

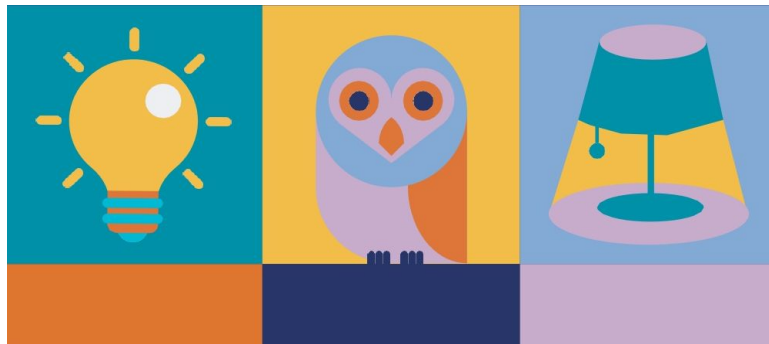
## Tuto@Mate - Wikipédia

lundi 12 juin 2023

Pierre-Yves.Beaudouin@u-bordeaux.fr, wikimédien en résidence à l'URFIST Bordeaux  
Delphine.Montagne@univ-pau.fr, ingénieure d'études à l'UMR TREE (UPPA/CNRS)

# Au programme

- Présentation générale de Wikipédia et Wikidata
  - Fonctionnement
  - Recherche
- Démonstration des outils pour extraire et importer les données
- Questions



# Un wikimédien, une wikimédienne en résidence ?

Recette :

- De la production, traitement et analyse des données (BAP D)
- De la diffusion des savoirs (BAP F)
- Avec des soupçons d'informatique/stat (BAP E) et de formation/valorisation de la recherche (BAP J)



# Introduction

# Pourquoi Wikipédia est important ?



irene fischer



Tous

Images

Actualités

Vidéos

Shopping

Plus

Outils

Environ 20 500 000 résultats (0,55 secondes)

[https://fr.wikipedia.org](https://fr.wikipedia.org/wiki/Irene_Fischer) > wiki > Irene\_Fischer ▼

**Irene Fischer - Wikipédia**

**Irene Fischer**, née Irene Kaminka le 27 juillet 1907 à Vienne et morte le 22 octobre 2009 à Boston, est une mathématicienne et géodésienne austro-américaine.

**A Deeper Investigation of the Importance of Wikipedia Links to the Success of Search Engines**

# Pourquoi Wikipédia est incontournable ?

The image is a screenshot of a Qwant search engine interface. At the top, there's a navigation bar with tabs: 'Tout', 'Applications', 'Documents', 'Web', and 'Plus'. Below this, a search bar contains the text 'irene fischer'. To the right of the search bar, there's a button that says 'Rechercher sur le Web'. Below the search bar, there's a section titled 'Meilleur résultat pour web' which shows a search icon and the text 'irene fischer' with a link to 'Afficher les résultats Web'. Below this, there's a section titled 'Rechercher sur le Web'. The main search results are displayed below the navigation bar. On the left, there's a sidebar with filters: 'Tous', 'Actualités', 'Images', 'Vidéos', 'Shopping', and 'Maps'. Below these, there's a dropdown for 'France' and a dropdown for 'Toute période'. The main results area shows a result for 'Irene Fischer — Wikipédia' with a link to 'https://fr.wikipedia.org/wiki/Irene\_Fischer'. Below this, there's a snippet of text: 'Irene Fischer, née Irene Kaminka le 27 juillet 1907 à Vienne et morte le 22 octobre 2009 à Boston, est une mathématicienne et géodésienne austro - américaine. Pendant l'âge...'. To the right of this snippet, there's a small portrait of Irene Fischer. Below the portrait, there's a section titled 'Irene Fischer' with the text 'Mathématicienne autrichienne' and a snippet of text: 'Irene Fischer, née Irene Kaminka le 27 juillet 1907 à Vienne et morte le 22 octobre 2009 à Boston, est une mathématicienne et'. To the right of this section, there's a list of links: 'Carrière', 'Hommages et Distinctions', 'Vie Privée', and 'Principales Publications'. At the bottom right, there's a button that says 'Ouvrir les résultats dans un navigateur' and a link to 'Afficher toutes les sections'.

**Qwant** irene fischer

Tous Actualités Images Vidéos Shopping Maps

France Toute période

fr.wikipedia.org > wiki > Irene\_Fischer

**Irene Fischer — Wikipédia**

Irene Fischer, née Irene Kaminka le 27 juillet 1907 à Vienne et morte le 22 octobre 2009 à Boston, est une mathématicienne et géodésienne austro - américaine. Pendant l'âge...

**Irene Fischer**

Mathématicienne autrichienne

Irene Fischer, née Irene Kaminka le 27 juillet 1907 à Vienne et morte le 22 octobre 2009 à Boston, est une mathématicienne et

**Irene Fischer — Wikipédia**

[https://fr.wikipedia.org/wiki/Irene\\_Fischer](https://fr.wikipedia.org/wiki/Irene_Fischer)

Irene Fischer, née Irene Kaminka le 27 juillet 1907 à Vienne et morte le 22 octobre 2009 à Boston, est une mathématicienne et géodésienne austro - américaine. Pendant l'âge d'or du projet Mercury et de la mission lunaire Apollo, Fischer est devenue l'une des deux femmes scientifiques de renommée internationale dans le domaine de la géodésie. **Wikipédia (FR)**

**Ouvrir les résultats dans un navigateur**

[Afficher toutes les sections](#)

# Pourquoi Wikipédia est incontournable ?

Dis Siri,  
Qui est Irene Fischer ?

«Alexa,  
Qui est Irene  
Fischer ?



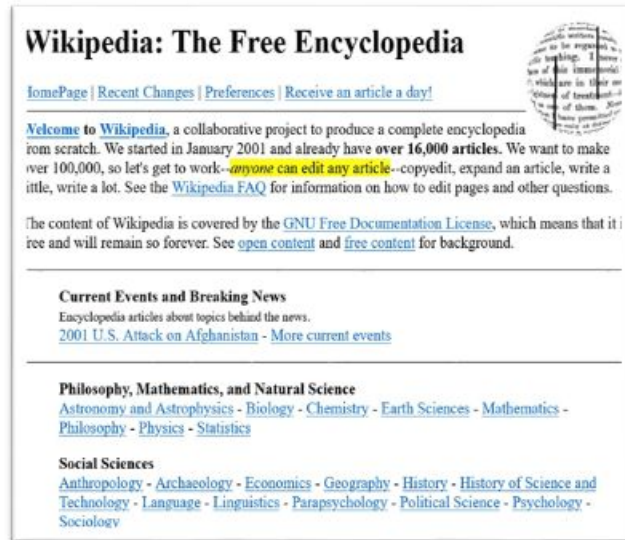
## Irene Fischer

**Irene Fischer**, née **Irene Kaminka** le 27 juillet 1907 à Vienne et morte le 22 octobre 2009 à Boston, est une mathématicienne et géodésienne austro-américaine. Pendant l'âge d'or du projet Mercury et de la mission lunaire Apollo, Fischer est devenue l'une des deux femmes scientifiques de renommée internationale dans le domaine de la géodésie. Elle joue un rôle déterminant dans la mise en place du système géodésique mondial (norme utilisée notamment par les GPS). Son *datum Mercury* (ou ellipsoïde de Fischer 1960 et 1968)<sup>1,2</sup>, ainsi que ses travaux sur la parallaxe lunaire, jouent un rôle déterminant dans la conduite de ces missions.

# Aux origines : un projet d'universitaires



Nupedia.com  
3 décembre 2000



Wikipedia.com  
3 décembre 2001



# Quelques statistiques



**WIKIPÉDIA**  
L'encyclopédie libre

**Wikipédia  
en français**

**350**

NOUVEAUX ARTICLES  
CRÉÉS PAR JOUR

**2 460 000**

ARTICLES EN  
OCTOBRE 2022






**17 000**

CONTRIBUTEURS  
ACTIFS (au moins  
une fois par mois)

**800**

CONTRIBUTEURS TRÈS  
ACTIFS

# Cinq principes pour une encyclopédie libre

-  • **Wikipédia est une encyclopédie**
-  • **Wikipédia recherche la neutralité de point de vue**
  - On cite ses sources
-  • **Wikipédia est publiée sous licence libre**
  - CC-BY-SA 3.0
-  • **Wikipédia suit des règles de savoir-vivre**
  - Par la discussion, le consensus
-  • **Wikipédia n'a pas d'autres règles fixes**

# Le “mouvement Wikimedia” : démêlage



- Les wikipédiens et wikipédiennes qui contribuent régulièrement aux projets : rédaction, modération
- Wikimedia Foundation : hébergement et responsabilité légale
- Wikimédia France : réseau d'acteurs associatifs qui soutiennent la mission du mouvement Wikimedia

Wikipédia & Wikidata

# Une page Wikipédia



Nombreuses pages d'aides, forum des nouveaux/nouvelles, système de parrainage

## Jacqueline Beaujeu-Garnier

Article Discussion

Lire Modifier Modifier le code Voir l'historique

**Jacqueline Beaujeu-Garnier**, née le 1<sup>er</sup> mai 1917 à Aiguille (Haute-Loire) et décédée le 28 avril 1995 à Paris, est une géographe française.

Spécialiste de géographie urbaine et de géographie du peuplement, elle est l'une des principales femmes géographes françaises du xx<sup>e</sup> siècle, à l'influence à la fois nationale et internationale.

### Biographie

Fille de gendarme, Jacqueline Garnier suit les cours de géographie d'André Cholley à la Sorbonne à la fin des années 1930<sup>[1]</sup>. Agrégée féminine d'histoire en 1941, tout juste mariée avec l'historien antiquisant Jean Beaujeu, elle est d'abord assistante de géographie à la Sorbonne pendant la guerre. Elle soutient sa thèse en 1947 sur la **géomorphologie** des marges du **Morvan** (géographie physique) ainsi que sur la **géographie humaine** et **régionale** d'une vallée des Alpes autrichiennes, devenant la première femme docteur d'état en géographie en France.

### Carrière

Brièvement nommée maîtresse de conférences à Poitiers, elle devient professeure de géographie à Lille en 1948, devenant la première femme française à atteindre cette fonction dans la discipline<sup>[2]</sup>, au même moment que Germaine Veyret-Verner (1913-1973) à Grenoble<sup>[3]</sup>. En 1960, elle est nommée à l'université de Paris, où elle reste 25 ans, enseignant à l'**institut de géographie**

Jacqueline Beaujeu-Garnier



### Références

- ↑ Nicolas Ginsburger, « Le quart féminin des géographes : dynamiques et limites de la féminisation dans la géographie universitaire française et internationale (1928-1938) », *Revue d'histoire des sciences humaines*, n<sup>o</sup> 28, septembre 2016, p. 213-248
- ↑  « Beaujeu-Garnier », Jacqueline Beaujeu-Garnier (1917-1995) », 50 ans de l'ULP, 1969 (lire en ligne) <sup>[archivé]</sup>.
- ↑ Nicolas Ginsburger, « Portrait en groupe de femmes-géographes. La féminisation du champ disciplinaire au milieu du xx<sup>e</sup> siècle, entre effets de contexte et de structure (1938-1960) », *Annales de géographie*, n<sup>o</sup> 713 (2017), p. 107-133
- ↑ Nicolas Ginsburger, « Femmes en géographie au temps des changements. Féminisation et féminisme dans le champ disciplinaire français et international (1960-1990) », *L'Espace géographique*, juillet-août-septembre 2017, p. 236-263
- ↑  (en) Philippe Pinchemel, « Obituaries: Jacqueline Beaujeu-Garnier », *The Geographical Journal*, vol. 161, no. 3 (Nov., 1995), p. 354-355 (accessible sur JSTOR) <sup>[archivé]</sup>.
- ↑ Jacqueline Beaujeu-Garnier, *Géographie de la population*, par Jacqueline Beaujeu-Garnier… M.-T. Génin, coll. « Géographie économique et sociale », 1969 (lire en ligne) <sup>[archivé]</sup>.
- ↑ Jacqueline Beaujeu-Garnier, *Géographie urbaine*, « Colin, coll. « U », 1995 (ISBN 978-2-200-21103-6, lire en ligne) <sup>[archivé]</sup>.
- ↑ Jacqueline Beaujeu-Garnier, *Les îles britanniques*, Centre de documentation universitaire, coll. « Les Cours de Sorbonne », 1960 (lire en ligne) <sup>[archivé]</sup>.
- ↑ Jacqueline Beaujeu-Garnier et Institut de géographie, *Atlas du Nord de la France*, Berger-Levrault, 1961 (lire en ligne) <sup>[archivé]</sup>.
- ↑ Jean Association universitaire de recherches géographiques et cartographiques et Jacqueline Beaujeu-Garnier, *Atlas de Paris et de la région parisienne*, Berger-Levrault, coll. « Collection des atlas régionaux français », 1967 (lire en ligne) <sup>[archivé]</sup>.
- ↑ Jacqueline Beaujeu-Garnier et Annie Delboz, *La géographie du*

### Annexes

### Sources











- (en) Philippe Pinchemel, « Obituaries: Jacqueline Beaujeu-Garnier 1917-1995 », in *The Geographical Journal*, Vol. 161, No. 3 (Nov., 1995), p. 354-355 (accessible sur JSTOR) <sup>[archivé]</sup>.

### Bibliographie

- (en) Hugh Clout, *Jacqueline Beaujeu-Garnier, 1917-1995*, Continuum, Londres, 2009, p. 131-146 (tiré à part de  : *Geographers : biobibliographical studies*, vol. 28)
- Nicolas Ginsburger, « Le quart féminin des géographes : dynamiques et limites de la féminisation dans la géographie universitaire française et internationale (1928-1938) », *Revue d'histoire des sciences humaines*, septembre, n<sup>o</sup> 29, 2016, 213-248.
- Nicolas Ginsburger, « Portrait en groupe de femmes-géographes. La féminisation du champ disciplinaire au milieu du xx<sup>e</sup> siècle, entre effets de contexte et de structure (1938-1960) », *Annales de géographie*, n<sup>o</sup> 713, 1/2017, p. 107-133

# Comprendre Wikidata

- Base de connaissances sur le monde
- Ouverte en 2012
- Données organisées et catégorisées de manière standardisée (web sémantique)
- Licence Creative Commons (CC0 1.0)

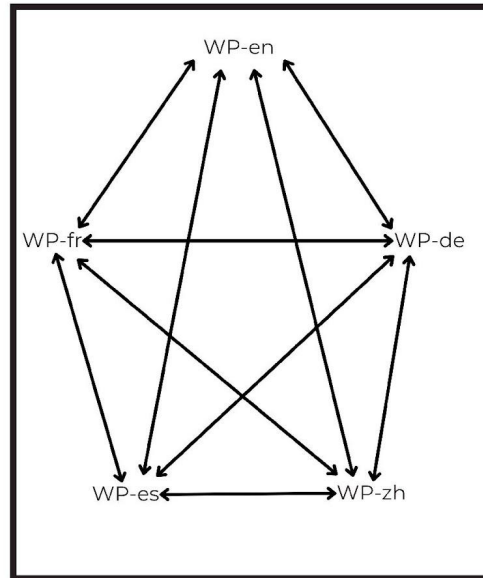
Jacqueline Beaujeu-Garnier	
	
Fonction	
Présidente Société de géographie	
1983-1995	
◀ Roger Blais (d) Jean Bastié ▶	
Biographie	
Naissance	1 <sup>er</sup> mai 1917  Aiguilhe (Haute-Loire) 
Décès	28 avril 1995  (à 77 ans) 14 <sup>e</sup> arrondissement de Paris 
Nom de naissance	Jacqueline Garnier 
Nationalité	française 
Activités	Géographe, professeure d'université 
Conjoint	Jean Beaujeu 
Autres informations	
A travaillé pour	Université Paris 1 Panthéon-Sorbonne Université de Lille 

# Comment relier les articles entre eux ?

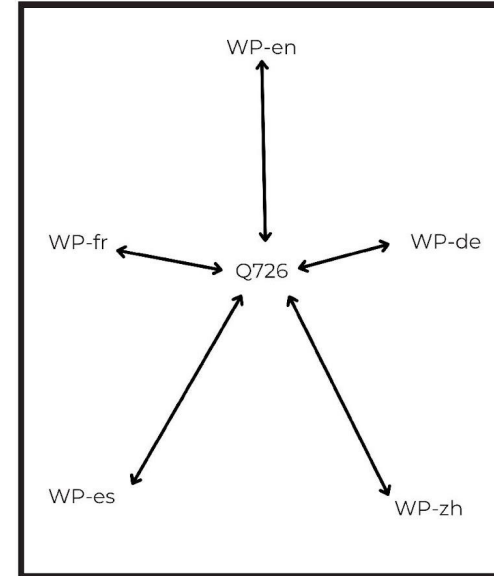
en : horse  
fr : cheval  
de : pferd  
es : caballo  
zh : 马

Wikidata =  
centralisation  
des données

Avant Wikidata



Après Wikidata 

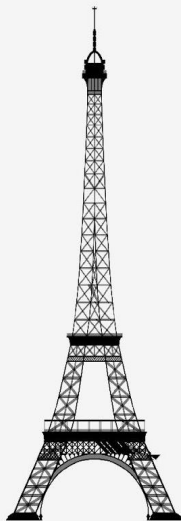


## SUJET : TOUR EIFFEL (Q243)

Propriété  
→ matériau (P186)

Valeur  
acier  
(Q11427)

Valeur  
fer forgé  
(Q860451)



Propriété  
→ hauteur (P2048)

Valeur  
324 mètres  
(au sommet)

Valeur  
300 mètres  
(architecture)

**[MOOC]**  
Wikidata



# Structuration des données dans Wikidata

The diagram illustrates the Wikidata data structure for the entity **Douglas Adams (Q42)**. It shows how various pieces of information are organized into different components, each with a specific label and function.

**Identifiant de l'élément:** **Douglas Adams (Q42)**

**Libellé:** **Douglas Adams**

**Description:** écrivain anglais de science-fiction  
Douglas Noel Adams | Douglas Noël Adams  
► Plus de langues

**Alias:** Douglas Noel Adams | Douglas Noël Adams

**Déclarations**

**Propriété:** **scolarité**

**Valeur:** **St John's College**

**Qualificatifs:**

date de fin	1974
principale spécialité d'étude	littérature anglaise
diplôme	baccalauréat en arts
date de début	1971

**Rang:**

**Groupe de déclarations:**

**Références dépliées:**

▼ 2 références

affirmé dans	Encyclopædia Britannica Online
URL de la référence	http://www.nndb.com/people/731/000023662/
langue originale	anglais
date de consultation	7 décembre 2013
éditeur	NNDB
titre	Douglas Adams (anglais)

+ ajouter une référence

**Références repliées:**

Brentwood School

date de fin	1970
date de début	1959

► 0 références

+ ajouter

Un terrain pour la recherche

# Exemples de recherches

## Pourquoi Wikipédia est incontournable ?

The screenshot shows a search engine interface with the query 'Irene Fischer'. The top result is from Wikipedia, titled 'Irene Fischer — Wikipédia'. The snippet describes her as a mathematician and geophysicist, born in 1907 and died in 2009. Below the search results, there is a section for 'Irene Fischer' with a portrait and a brief biography. The interface includes navigation tabs like 'Tout', 'Applications', 'Documents', 'Web', and 'Plus'.

## Quelques statistiques



Wikipédia  
en français

**350**

NOUVEAUX ARTICLES  
CRÉÉS PAR JOUR

**2 460 000**

ARTICLES EN  
OCTOBRE 2022

**17 000**

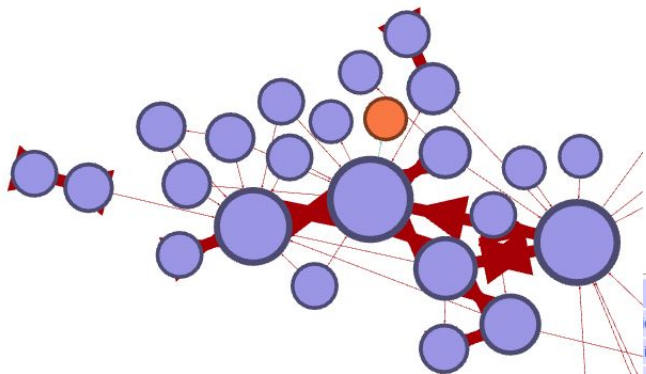
CONTRIBUTEURS  
ACTIFS (au moins  
une fois par mois)

**800**

CONTRIBUTEURS TRÈS  
ACTIFS

Co-by-SA 4.0 Commons Amélie Charles WMFr

# Exemples de recherches (2)



Graphe entre articles

## Biais géographique

🇨🇮 Élections en Côte d'Ivoire <span>[masquer]</span>	
départementales	1960 • 1965 • 1970 • 1975 • 1980 • 1985 • 1990 • 1995 • 2000 • 2010 • 2015 • 2020 • 2025
législatives	1957 • 1959 • 1960 • 1965 • 1970 • 1975 • 1980 • 1985 • 1990 • 1995 • 2000 • 2011 • 2016 • 2021 • 2026
municipales (indirectes)	2018 • 2023
Régionales	2013 • 2018 • 2023
Municipales	1960 • 1965 • 1970 • 1975 • 1980 • 1985 • 1990 • 1995 • 2001 • 2011 • 2013 • 2018 • 2023
Référendums	1958 • 2000 • 2016
🌍 Élections en Afrique <span>[masquer]</span>	
États souverains	Afrique du Sud • Algérie • Angola • Bénin • Botswana • Burkina Faso • Burundi • Cameroun • Cap-Vert • Centrafrique • Comores • République du Congo • République démocratique du Congo • Côte d'Ivoire • Djibouti • Égypte • Érythrée • Eswatini • Éthiopie • Gabon • Gambie • Ghana • Guinée • Guinée-Bissau • Guinée équatoriale • Kenya • Lesotho • Liberia • Libye • Madagascar • Malawi • Mali • Mauritanie • Maurice • Maroc • Mozambique • Namibie • Niger • Nigeria • Ouganda • Rwanda • Sao Tomé-et-Principe • Sénégal • Seychelles • Sierra Leone • Somalie • Soudan • Soudan du Sud • Tanzanie • Tchad • Togo • Tunisie • Zambie • Zimbabwe



# Exemples de recherches (3)

Qui construit la connaissance ?

Structuration de WP (différentes « classes » de contributeurs/contributrices) ?

[v](#) · [m](#)

Statuts techniques des contributeurs [\[masquer\]](#)

 Administrateur ·  Administrateur d'interface ·  Bureaucrate ·  Steward ·  Masqueur de modifications ·  Agent EBRC ·  Péon

 Révocateur ·  Bot ·  Vérificateur d'adresses IP ·  Importateur (transwiki) ·  Modificateur de filtre anti-erreur ·  Créateur de comptes

Liste des droits de groupe

# Faire de la recherche sur Wikipédia

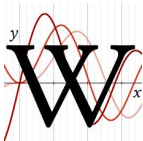
- Questions d'éthique
  - Wikipédia n'est pas un cobaye ;
  - Les personnes qui participent ne sont pas des cobayes ;

⇒ pas d'expérience sociale sur l'encyclopédie

**Pourquoi e...ment j'ai créé un  
canular sur Wikipédia**



# Suivre l'actualité de la recherche sur Wikimedia



- Wikimedia Research Newsletter
  - Mail (<https://lists.wikimedia.org/postorius/lists/research-newsletter.lists.wikimedia.org/>)
  - Mastodon
  - Twitter
- Une liste de diffusion (anglais) : [wiki-research-l@lists.wikimedia.org](mailto:wiki-research-l@lists.wikimedia.org)
- Un colloque chaque année, le wiki workshop

## Wiki Workshop 2023

A forum bringing together researchers exploring all aspects of Wikimedia projects. Held virtually as a standalone event, May 11, 2023.

# + Ressources complémentaires

Une synthèse bibliographique sur Wikipédia :

[Twenty years of Wikipedia in scholarly publications: a bibliometric network analysis of the thematic and citation landscape](#)

Revue de la littérature concernant Wikidata :

[A systematic review of Wikidata in Digital Humanities projects](#)

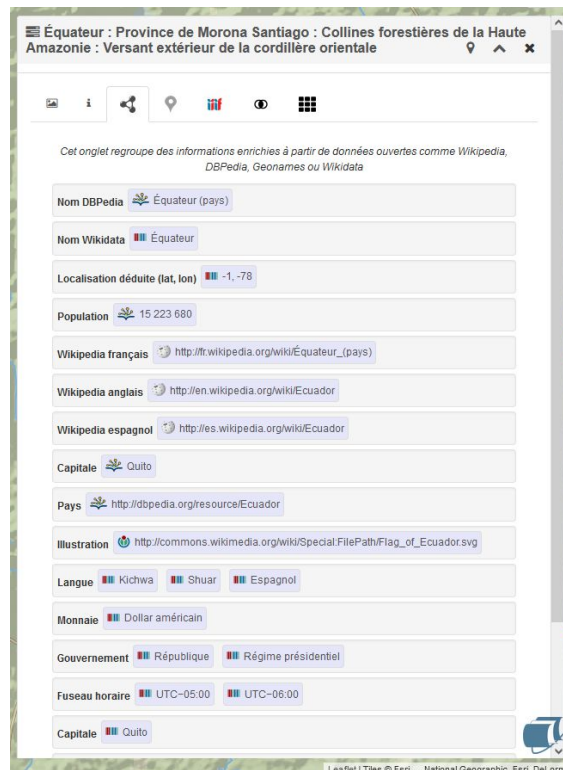




# Intérêt de Wikipédia et Wikidata dans la recherche

- Comme terrain de recherche
- Extraire des données
- Enrichir un jeu de donnée existant
  - Ajout d'informations structurées
  - Ajout de coordonnées géographiques
  - Multilinguisme
- Compléter des jeux de données  
("aligner des données")

Enrichissement de données sur le  
site de cartographie Navigae



Outils

# Panorama des solutions

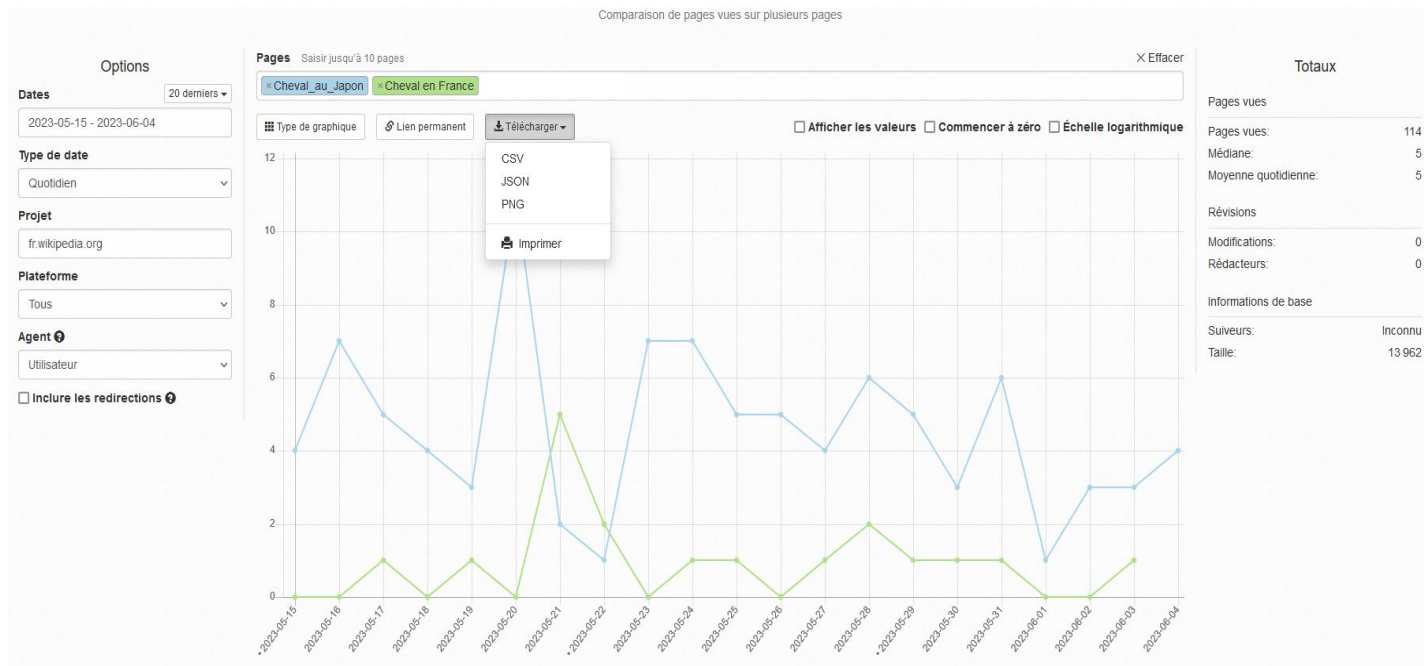


Kati Szilágyi, CC BY-SA 4.0

1. Visualiser/extraire des données
  - a. Audience des articles Wikipédia (PageViews)
  - b. Pages Wikipédia (Excel)
  - c. Récupérer une liste d'articles (Petscan)
  - d. Faire des requêtes SQL pour récupérer beaucoup de données de Wikipédia (Quarry)
  - e. Faire des requêtes SPARQL pour récupérer les données structurées de Wikidata (Wikidata Query Service)
  
2. Enrichir Wikipédia / Wikidata
  - a. QuickStatement
  - b. R/Python
  - c. Pattypan

# Les outils de Wikipédia

- Téléchargement directement depuis la page grâce aux statistiques



# Les outils de Wikipédia (2)

Affiche le nombre de vues par jour, mois, année (depuis 2015)

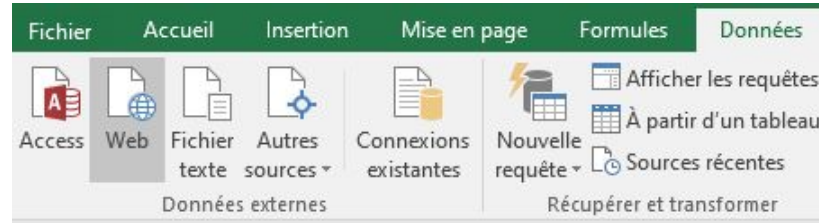
- Outil accessible à partir de l'onglet historique
- Ou directement <https://pageviews.wmcloud.org>
- Et le jeu de données <https://dumps.wikimedia.org>

Ne pas oublier de cocher deux cases :

- Inclure les redirections
- Commencer à zéro

# Wikipédia, une ressource balisée

Import dans un tableur grâce aux balises (Données/web)



# Obtenir une liste d'articles (PetScan)

Outil permettant d'interroger la base de données, sans connaître de langage de programmation

Utilisations :

- obtenir une liste d'articles Wikipédia (par ex. toutes les femmes économistes françaises)
- intégrer facilement cette sélection d'articles dans un autre outil, ex. MassViews



Aide disponible sur <https://meta.wikimedia.org/wiki/PetScan/fr>

# Usages plus poussés (Quarry)

Faire des requêtes SQL en ligne, tout en n'ayant pas à installer sur une machine les dumps.

Pour ne pas surcharger le serveur, il y a un temps d'exécution limité.



Aide disponible sur <https://meta.wikimedia.org/wiki/Research:Quarry>



## Exemples d'utilisations

- [Quelle est la page la plus modifiée de la Wikisource en suédois ?](#)
- [Quel est le quarante-deuxième contributeur le plus prolifique de Wikispecies ?](#)
- [Quelle est la durée moyenne entre deux éditions de l'article "Insomnie" entre 21 h et 7 h ?](#)
- [Liste des articles renvoyant vers persee.fr](#)
- [Pages du site insee.fr les plus mentionnées](#)
- [Sites web les plus mentionnés sur les articles des communes françaises](#)
- [Nombre d'articles supprimés par jour sur la Wikipédia en français](#)

## Récupérer des données structurées

- Quel était le neuvième pays le plus peuplé du monde en 1945
- En 2024, quelles œuvres d'art entreront dans le domaine public ?

Père	<a href="#">Michel Duflo</a> 
Conjoint	<a href="#">Abhijit Banerjee</a> (depuis 2015) 
Autres informations	
A travaillé pour	<a href="#">Massachusetts Institute of Technology</a> <a href="#">Collège de France</a> 
Membre de	<a href="#">Société d'économétrie</a> (2010) <a href="#">Conseil scientifique de l'Éducation nationale</a> (2018) <a href="#">Académie américaine des sciences</a> <a href="#">Académie américaine des arts et des sciences</a>

# Récupérer des données structurées : SPARQL

Interroger Wikidata grâce à des requêtes SPARQL via l'outil Wikidata Query Service (WDQS)

Limite : pour ne pas surcharger le serveur, il y a un temps d'exécution limité (time out fixé à 60 secondes).



Aide disponible sur [https://www.wikidata.org/wiki/Wikidata:Request\\_a\\_query/fr](https://www.wikidata.org/wiki/Wikidata:Request_a_query/fr)

[Support de cours sur SparQL donné à l'école de bibliothécaires documentalistes par Benoît Deshayes en 2023](#)

[Vidéos de Nicolas Vigneron](#) + chaîne Twitch (mardi 19h)

## Exemple 1. éphéméride des personnes enterrées au cimetière du Père-Lachaise

[https://commons.wikimedia.org/wiki/Commons:WikiProject\\_P%C3%A8re-Lachaise\\_Cemetery/Query?uselang=fr#On\\_this\\_day](https://commons.wikimedia.org/wiki/Commons:WikiProject_P%C3%A8re-Lachaise_Cemetery/Query?uselang=fr#On_this_day)

## Exemple 2. répartition par sexe / genre des personnes enterrées au cimetière du Père-Lachaise

<https://w.wiki/6p6y>

## Exemple 3. Localisation des tombes des personnalités sur la Wikipédia francophone

<https://w.wiki/6p7A>

# Modifier une série d'éléments Wikidata : QuickStatements

Modifier massivement Wikidata sans savoir faire fonctionner un robot

QuickStatements est un outil d'édition par lots de Wikidata

This is a new interface for QuickStatements V2. Just in case, the old interface is **here**.

Nouveau lot

numéro du lot

Voir les détails du lot

nom d'utilisateur

Voir les lots faits par un utilisateur

Numéro temporaire

Discuter ou reverter un lot temporaire (qui tourne dans le navigateur)



Aide disponible sur <https://www.wikidata.org/wiki/Help:QuickStatements>

# OpenRefine

Un outil si riche et complet...


Qu'il a fait l'objet d'un tuto  
MATE SHS en 2019 !

OpenRefine : un couteau suisse  
pour manipuler les données



A free, open source, powerful tool for  
working with messy data

Data cleaning avec OpenRefine : nettoyer et enrichir ses données, tuto@mate, 01/2019, stefan.gaget@cnrs.fr

 Aide disponible sur le forum d'OpenRefine

... Et 19 vidéos par le réseau bases de  
données en 2023



OpenRefine



TES PREMIERS PAS AVEC OpenRefine



# Avec R et Python

- R studio
  - WikidataR
  - WikipediR
  - GlittR (orienté web semantic)
- Python
  - Pywikibot
  - Wikipedia



*En général : les applications de web scrapping  
(rvest, BeautifulSoup).*

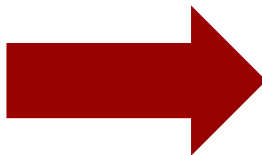
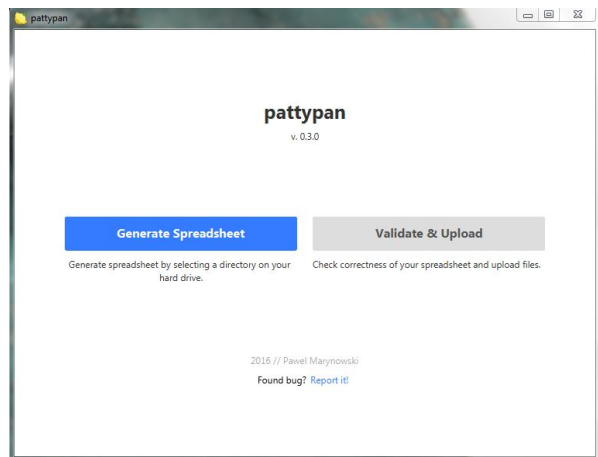
*Pour les stats : pageviews, wikipediatrend*



Aide disponible sur les listes de diffusion R, site internet Rzine  
Manuel Python sur MediaWiki

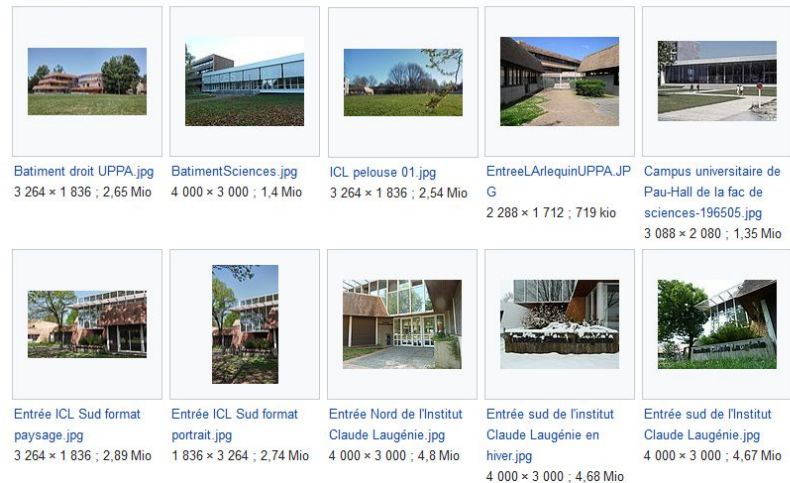
# Pattypan

Outil permettant d'importer des fichiers sur Wikimedia Commons



## Category:Université de Pau et des Pays de l'Adour/Pau

Une page de Wikimedia Commons, la médiathèque libre.  
Cette catégorie comprend 55 fichiers, dont les 55 ci-dessous.



Catégories (+): Universities and colleges in Pau, France (-) (±) (↓) (↑) | Université de Pau et des Pays de l'Adour (-) (±) (↓) (↑) | (+)



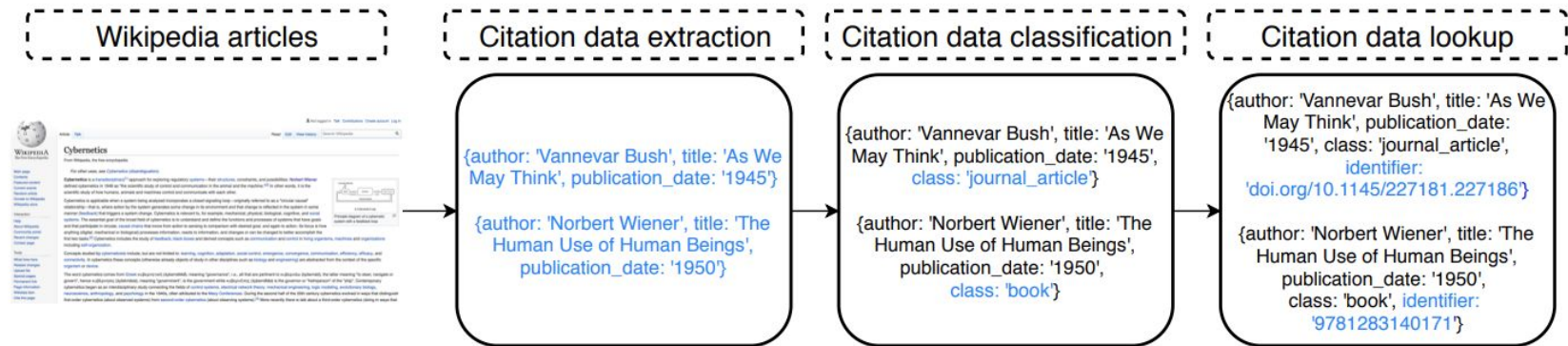
Aide disponible sur <https://commons.wikimedia.org/wiki/Commons:Pattypan>



Données



Figure 1: Example of citations in Wikipedia.



#### Data Dumps (details)

[Homepage – Download](#)

Dumps of all [Wikipedia](#) projects for backup, offline use, research, etc.

- Wiki content, revisions, metadata, and page-to-page and outside links
- XML and SQL format
- once/twice a month
- large file sizes
- The [dumps.wikimedia.org](#) domain also hosts [other data](#)

#### APIs (details)

- The [MediaWiki API](#) provides direct, high-level access to the data contained in MediaWiki databases over the web.
  - Meta info about the wiki and logged-in user, properties of pages (revisions, content, etc.) and lists of pages based on criteria
  - JSON, XML, and PHP's native serialization format
- The [Wikimedia REST API](#) provides e.g. [pageviews and aggregate edit stats](#) (see also [Pageviews API documentation](#)), and [page content in various formats](#).

#### Database access (Toolforge, PAWS, Quarry) (details)

The [Toolforge](#) hosting environment allows you to connect to shared server resources and query a copy of the Wikimedia project's content databases.

- Acts as a standard web server hosting web-based tools
- Command-line tools
- Account required

[PAWS](#) is a [Jupyter Notebook](#) environment within Toolforge that allows e.g. querying database replicas and APIs for analysis.

[Quarry](#) is a public web interface allowing SQL queries to database replicas.

#### Recent changes stream (details)

[Homepage](#)

Wikimedia broadcasts every [change](#) to every Wikimedia wiki using Server Sent Events over HTTP.

#### Analytics Dumps (details)

[Homepage](#)

Raw [pageviews](#), [unique device estimates](#), [mediacounts](#), etc.

- Delimited, usually: Project, (Page title,) Count
- Aggregated hourly or daily
- [pageviews complete](#) – [mediacounts](#) – [unique devices](#)

#### WikiStats (details)

[Homepage](#)

Reports based on data dumps and server log files.

- Unique visits, page views, active editors and more
- Intermediate CSV files available
- Graphical presentation



Explorer les sujets ^

Chercher ou parcourir les questions et en choisir une pour voir les réponses



## Vue générale mensuelle

All wikis



### Lectures

#### Total de pages vues

**25MRD**

Mai ↑ 4.12 % par rapport au mois précédent



**284MRD**

↑ 17.77 % par rapport à l'année précédente

12 Derniers Mois (Jun 2022 - May 2023)

#### Pages vues par pays

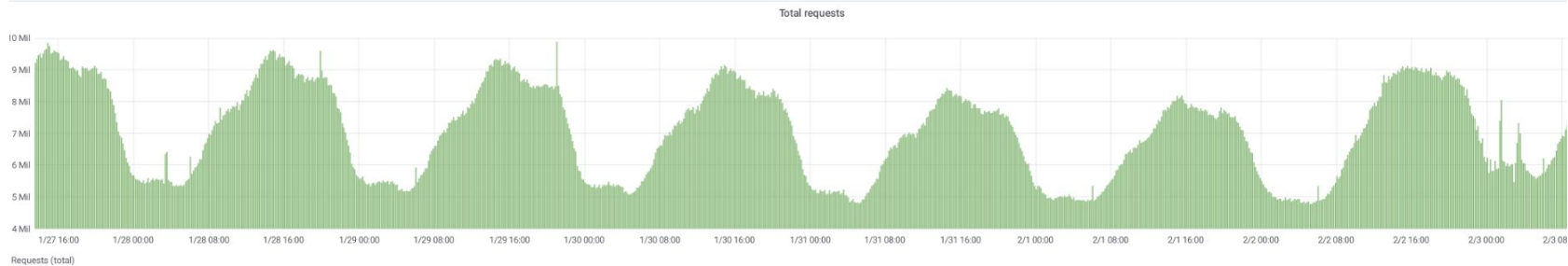
Pays avec le plus de vues pour mai

3MRD	United States of A
1MRD	Japan
943M	India
905M	United Kingdom

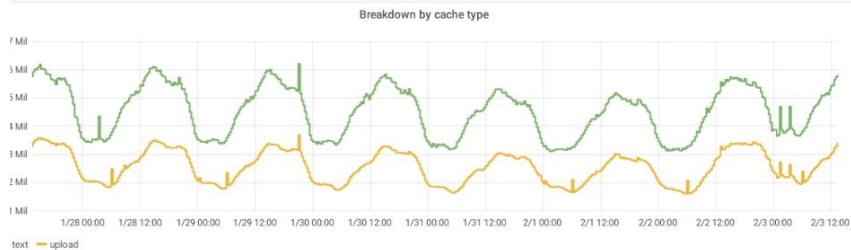


L'indicateur Appareils uniques n'est pas disponible pour tous les projets. Veuillez choisir un wiki spécifique.

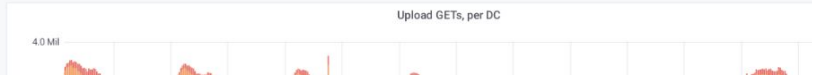
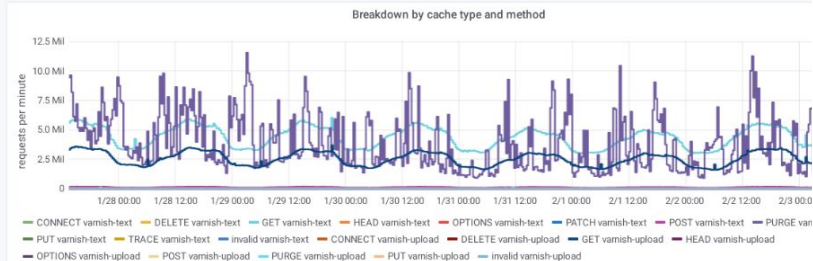
## Overview



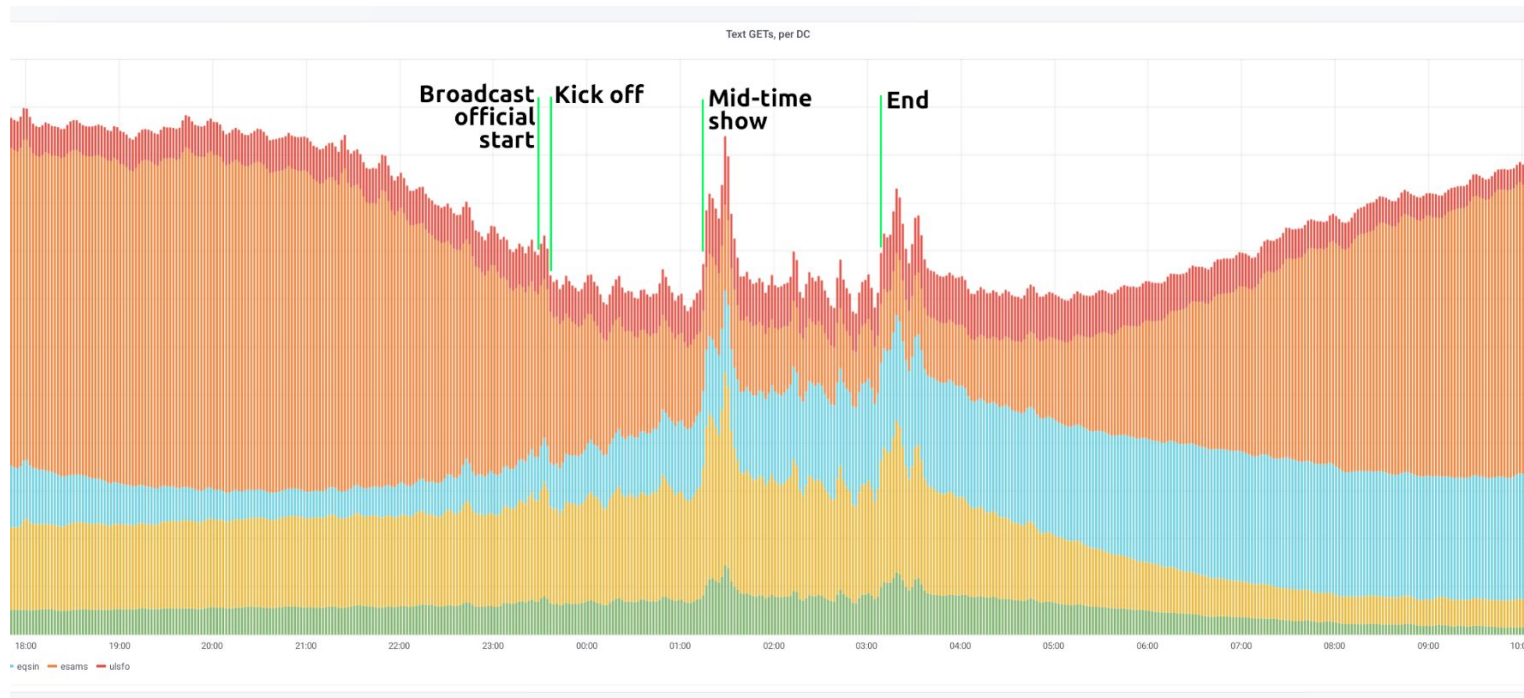
## Breakdown by cache type



## Breakdown by DC

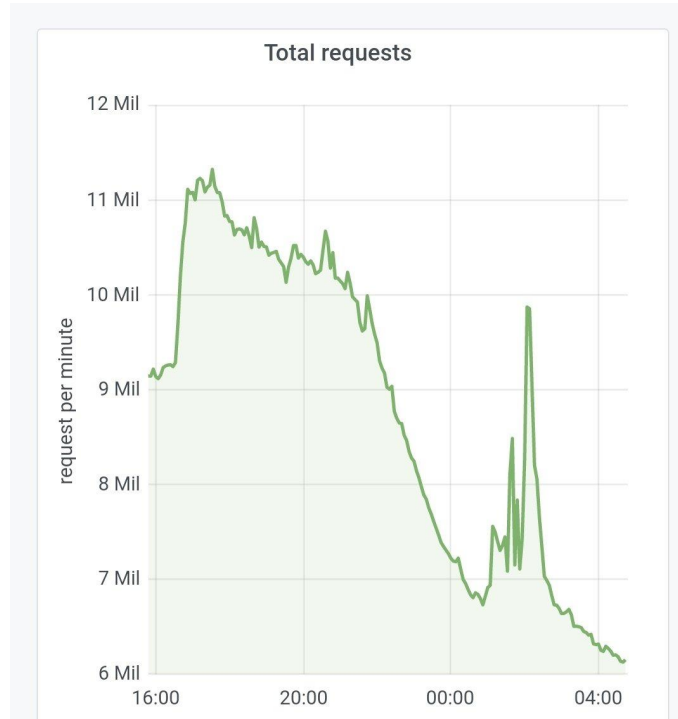


# Superbowl

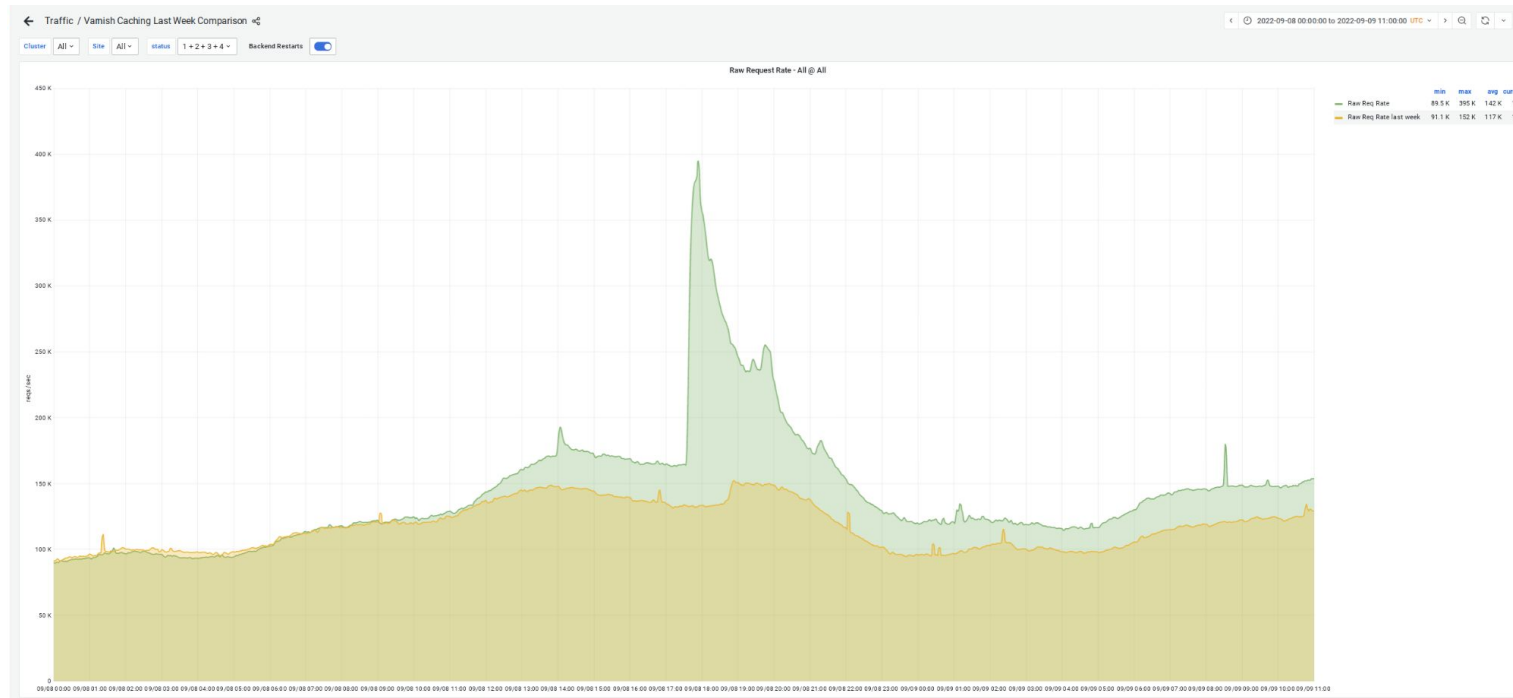


<https://twitter.com/jynus/status/1224331669194395650>

# Élection de Biden



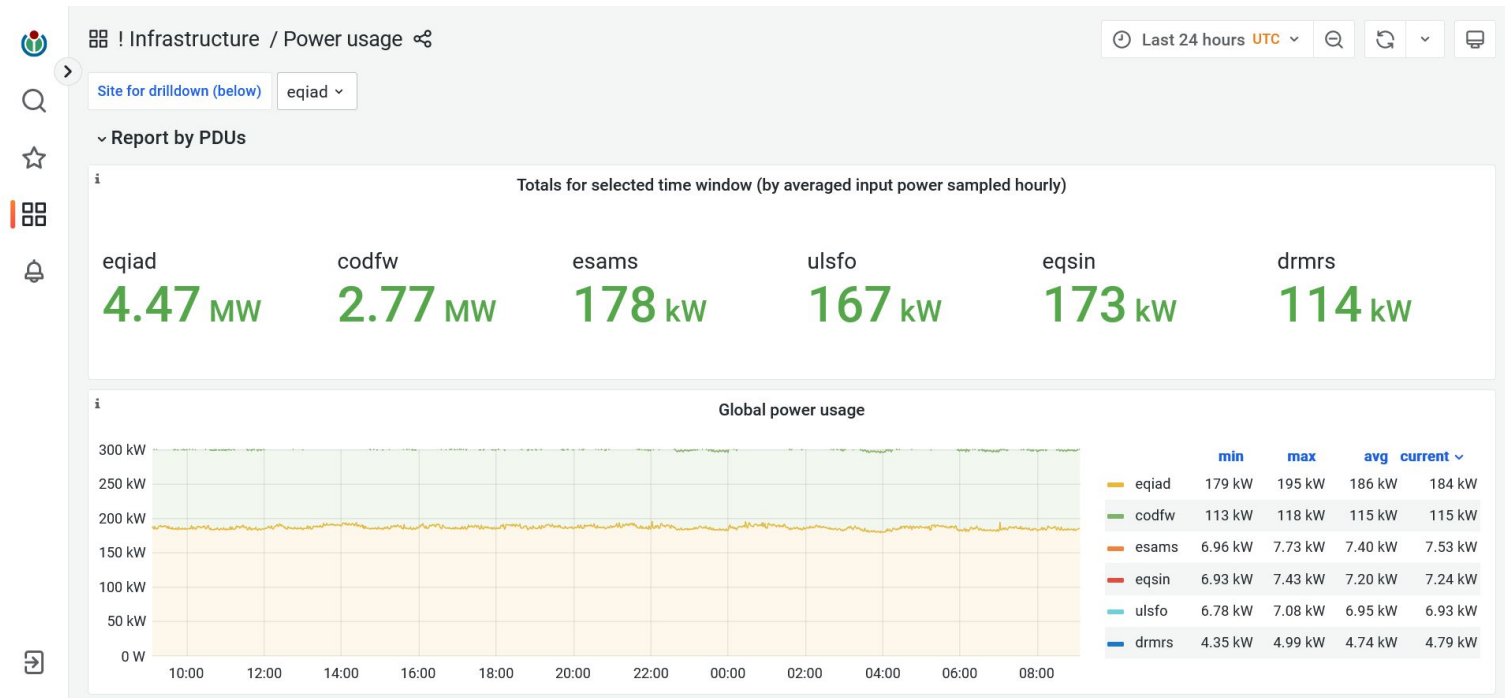
# Décès d'Elisabeth II



[https://grafana.wikimedia.org/d/O\\_OXJyTVk/home-w-wiki-status](https://grafana.wikimedia.org/d/O_OXJyTVk/home-w-wiki-status)



# Consommation électrique



## Contenu

	Articles	Total
<b>Pages</b>	2 527 395	12 587 993
<b>Modifications</b>	139 160 754	194 350 504
<b>Contenu actuel</b>		34 G
<b>Contenu avec historique</b>		3967 G

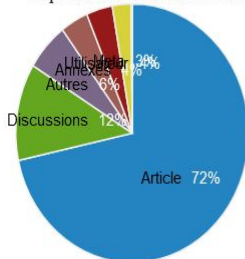
## Utilisateurs

	Actifs/mois	Total
<b>Utilisateurs</b>	2 108	1 217 780
<b>5+ modifs</b>	678	376 298
<b>100+ modifs</b>	102	37 515
<b>1000+ modifs</b>	10	8 185

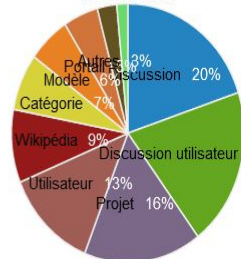
## Moyennes par jour

	Articles	Total pages
<b>Modifications</b>	20 164	28 592
<b>Créations</b>	275	1 604
<b>Contenu</b>	3 M	5,2 M

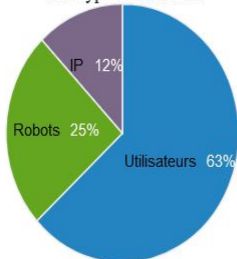
Répartition des modifications



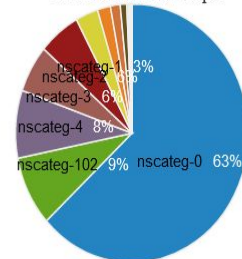
Hors articles



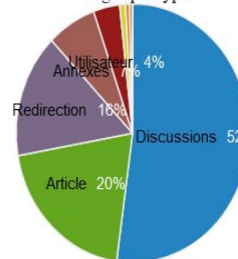
Par type d'utilisateur



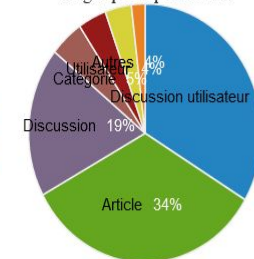
Contenu avec historique



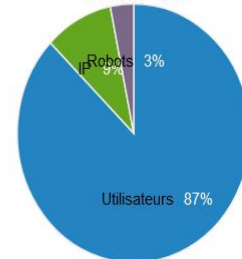
Pages par type



Pages par esp. de noms



Articles créés

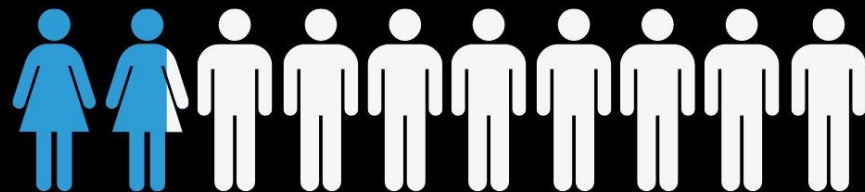
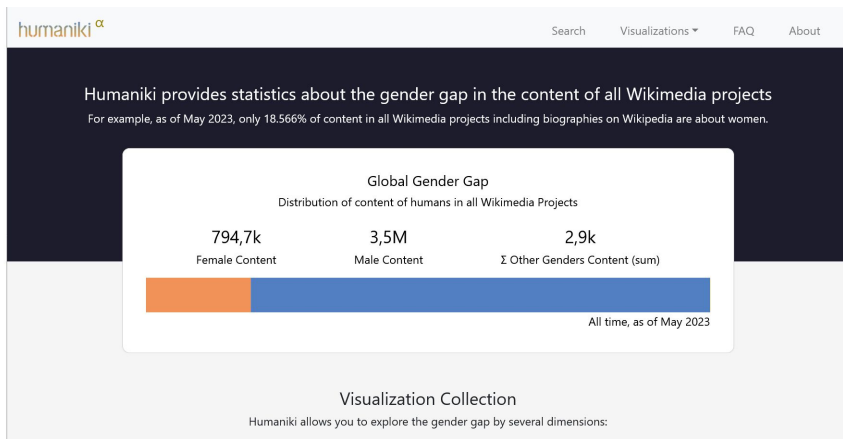


## Utilisateurs / jour



## Modifications des utilisateurs / jour





**18 % des biographies de la Wikipédia en langue française consacrées aux femmes**



Période	Biographies	% femmes
1800-1899	95 284	6-14 %
1900-1949	139 483	14-18 %
1950-1999	235 228	20-30 %
2000-2020	1 741	33-42 %

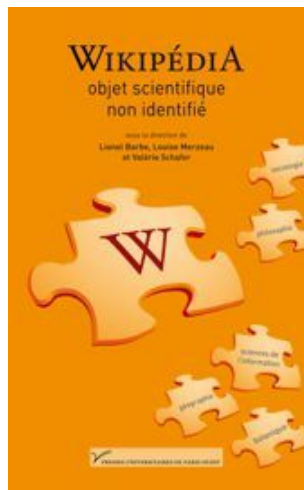
Se former

**[MOOC]**  
Wikipédia

**[MOOC]**  
Wikidata

- Wikipédia : vidéos sur [Wikimédia Commons](#) et [YouTube](#)
- Wikidata (2e session à la rentrée 2023)

# Livres



<https://inventaire.io/users/pyb/inventory>

# Se former et rencontrer la communauté

Permanence wikipédia/groupe local (cabale)/éditathon/wiki camp/wiki convention



Wiki-camp en 2017,  
Par Azniv Stepanian, CC BY-SA 4.0



Wiki-convention francophone en 2022  
Par Thesupermat, CC BY-SA 4.0



Cabale de Brest en 2015,  
Par Filyg, CC BY-SA 4.0

# Formations URFIST 2023



## Prochains événements :

- 13 juin. Webinaire sur Wikidata
  - 16 juin. Webinaire réseaux Médecis & Repères sur les revues scientifiques et Wikipédia
  - 20 juin. Bibliothèques et IA (ChatGPT) : quelles leçons retenir de la réception de Wikipédia ?
  - Wikicafés : 20 juin, 4 et 18 juillet
- 
- [https://fr.wikipedia.org/wiki/Projet:Wikifier\\_la\\_science](https://fr.wikipedia.org/wiki/Projet:Wikifier_la_science)
  - <https://sygefor.reseau-urfist.fr/>



# Merci de votre attention, place aux questions



Réalisé à partir de plusieurs présentations dont : Amélie Charles WMFr.

Présentation sous licence CC0.