

Wikipédia, wikidata et les données culturelles

Bastien Guerry

Novembre 2017

Les origines de Wikipédia

- ▶ **GNUpedia** : proposé en 1999 par Stallman et lancé en 2001.
- ▶ **Nupedia** : fondée par L. Sanger et J. Wales en 2000, fermée en 2003.
- ▶ **Wikipédia** : mis en place par L. Sanger en janvier 2001 initialement pour servir de sas d'entrée pour Nupédia.
- ▶ **Citizendium** : projet d'encyclopédie en ligne lancé en 2006 par L. Sanger.

Le but de Wikipédia

- ▶ Une encyclopédie libre et collaborative pour **collecter** et **diffuser** l'ensemble des connaissances disponibles.
- ▶ Wikipédia n'est pas un espace de **production** de la connaissance, elle reste une source secondaire.

Les principes fondateurs de Wikipédia

- ▶ Wikipédia est une encyclopédie
- ▶ Wikipédia recherche la neutralité de point de vue
- ▶ Wikipédia est publiée sous licence libre
- ▶ Wikipédia est un projet collaboratif qui suit des règles de savoir-vivre
- ▶ Wikipédia n'a pas d'autres règles fixes : **be bold** !

Fonctionnement basique de Wikipédia

- ▶ La modération **a posteriori** (exemple).
- ▶ Les administrateurs n'ont qu'un rôle **technique**, pas **éditorial**.

Quelques chiffres sur la wikipédia francophone

Chiffres à la date du 8 novembre 2017 :

- ▶ Comptes : 2 930 291
- ▶ Contributeurs actifs : 16720 avec 1 modif < 30 jours
- ▶ Contributeurs actifs : 5000 avec 5 modif < 30 jours
- ▶ Contributeurs actifs : 800 avec 100 modif < 30 jours
- ▶ Administrateurs : 163
- ▶ Articles : 1 926 396

Wikipédia est le cinquième site le plus visité ([source](#)).

Où sont hébergées les données ?

Aux États-Unis, par la Wikimedia Foundation.

Quelle est la différence entre Wikipédia et Wikimedia ?

- ▶ **Wikipédia** est l'un des projets Wikimedia.
- ▶ **Wikimedia** est le nom du mouvement réunissant les salariés de la fondation américaine et des « chapitres » régionaux.

Quels sont les autres projets ?

- ▶ Wikimedia Commons
- ▶ Wikidata
- ▶ Wiktionnaire
- ▶ Wikisource
- ▶ Wikispecies
- ▶ Wikiquote
- ▶ Wikivoyage
- ▶ ...
- ▶ Voir la [liste des projets](#).

Comment éditer Wikipédia ?

[Démonstration]

Quelques concepts (1/2)

- ▶ Les wikiprojets (ou projets) : [Monuments historiques](#)
- ▶ Les portails : [Monuments historiques](#)
- ▶ Les catégories : `[[Catégorie:Base de données sur Internet]]`
- ▶ Les modèles : `{{Portail|informatique|monuments historiques}}`

Quelques concepts (2/2)

- ▶ **Infobox** : encarts avec des données structurées
- ▶ **Liens interwikis** : **lien** entre deux projets, par exemple entre le wiktionnaire et wikipédia.
- ▶ **liens interlangues** : **lien** d'un projet (ex: wikipédia) dans une langue vers le même projet (wikipédia) dans une autre langue.

Exemple de contenu littéral d'une infobox

```
{{Infobox Jeu de données
| nom =
| producteur =
| couverture géographique = France
| couverture temporelle =
| unité statistique =
| variables =
| licence = <!-- wikidata: Property:P275 -->
| url = <!-- wikidata: Property:P856 -->
}}
```

Statistiques sur le projet « Monuments historiques »

[https://fr.wikipedia.org/wiki/Projet:
Monuments_historiques](https://fr.wikipedia.org/wiki/Projet:Monuments_historiques)

Le lien wikidata sur les pages de Wikipédia

Voir dans la colonne de gauche : [Victor Hugo](#)

Problèmes à résoudre

Problème #1 Comment tirer parti des connaissances semi-structurées de Wikipédia ?

- ▶ Les infoboxes
- ▶ Les liens interwikis
- ▶ Les liens interlangue
- ▶ ...

Problème #2 Comment interroger ces données ?

Web sémantique ?

- ▶ Langage naturel : Sujet / Verbe / Complément

Les acceptions de « prédicat »

- ▶ Aristote : le prédicat est ce que l'on dit du sujet. Par exemple : Socrate **mange des pommes** ou Socrate **est un homme**. C'était la notion de grammaire en jeu dans le débat de 2017 sur l'évolution des programmes d'enseignement (voir ce [tweet](#)).
- ▶ Philosophie scolastique : le prédicat désigne une **qualité** qui peut être attribué à un sujet via une **copule**, le verbe « être ». Par exemple : Socrate est **mortel**. Trois notions se confondent : prédicat, qualité et « attribut du sujet ».
- ▶ Logique : après le calcul des propositions (si p et q sont vraies, alors "p et q" est vrai, etc.) apparaît le calcul des Prédicats :
« Tout x est gentil » : $\forall x \text{ gentil}(x)$ et « Il existe un x tel que pour tout y, x est ami avec y » : $\exists x \forall y \text{ amis}(x,y)$
 - ▶ "est gentil" = un prédicat unaire
 - ▶ "sont amis" = un prédicat binaire

Un « prédicat » en web sémantique

- ▶ En web sémantique, une affirmation est un **prédicat binaire** (voir cette [présentation du RDF](#))

Par exemple : `est_Construit_En(Q-ndp,Q-ndp2)`

On parle aussi de « triplet », et on retrouve l'écriture plus naturelle : L'église (Sujet) est construite en (Prédicat) 1912 (Objet). Le prédicat associe le sujet et l'objet.

- ▶ Wikidata utilise un vocabulaire différent : au lieu de Sujet / Prédicat / Objet, nous avons Item (Q) / Property (P) / Value.

Par exemple, l'affirmation : `{Einstein(Q), occupation(P), scientifique(V)}` dit que l'occupation d'Einstein est d'être un scientifique.

Exemple de code HTML

```
<html>
<head>
  ...
</head>
<body>
  ...
  <header1>The Trouble with Bob</h1>
  <paragraphe>Date: 2011-09-10</paragraphe>
  ...
</body>
```

Exemple de code HTML « sémantisé »

```
<html>
<head>
  ...
</head>
<body vocab="http://purl.org/dc/terms/">
  ...
  <h2 property="title">The Trouble with Bob</h2>
  <p>Date: <span property="created">2011-09-10</span></p>
  ...
</body>
```

Wikidata est constituée de **déclarations** sur des **entités**

Définitions :

- ▶ Entité (item)
- ▶ Libellé (label)
- ▶ Propriété (properties)
- ▶ Affirmation (affirmation)
- ▶ Déclaration (declaration)
- ▶ Qualificatif (qualifier)

Différence entre **affirmation** ("claim") et **déclaration** ("statement") : une déclaration contient aussi des références venant à l'appui de l'affirmation et des "rangs" (ranks) pour dire s'ils sont préférés, normaux ou dépréciés.

Voir le [glossaire](#) des termes utilisés pour Wikidata.

Quelle différence avec des données « plates » ?

Comparaison entre les bases Mérimée, Palissy et les données équivalentes présentes dans Wikidata.

L'évolution de Wikidata

- ▶ Par l'ajout d'*entités*
- ▶ Par l'ajout de *propriétés*
- ▶ Survol de l'évolution des propriétés

L'accès aux données de Wikidata

Exemple : rechercher les propriétés relatives aux dates.

- ▶ Via l'API
- ▶ Via le SparQL endpoint
- ▶ Chats avec photos
- ▶ Monuments historiques (Mérimée) de Loire-Atlantique
- ▶ Documentation

Les données culturelles sur Wiki Loves Monuments

Le projet Wiki Loves Monuments :

<https://wikilovesmonuments.fr>

Un outil d'exploration et d'édition des monuments renseignés lors des concours WLM.

Les données culturelles sur Wikipédia

Exemple : Le wikiprojet [monument historique](#).

Les données culturelles sur Wikidata

- ▶ Représentent "80% des données" (cf. présentation Wikidatacon 2017)
- ▶ Exemple de mise en forme des données sur [reasonator](#).
- ▶ Réutilisation : <http://histropedia.com/timeline>
- ▶ Réutilisation : <http://www.zone47.com/crotos/>
- ▶ Voir la [liste des données ayant un identifiant Palissy](#).
- ▶ [Sum of all paintings](#)

Exemple de requête Wikidata sur des données culturelles

```
# Paintings made on places that are
# nota work location of Van Gogh
SELECT ?item ?inception ?location ?image
WHERE {?item wdt:P31 wd:Q3305213 .
       ?item wdt:P170 wd:Q5582 .
       ?item wdt:P571 ?inception .
       OPTIONAL { ?item wdt:P18 ?image }
       ?item wdt:P1071 ?location .
       MINUS { wd:Q5582 wdt:P937 ?location } .
       MINUS { wd:Q5582 wdt:P937 ?superlocation .
               ?location wdt:P131 ?superlocation} .
}
```

Quel est l'intérêt de Wikidata pour le MC ?

- ▶ Wikidata permet d'exposer les données du MC plus largement.
- ▶ Wikidata permet de faire des recherches (de la recherche ?) : exemple, les peintres **les plus prolifiques**.
- ▶ Wikidata peut être utilisé pour obtenir des traductions des labels existants.

Quel est l'intérêt des données du ministère de la culture pour Wikidata ?

- ▶ Le ministère de la culture peut contribuer à l'ajout d'entités.
- ▶ Le ministère de la culture peut contribuer à l'ajout de propriétés.
- ▶ Le ministère de la culture peut enrichir les données existantes (ex: Palissy).

Ressources et droits

- ▶ https://fr.slideshare.net/_Emw/an-ambitious-wikidata-tutorial
- ▶ https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/queries/examples
- ▶ https://www.mediawiki.org/wiki/Wikidata_Query_Service/User_Manual

Cette présentation est disponible sous licence **Creative Commons by-sa 4.0**.