

wikipedia
dbpedia



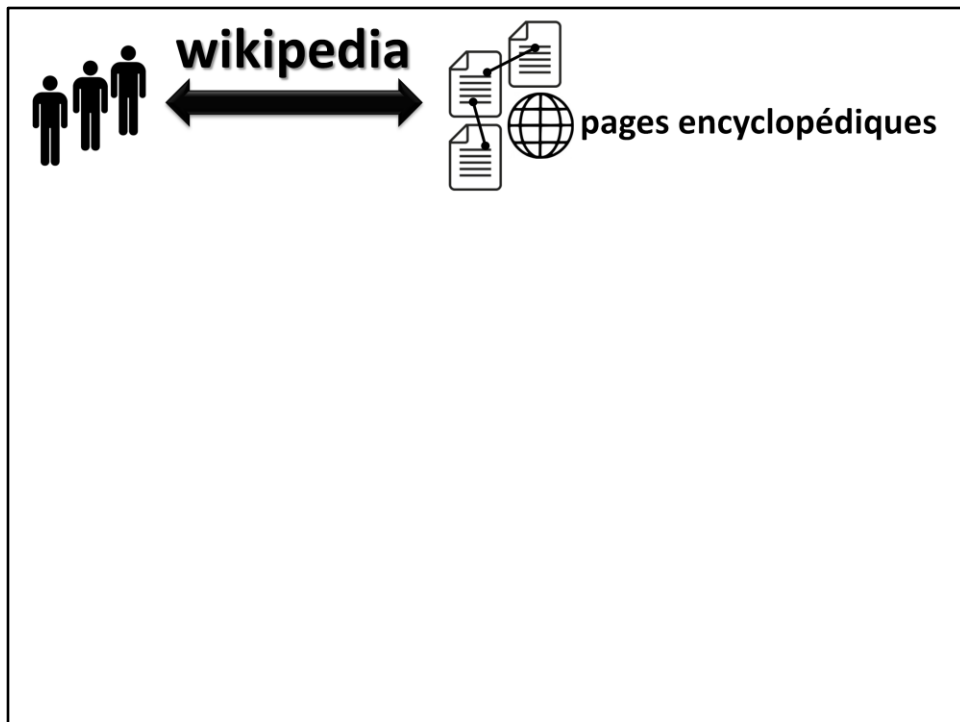
Fabien Gandon, <http://fabien.info> @fabien_gandon



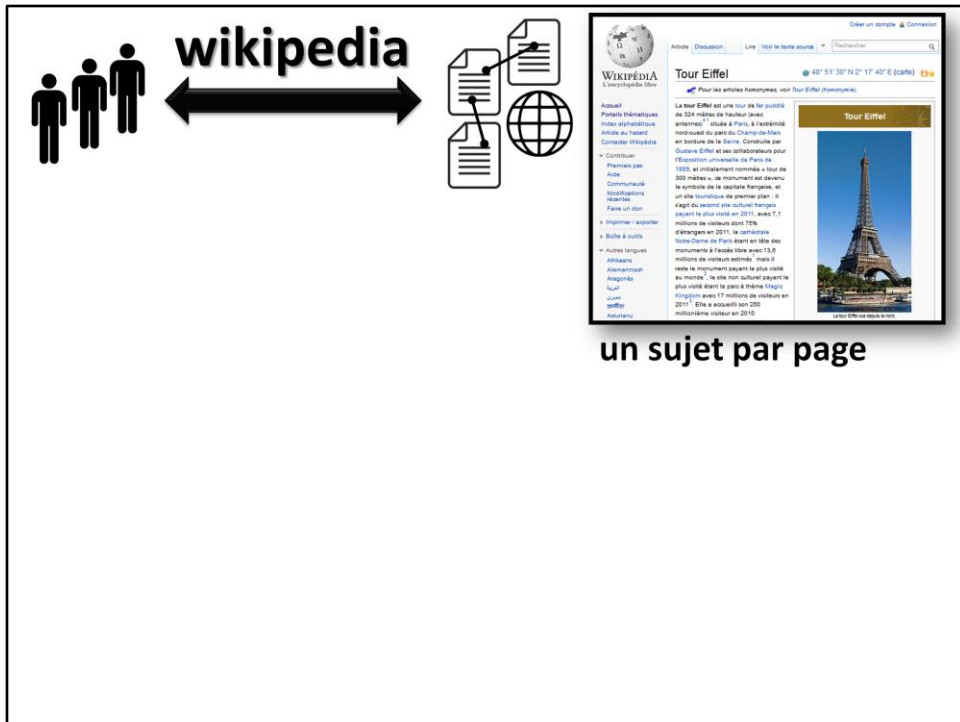
Bonjour à tous.

Il me revient donc de présenter le premier projet de la convention Sémanticpédia qui vient d'être signée.


Il s'agit du projet DBpédia.fr c'est-à-dire le chapitre francophone du projet international DBpedia qui vise à publier sur le web un maximum de données disponibles dans l'encyclopédie Wikipédia et ceci dans des formats standards et ouverts utilisables par les logiciels.




Comme vous le savez Wikipédia est une encyclopédie libre et gratuite sur le web. Le nom de « wiki-pédia » vient du fait que cette encyclopédie repose sur le principe des wikis qui permet à une communauté d'utilisateurs de créer et de maintenir les pages d'un site web. Et cette communauté est impressionnante puisqu'avec plus de 400 millions d'utilisateurs mensuels Wikipédia est aujourd'hui le cinquième site web le plus visité au monde avec plus de 23 millions d'articles dans de multiples langues.



Typiquement une page encyclopédique de wikipedia est dédiée à un sujet par exemple la Tour Eiffel



wikipedia



**beaucoup
d'informations**

WIKIPÉDIA
L'encyclopédie libre

Accueil
Portails
Inter alphabétique
Index alphabétique
Accès au hasard
Compter Wikipédia

Contribuer
Premiers pas
Aide
Communauté
Modifications
Reverts
Faire un don

Imprimer / exporter

Boîte à outils

Autres langues

Arabe
Arménien
Azerbaïdjanais
Bulgare
Catalan
Chinois
Coréen
Croatien
Danois
Allemand
Espagnol
Espéranto
Estonien
Finnais
Français
Géorgien
Grec
Hebreu
Hindouïste
Indonésien
Japonais
Coréen
Kazakh
Koryéen
Letton
Lituanien
Luxembourgeois
Malgache
Mandarin
Moldave
Néerlandais
Norvégien
Ouzbék
Pakistanaise
Persan
Polonais
Portugais
Roumain
Russe
Serbe
Slovaque
Slovenne
Espéranto
Suédois
Tadjik
Tatar
Tchaïche
Tibétain
Ukrainien
Vietnamien
Yiddish

Tour Eiffel

48° 51′ 30″ N 2° 17′ 40″ E (carte)

Le tour Eiffel est une tour de fer peinte de 324 mètres de hauteur avec antennes^[réf. souhaitée] située à Paris, à l'extrémité nord-ouest du parc du Champ-de-Mars en bordure de la Seine. Conçue par Gustave Eiffel et ses collaborateurs pour l'Exposition universelle de Paris de 1889, et initialement nommée à tour de 300 mètres, ce monument est devenu le symbole de la capitale française, et un site touristique de premier plan. Il s'agit du second site culturel français ayant le plus visité en 2011, avec 5,1 millions de visiteurs dont 79% d'étrangers en 2011, le troisième lieu de Paris étant en elle des monuments à l'exception des sites de 13,6 millions de visiteurs annuels, mais il reste le monument payant le plus visité au monde. Le site non culturel payant le plus visité étant le parc à thème Valréa. Fréquenté avec 17 millions de visiteurs en 2011, Elle a accueilli en 2009 millionième visiteur en 2010.

D'une hauteur de 324 mètres^[réf. souhaitée] à Paris, le tour Eiffel est resté le monument le plus élevé du monde pendant 41 ans. Le record tenez du troisième étage, situé à 276,10 m, est la plus haute plateforme d'observation accessible au public en France européenne et la plus haute d'Europe, tant que celle de la Tour Copernic à Moscou s'élevait à 300 m d'altitude nommée au public, suite à l'incendie survenu en 1920. Le hauteur de la tour a été plusieurs fois augmentée par construction de nombreuses antennes, pour atteindre à 324 m. Culmine à 277 m depuis le 8 mai 2011 avec un antenneur TWT. Utilisée dans le passé pour de nombreuses expériences scientifiques, elle est aujourd'hui d'instrument de programmes astronomiques et télévisés.

Selon une étude italienne de 2012, qui a classé le site des monuments d'Europe les plus visités, le Tour Eiffel aurait 424 millions de visites, son équivalent la Colisée (87 millions de visites) et la Sagrada Família (90 millions de visites).

Statistiques (France)

- 1 Population générale
- 1.1 Données techniques
- 1.2 Description de la tour étage par étage
 - 1.2.1 La base
 - 1.2.1.1 Des fondations non solides
 - 1.2.1.2 Sous-basements
 - 1.2.1.3 Les ans
 - 1.2.2 Le premier étage
 - 1.2.3 Le deuxième étage

Coordonnées : Champ de Mars (Paris), ^[?]
48° 51′ 30″ N, 2° 17′ 40″ E

Empreinte au sol : Carré de 123 m

Hauteur :
- Hauteur du 1^{er} étage : 57,63 m
- Hauteur du 2^e étage : 115,73 m
- Hauteur du 3^e étage : 276,10 m
- Total avec antenne : 324 m (depuis 2011)

Poids :
- Total : 10 100 t
- Chapeau métallique : 7 200 t

Matériau : Fer peinte des aciers de France (Eiffel)

Nombre d'étages : 13 (33 plates métalliques et 7 500 000 rivets)

Coloris : « gironnet Eiffel » (depuis 1981)

Historique

Concepteur : Gustave Eiffel & C^{te}

Ingénieurs : Maurice Koechlin, Émile Nouguier

Architecte : Stephen Sauvestre

Date(s) de :
- Construction : 1884 - 1887
- Inauguration : 2 avr 1889 et 3 juin 1889

Propriétaire : Paris

Exploitant :
- 1881 - 2005 : Société nouvelle d'exploitation de la tour Eiffel (SNTE)
- 2006 - 2010 : Société d'exploitation de la tour Eiffel (SETE)
- 2010 - présent : EPIC

Personnel :
- Fonctionnaires : 400
- Salariés : 4 810 (2010)

Une page peut avoir un contenu très riche, par exemple ici en détaillant toute l'histoire et les caractéristiques du monument.

wikipedia

beaucoup de données

Tour Eiffel

Coordonnées Champ de Mars (Paris, ^{fr})
 @ 48° 51′ 30″ N, 2° 31′ 40″ E

Élévation au sol Carré de 125 m

Hauteur

- Hauteur au 1^{er} étage 57,63 m
- Hauteur au 2^e étage 115,73 m
- Hauteur au 3^e étage 276,13 m
- Hauteur au sommet 324 m (niveau 2000)

Poids

- Total 10 100 t
- Charges mécaniques 7 200 t

Matériaux Fer puddlé des aciéries de Fives-Lille
 13 235 tonnes métalliques

Nombre d'éléments 2 500 000 rivets

Coût 4 200 000 francs (niveau 1900)

Entrepreneur Gustave Eiffel & C^{ie}

Ingénieurs Maurice Koechlin, Émile Nouguier

Architecte Stephen Sauvestre

Dates clés

- Conception 1884 - 1887
- Construction 1887 - 1889
- 2 ans, 2 mois et 3 jours
- Inauguration 31 mars 1889

Propriétaire Paris

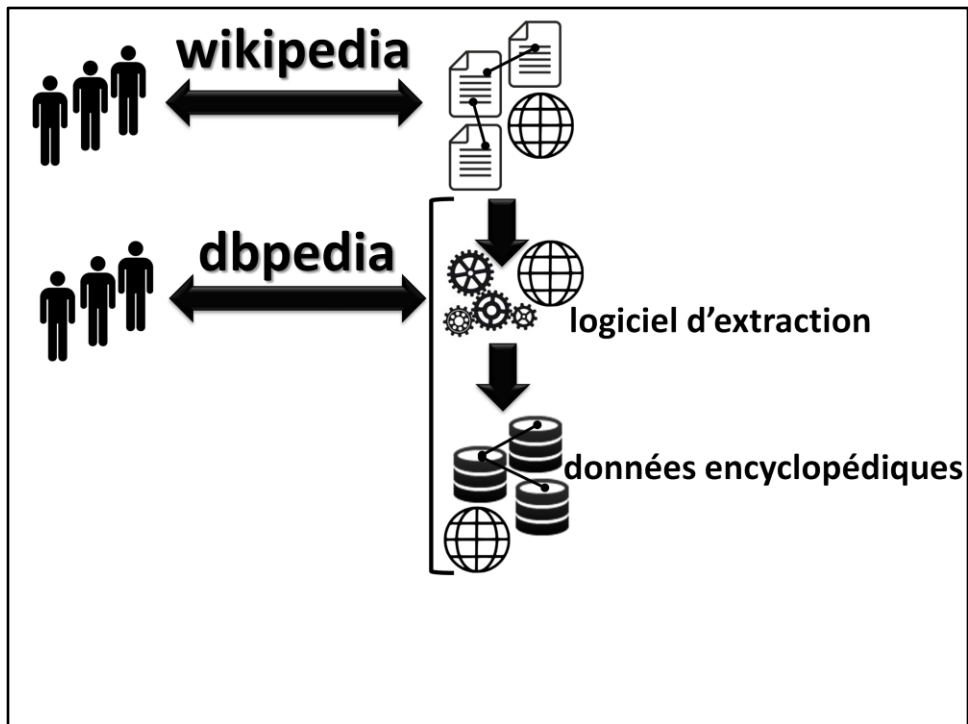
Opérateur

- 1881 - 2005 Société nouvelle d'exploitation de la tour Eiffel (SNEF)
- 2008 - 2010 Société d'exploitation de la tour Eiffel (SETE)

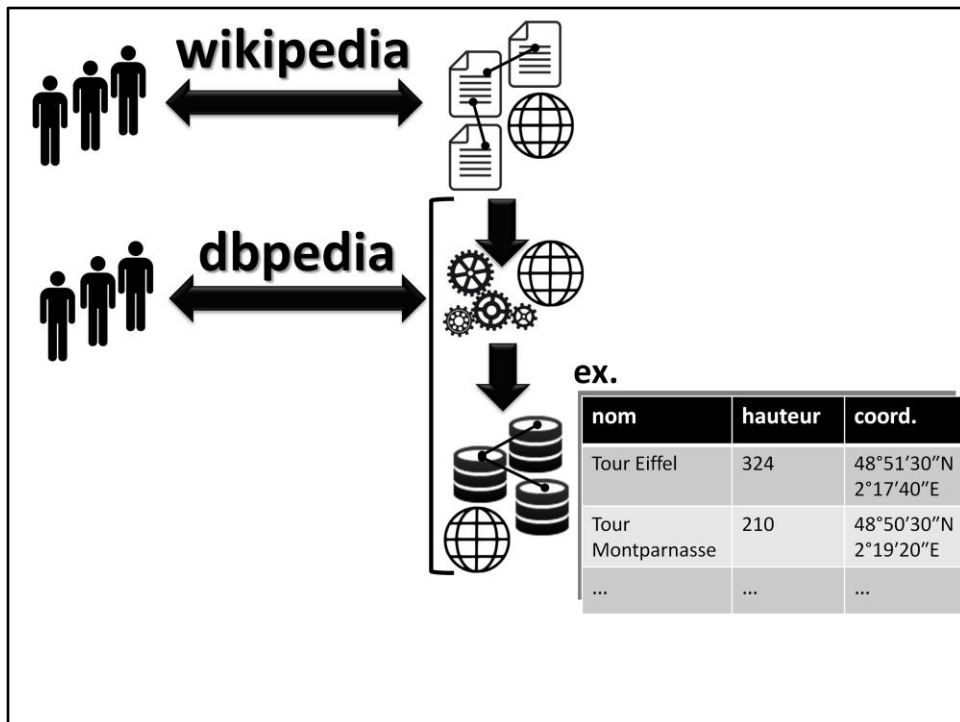
Personnel + 500

Équipement

Plus ce contenu est riche et plus il a de chances de receler des données, qu'il s'agisse par, exemple, de la hauteur ou des coordonnées GPS du monument.

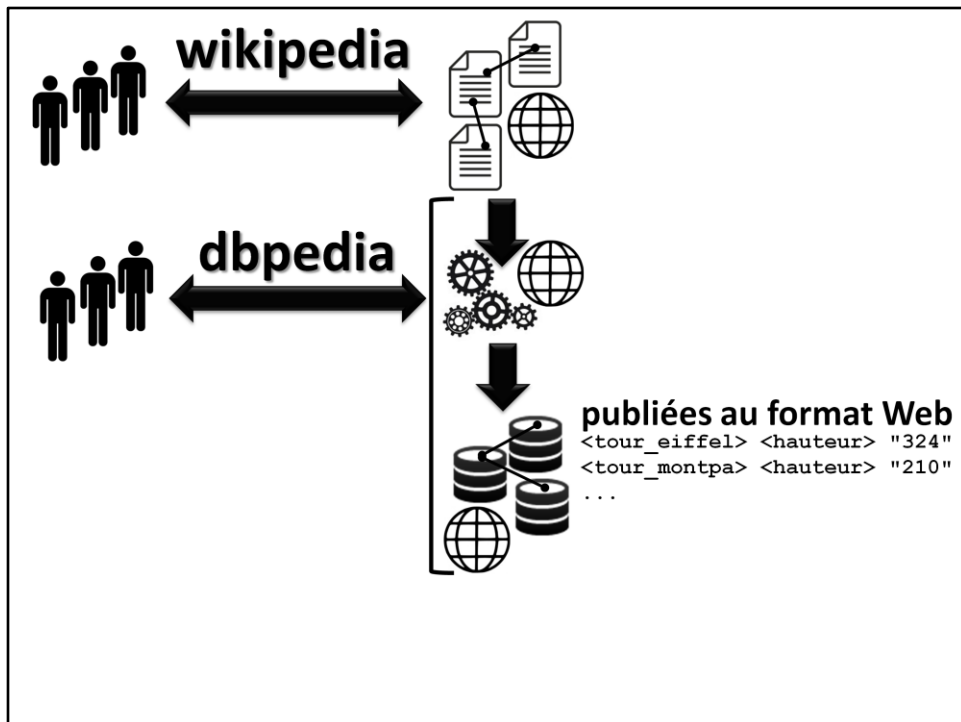


L'idée de Dbpédia fut donc d'extraire ces données en se basant dans un premier temps essentiellement sur le chapitre anglophone de Wikipédia.

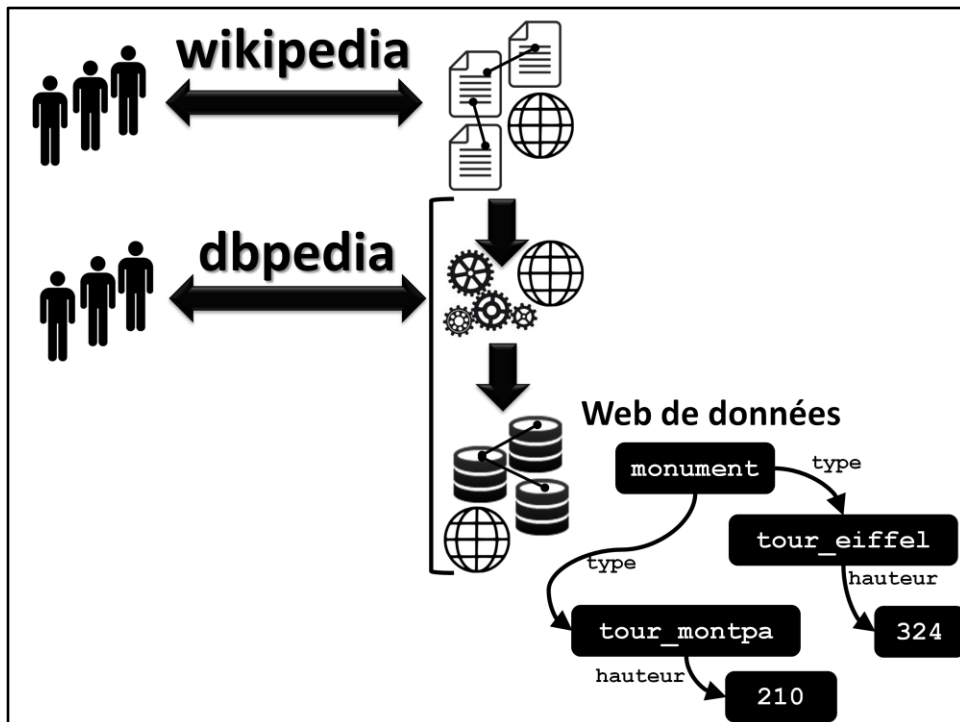


Ces données extraites sont stockées et maintenues de façon structurée permettant des traitements automatiques.

Ainsi, la liste des monuments avec leurs données structurées peut permettre de classer ces derniers par hauteur ou de calculer un itinéraire touristique dans une ville.

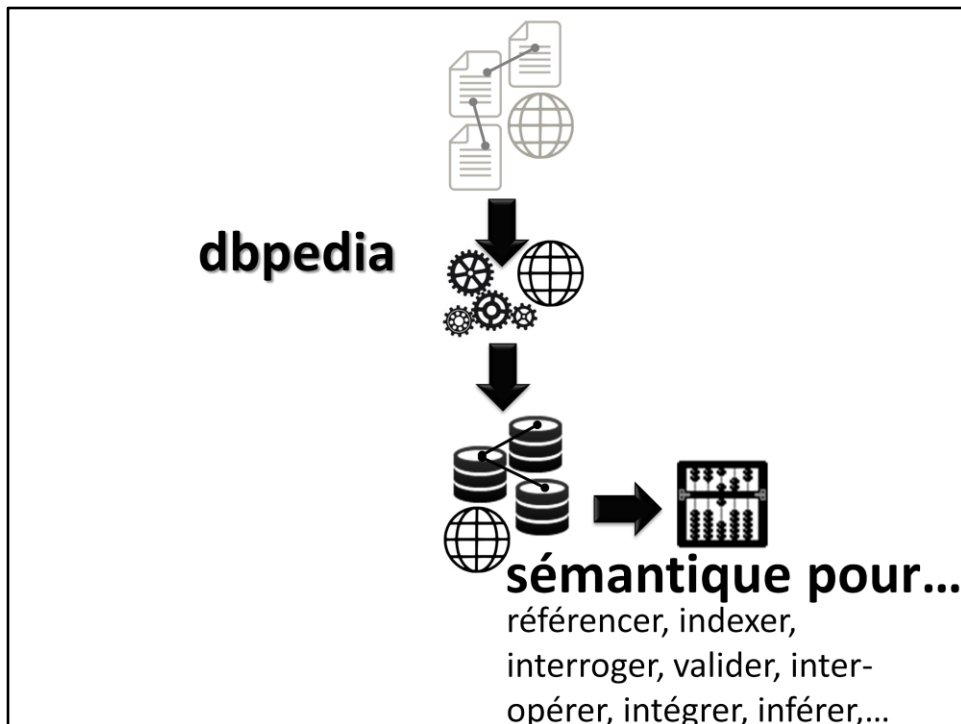


Ces données extraites sont ensuite publiées sur le web cette fois-ci non pas sous la forme de pages lues par des utilisateurs comme dans l'encyclopédie Wikipédia mais dans des formats de données standards et ouverts pour les mettre à disposition des programmes et machines connectées à internet.



Ces données viennent alors augmenter ce que l'on appelle le web de données et le web sémantique c'est-à-dire un web où on ne relie plus uniquement des pages mais bien des entités et leurs données directement exploitables par tous les outils qui ont accès au Web.

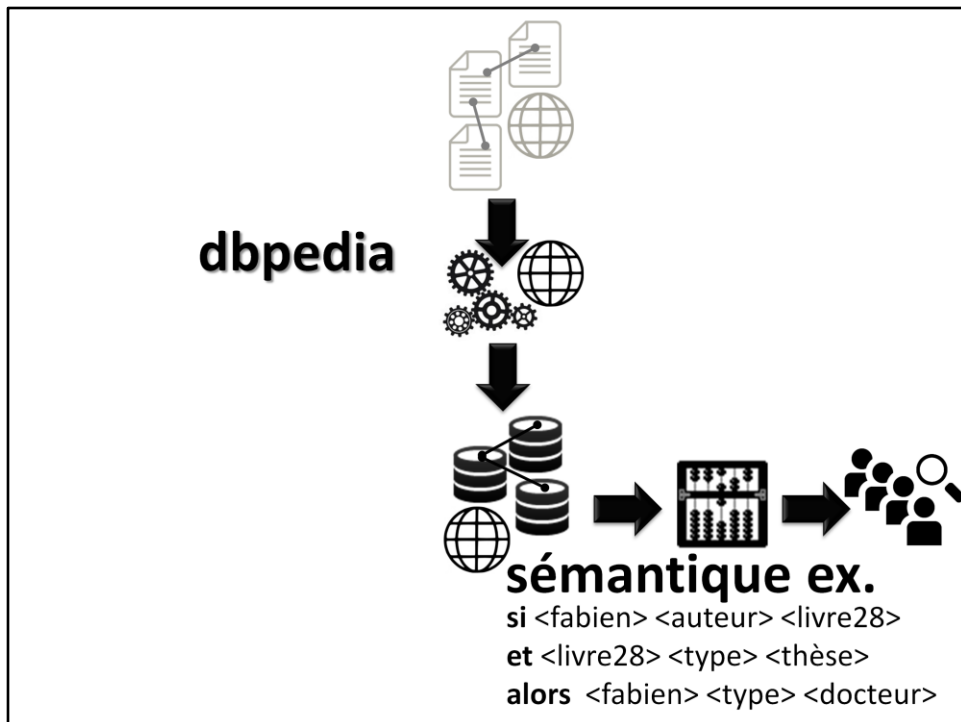
Liées et publiées sur le web, les possibilités de recherche, de parcours, d'accès et d'utilisation de ces données sont démultipliées.



On parle de « web de donnée » et de données liées car les entités, leurs données et les vocabulaires utilisés dans leurs descriptions utilisent des adresses web pour identifier non seulement les sujets, mais aussi les relations et les attributs qu'elles fournissent.

On parle de « web sémantique » par ce que ces vocabulaires, ou schémas, et les relations qu'entretiennent leurs termes sont eux aussi publiés et formalisés selon des langages standard du web.

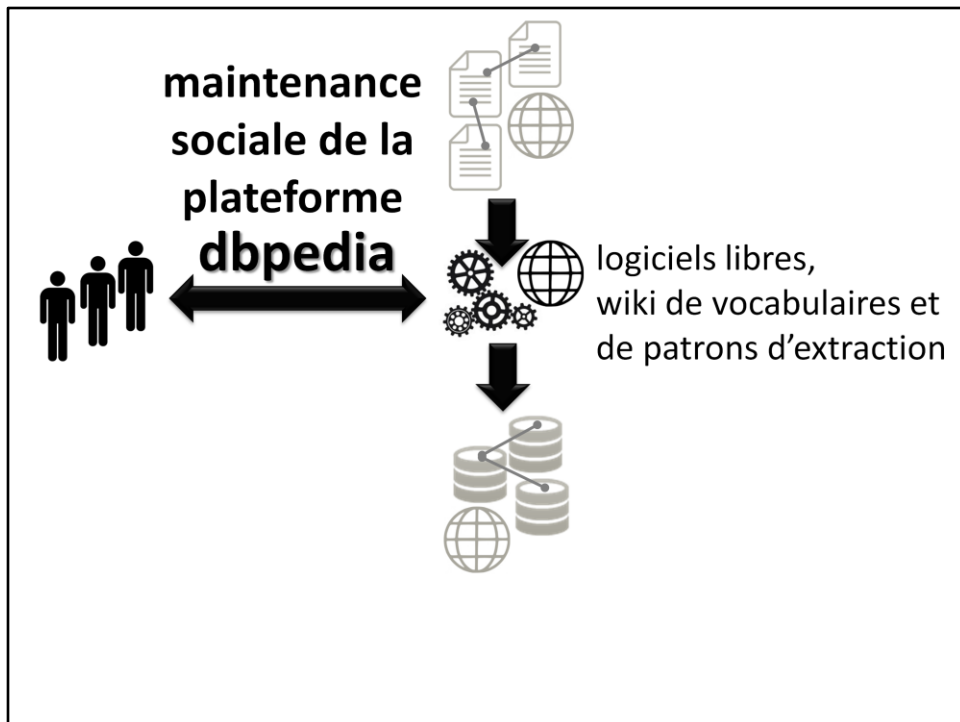
Cette sémantique partagée sur le web permet par exemple de référencer de façon unique un objet ou une caractéristique de cet objet, d'indexer des collections d'objets en partageant les références communes, d'interroger de façon très précise les données et leurs schémas, de valider des descriptions en vérifiant des règles de cohérence, d'assurer l'interopérabilité de systèmes distants en reposant sur un vocabulaire partagé ou des vocabulaires alignés, d'intégrer des jeux de données d'origines différentes, de raisonner sur les données et d'inférer de nouvelles données.



Par exemple, imaginons qu'une donnée indique que je suis auteur du livre 28 et qu'une autre donnée indique que cet ouvrage est une thèse.

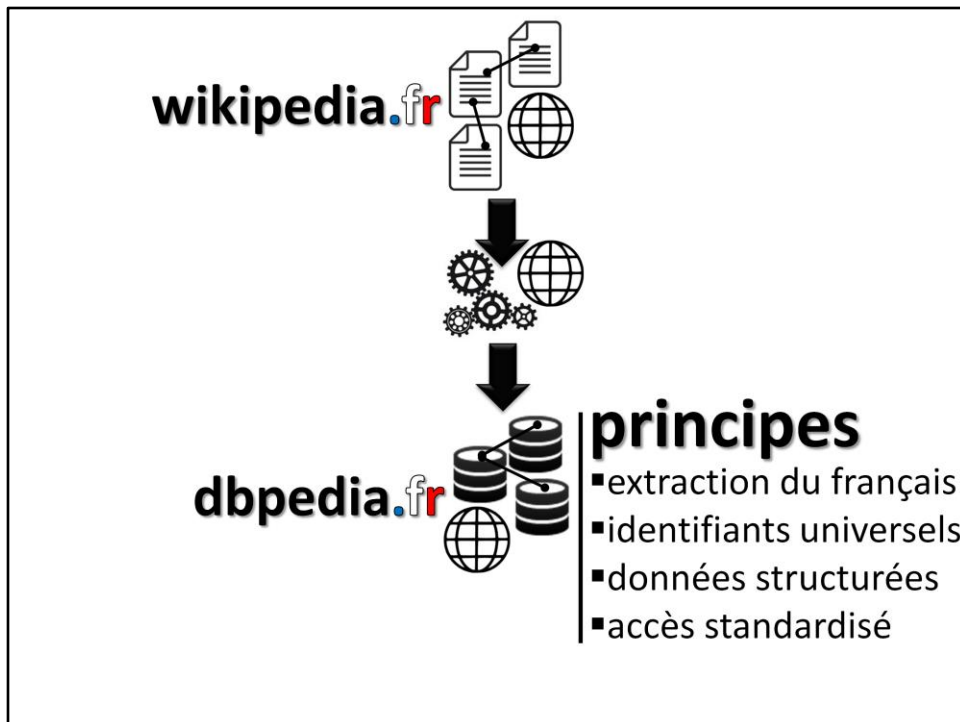
En raisonnant sur les données et leurs schémas, un logiciel pourra en déduire que je suis docteur et par conséquent inclure mon nom dans une requête demandant la liste des docteurs, même si cette donnée n'était pas initialement fournie.

Et il ne s'agit là que d'un exemple d'inférence parmi d'autres qui peuvent être échangées sur le web sémantique et dont les résultats viennent augmenter le web de données et se propager à son échelle.



Insistons enfin sur le fait que la maintenance des logiciels et des plateformes de Dbpédia est elle-même le résultat de l'action collective d'une communauté ouverte. Non seulement la plateforme est libre et les codes sources sont ouverts, mais de plus les collections de règles d'extractions tout comme les vocabulaires utilisés pour structurer les données extraites sont maintenus de façon collaborative là encore dans des wikis.

Ainsi la configuration et la maintenance d'une plateforme Dbpédia font elles aussi l'objet d'une activité sociale à laquelle chacun peut participer par exemple pour ajouter ou modifier des règles d'extractions et ainsi augmenter ou améliorer les données extraites.



Wikipédia est disponible dans de très nombreuses langues.

Le chapitre français lancé en 2001 contient aujourd'hui plus d'un million trois-cent mille articles.

Ce qui en fait le troisième chapitre de ce point de vue.

Ces articles en français regorgent eux aussi de données et notamment de données culturelles, qui n'étaient que très partiellement extraites par la version initiale de DBpédia focalisée sur le chapitre anglophone de wikipédia.

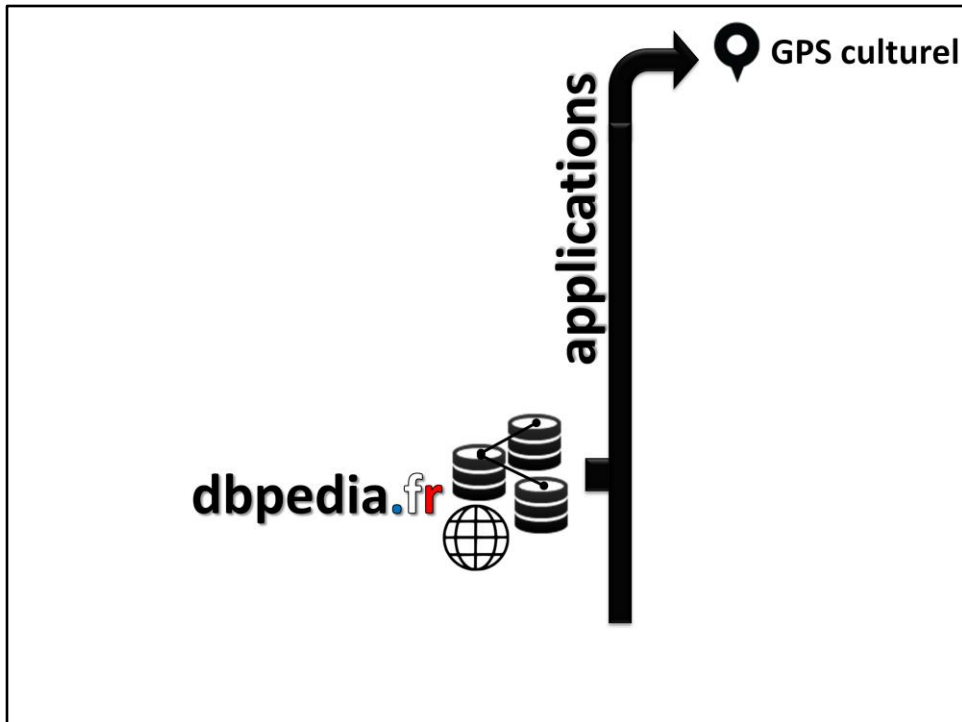
Ce premier projet de la plate-forme Semantipédia a donc permis de concevoir, développer et publier le chapitre francophone de Dbpédia.

Loin cependant de se borner à rattraper un retard francophone, ce fut l'occasion d'innover en proposant un DBpedia tirant parti des dernières évolutions techniques et des retours d'usages analysés par notre équipe.

Plus précisément nous avons repris et augmenté les extractions de données pour le Français, la publication d'identifiants universels pour tous les sujets couverts, la structuration des données et leur mise à disposition selon les standards du web.

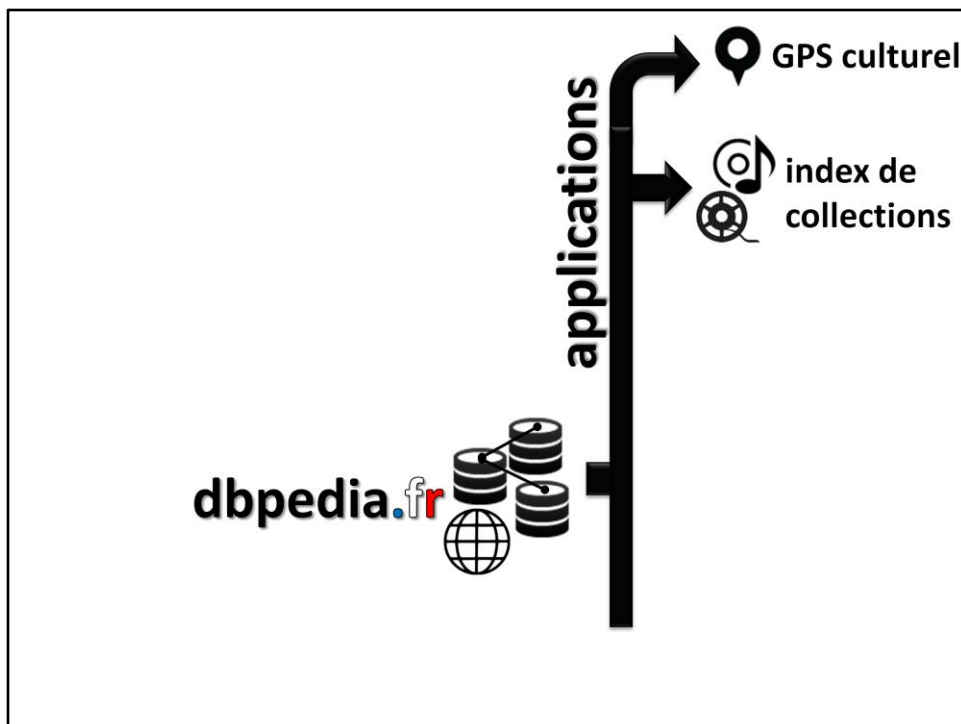
Plus de 130 millions de données sont maintenant extraites et publiées pour le chapitre francophone ce qui nous place aujourd'hui en deuxième position derrière les données extraites du chapitre anglophone.

Notons bien qu'un enjeu de la plate-forme Sémantipédia est de structurer la contribution française et de la pérenniser dans ses services d'extractions des données, dans ses schémas d'identification des sujets et dans ses moyens d'accès ouverts.

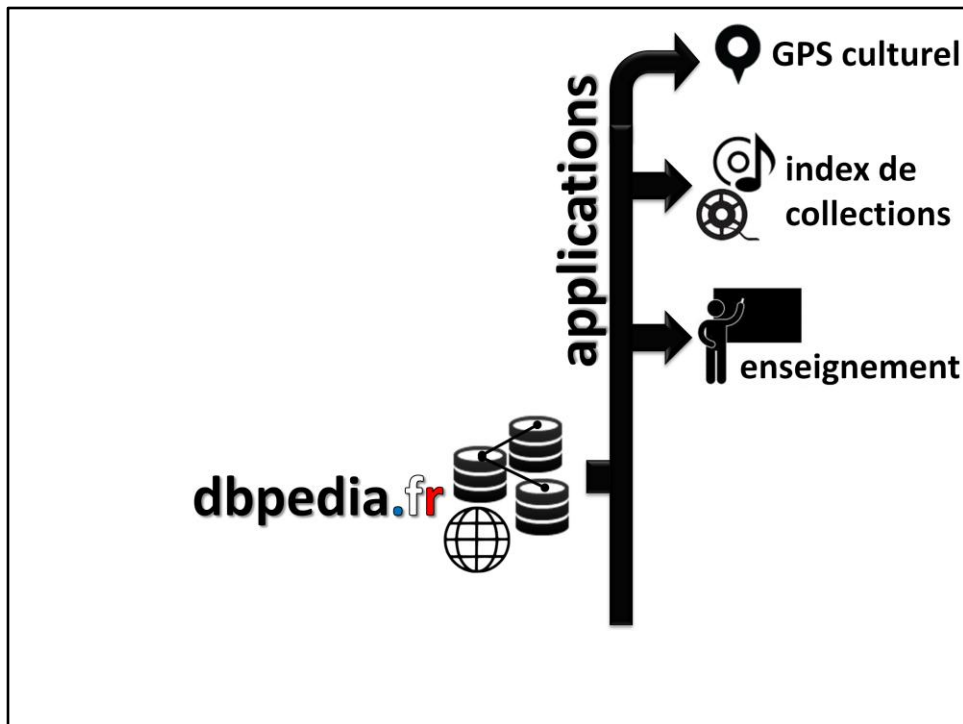


Comme vous le verrez dans les démonstrations qui suivront, ces données une fois publiées permettent de multiples applications. On peut imaginer

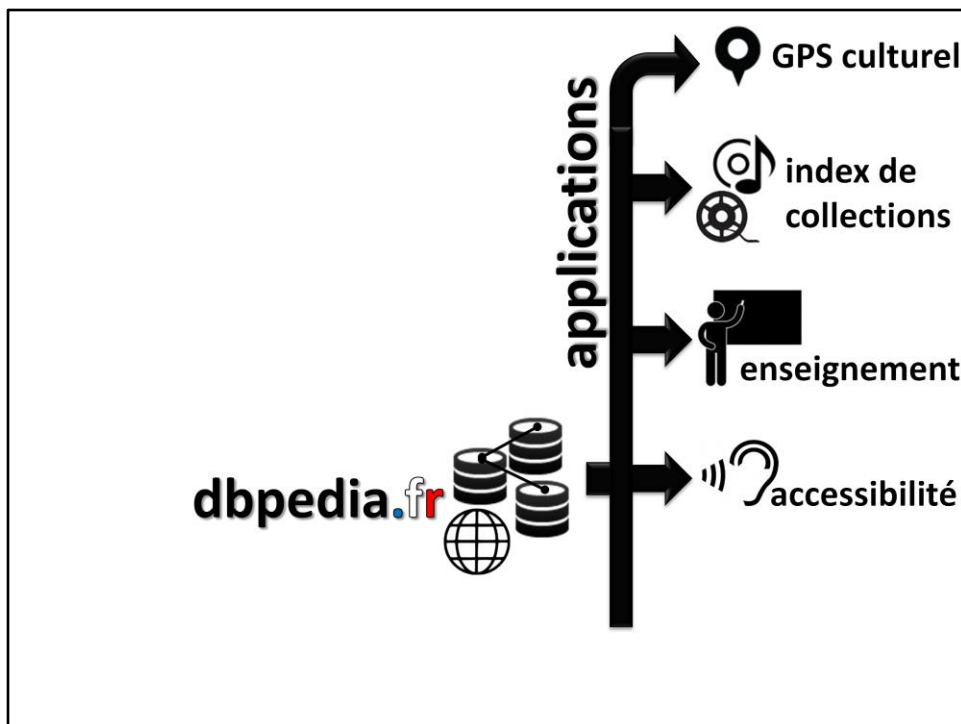
- Des GPS culturels qui nous guident et nous informent en même temps en se renseignant grâce au Web de données sur les lieux que nous traversons



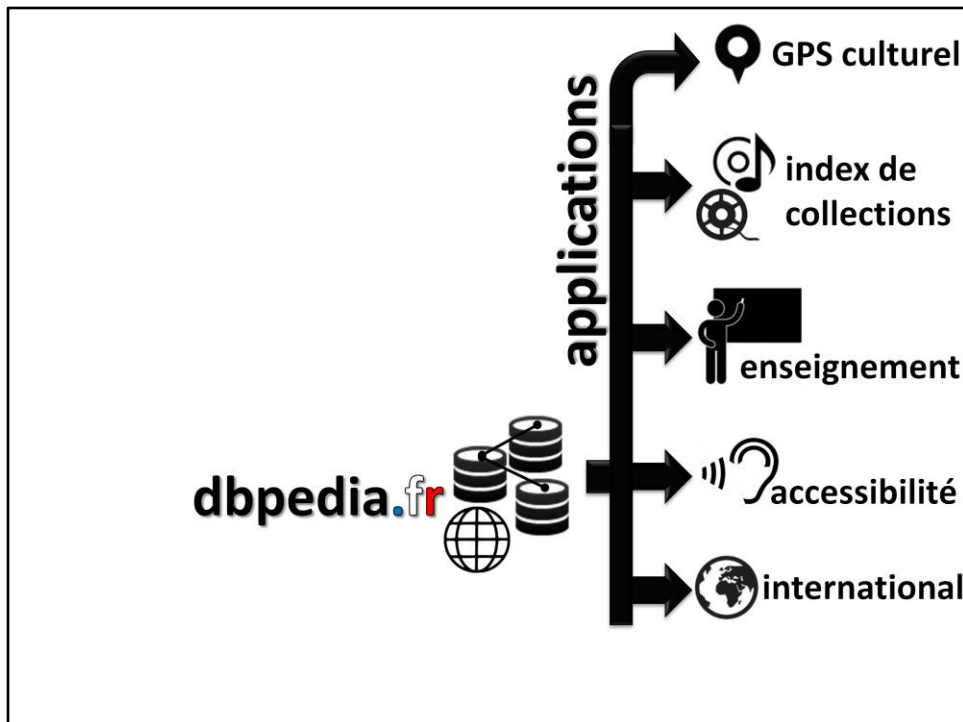
- De nouveaux portails pour organiser, accéder, naviguer et enrichir la description et les liens dans des collections culturelles pouvant avoir des sources très différentes



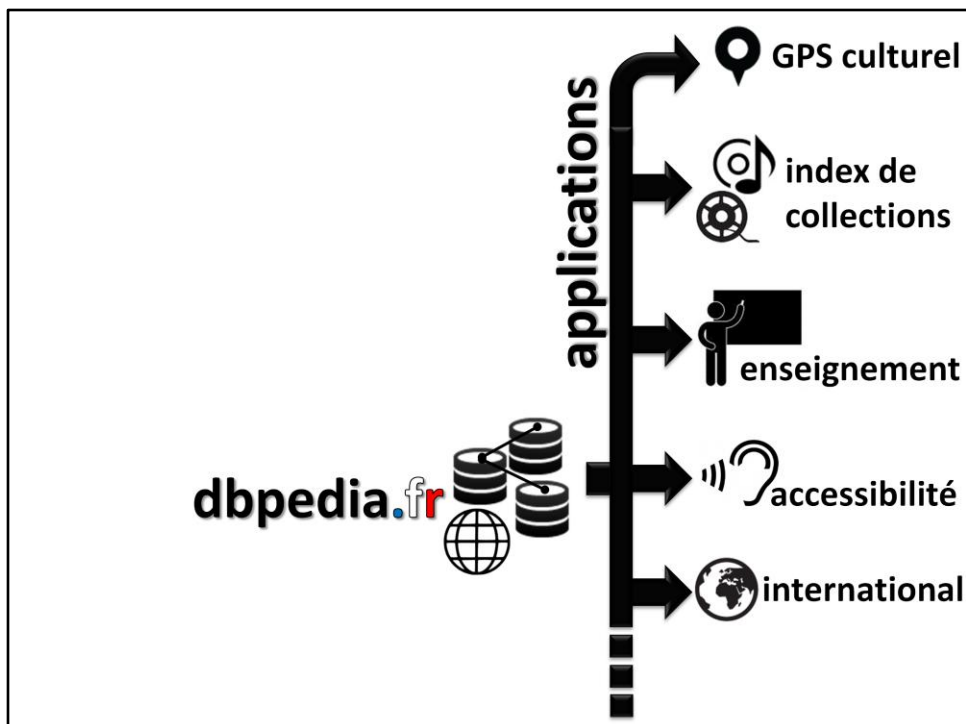
- De nouveaux logiciels interactifs pour l'enseignement par exemple pour fournir de multiples parcours de l'histoire de l'art



- Des interfaces et applications adaptant l'accès à ces données aux handicaps et aux contextes des utilisateurs

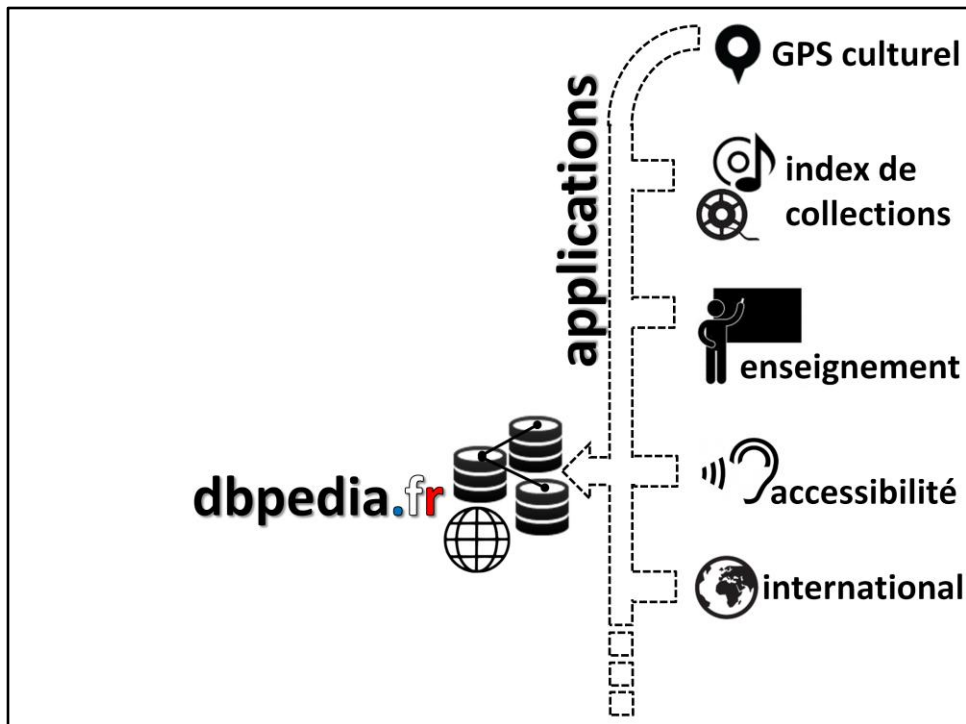


- L'internationalisation de ces données et leur intégration aux données d'autres aires linguistiques et d'autres cultures dans des applications pour des utilisateurs étrangers par exemple dans le tourisme



En fait il est en pratique impossible d'imaginer toutes les applications potentielles de ces données.

Et c'est d'ailleurs là que repose l'enjeu de l'ouverture de ces données qui doivent soutenir une innovation ouverte et des usages inédits, notamment dans le domaine de la culture.



Il reste aussi des défis à relever pour ce projet et pour les projets à venir dans le cadre de la convention Sémanticpédia.

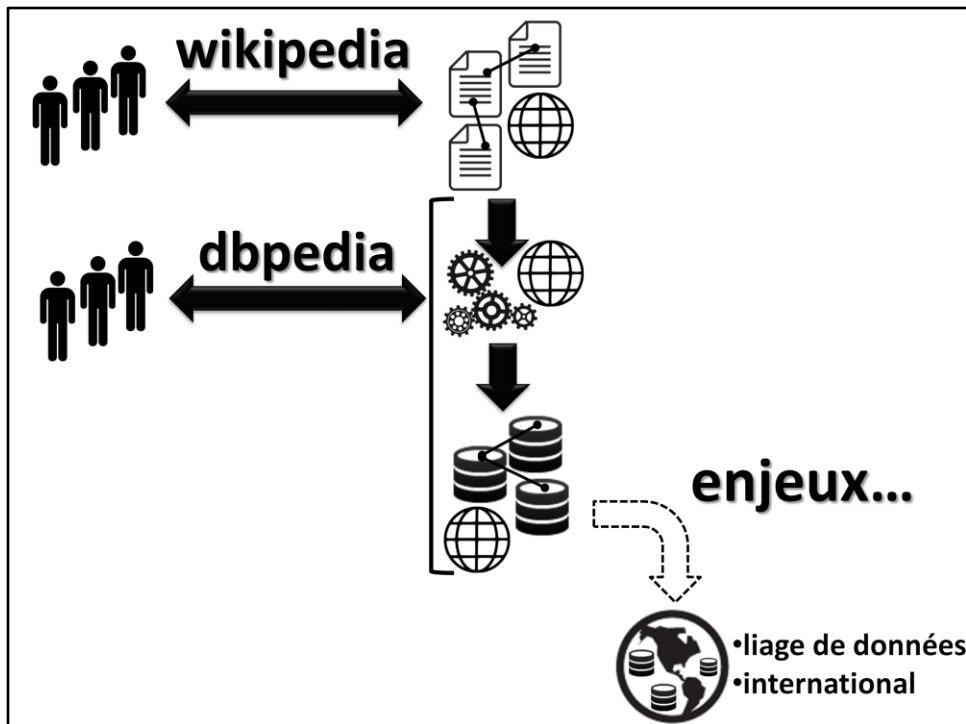
Par exemple, il y a un enjeu à trouver comment permettre à ces applications de participer au cycle de vie complet des données notamment en les enrichissant et en les maintenant à jour.

Comment un GPS culturel pourrait-il aussi permettre de compléter les données manquantes ?

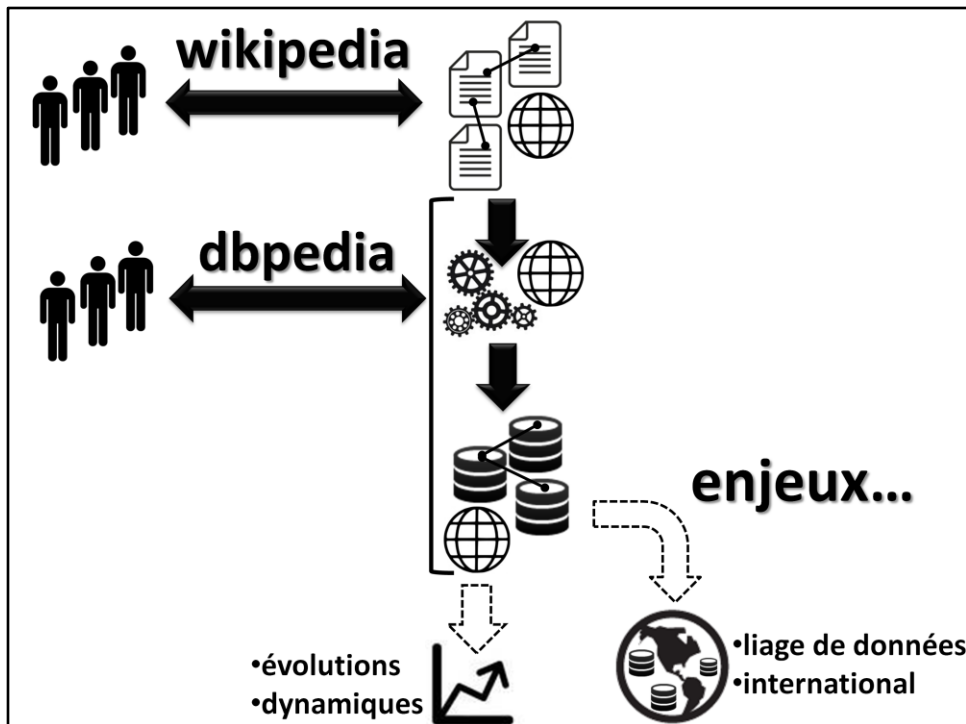
Comment un photographe pourrait-il enrichir la base en versant des photos de son choix ?

Comment un expert d'une institution culturelle pourra-t-il corriger une erreur qu'il a trouvée ?

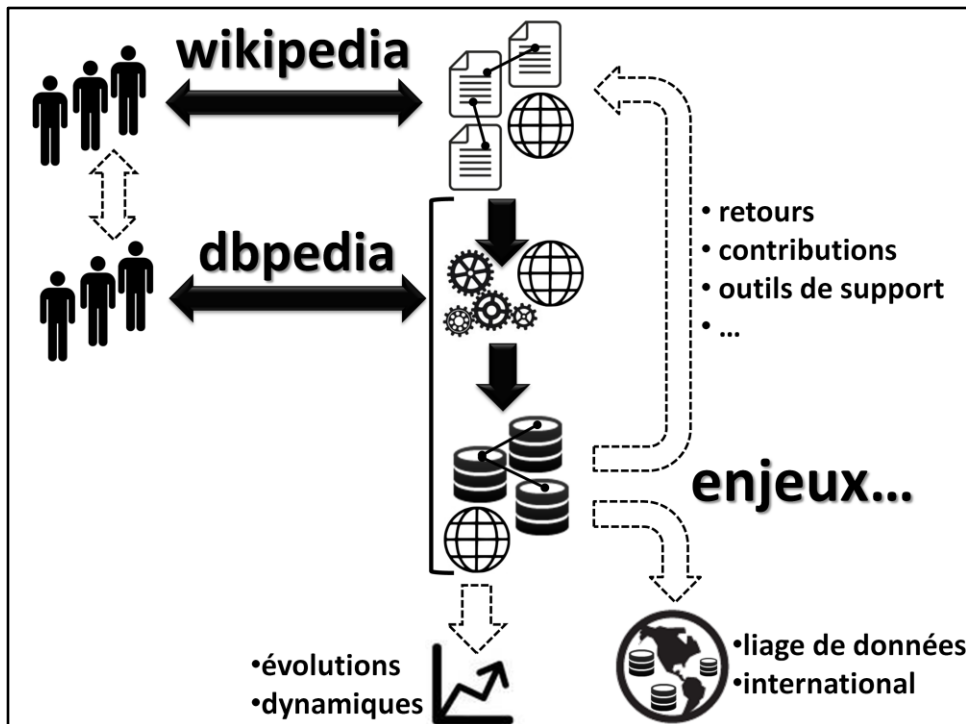
Et de façon générale comment passer de l'utilisateur à l'utilis-acteur des données de la culture.



Un autre enjeu serait aussi de lier ces données au reste des données du web, de croiser les bases en ligne pour qu'elles s'enrichissent mutuellement et d'internationaliser ces données pour qu'à travers ce processus les données francophones de la culture Française soient davantage disponibles dans d'autres langues et accessibles à d'autres cultures et d'autres publics.



Ces données changent aussi en permanence et il y a un enjeu non seulement à soutenir leur cycle de vie mais aussi à analyser ces changements comme des données portant elles-mêmes sur les évolutions, les tendances et toutes les dynamiques qui touchent aux thématiques couvertes dans l'encyclopédie.



Enfin il y a un véritable enjeu de pérennisation à mettre en place un cercle vertueux entre l'encyclopédie Wikipédia et les données extraites et publiées par Dbpedia. Il s'agit notamment de permettre de réintégrer des retours, des contributions et des résultats de calculs sur les données au sein des pages de l'encyclopédie dont elles sont issues.

Et ceci pour enrichir et améliorer son contenu, aider à détecter des manques, des erreurs aussi bien dans les pages que dans les données extraites ou dans les données calculées.

Il s'agit à terme de fournir des nouveaux outils de support à la communauté de Wikipédia qui est, à l'origine, la source qui permet toutes les applications que nous venons d'envisager et dont nous allons voir maintenant quelques exemples.

wikipedia
dbpedia



Fabien Gandon, <http://fabien.info> @fabien_gandon



Merci.