

# Webinaire Open Data

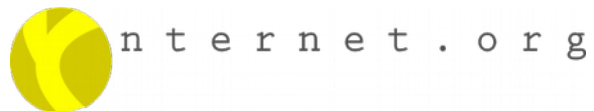


# WIKIDATA

**"Wikidata, le Wikipédia que vous utiliserez bientôt tous"**  
[Article paru](#) dans le Nouvel Observateur (L'Obs).

Mercredi 20 mai 2020

Organisé par Yinternet.org



# Florence Devouard (Q266722)

Username on Wikipedia & Wikidata : Anthere



Formerly Chair of Wikimedia Foundation



Cofunder and VP Wikimedia France



Funder Wiki in Africa



Board member Open Food Facts



VP Yinternet.org



Twitter: @anthere  
Email : [fdevouard@anthere.org](mailto:fdevouard@anthere.org)


# Wikimedia galaxy



TOM SIMONITE BUSINESS 02.18.18 06:00 AM

**SHARE**

INSIDE THE ALEXA-FRIENDLY WORLD OF WIKIDATA



TYLER GROSS

**HUMANS PRICKED BY** info-hunger pangs used to hunt and peck for scraps of trivia on the savanna of the internet. Now we sit in screen-glow-flooded caves and grunt, “Alexa!” Virtual assistants do the dirty work for us. Problem is, computers can’t really speak the language.

**Alexa, Ask Wikidata! Voice interaction with knowledge graphs using Amazon Alexa**

Peter Haase, Andriy Nikolov, Johannes Trame, Artem Kozlov, Daniel M. Herzig

metaphacts GmbH, Walldorf, Germany  
{an, ph, jt, ak, dh}@metaphacts.com

**Abstract.** Voice-enabled user interfaces have become a popular means of interaction with various kinds of applications and services. In addition to more traditional interaction paradigms such as keyword search, voice interaction can be a convenient means of communication for many groups of users. Amazon Alexa

On Wire, for PA Alexa

**Siri Erroneously Told People Stan Lee Was Dead**

Beth Elderkin  
7/03/18 2:45pm • Filed to: STAN LEE

17.4K 48 5



Stan Lee at the premiere of Spider-Man: Homecoming.  
Photo: Alberto E. Rodriguez (Getty Images)

For a few brief moments, comic book fans around the world were shocked to hear some tragic news. But luckily, the panic didn’t last long. As first reported by CinemaBlend, Siri spent a little time this week telling people that Stan Lee had died on July 2. Why would a computer program falsely report a famous figure’s death? You can blame a Wiki user for that.

Sources tell us the problem can be traced back to revisions in Lee’s Wikidata. If you look at the [profile’s recent history](#), Wiki user “&beer&love” changed Lee’s Wiki data to include a “date of death.” Since Siri [pulls data automatically from Wiki pages](#), without edit or modification, the program temporarily included the false death claim. The timing of the inquiry just happened to coincide with the false information being present at that point.

Mistake on WikiData found its way on Siri...

# Time for the tough question...

So we know the mayor of Barcelone is a female...

But can you list the biggest cities in the world with female mayor ?

Try it : <https://tinyurl.com/rpg6olv>



Wikidata is an open, multilingual structured knowledge base that can be read and edited by both humans and machines.

It provides the power of searching, sorting and querying

It is a place where humans and bots can share knowledge on more equal terms.



The sum of all **human knowledge**



**WIKIDATA**

The sum of all **data and sources** backing human knowledge

# Wikidata contains items...

An item may be an object, a topic, a person, a concept, a location...

Each item has a unique identifier

For example,

- Barcelona is Q1492
- Education is Q8434
- Women is Q467

Each item is language neutral



Ada (Q346047)

Ada Lovelace (Q7259)

Ada Colau (Q4779594)

Spain (Q29)

female (Q6581072)

Mayor of Barcelona (Q12155639)

2

given name  
P735

given name  
P735

sex or gender  
P21

number of children  
P1971

country of citizenship  
P27

sex or gender  
P21

coordinate location  
P625

country  
P17

position held  
P39



41°22'57"N, 2°10'37"E

Qnnn      Pxxx      Qmmm  
Thing      instance of      other thing

For example Q4779594      P31      Q5

Means « Ada Colau is a human »

And also means « Ada Colau est un être humain »

And it might mean « Ada Colau es un humano »

Rather unknown... yet...

# Wikidata is BIG

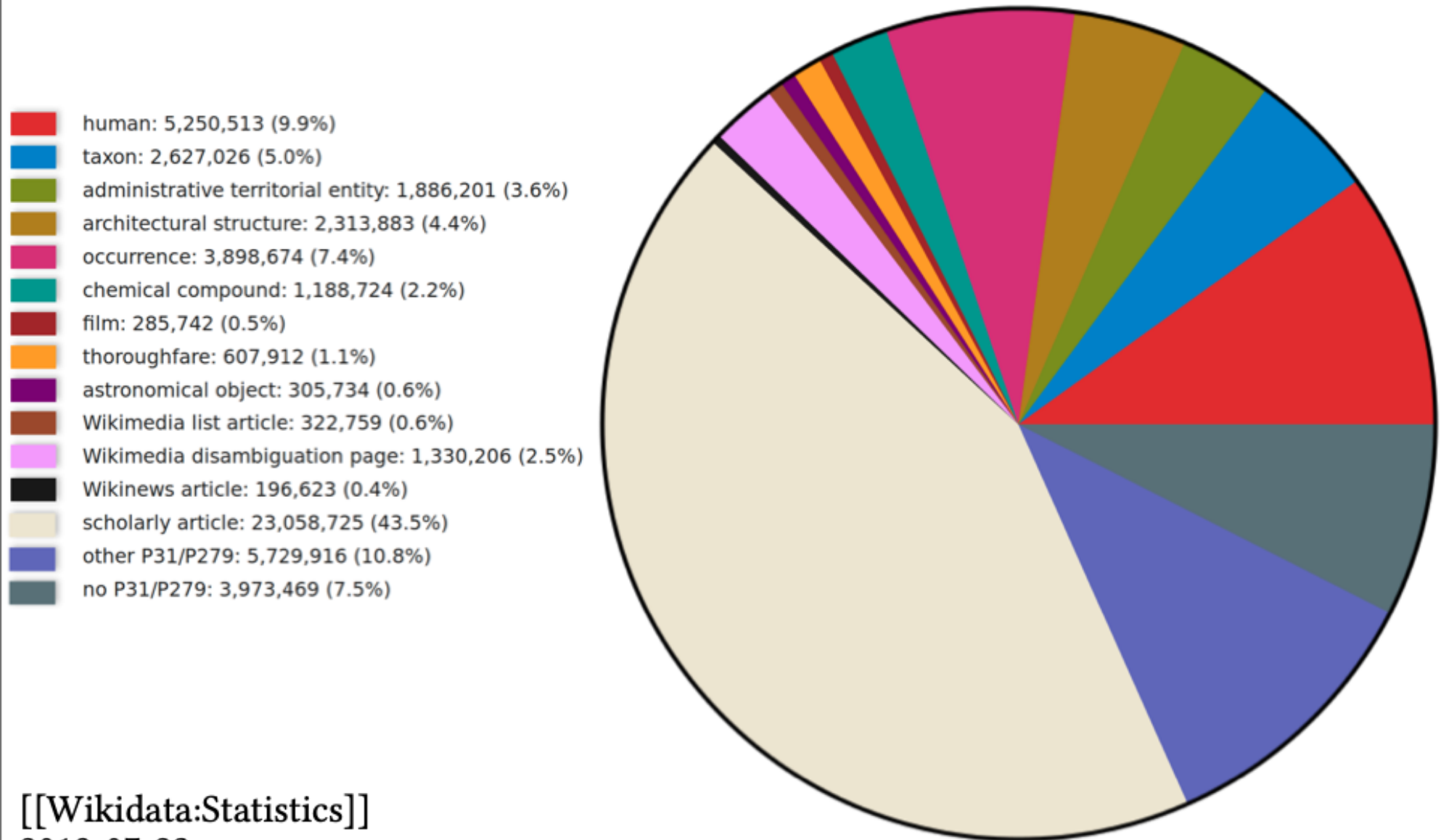
~~67 millions~~ of unique elements in Nov 2019

**86 millions** of unique elements in May 2020

**over 1 billion edits** have been made  
since 2012 (creation)

**12000 - 20000 active editors** in 300 languages

# WHAT DO WE HAVE? IN WIKIDATA?



[[Wikidata:Statistics]]

2019-07-23

# Tools to improve Wikidata

## Case 1 : Mix'n'match

This tool can list entries of some external databases, and allows users to match them against Wikidata items. REDLISTS on STEROIDS !

<https://tools.wmflabs.org/mix-n-match/#/>

The screenshot shows the 'Mix'n'match' tool interface. At the top, there are navigation links: 'Tools', 'Git', 'Discuss', and a logo. Below this is a breadcrumb trail: 'Catalogs | Recent Changes | Disambiguation links | Same name | Creation candidates | Search | Permalink'. The main content area displays the following information:

- Catalog ID:** 11L3
- Catalog Name:** the articles of faith, confessions
- Catalog description:**

Below the catalog information is a search input field with the placeholder text 'Qxxx or yy.wikipedia'. To the right of the input field are four buttons: 'Set Q' (blue), 'No Wikidata entry' (orange), 'N/A' (red), and 'Skip' (grey). Below the search field, the results are titled 'en.wikipedia search results' and are presented in a table with the following columns: Article Title, Description, and Wikidata ID.

Article Title	Description	Wikidata ID
<a href="#">George, Margrave of Brandenburg-Ansbach</a>	certain <b>articles of faith</b> and <b>confession</b> to be drawn up by Luther; <b>the</b> commission was executed in <b>the</b> seventeen <b>articles of Schwabach</b> on <b>the</b> basis of <b>the</b> fifteen	Q384941
<a href="#">Creed</a>	" <b>Articles of Faith</b> " redirects here. For other uses, see <b>Articles of Faith</b> (disambiguation). This article is about a statement of belief. For other uses	Q183770
<a href="#">Reformed confessions of faith</a>	Reformed <b>confessions of faith</b> are <b>the confessions of faith</b> of various Reformed churches. These documents express consensus on doctrine for <b>the</b> church adopting	Q2992463
<a href="#">1689 Baptist Confession of Faith</a>	<b>The</b> 1689 Baptist <b>Confession of Faith</b> , also called <b>the</b> Second London Baptist <b>Confession</b> , was written by Particular Baptists, who held to a Calvinistic Soteriology	Q965897
<a href="#">French Confession of Faith</a>	<b>The</b> French <b>Confession of Faith</b> (1559) or <b>Confession</b> de La Rochelle or Gallic <b>Confession of Faith</b> or La Rochelle <b>Confession of Faith</b> is a Reformed confession	Q1125186
<a href="#">Book of Confessions</a>	Creed, <b>the</b> Scots <b>Confession</b> , <b>the</b> Heidelberg Catechism, <b>the</b> Second Helvetic <b>Confession</b> , <b>the</b> Westminster <b>Confession of Faith</b> , <b>the</b> Shorter Catechism, <b>the</b> Larger	Q4943008
<a href="#">Harmony of the Confessions of Faith</a>	<b>articles:</b> Catechism and Reformed Christian <b>confessions of faith</b> Published in 1581, <b>the</b> Harmonia confessionum fidei (Harmony of Confessions of Faith)	Q5659221
<a href="#">Westminster Confession of Faith</a>	<b>The</b> Westminster <b>Confession of Faith</b> is a Reformed <b>confession of faith</b> . Drawn up by <b>the</b> 1646 Westminster Assembly as part of <b>the</b> Westminster Standards	Q815013
<a href="#">Dordrecht Confession of Faith</a>	<b>The</b> Dordrecht <b>Confession of Faith</b> is a statement of religious beliefs adopted by Dutch Mennonite leaders at a meeting in Dordrecht, <b>the</b> Netherlands, on	Q957109

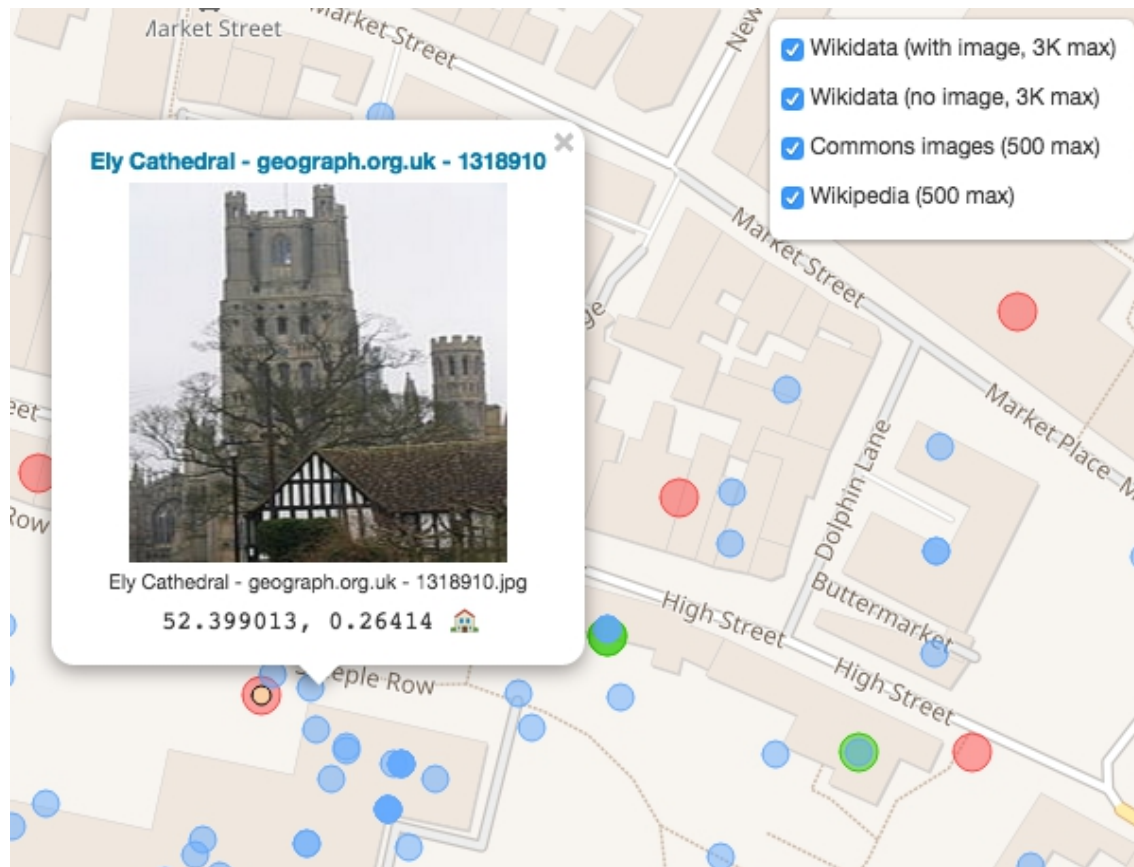
At the bottom of the page, there are links for 'Google-search Wikipedias | Wikidata | Wikisource'.

Wikidata aligns with >2500 databases and catalogues

# Tools to improve Wikidata

## Case 2 : WikiShootMe

This is a tool to show Wikidata items, Wikipedia articles, and Commons images with coordinates, all on the same map. It facilitates addition of images to Wikimedia projects. <https://tools.wmflabs.org/wikishootme>



# Tools to improve Wikidata

## Case 3 : WikiData games

A set of **games** to quickly add statements to **Wikidata**.

<https://tools.wmflabs.org/wikidata-game/>

<https://tools.wmflabs.org/wikidata-game/distributed/>

### Birthday Conflicts

Confirm/fix wikidata date of birth records.

We found the birthday information may differ between wikidata and wikipedia pages (currently only en, fr, and de wikipedia were analyzed). Click "Keep wikidata" to confirm the wikidata record. Click the green buttons to change the wikidata record to corresponding wikimedia records. If you're not sure, click "Skip". This game only changes wikidata records. It's highly recommend to correct the wrong wikimedia pages directly (with reliable reference), to make the records consistent across different wiki pages. Bug reports and feedback should be sent to [User:Chaoyuel](#).

#### Anastasia Acosta [Q442543]

Anastacia Acosta Anastasia Acosta Rodríguez Αναστασία Ακωστής Анастасия

[Auto](#) | [de](#) | [el](#) | [en](#) | [eo](#) | [es](#) | [fr](#) | [it](#) | [pt](#) | [ru](#) | [yi](#)

Anastasia Acosta est une acteur de télévision, mannequin et acteur de cinéma Costa Rica.  
Elle est née le 24 avril 1975 à San José.

Wikidata record: 1975-04-24

EN Wikipedia: 1974-10-18

DE Wikipedia: 1975-04-24

[Keep wikidata](#) [1974-10-18](#) [skip](#)

#### Justina Robson [Q442526]

ジャスティナ・ロブソン

[Auto](#) | [commons](#) | [de](#) | [en](#) | [en.wikiquote](#) | [fi](#) | [fr](#) | [ja](#) | [sv](#)

Justina Robson est une écrivain et écrivain de science-fiction britannique.  
Elle est née le 11 juin 1968 à Leeds.  
Elle a étudié à université d'York.



Wikidata record: 1968-06-11

FR Wikipedia: 1968-06-11

DE Wikipedia: 1968

# Wikidata and queries

At least 26 different tools allow to query the data in different ways.

**Wikidata Query Service** lets users run queries on the data contained in Wikidata.

The service uses **SPARQL** as the query language.

List of examples :

[https://www.wikidata.org/wiki/Wikidata:SPARQL\\_query\\_service/queries/examples](https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/queries/examples)

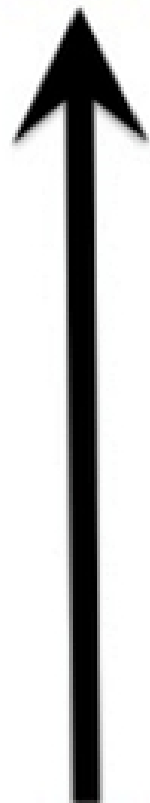
Eg 3.13.1 Overall causes of death ranking

<https://query.wikidata.org>













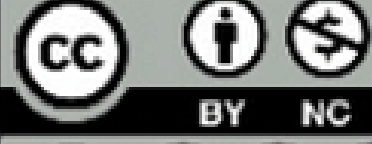









Try it !



Most Free



Least Free

 PUBLIC DOMAIN		Can someone use it commercially?	Can someone create new versions of it?
 Attribution			
 Share Alike			Yup, AND they must license the new work under a Share Alike license.
 No Derivatives			
 Non-Commercial			Yup, AND the new work must be non-commercial, but it can be under any non-commercial license.
 Non-Commercial Share Alike			Yup, AND they must license the new work under a Non-Commercial Share Alike license.
 Non-Commercial No Derivatives			

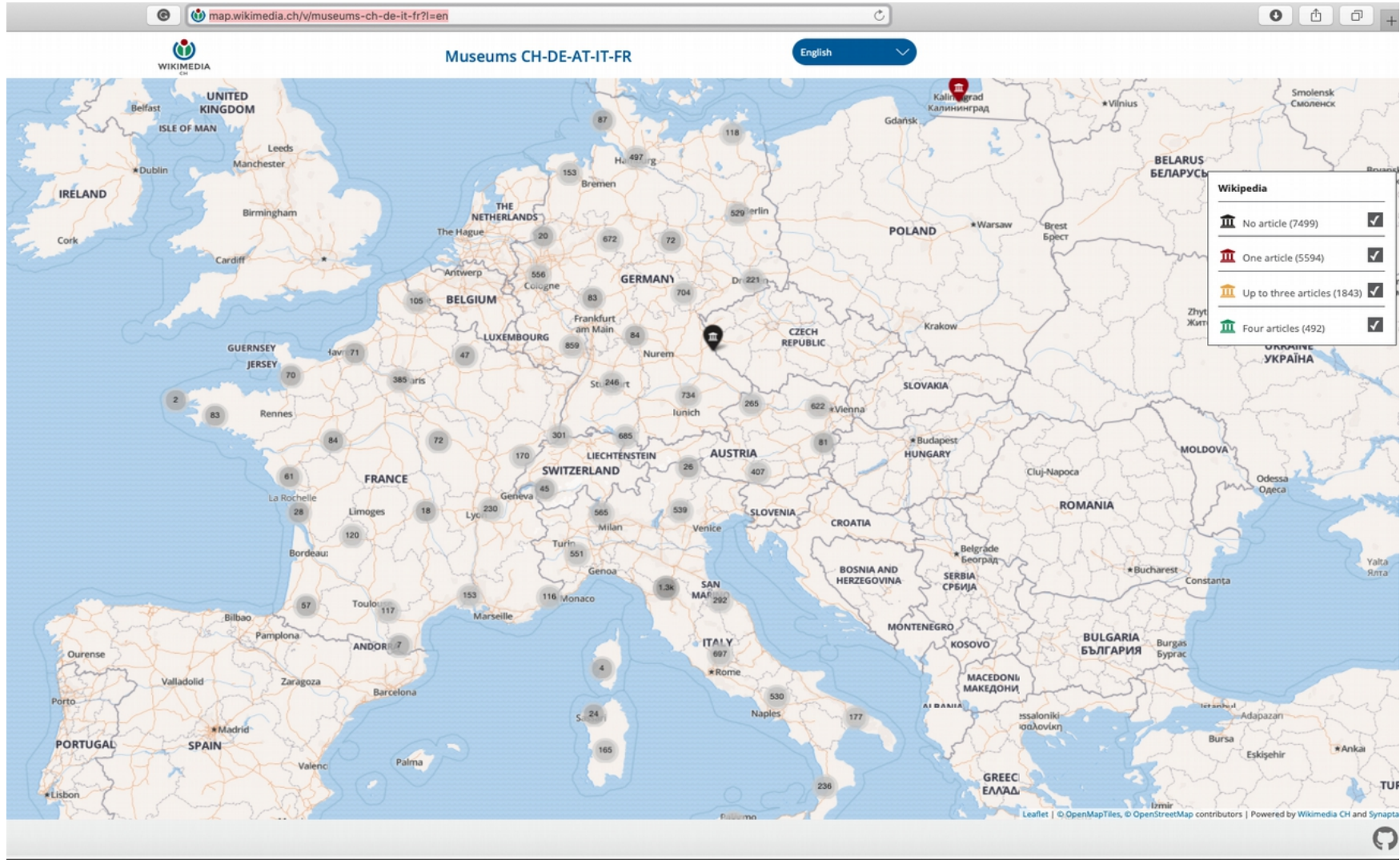
# Linked data versus Open data

"Open Data" : is accessible to anyone (e.g. without monetary cost to access, or not behind a wall) with a permissive license on reuse (e.g. public domain or CC0)

"Linked data" : data that is machine readable, semantic data, that a machine can 'understand'.

Wikidata for wikimedia  
projects

# Wikidata to identify content gaps



- <http://map.wikimedia.ch/v/museums-ch-de-it-fr?l=en>

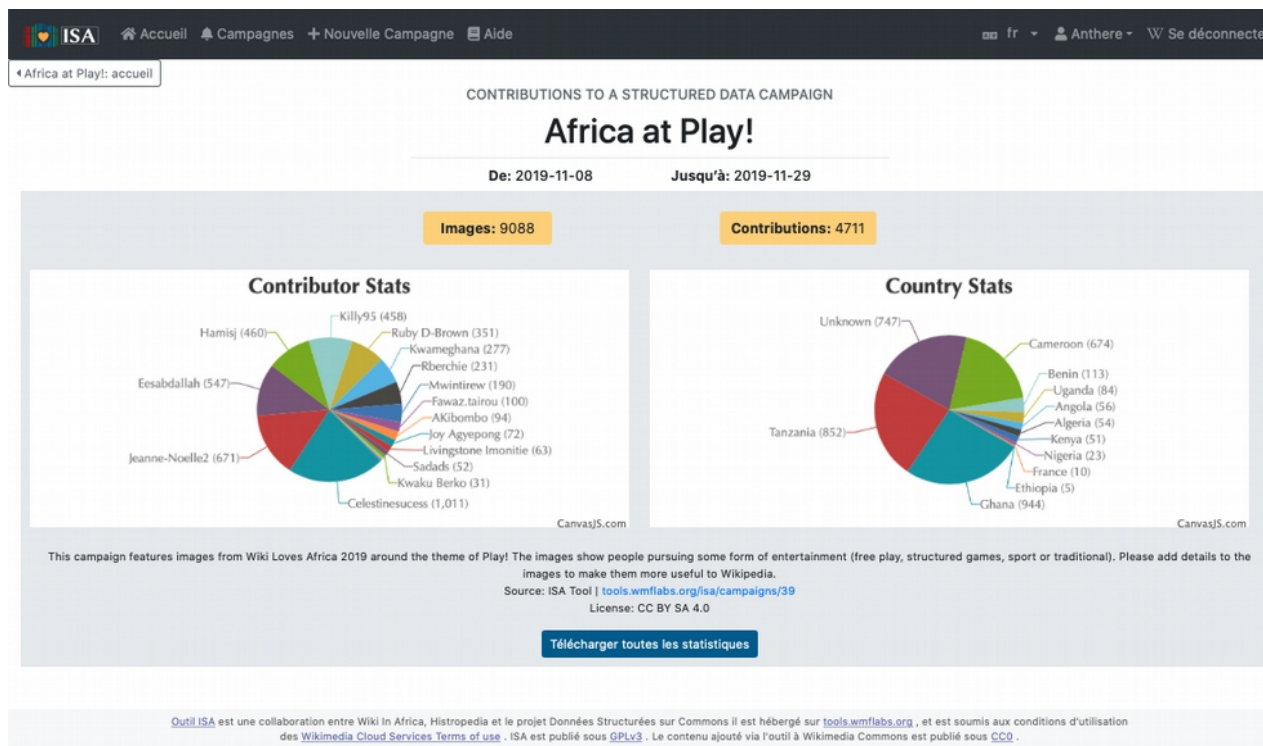
# ISA Tool, a tool to add structured data to images



## ISA TOOL :

A pilot tool to easily add structured data to Wikimedia Commons

<https://tools.wmflabs.org/isa/campaigns/42>



# Wikidata, a support to Wikimedia Commons

- Infoboxes on Wikimedia Commons pages

<https://commons.wikimedia.org/wiki/Barcelona>

- Structured data to better describe images

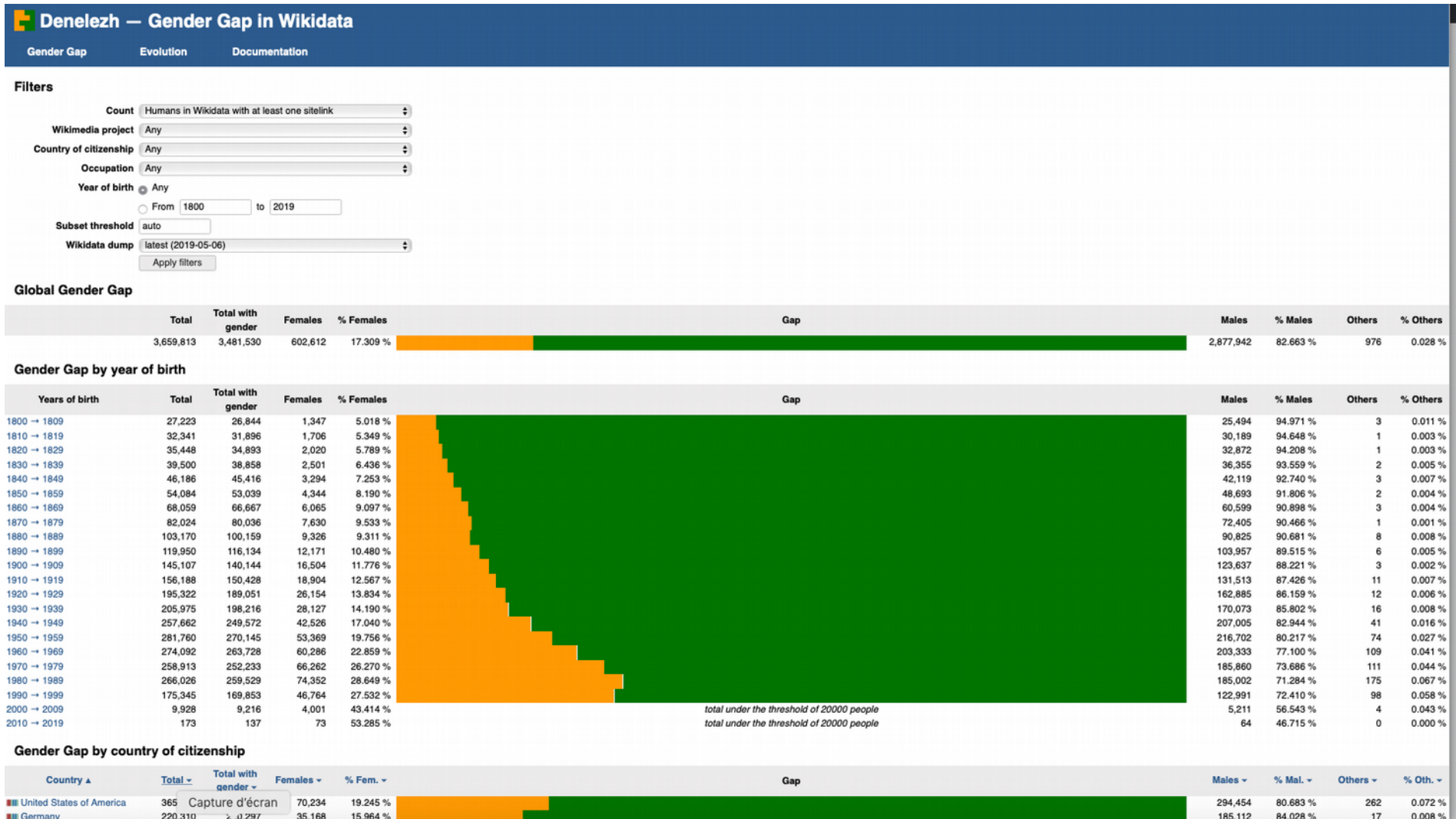
[https://commons.wikimedia.org/wiki/File:Münster,\\_LVM,\\_Skulptur\\_–Körper\\_und\\_Seele\\_–\\_2016\\_–\\_5920-6.jpg](https://commons.wikimedia.org/wiki/File:Münster,_LVM,_Skulptur_–Körper_und_Seele_–_2016_–_5920-6.jpg)



# Third party services using Wikidata

## Denelezh

a tool to explore the gender gap  
<https://www.denelezh.org>



# Listeria

Use Wikidata to create todo lists for Wikipedia

Let's have a look on two examples...

[https://en.wikipedia.org/wiki/Category:Lists\\_based\\_on\\_Wikidata](https://en.wikipedia.org/wiki/Category:Lists_based_on_Wikidata)

[https://en.wikipedia.org/wiki/User:Camelia.boban/WikiGap\\_business\\_women](https://en.wikipedia.org/wiki/User:Camelia.boban/WikiGap_business_women)

Exemple : Liste de biographies de femmes algériennes dont il existe au moins un article wikipédia, mais pas en français

[https://fr.wikipedia.org/wiki/Projet:Wiki\\_Loves\\_Women/Articles\\_manquants\\_par\\_pays/Algérie](https://fr.wikipedia.org/wiki/Projet:Wiki_Loves_Women/Articles_manquants_par_pays/Algérie)

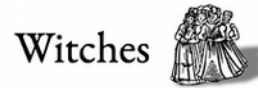
Exemple : liste des occupations relatives au cinéma, avec libellé multi-genres

[https://en.wikipedia.org/wiki/Wikipedia:WikiProject\\_Wiki\\_Loves\\_Women/Occupations/fr](https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Wiki_Loves_Women/Occupations/fr)



Wikidata for the third  
parties

# Witches map by University of Edinburgh



Witches

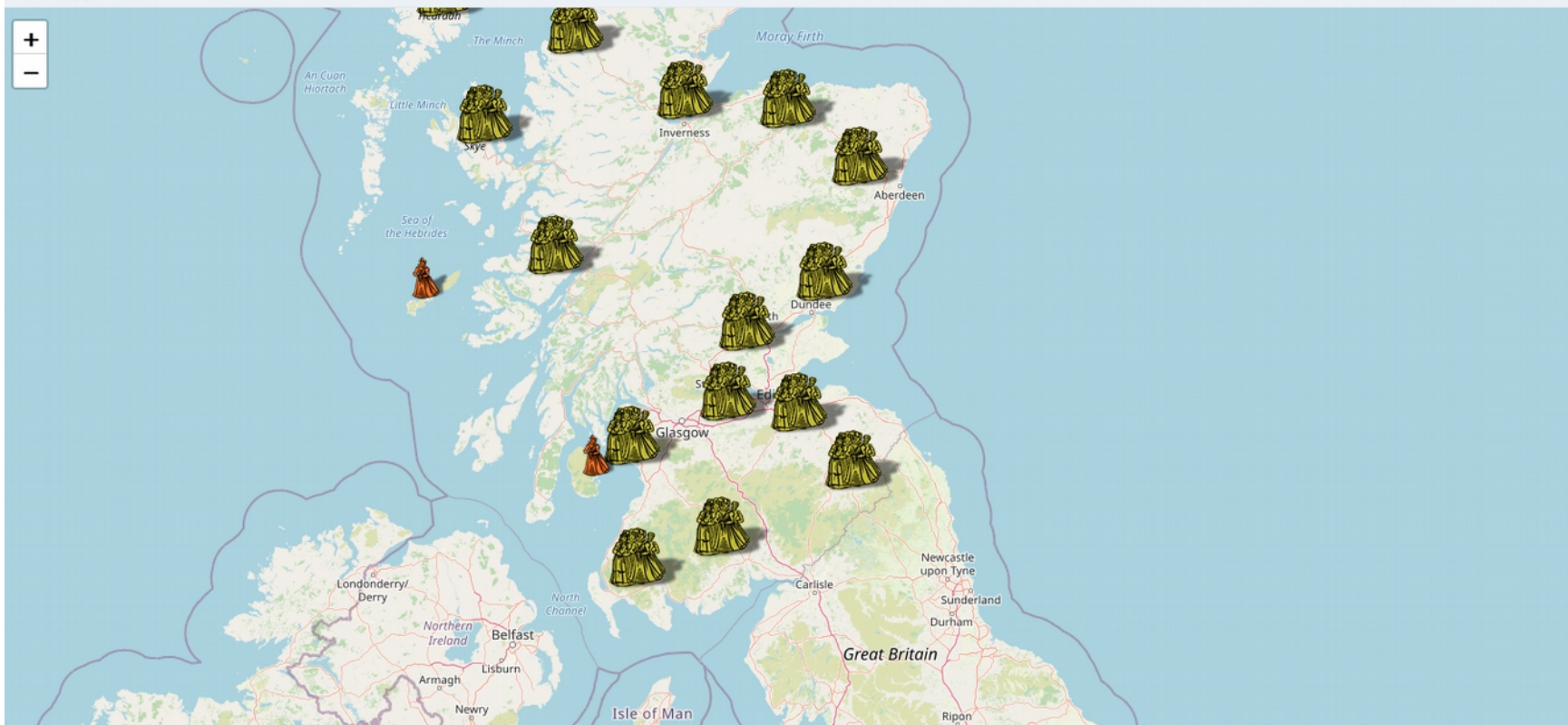
Residences Detention Locations Trial Locations Death Locations People Associated Extra Visualisations About Blog



THE UNIVERSITY  
of EDINBURGH

Places of Residence for Accused Witches (total named accused witches: 3141) (i)

Filters ▾



<https://witches.is.ed.ac.uk/>

# Wikidata for companies IA and Natural language

## Kairntech : quand l'IA part à la conquête du langage naturel

19 février 2019 | Arrivées / départs | Tarmac, L'écho des startups du Tarmac, Tarmac incubator



Point critique dans le développement de l'IA et du machine learning : la **disponibilité de jeux de données d'entraînement.**

Fréquentes dans le monde académique, ces jeux de données n'existent pratiquement pas dans l'entreprise où 80% des données sont non structurées (texte, audio) ce qui rend inexploitable les algorithmes d'IA.

### De la Vision au Langage...

L'IA a franchi un véritable cap en 2012 dans le domaine de la Vision avec de nombreuses applications industrielles à présent implémentées dans les secteurs de l'imagerie médicale, du transport... Dans le domaine du Langage, l'année 2018 a marqué un tournant pour l'IA grâce à de nombreuses avancées technologiques : la performance accrue des algorithmes et des modèles de langage pré entraînés, l'augmentation exponentielle de la puissance de calcul...

Reste un point critique dans le développement de l'IA et du machine learning : la disponibilité de jeux de données d'entraînement. En effet, toute IA a besoin d'exemples et de contre exemples pour apprendre et devenir auto-apprenante. Contrairement au monde académique, ces jeux de données dans le domaine du langage n'existent pratiquement pas dans l'entreprise où 80% des données sont non structurées (texte, audio...) ce qui rend inexploitable les algorithmes d'IA.

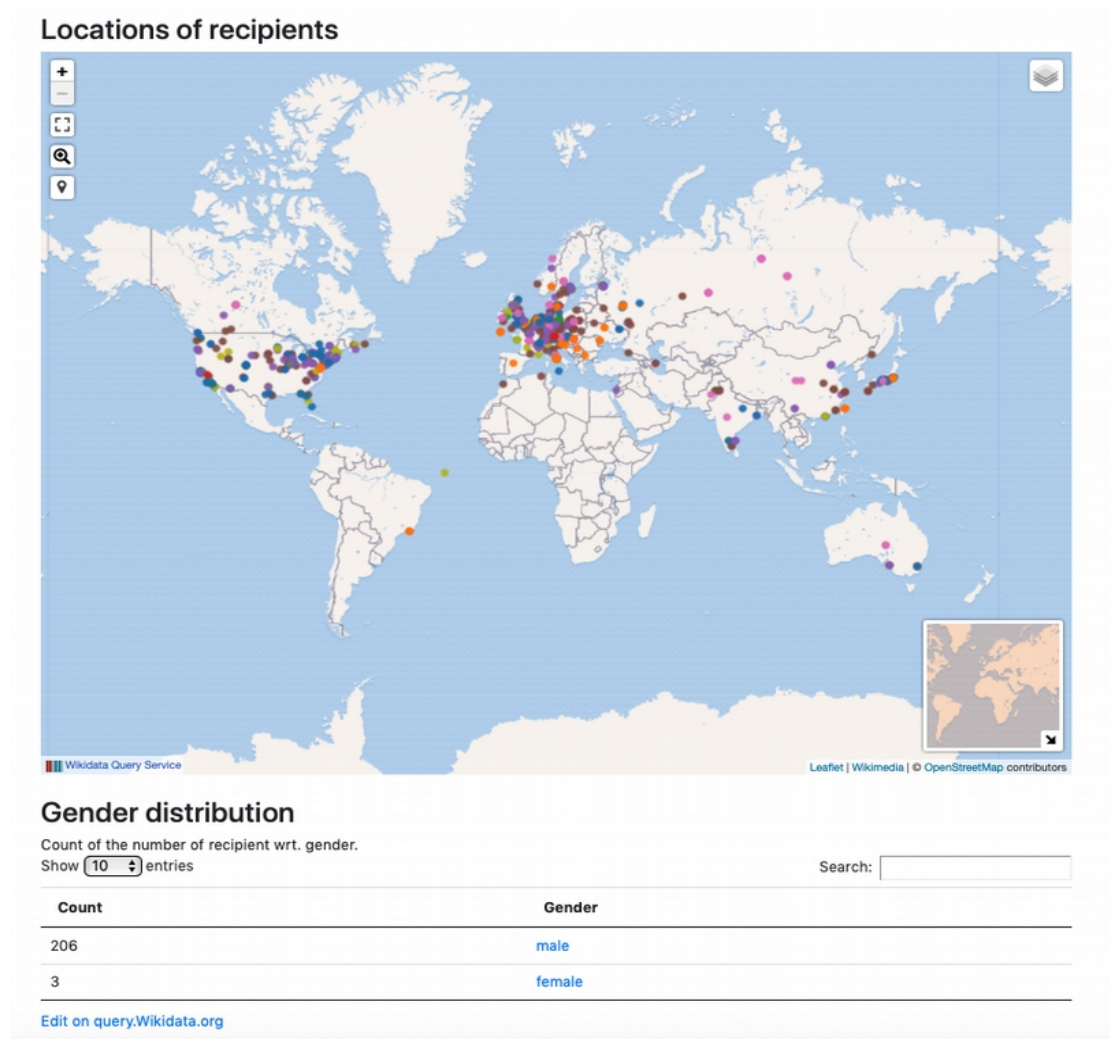
### ...l'importance des jeux de données pour l'IA

C'est sur cette faille qu'Olivier Terrier et Vincent Nibart, forts de plus de 20 ans d'expérience en R&D (IA, traitement du langage naturel et développement logiciel), entendent bien surfer en proposant à leurs futurs clients d'industrialiser la création de leurs jeux de données pour leur appliquer des services IA à forte valeur ajoutée.

# Third party services using Wikidata

## Scholia:

a search and display engine for academics based on Wikidata  
<https://tools.wmflabs.org/scholia/>



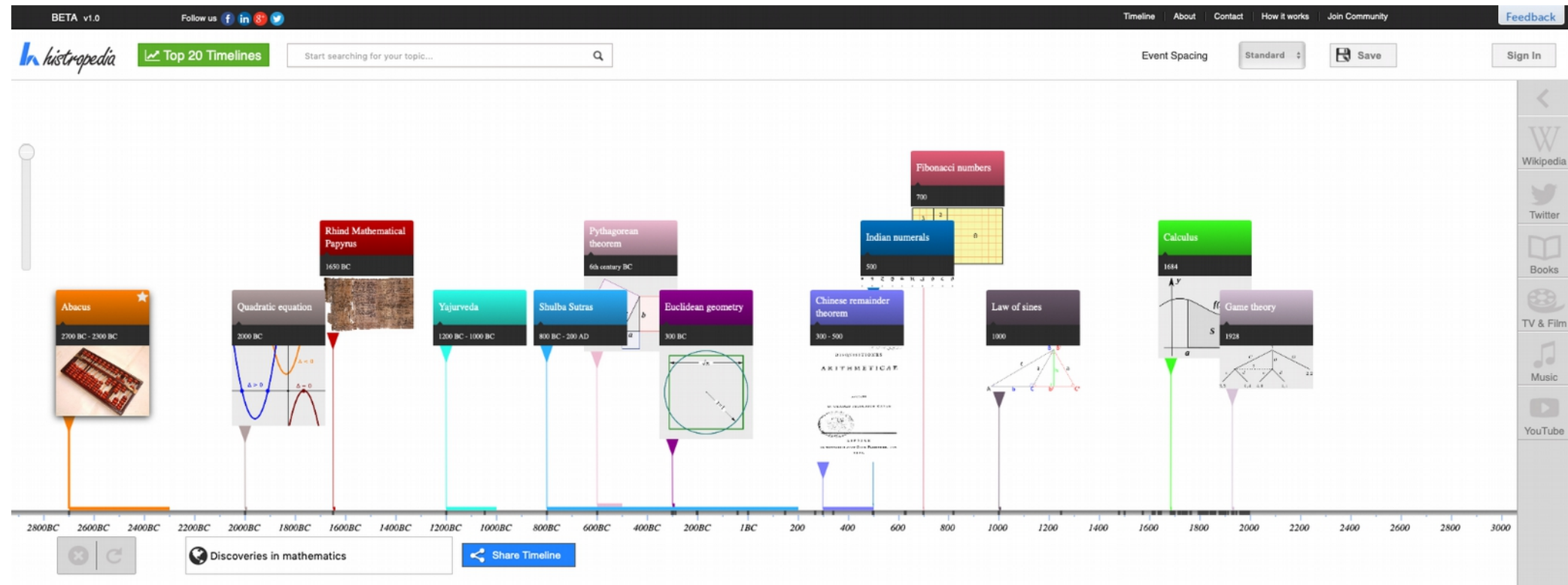
Nobel Prize in Physics  
([Q38104](#))

# Third party services using Wikidata

## Histropedia :

timeline application based on Wikidata

<http://histropedia.com>



# Third party services using Wikidata

## Crotos :

a search and display engine for visual artworks based on Wikidata and using Wikimedia Commons files.

<http://zone47.com/crotos/>

The screenshot shows the Crotos search engine interface. At the top left is the Crotos logo, a winged figure on a horse, with the text 'Crotos Cosmos Callisto Lab'. To the right is a search bar containing the text 'estaque' and a 'Search' button. Further right are settings for 'Language English' and 'Images/page 20'. Below the search bar, it says '12 Results'. A filter dropdown is set to 'Everything'. A page indicator shows 'Page: 1'. The main area displays 12 search results, each with a thumbnail image, a title, and the artist/collector. The results are:

- Oficiais no Mar, Destaque para o Morro do Pico - Thomas Lyde Hornbrook - São Paulo
- L'Estaque, Melting Snow - Paul Cézanne - private collection
- The Pine tree at Estaque - Paul Cézanne - Jean Walter-Paul Guillaume Collection
- The Sea at l'Estaque - Paul Cézanne - musée Picasso
- The Road Bridge at L'Estaque - Paul Cézanne - Finnish National Gallery
- Rocky Crags at L'Estaque - Pierre-Auguste Renoir - Museum of Fine Arts, Boston
- Morning View of L'Estaque Against the Sunlight - Paul Cézanne - Israel Museum
- L'Estaque - Paul Cézanne - Musée d'Orsay
- Houses in Provence: The Riaux Valley near L'Estaque - Paul Cézanne - National Gallery of Art
- The Gulf of Marseilles Seen from L'Estaque - Paul Cézanne - Metropolitan Museum of Art
- The Bay of Marseilles, Seen from L'Estaque - Paul Cézanne - Art Institute of Chicago
- The Terrace, L'Estaque - Albert Marquet - Statens Museum for Kunst

At the bottom, there is a 'Page: 1' indicator and a footer with logos for Wikidata, Wikimedia Commons, and Semantipedia, along with the text '158 383 artworks with image' and 'Last update: 12 / 04 / 2019'.

# Wikibase and its registry

**Wikibase** is the software behind Wikidata.

It consists of

- \* a set of extensions to the MediaWiki software for storing and managing data (**Wikibase Repository**)
- \* and for embedding data on other wikis (**Wikibase Client**)

[http://wikibase-registry.wmflabs.org/wiki/Main\\_Page](http://wikibase-registry.wmflabs.org/wiki/Main_Page)

<https://wikibase-registry-query.wmflabs.org/#%23defaultView%3AT>