

Service du  
numérique



MINISTÈRE  
DE LA CULTURE

*Liberté  
Égalité  
Fraternité*

# Pas à pas vers le Graphe Culture

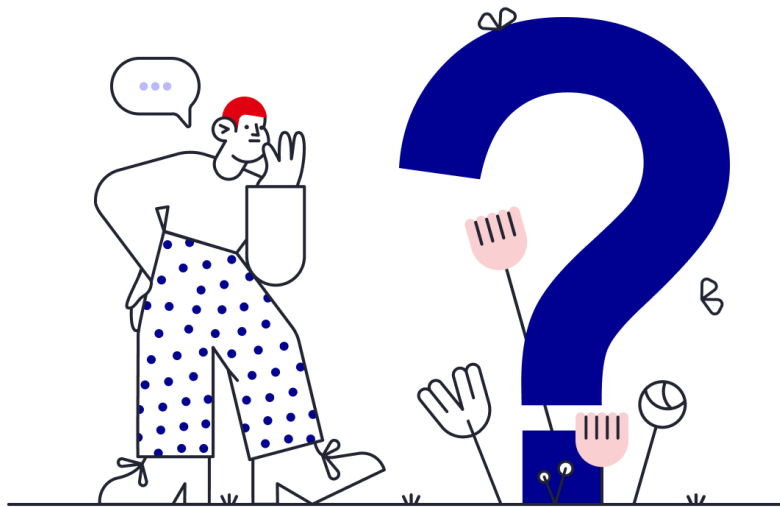
Marie-Véronique LEROI

SemWeb Pro 2022





# Une stratégie pour la sémantisation des données culturelles



# Les espaces communs européens de données



**MINISTÈRE  
DE LA CULTURE**

*Liberté  
Égalité  
Fraternité*





# Feuille de route « Données et contenus culturels »

<https://www.culture.gouv.fr/content/download/295347/3378042?version=15>

Améliorer l'action  
publique grâce à la  
donnée

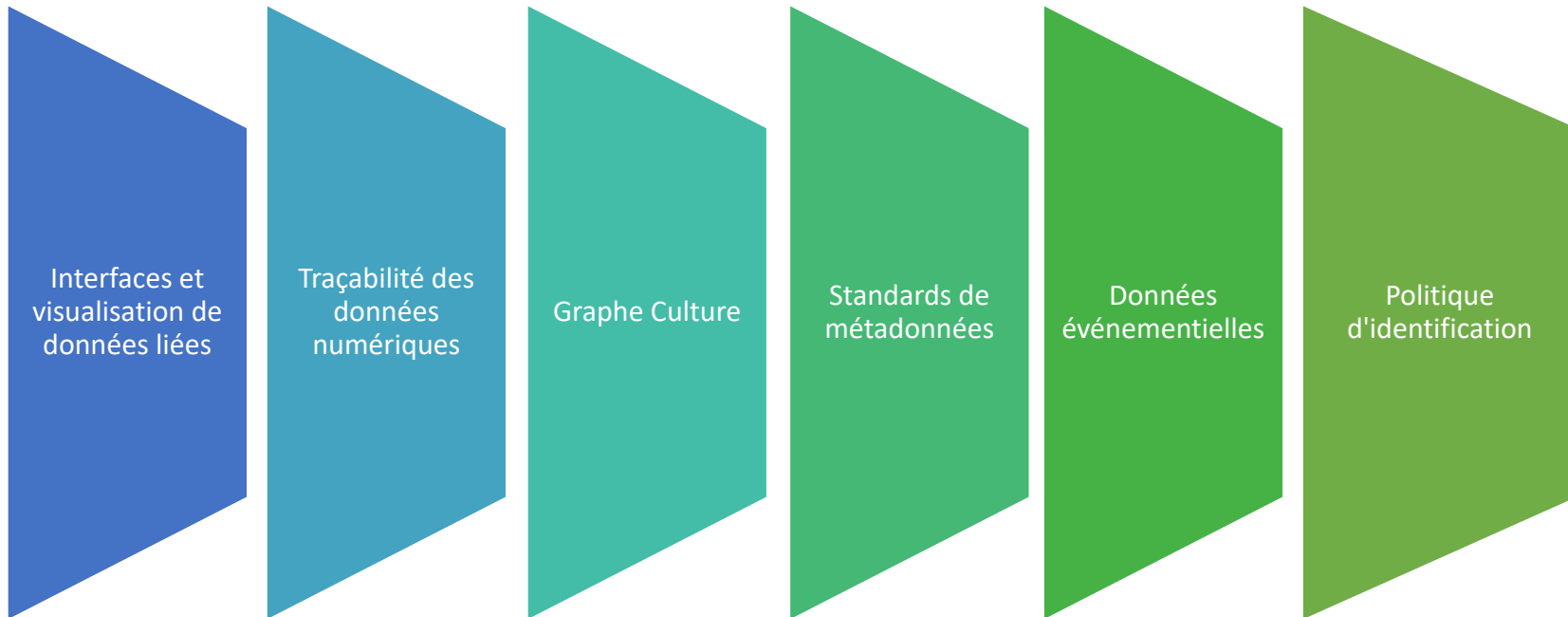
Faciliter et valoriser  
l'ouverture, le partage et  
l'exploitation des  
données

Développer le capital  
humain : acculturation  
et gestion des  
compétences

Mobiliser et évaluer :  
gouvernance et pilotage



# Une stratégie pour la sémantisation des données culturelles en cours de refonte



# Stratégie « Métadonnées culturelles et transition Web 3.0 » (2014)



**MINISTÈRE  
DE LA CULTURE**

*Liberté  
Égalité  
Fraternité*

## Principe ?

- Appliquer les principes du Web des données liées aux données culturelles
- Déployer les standards et les technologies sémantiques chez les acteurs culturels

## Comment ?

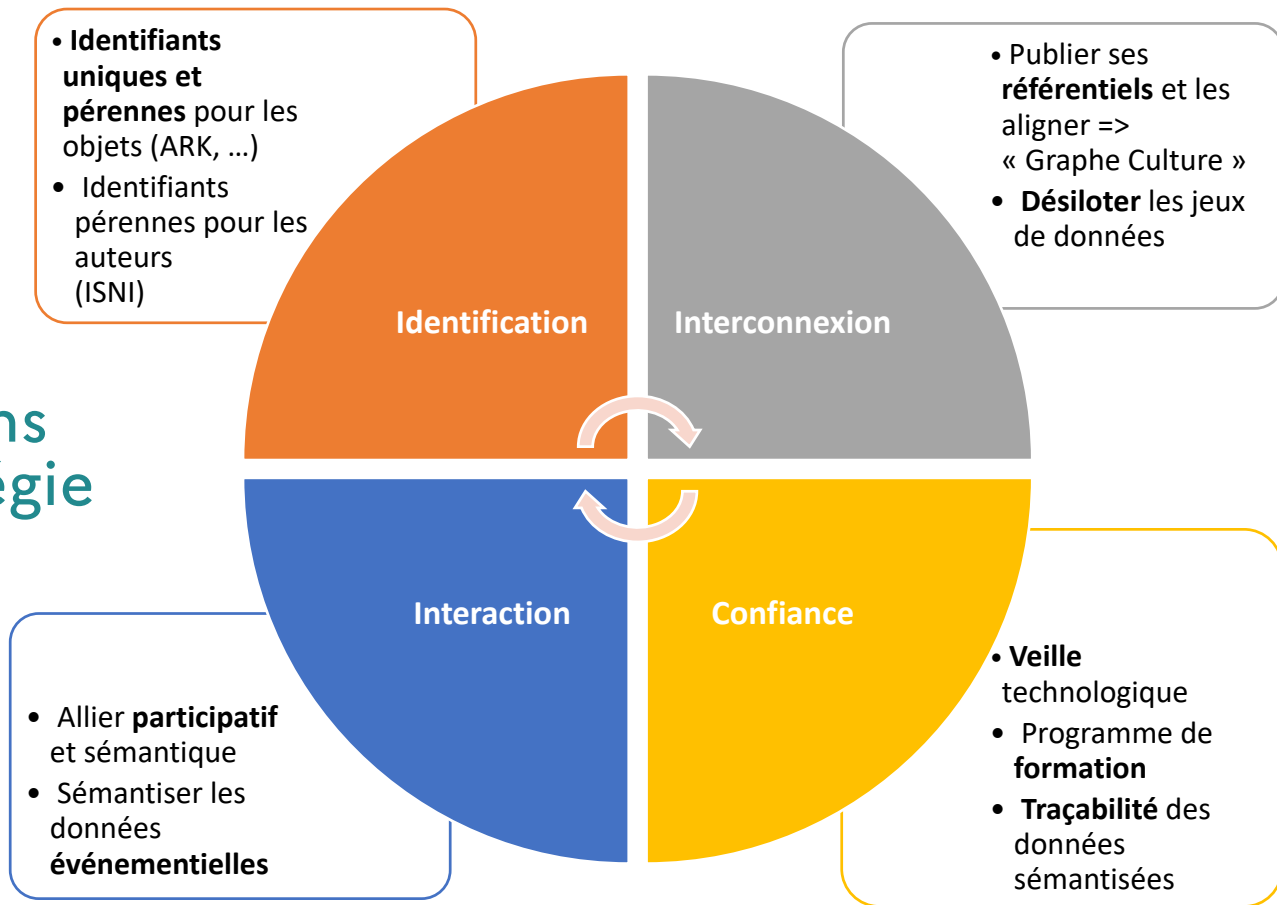
- Groupes de travail impliquant des personnes référentes de chaque secteur culturel
- Aspects techniques, technologiques mais aussi organisationnels
- L'administration centrale mais aussi les services déconcentrés et les établissements publics
- 9 actions opérationnelles pour l'ouverture et le partage des données publiques culturelles

## Pourquoi ?

- Constituer le « Graphe Culture »
- Doter le Ministère de la Culture de son propre graphe de connaissances

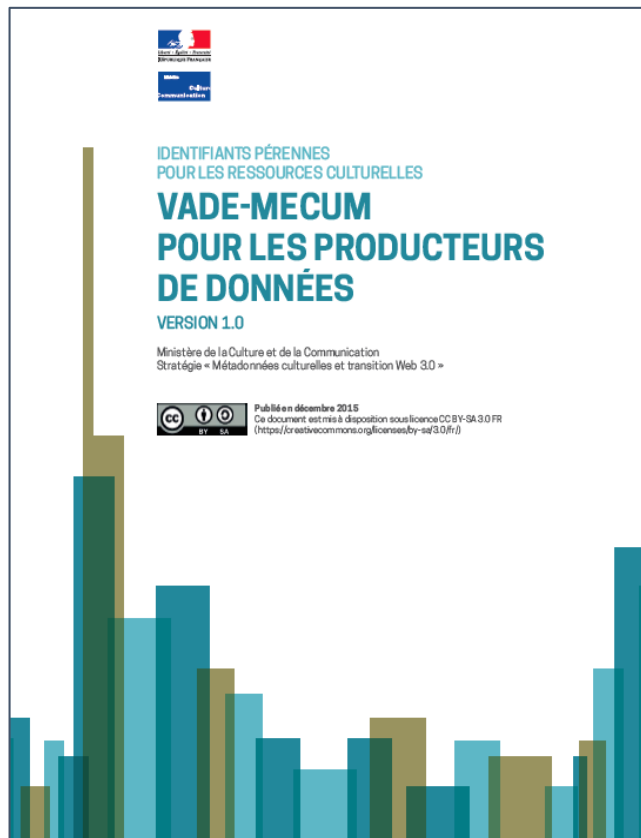


## Les actions de la stratégie





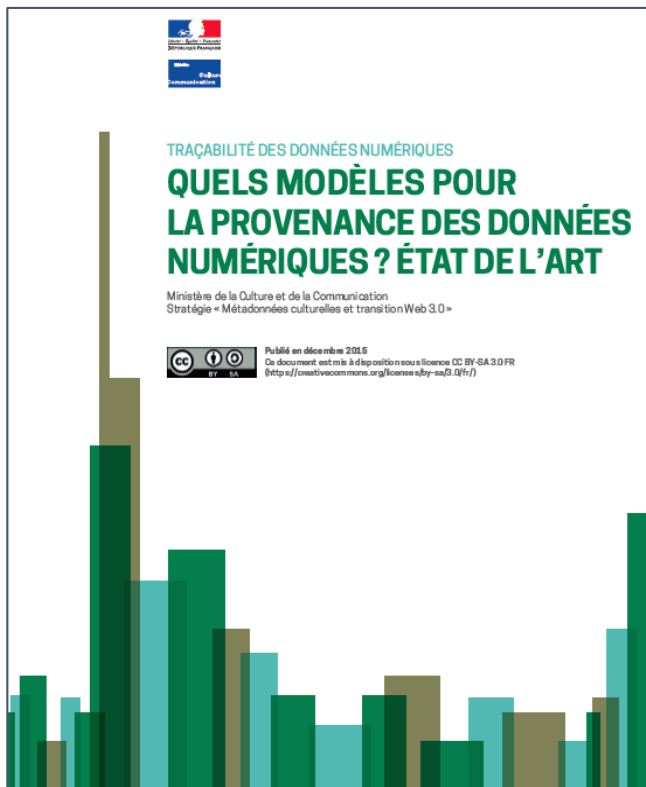
# Publications







# Publications



# Des projets de sémantisation



MINISTÈRE  
DE LA CULTURE

*Liberté  
Égalité  
Fraternité*



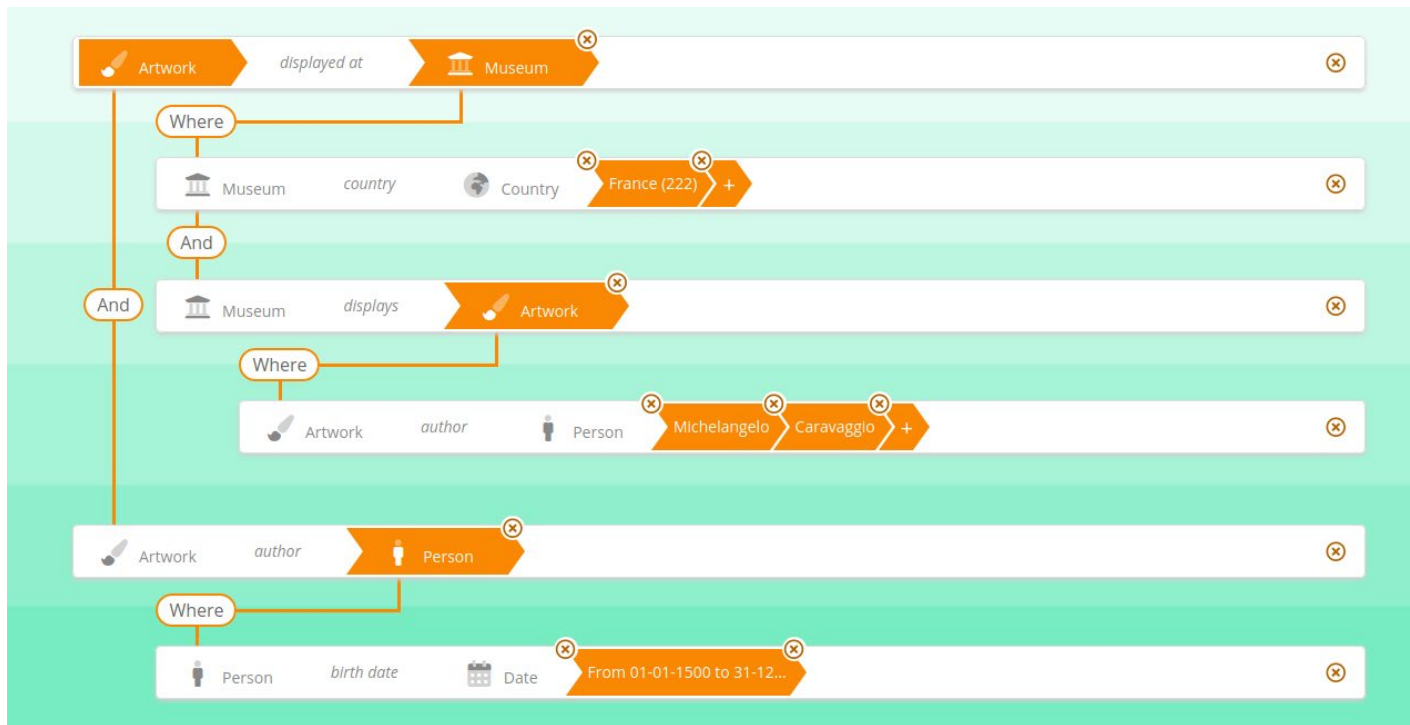
{ BnF

Mucem



# Des briques essentielles

SPARNATURAL





# Des briques essentielles



## Le projet

DBpedia en français est le chapitre francophone de DBpedia, il s'inscrit dans l'effort d'internationalisation de DBpedia dont le but est de maintenir des données structurées extraites de différents chapitres de Wikipedia.

Le développement de DBpedia en français est un projet mené dans le cadre de la convention entre le Ministère de la Culture et Inria et porté par l'équipe Wimmics commune à l'Inria et au laboratoire I3S (UCA / CNRS).

Service du  
numérique



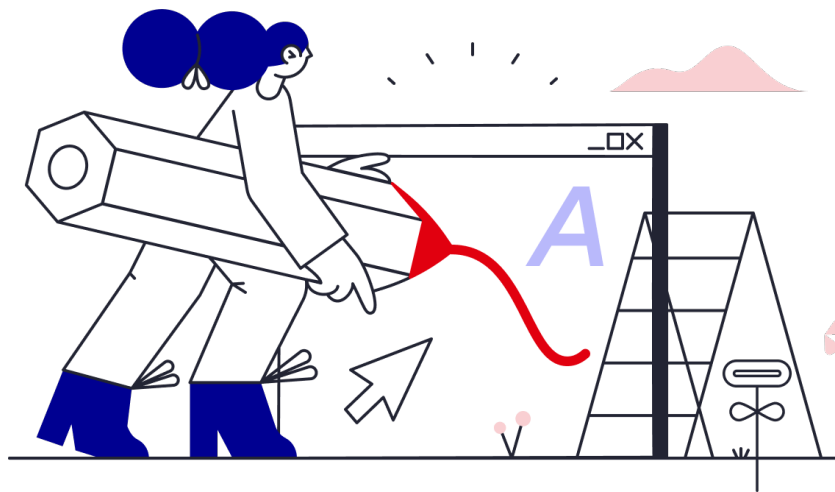
MINISTÈRE  
DE LA CULTURE

*Liberté  
Égalité  
Fraternité*

# Le projet DBpedia en français



Fabien GANDON  
Franck MICHEL  
Célian RINGWALD





# SemanticPedia



## ▪ Un partenariat signé en 2012

- Ministère de la Culture
- Inria
- Wikimedia

## ▪ L'alignement a permis

- Lancement de Dbpedia en français
- Une preuve de concept





MINISTÈRE  
DE LA CULTURE

Liberté  
Égalité  
Fraternité

## ■ Une première expérimentation au ministère de la Culture:

- Alignement des vocabulaires de la base Joconde avec DBpedia (volet sémantique de Wikipedia)
- Co-pilotage SMF et DGLFLF
- **L'alignement a permis**
  - l'Intégration des cartouches de Wikipedia
  - une interface de consultation multilingue
  - D'autres modes de navigation dans la base Joconde

Français English Italiano Español Deutsch Português العربية Русский 简体中文 日本語 Català Euskara Brezhoneg Occitan

JocondeLab » Accueil

Bienvenue sur JocondeLab

À propos

Tout savoir sur  
l'expérimentation  
JocondeLab.

Carte du monde

Découvrez le monde à  
travers les collections des  
musées de France.

Chronologie

Découvrez les collections  
des musées de France au  
fil du temps (période ou  
date précise).

Mots-clés

Choisissez un artiste, un  
sujet représenté, un  
domaine ou une époque  
d'un simple clic.

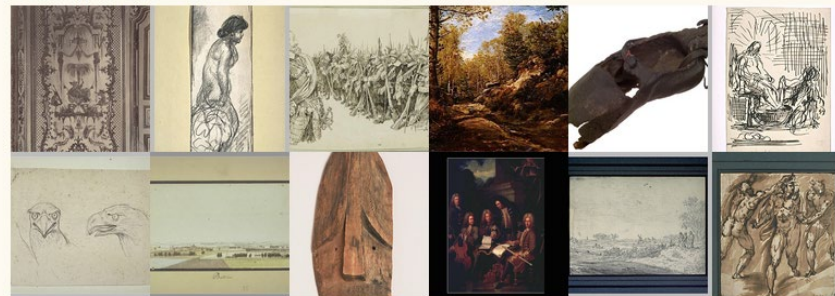
Contribuer

Aidez-nous à décrire le  
sujet d'une œuvre à l'aide  
de Wikipédia.

Galerie de créations

Les travaux des étudiants  
des GOBELINS.

298597 œuvres des musées de France



Mentions légales • Crédits •



Ce site vous est proposé par



et






MINISTÈRE  
DE LA CULTURE

Liberté  
Égalité  
Fraternité

- **Exemple de recherche**
  - On trouve une œuvre représentant le personnage d'Esmeralda dans le roman « Notre-Dame de Paris » de Victor Hugo
  - Les éléments en bleu sont les index des vocabulaires de la base Joconde alignés avec DBpedia
- **L'alignement a permis**
  - Des résultats de recherche enrichis
  - Une traduction automatique des index en bleu, disponibles dans les 14 langues du site.

JocondeLAB » Détail d'une notice

Accueil À propos Carte du monde Chronologie Mots-clés Contribuer Galerie de créations



**Artiste(s) :** Anonyme

**Titre :** Portrait de femme

Domaine(s) : Textile

Désignation : **tableau (octogonal)** | tableau tissé

Sujet représenté : Botte (chaussure) Chèvre Esmeralda (personnage) Femme Feuille Lit (meublier) Tambourin (sur cadre) cuirlande orn

Source : **Notre-Dame de Paris (roman)**

Datation : XIXe siècle

Techniques : sole, tissé

Mesures : hauteur en cm 14.1 ; largeur en cm 7.3 ; hauteur hors tout en cm 86 ; largeur hors tout en cm 69.5 ; épaisseur hors tout en cm 3.5

Conservé à : Saint-Etienne ; musée d'Art et d'Industrie

Copyright notice : © Saint-Etienne, musée d'Art et d'Industrie, © Service des musées de France, 2011

Crédits photographiques : © SARA

[Cette œuvre sur le portail Joconde](#)

Ajoutez vos mots-clés Les termes proposés proviennent de Wikipédia.





# Projet Dbpedia en français

- Lot 1 : Pérenniser et augmenter la visibilité du chapitre Francophone de DBpédia
  - > Automatiser les tâches d'administration et de maintenance
  - > Amélioration des moyens de dissémination et d'adoption
  - > Déploiement en production de DBpédia Live et DBpédia Historique
- Lot 2 : Évaluation de la fertilisation croisée de DBpédia avec les autres projets de la fondation Wikimedia (en particulier Wikidata)
- Lot3 : Développement de nouveaux axes de recherches
  - > Application des techniques de TAL au texte des articles de Wikipédia
  - > DBpédia Spotlight Français



# Impact sur les outils DBpedia



Entités nommées



Annotation  
d'entités nommées

Check out the DBpedia Databus NOW!

```
#####  
# # ## ##### ## ##### # # ####  
# # # # # # # # # # # #  
# # # # # # # ##### # # ####  
# # ##### # ##### # # # #  
# # # # # # # # # # # # #  
##### # # # # # ##### #### ####
```

Catalogue et versionning

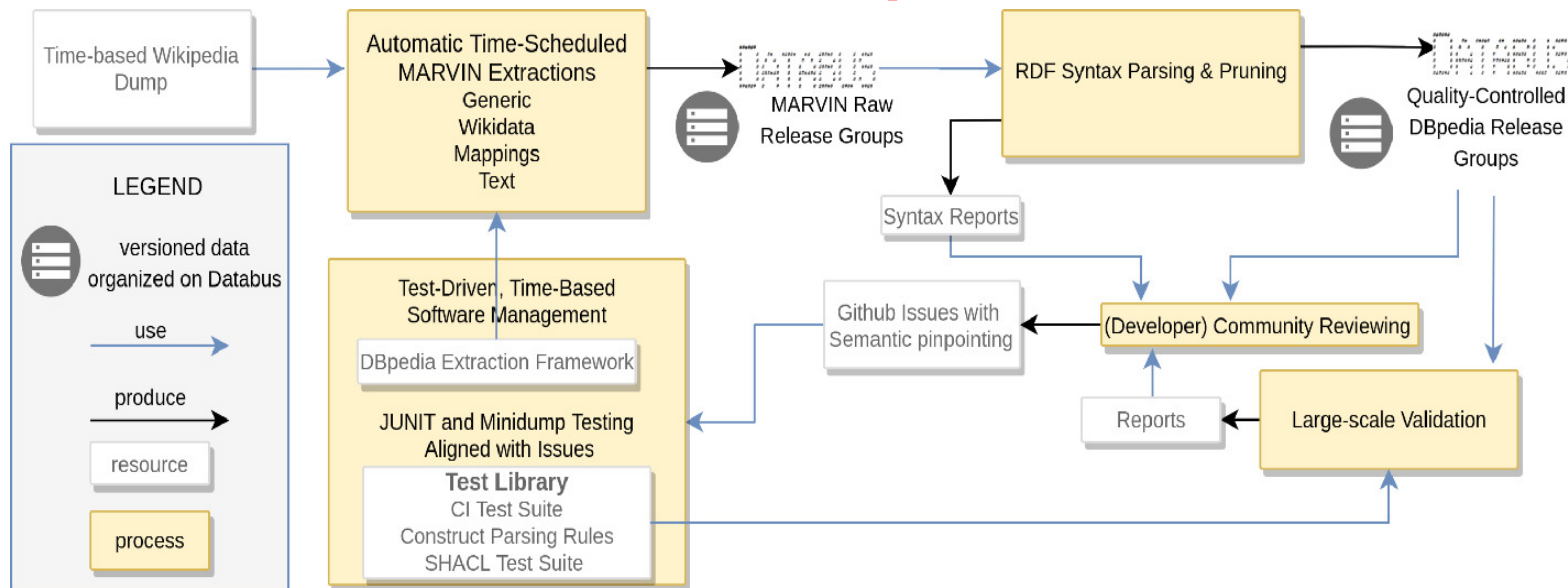


Extraction instantanée de  
Wikipedia et autres



Historique  
des articles

# Actualisation de DBpedia



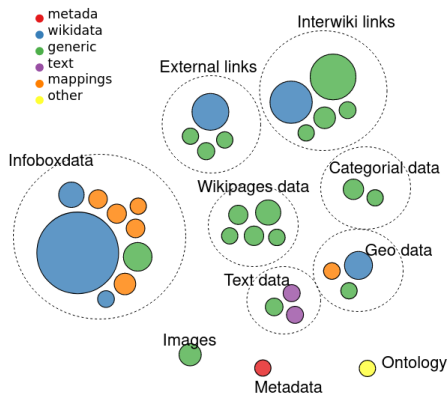
- Un nouveau cycle de publication
  - > Contrôle qualité des données
  - > Suivi des extractions
  - > Publication automatique sur le DataBus

# Raffinage SPARQL



MINISTÈRE  
DE LA CULTURE

Liberté  
Égalité  
Fraternité

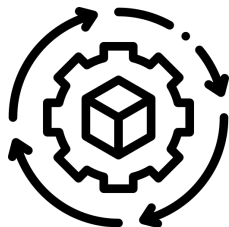


1. Meilleur suivi de la source des données : structuration sous forme de graphes nommés
2. Amélioration de la lisibilité des données : réification des données géographiques
3. Intégration des données Wikidata
4. Redirection des labels multilingues vers les entités françaises (17 langues)
5. Calcul de statistiques intermédiaires
6. Nettoyage des données inutiles (wikidata sans plus-value, labels inutilisés)
7. Calcul de statistiques descriptives
8. Enregistrement et publication des jeux de données enrichis sur le databus



# Publication itérative

- 1. 2016 : 130 millions de triplets (données extraites via le DIEF)
- février 2022 : 250 millions de triplets (données issues du databus + stats+ métadonnées)
- mars 2022 : 681 millions de triplets (+wikidata non nettoyées)
- **septembre 2022 : 695 millions de triplets (images + abstracts + multilang) > non publiés**
- octobre 2022 : ~ 330 millions de triplets (nettoyage wikidata)





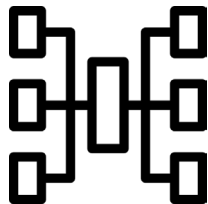
# De nouveaux outils pour intégrer Wikidata

2019

Frey, J., Hofer, M., Obraczka, D., Lehmann, J., and Hellmann, S. (2019) DBpedia FlexiFusion the Best of Wikipedia > Wikidata > Your Data.. In *ISWC (2)* (Ghidini, C., Hartig, O., Maleshkova, M., Svátek, V., Cruz, I. F., Hogan, A., Song, J., Lefrançois, M., and Gandon, F., Eds.), pp. 96–112, Springer.[URLBibTeXEndNoteBibSonomy](#)

2018

Ismayilov, A., Kontokostas, D., Auer, S., Lehmann, J., and Hellmann, S. (2018) Wikidata through the Eyes of DBpedia, *Semantic Web Journal* (Hogan, A., Ed.) 9, 493–503.[BibTeXEndNoteDOIIBibSonomy](#)





# Comparaison des descripteurs des entités de type Event



WIKIDATA



2 propriétés communes attenantes à des lieux

31 propriétés atypiques  
plutôt liées à des évènements  
historiques

<http://dbpedia.org/ontology/follows>  
<http://dbpedia.org/ontology/followedBy>  
<http://dbpedia.org/ontology/foundedBy>

54 propriétés atypiques  
plutôt liées à des évènements  
sportifs

<http://dbpedia.org/ontology/followingEvent>  
<http://dbpedia.org/ontology/previousEvent>  
<http://dbpedia.org/ontology/sport>



# Comparaison des descripteurs des entités de type Organisation



WIKIDATA



30 Propriétés communes attenantes à des lieux + créateur et employés  
« illustres »

49 propriétés atypiques

<http://dbpedia.org/ontology/namedAfter>

<http://dbpedia.org/ontology/award>

<http://dbpedia.org/ontology/follows>

51 propriétés atypiques

<http://xmlns.com/foaf/0.1/homepage>

<http://dbpedia.org/ontology/nationality>

<http://dbpedia.org/ontology/legalForm>





# Comparaison des descripteurs des entités de type Person



WIKIDATA



20 propriétés communes basiques

13 propriétés atypiques

<http://dbpedia.org/ontology/almaMater>

<http://dbpedia.org/ontology/employer>

<http://dbpedia.org/ontology/placeOfBurial>

104 propriétés atypiques

<http://dbpedia.org/ontology/nationality>

<http://dbpedia.org/ontology/profession>

<http://dbpedia.org/ontology/currentTeam>

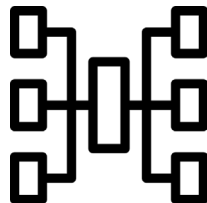



# Un SPARQL endpoint pour les explorer tous

Dataset  Mode  Output

Query 1

```
1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
3
4 SELECT * WHERE {
5     ?s ?p ?o
6 }
7 LIMIT 10
```



Service du  
numérique



**MINISTÈRE  
DE LA CULTURE**

*Liberté  
Égalité  
Fraternité*

*Inria*



Marie-Véronique Leroi : [marie-veronique.leroi@culture.gouv.fr](mailto:marie-veronique.leroi@culture.gouv.fr)

Fabien GANDON : [fabien.gandon@inria.fr](mailto:fabien.gandon@inria.fr)

Célian RINGWALD : [celian.ringwald@inria.fr](mailto:celian.ringwald@inria.fr)

**METADATA**