Webinaire Open Data

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

Mercredi 20 mai 2020

Organisé par Ynternet.org

nternet.org

Florence Devouard (Q266722)

Username on Wikipedia & Wikidata : Anthere

Formely Chair of Wikimedia Foundation

Cofunder and VP Wikimedia France

Funder Wiki in Africa

Board member Open Food Facts

VP Ynternet.org

Twitter: @anthere
Email : fdevouard@anthere.org

Wikimedia galaxy

SHARE

SHARE

TWEET

COMMENT

EMAIL

TOM SIMONITE BUSINESS 02.18.19 06:00 AM

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...

• @Florence Devouard – cc by sa 4.0

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

Florence Devouard – cc by sa 4.0

WIKIDATA

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.

Florence Devouard – cc by sa 4.0

+

WIKIDATA

The sum of all **human knowledge**

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

Florence Devouard – cc by sa 4.0

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

Florence Devouard – cc by sa 4.0

Ada (Q346047)

given name

P735

Ada Lovelace (Q7259)

2

number of children

P1971

given name

P735

sex or gender

P21

Ada Colau (Q4779594)

country of citizenship

P27

sex or gender

P21

Spain (Q29)

female (Q6581072)

coordinate location

P625

country

P17

position held

P39

Mayor of Barcelona (Q12155639)

41°22'57"N, 2°10'37"E

Florence Devouard – cc by sa 4.0

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 »

Florence Devouard – cc by sa 4.0

Rather unknown... yet... Wikidata is BIG

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

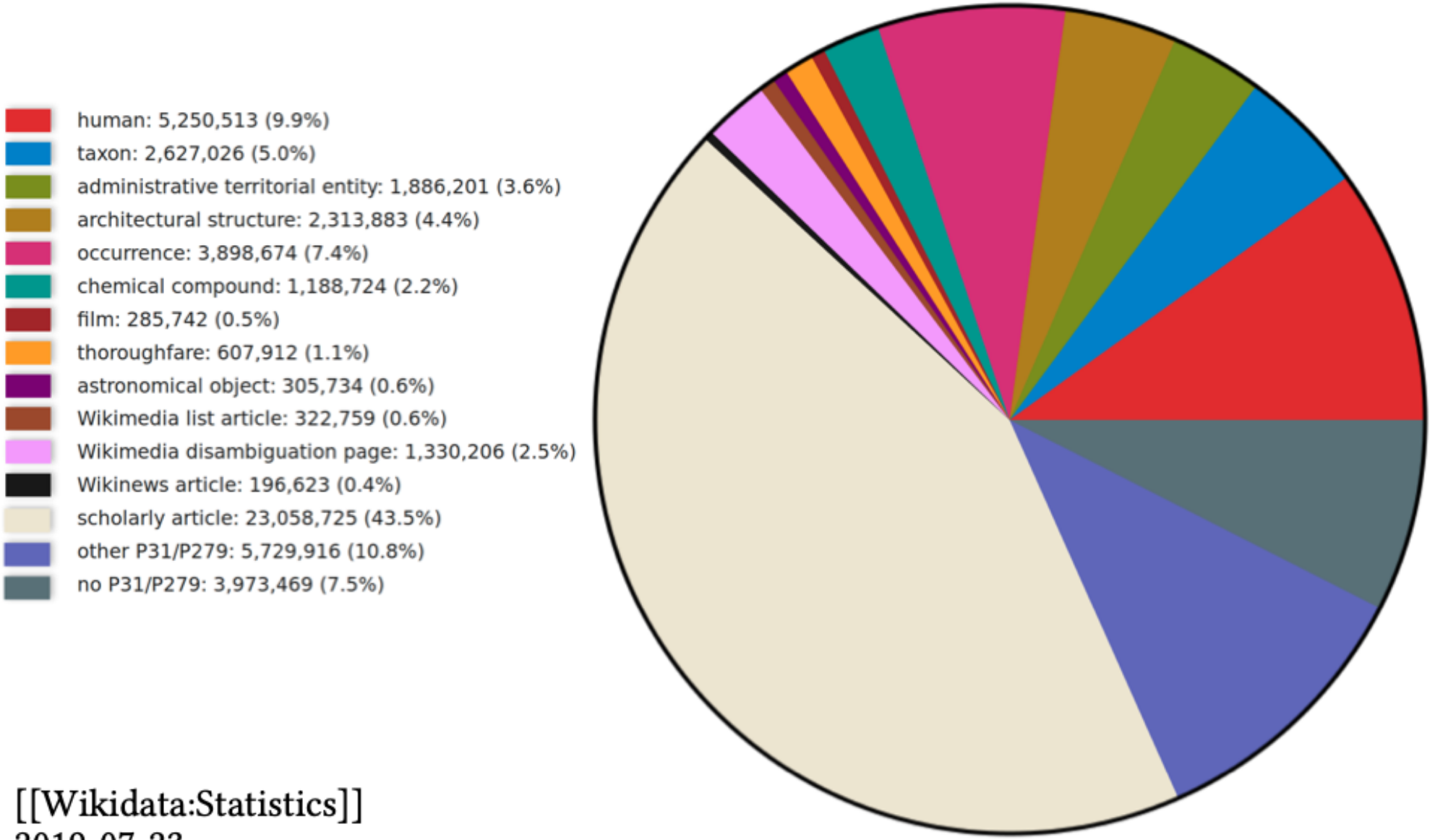
86 millions of unique elements in May 2020

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

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

Source : https://stats.wikimedia.org/v2/#/wikidata.org

What do we have in Wikidata?



[[Wikidata:Statistics]]
2019-07-23

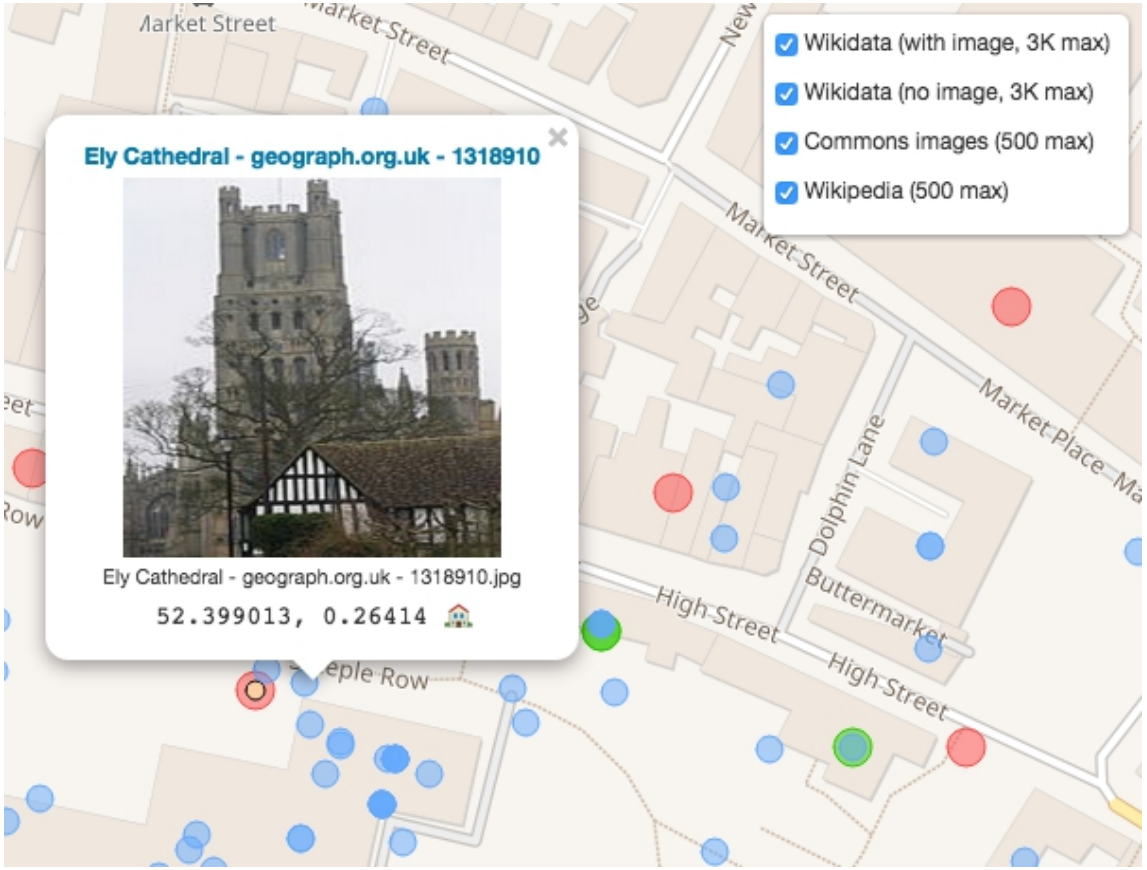
Tools to improve Wikidata

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/#/

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



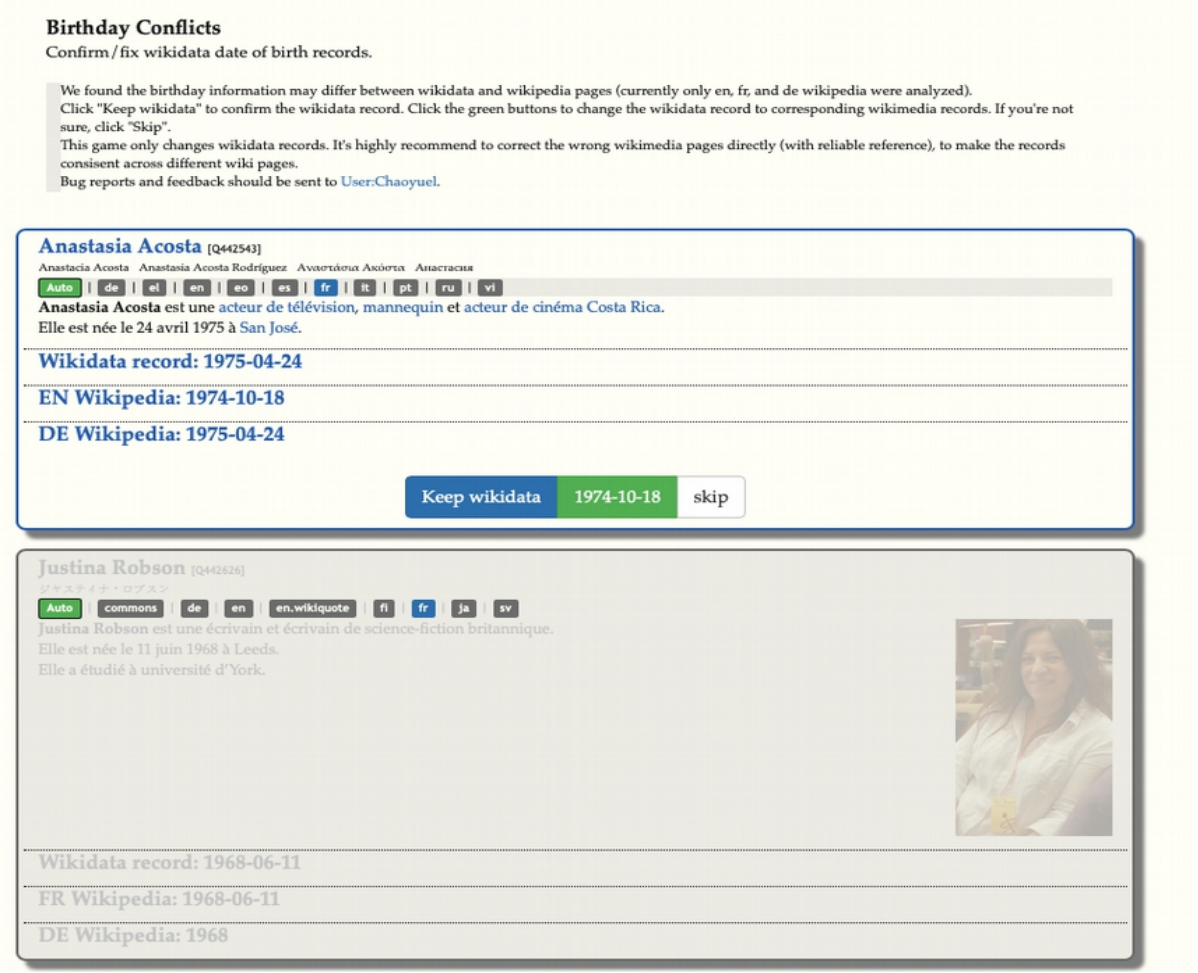
Florence Devouard – cc by sa 4.0

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/



Florence Devouard – cc by sa 4.0

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.

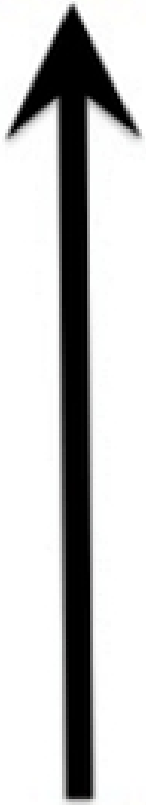
List of examples :
https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/queries/examples

Eg 3.13.1 Overall causes of death ranking

https://query.wikidata.org

Try it !

Most Free



Least Free

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

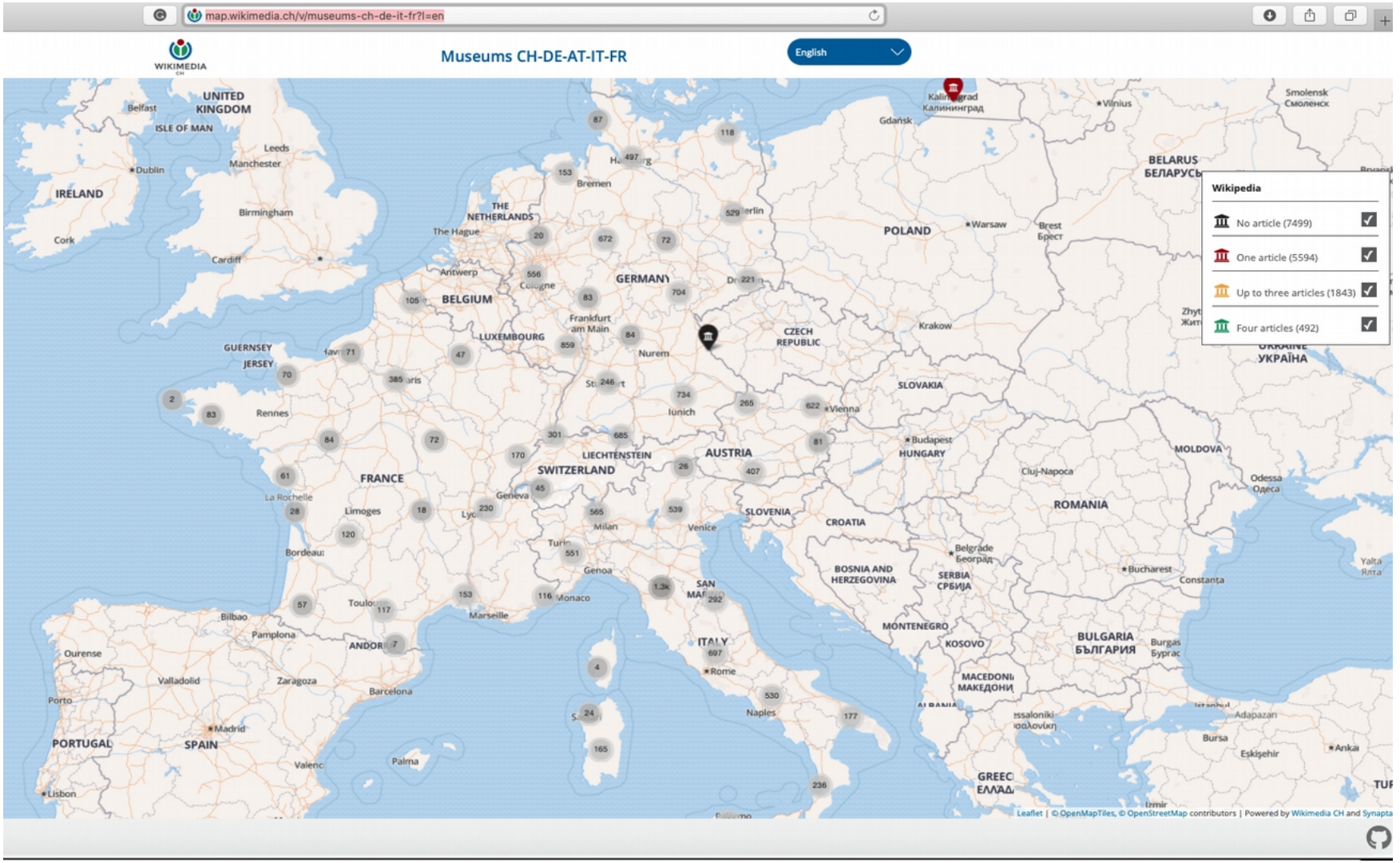
Linked data versus Open data

"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

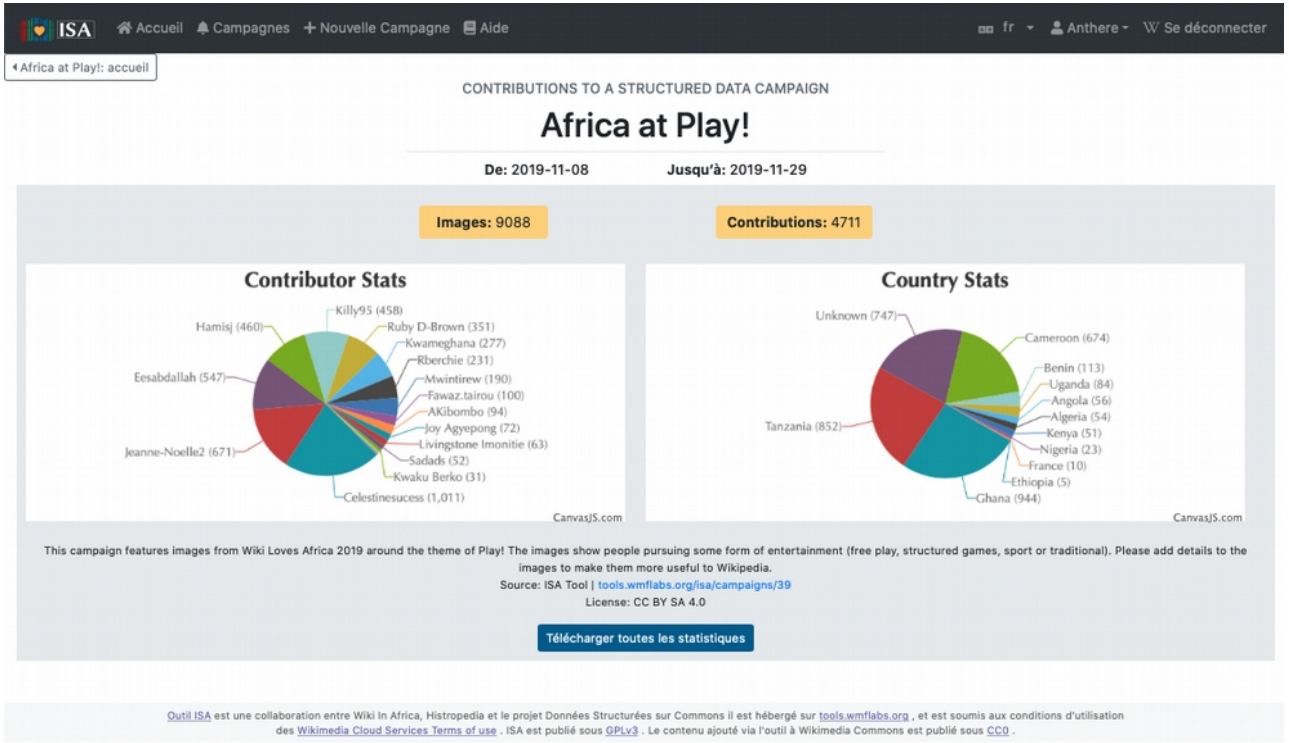


- http://map.wikimedia.ch/v/museums-ch-de-it-fr?l=en

ISA Tool, a tool to add structured data to images

ISA TOOL :
A pilot tool to add easily add structured data to Wikimedia Commons

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



Florence Devouard – cc by sa 4.0

Wikidata, a support to Wikimedia Commons

- Infoboxes on Wikimedia Commons pages

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

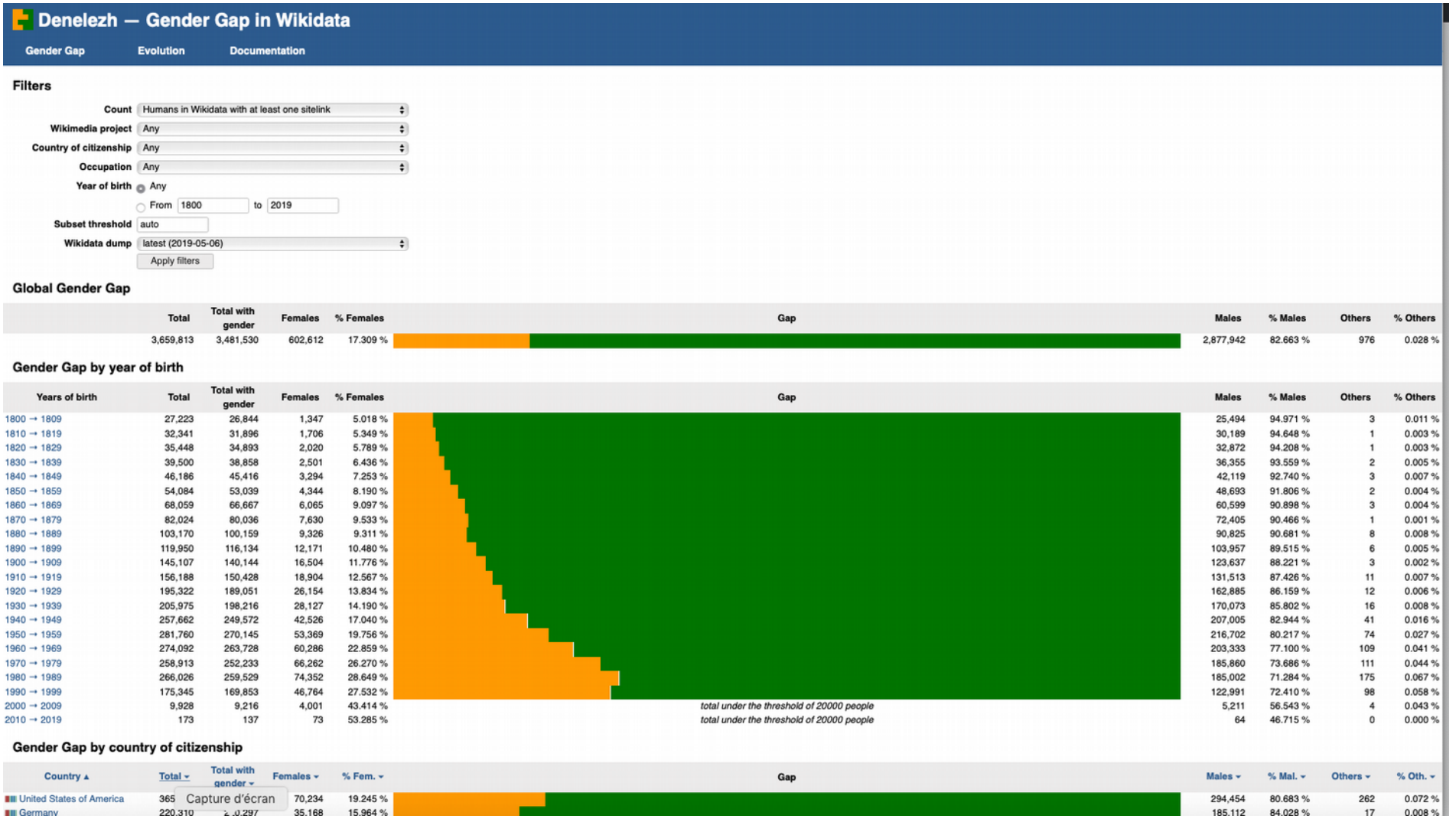
- Structured data to better describe images

  https://commons.wikimedia.org/wiki/File:Münster,_LVM,_Skulptur_-Körper_und_Seele-_--_2016_--_5920-6.jpg

Florence Devouard – cc by sa 4.0

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/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/Algé

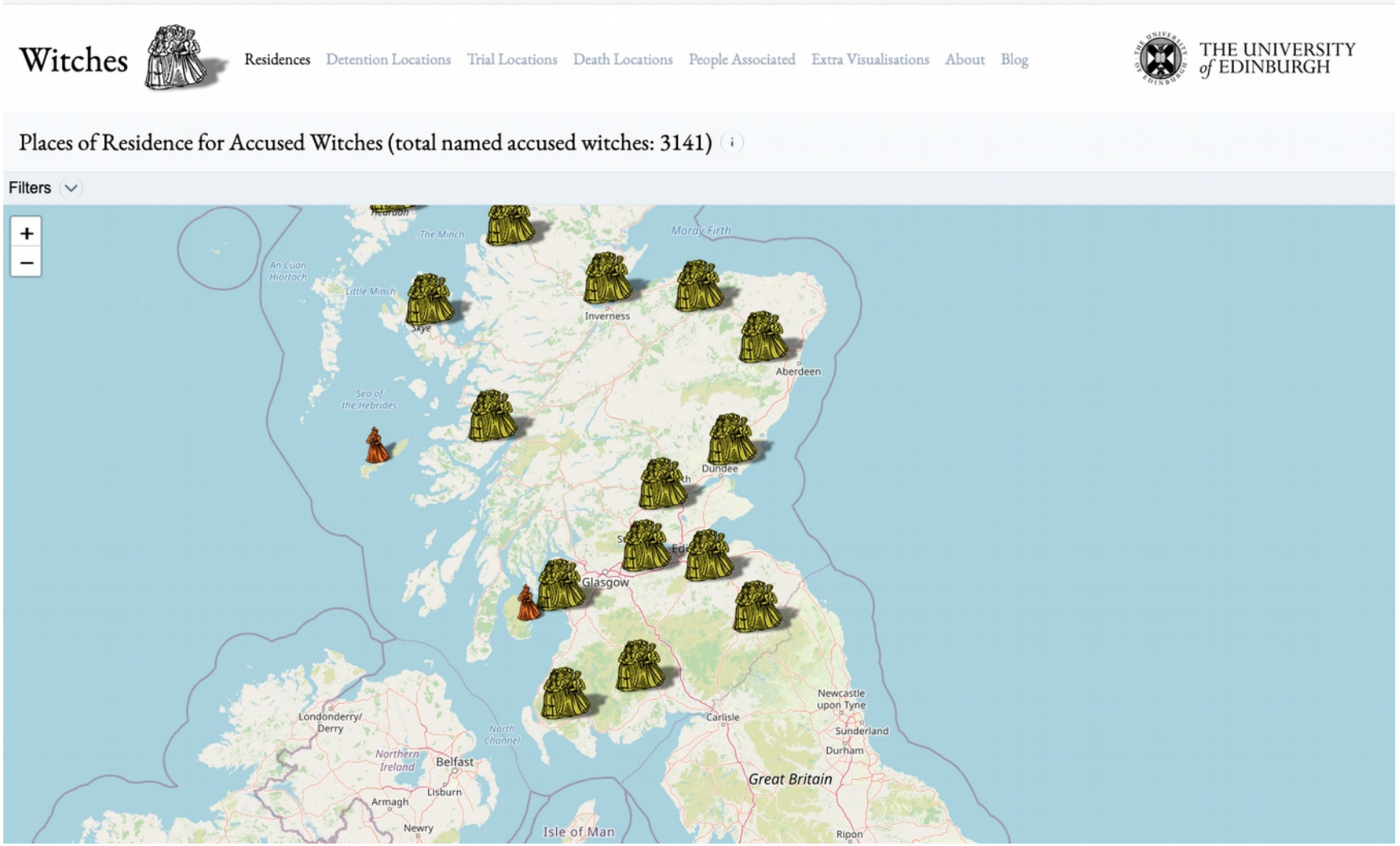
Exemple : liste des occupations relatives au cinéma, avec libellé multi-genres

https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Wiki_Loves_Women/Occupations/fr

Florence Devouard – cc by sa 4.0

Wikidata for the third parties

Witches map
by University of Edinburgh



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

Florence Devouard – cc by sa 4.0

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

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 entrainés, 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'entrainement. 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 inexploitables 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.

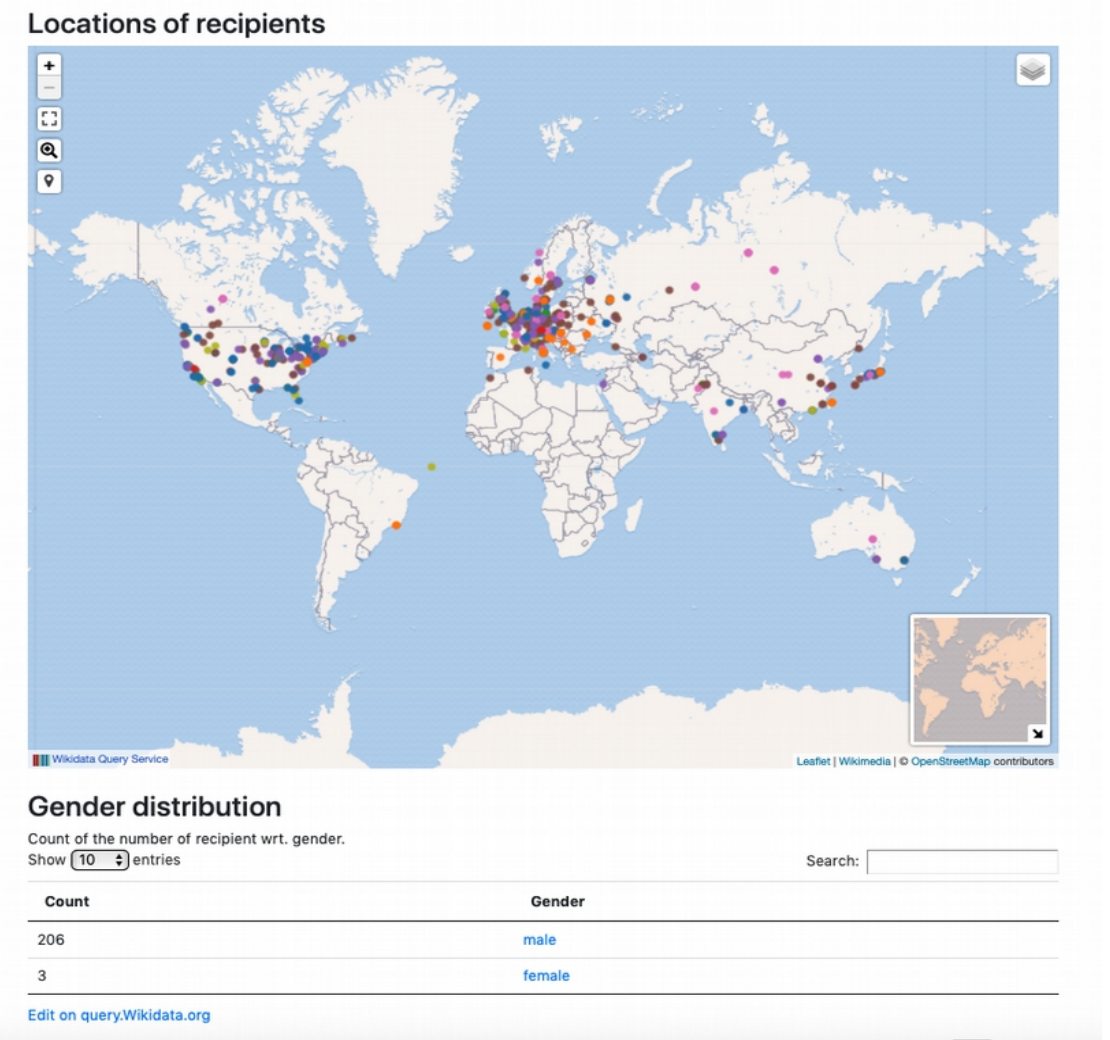
Point critique dans le développement de l'IA et du machine learning : la **disponibilité de jeux de données d'entrainement**.

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

Florence Devouard – cc by sa 4.0

Third party services using Wikidata

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

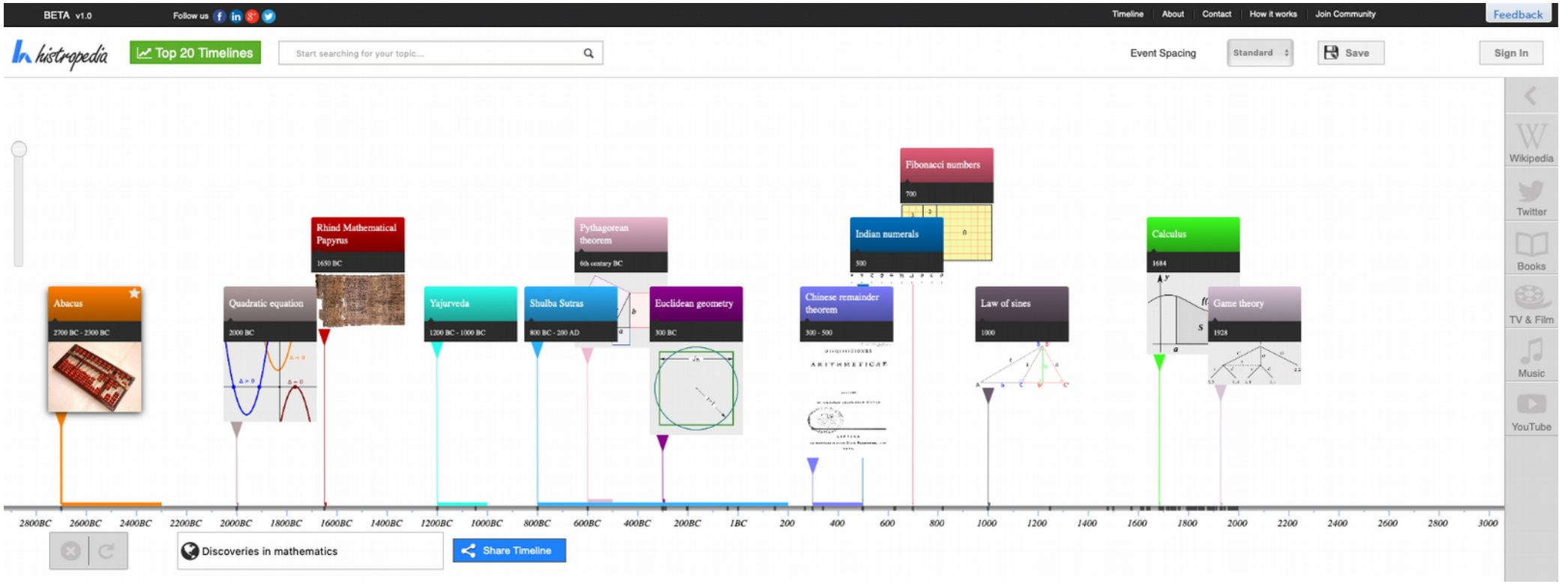


Nobel Prize in Physics (Q38104)

Florence Devouard – cc by sa 4.0

Third party services using Wikidata

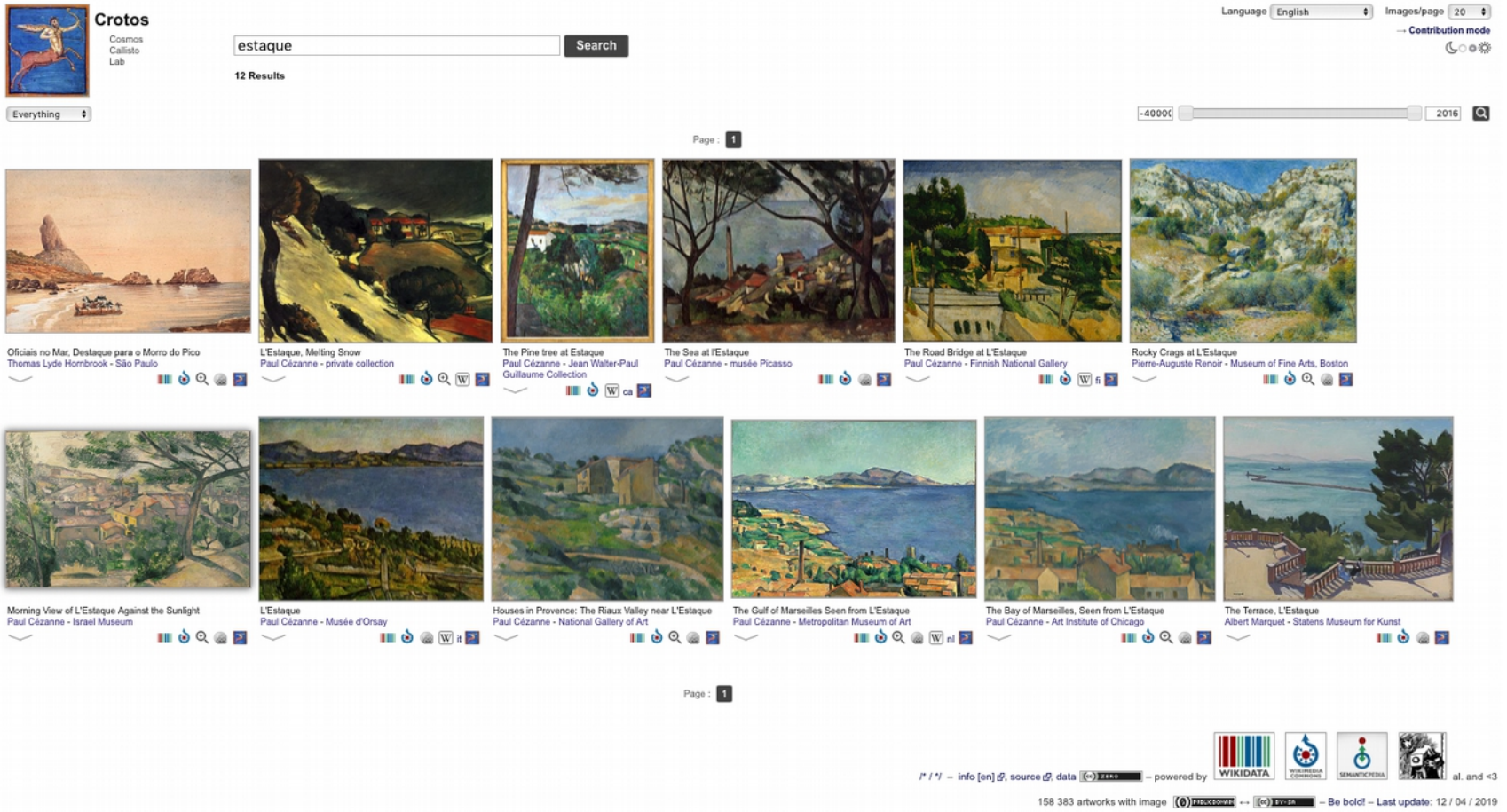
Histropedia :
timeline application based on Wikidata
http://histropedia.com



Florence Devouard – cc by sa 4.0

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/



Florence Devouard – cc by sa 4.0

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

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