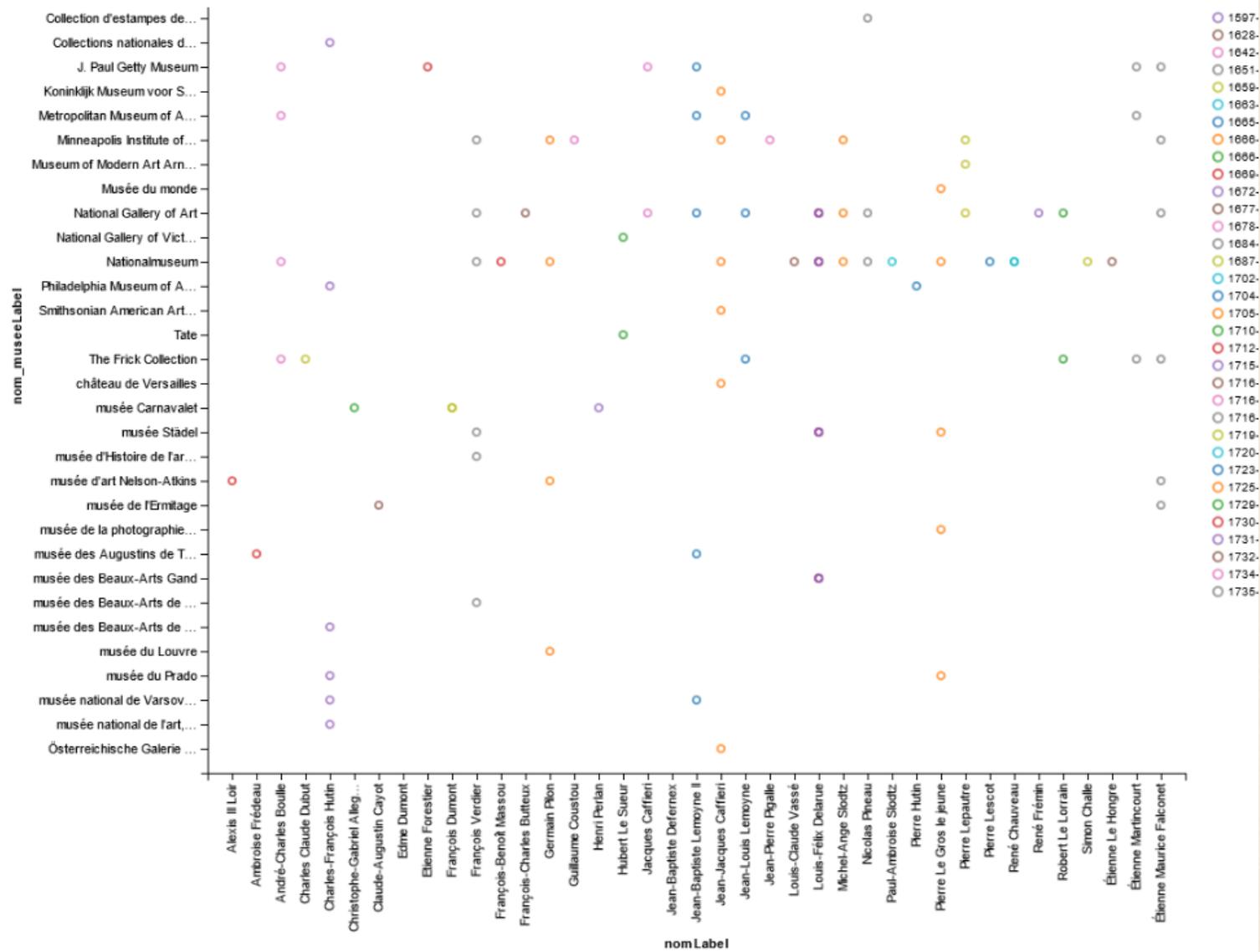
color ▼ **A** date naissance ✕

shape

detail

text

nomLabel	nomLabel
François Guernon dit Bel...	
François-Alexandre Verdi...	
François-Benoît Massou	
François-Nicolas Delaist...	
Félix Lecomte	
Germain Pilon	
Guillaume Coustou the Yo...	
Henri Perlan	
Hubert Le Sueur	
Isidore Hippolyte Nicola...	
Jacques Caffieri	
Jacques-Augustin Dieudon...	
Jacques-Edme Dumont	
Jacques-Édouard Gatteaux	
Jean Guillaume Moitte	
Jean Roller	
Jean-Antoine Alavoine	
Jean-Baptiste Defernex	
Jean-Baptiste Lemoyne	
Jean-Baptiste Stouf	



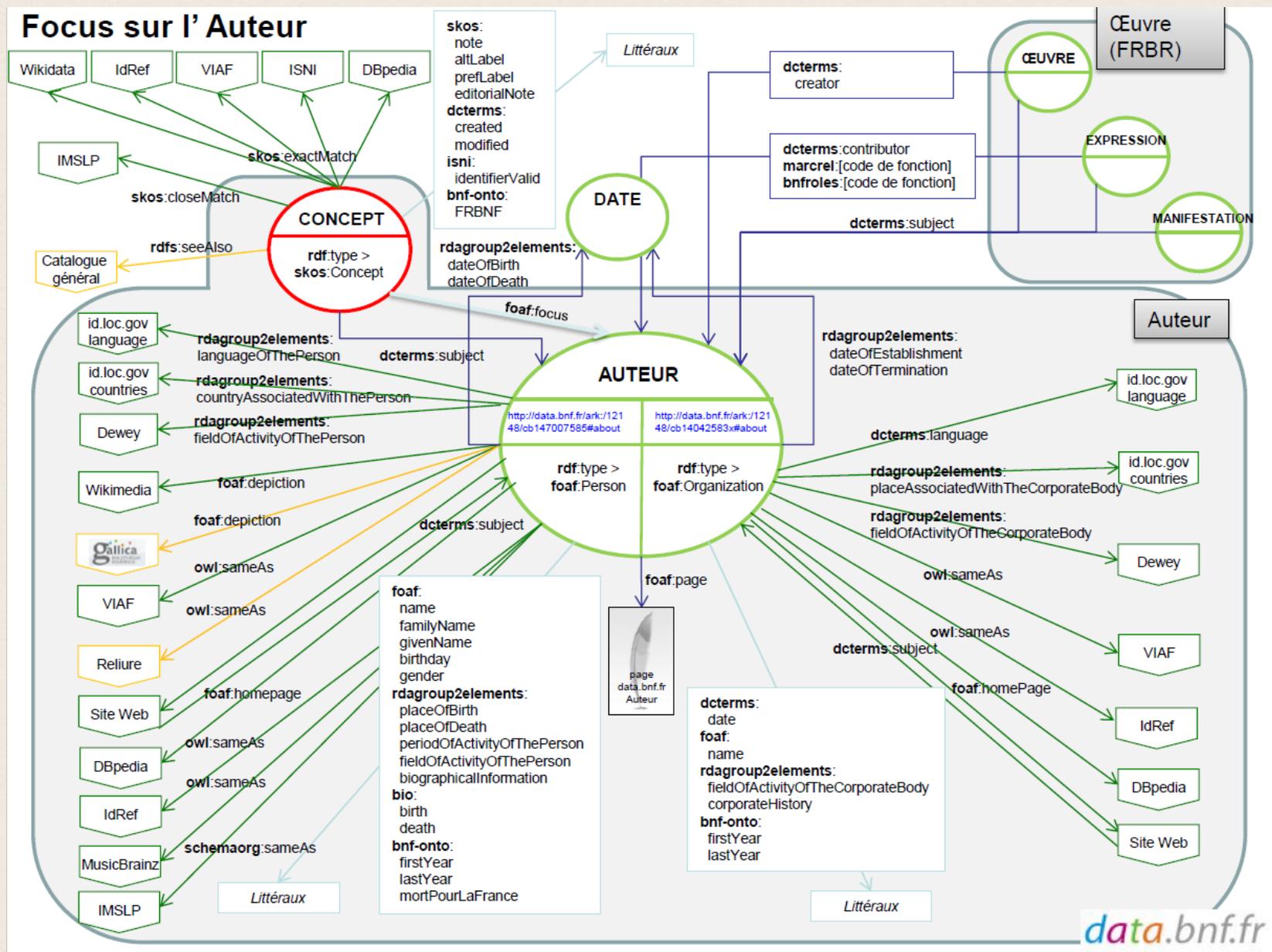
DATA.BNF



DATA.BNF

- | | |
|---|---|
| <ul style="list-style-type: none">• Projet de 2011• Base de données sémantique• Fiches de référence sur :<ul style="list-style-type: none">✓ Les œuvres✓ Les auteurs✓ Les thèmes du catalogue BnF✓ Les lieux | <ul style="list-style-type: none">✓ Les dates✓ Les spectacles✓ Les périodiques• Liens vers les documents numérisés sur Gallica• Données librement accessibles à condition d'en donner la source• SPARQL endpoint |
|---|---|

Focus sur l'Auteur



Le modèle de données data.bnf

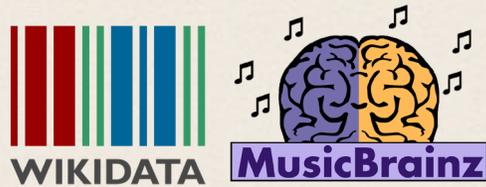
MASHUP DE DONNÉES

- Création originale mise au point à partir de l'assemblage de créations existantes.
- « The real power of semantic Web will be realized when people create many programs that collect Web content from **diverses sources**, **process the information** and **exchange** the results with other programs. »

in Tim Berners Lee, Ora Lassila, James Hendler,
« Semantic Web », Scientific american, 2001

APPORTS DES TECHNOLOGIES DU WEB SÉMANTIQUE

La place centrale
des référentiels



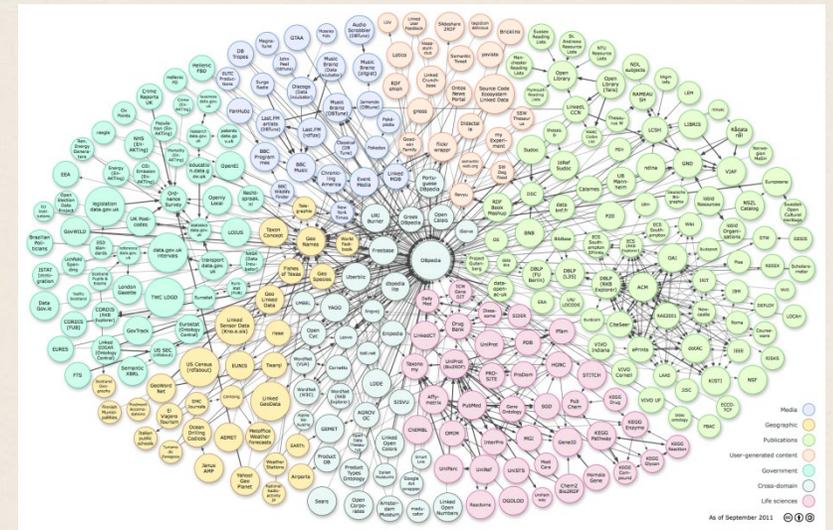
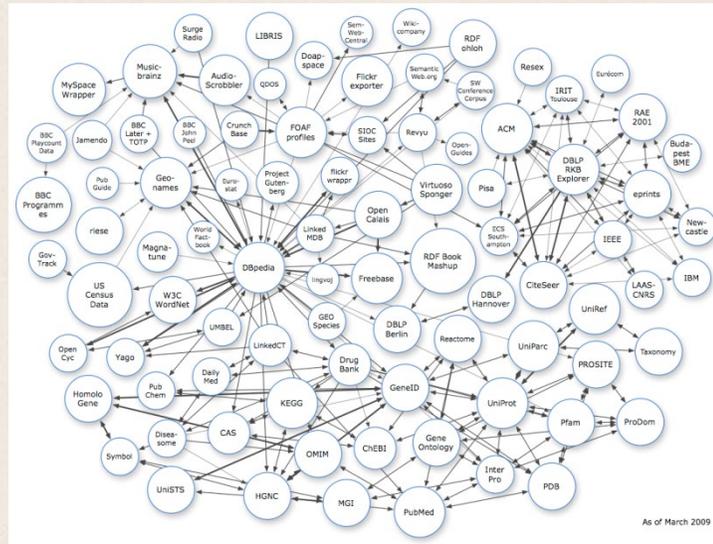
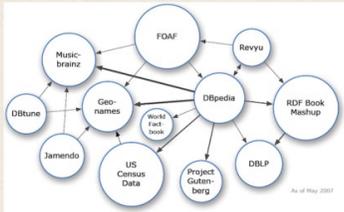
Remise en cause des
modèles existants



Libérer les données
des usages



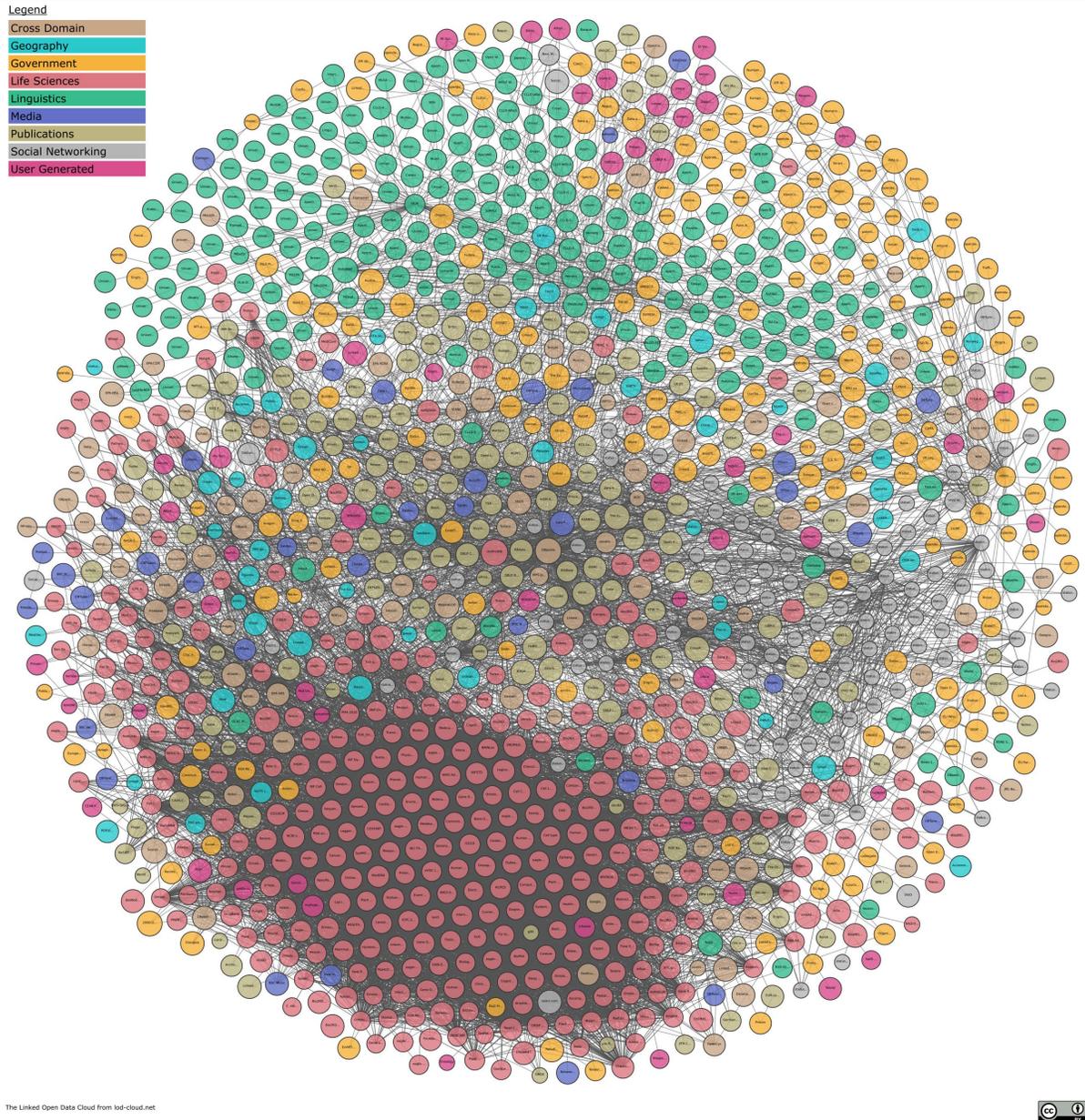
APPLICATION DES TECHNOLOGIES DU WEB SÉMANTIQUE POUR PARTAGER DE LA DONNÉE SUR LE WEB



2007

2009

2011



The Linked Open Data Cloud from lod-cloud.net



2021

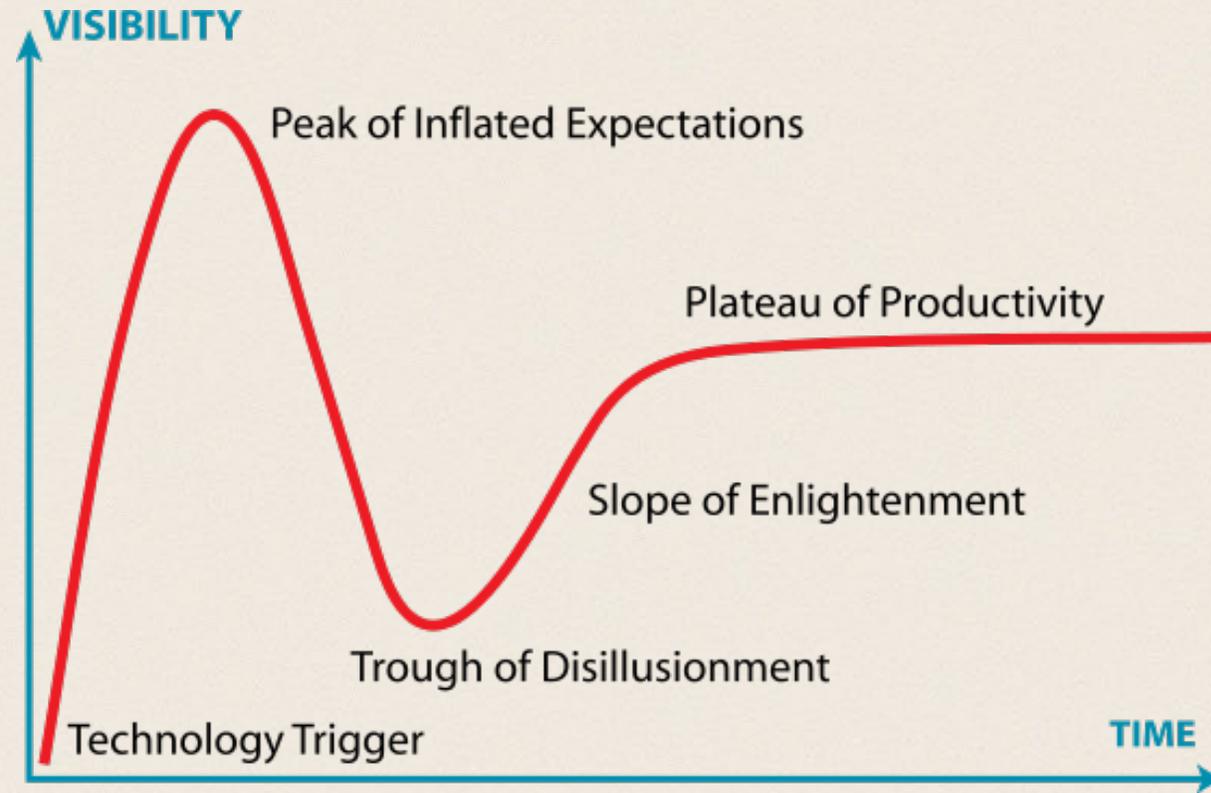
LES LIMITES DU WEB DE DONNÉES

L'exercice consiste ici à récupérer les données exposées selon les principes du Linked Data ou à travers un sparql endpoint

Ici nous nous sommes volontairement limité à une récupération simple des données. Les technologies du web de données posent question et ne tiennent plus vraiment leurs promesses.

« si vous souhaitez disposer d'une application disponible, rapide et scalable, vous devez récupérer les données de manière asynchrone, les traiter puis les stocker dans une base pour les exploiter localement. **La promesse de décentralisation est largement écornée** et cela oblige à mettre en place des mécanismes lourds pour effectuer la mise à jour des données qui ne peut se faire en temps réel. » Gautier Poupeau, architecte de données à l'Ina

UNE TECHNOLOGIE QUI NE S'EST PAS GÉNÉRALISÉE



EXERCICE

1. requêter les hommes politiques français et les faire apparaître comme label
2. requêter leurs pères et les faire apparaître comme label
3. requêter les dates de naissance et les faire apparaître puis filtrer les hommes politiques nés après la date de votre choix

POUR ALLER PLUS LOIN...



LISTE DES RÉFÉRENCES UTILES

- Le tutoriel SPARQL de wikidata :
https://www.wikidata.org/wiki/Wikidata:SPARQL_tutorial (en.)
- Querying Wikidata with SPARQL for Absolute Beginners :
<https://www.youtube.com/watch?v=kJph4q0Im98> (en.)
- Wikimedian in Residence Channel :
<https://media.ed.ac.uk/channel/Wikimedian+in+Residence+Channel/51020161> (en.)
- Les recommandations du SPARQL working group du W3C :
https://www.w3.org/2009/sparql/wiki/Main_Page (en.)
- Le modèle de données de data.bnf.fr : <https://data.bnf.fr/fr/semanticweb>

LISTE DES RÉFÉRENCES UTILES

- Pour une réflexion sur le web sémantique et ses usages :
<http://www.lespetitescases.net/index355>
- Les ressources pédagogiques de l'ANF « Initiation au SPARQL et à la diffusion de données en RDF » (2020) :
<http://rbdd.cnrs.fr/spip.php?article337>
 - ✓ Notamment : FRAN CART, Thomas, [sans date]. [Découvrir les données de data.bnf.fr à travers SPARQL](#). In : . 2020-12-08.
- Tutoriel « pas-pas » pour créer un corpus avec wikidata :
https://github.com/Humanistica/ArtDesignDH/blob/master/tutoriels/tutoriel_cree_corpus_avec_wikidata.md