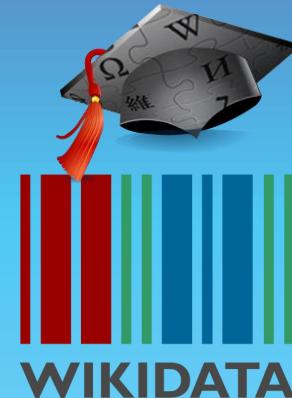


An introduction to Wikidata



SCHOOL OF
ADVANCED STUDY
UNIVERSITY
OF LONDON



UNIVERSITÄT
LEIPZIG

SunoikisisDC, December 2022

Shani Evenstein Sigalov

Shani.even@gmail.com



The Team for this session

Monica Berti

University of Leipzig

Gabriel Bodard
University of London

Shani Evenstein Sigalov

University of London; Tel Aviv University

Educator, lecturer, researcher & Free Knowledge advocate

Interested in the intersection of Education, Technology, Innovation & Openness.

What's on the menu today?

Part 1: Intro to Wikidata

Part 2: Data Extraction & Visualization

Part 3: Exploring Uses

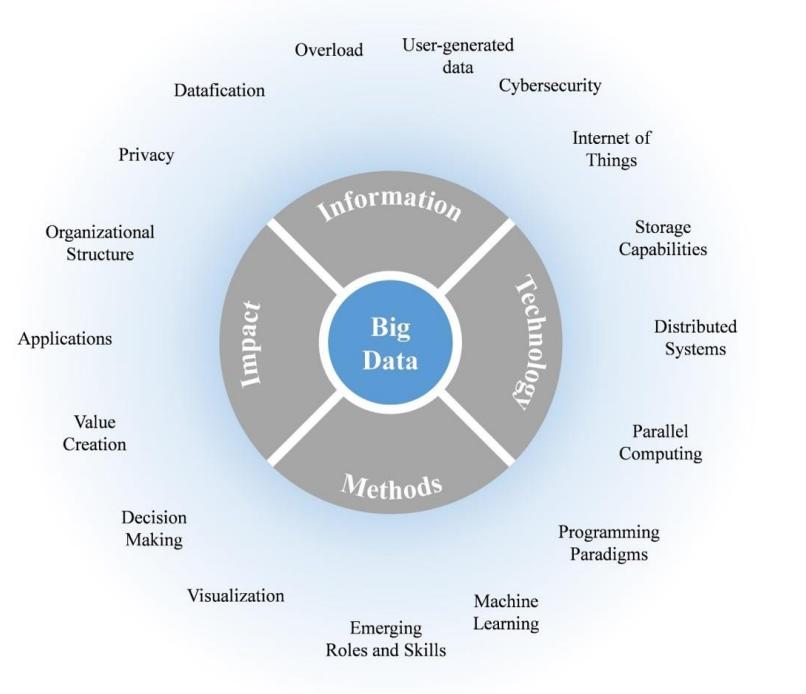
Part 4: Summing up

Part 5: Q & A

Part 1: A (short) Intro to Wikidata

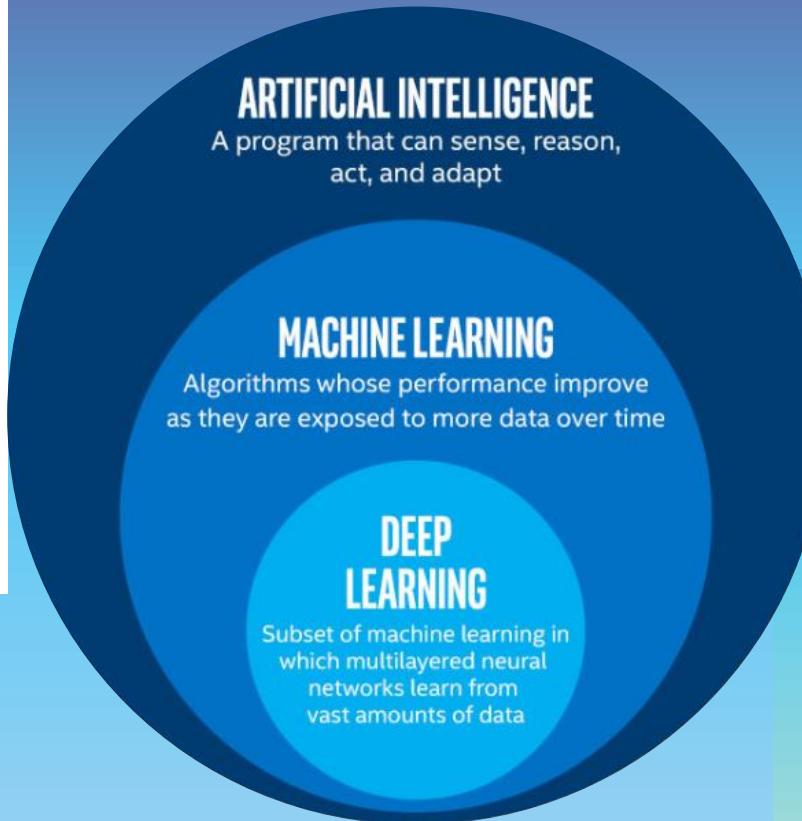


Big Data

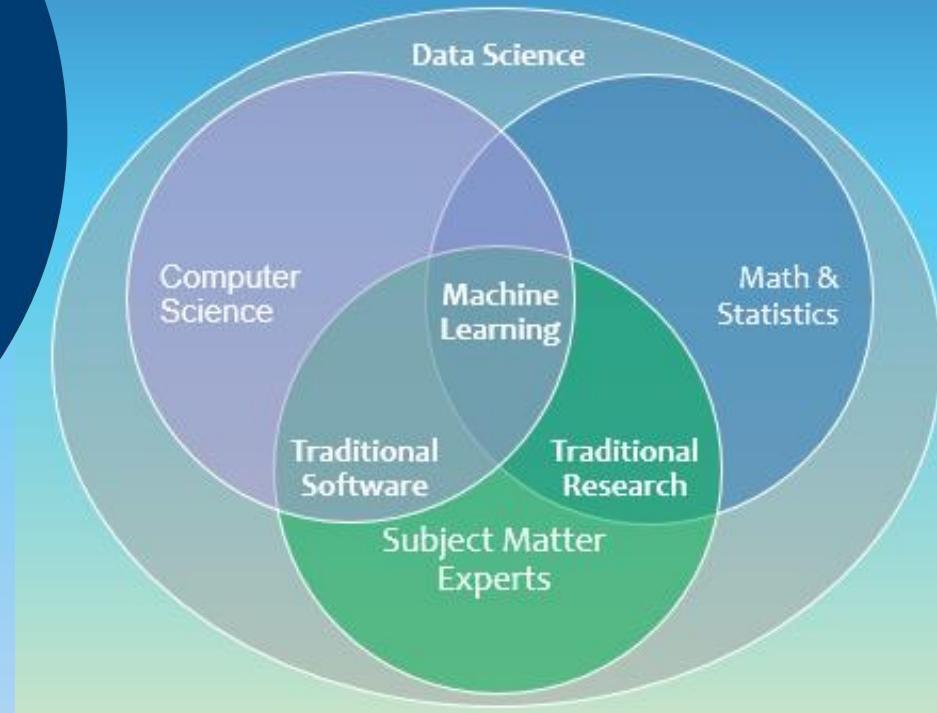


Leibniz, CC-BY-SA 4.0

AI



Data Science



No Big Data nor Data Science without Data

But there is no Big Data nor
Data Science without Data.

That is why Wikidata
is an **essential** part of this
ecosystem.

But first thing's first –

What's Wikidata, anyway?

Wikidata is part of Wikimedia ecosystem,
launched in 2012.

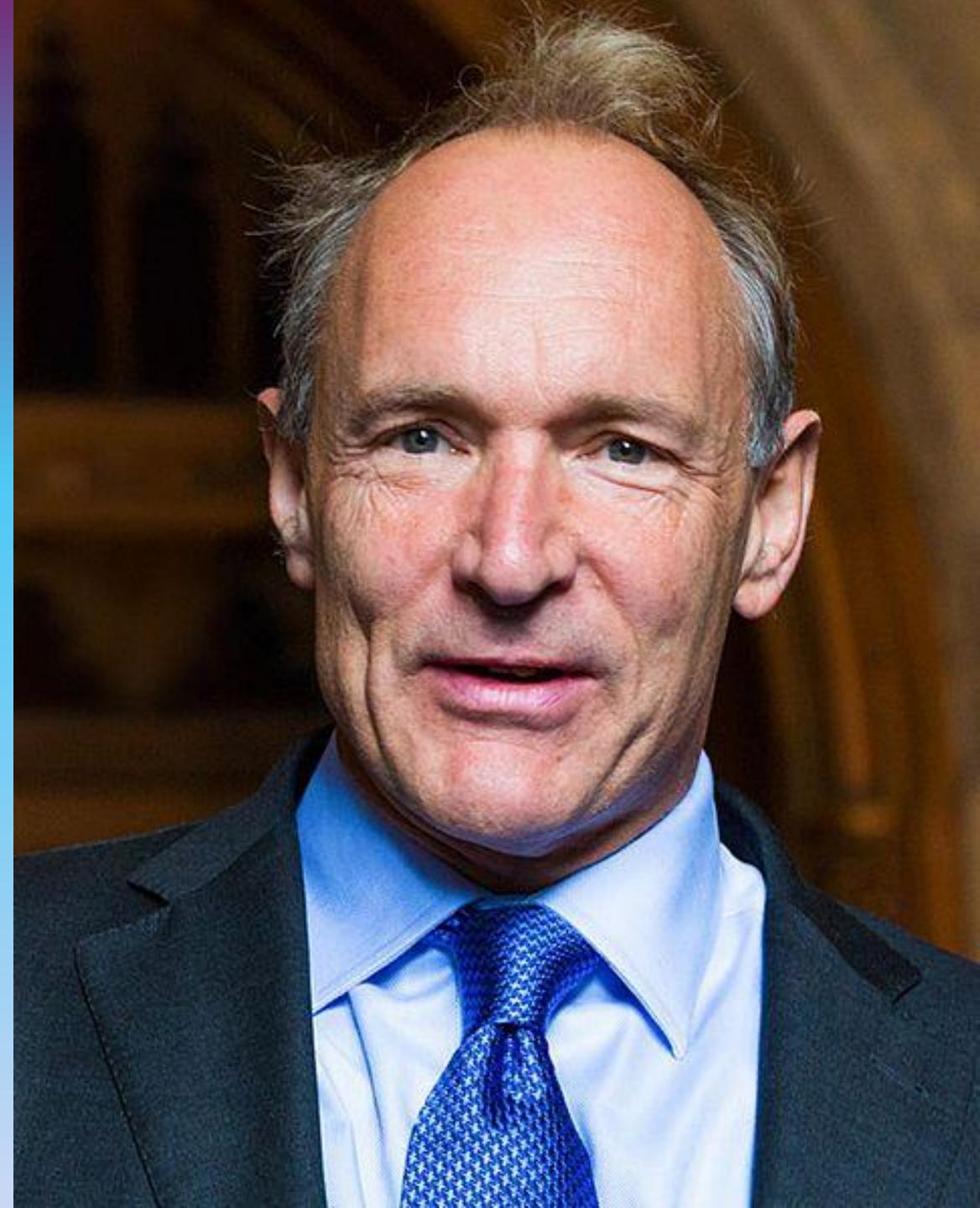


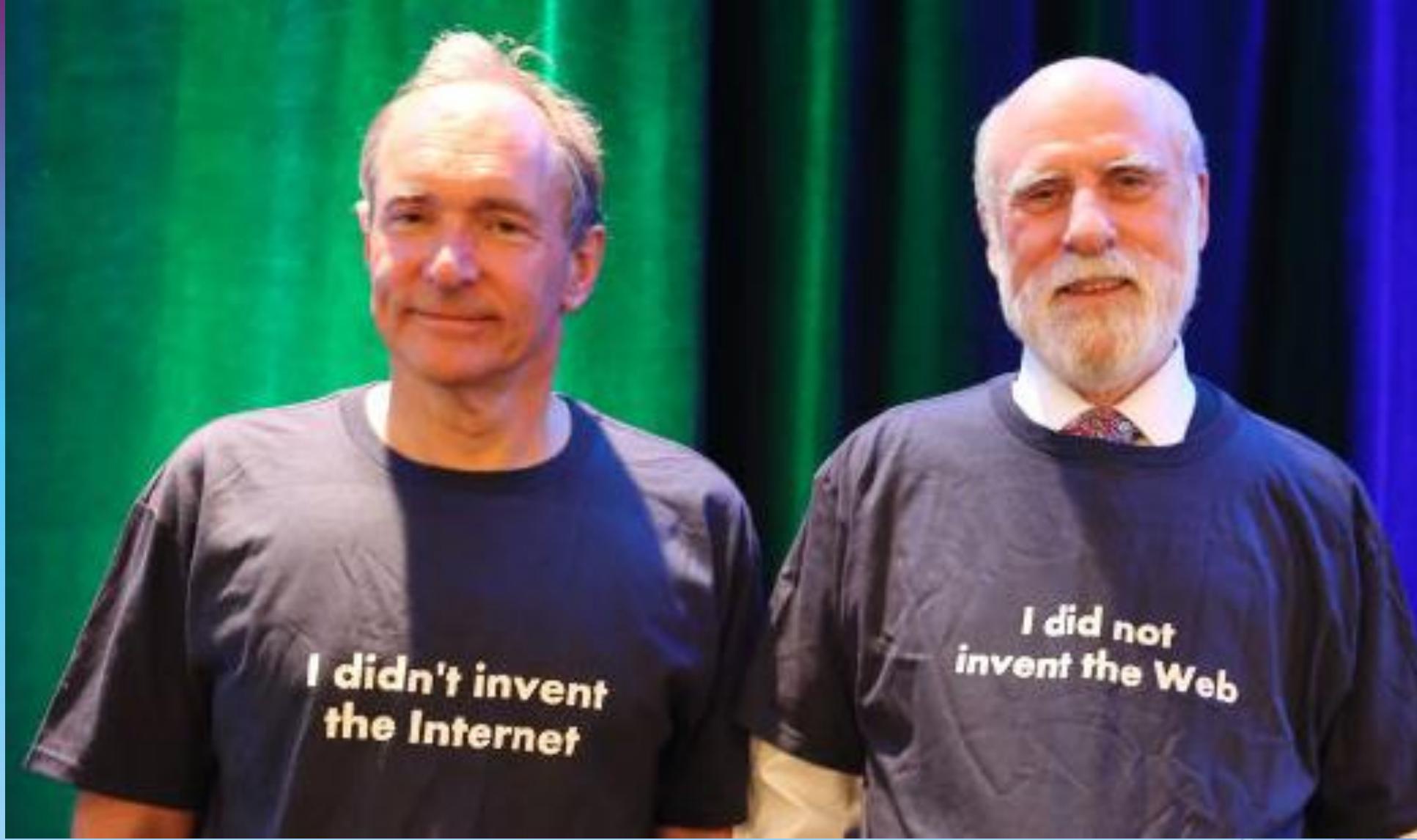
A database of structured, linked data
Multilingual and read by humans & machines
Open to all, under a free license

But first thing's first –

What's Wikidata, anyway?

A man with a
vision:
The Semantic
Web dream
(realized?)





Tim Berners-Lee (left) and Vinton Cerf (right) taken by the W3C during the 25th anniversary of the Web

[Watch: Vint Cerf – Will Computers Ever Think Like Human Beings?](#)

Answering a Need

- Dated data; repetitive work
- Inflexible ways of lateral queries of knowledge
(Categories and their limitations)
- No central “hub” for metadata from all GLAMs
- A manifestation of the “Semantic Web” dream

Wikidata under CC-0:

The largest OER “big data” repository
humanity created

Over 100 million items
and counting..

4 main Interactions

* **Data curation (knowledge creation):**

A central repository to curate / salvage and enrich datasets

A way to teach literacies, capacities & skills for life-long learning

* **Data extraction:**

Answering questions difficult or impossible to answer before

Visually exploring data in an engaging way

Exposing relationships and connections not apparent / visible / known before

* **Data creation:**

Autogenerating new digital objects based on scattered, external data

Autogenerating content on other Wiki projects

* **Research and Teaching**

Let's explore!

www.wikidata.org

Q42

Examples of interesting items

1. Felix Jacoby (scholar)

<https://reasonator.toolforge.org/?q=Q66145&lang=en>

2. Herodotus (ancient Greek historian)

<https://reasonator.toolforge.org/?q=Q26825&lang=en>

3. Ernst Meyer, German historian of classical antiquity

<https://www.wikidata.org/wiki/Q1359495> (descriptions uneven; no Italian translation)

4. Alexandros of Aphrodisias, philosopher

<https://www.wikidata.org/wiki/Q317146> (descriptions uneven)

Part 2: Data Extraction & Visualization



The beauty of openness:

Internal & External parties use Wikidata

Including: Siri, Alexa, and other AI-based agents.

Examples for Data Extraction & visualization

Including external tools

Women Physicists

<http://tinyurl.com/yxswh5t2>

Histropedia

BETA v1.0 Follow us:

Timeline About Contact Help Join Community

Event density: High | Low | Sort: Sign In

Start reviewing the other items...

A horizontal timeline showing various historical events and artworks from 1400 to 1525. The timeline features several paintings, including the Mona Lisa and the Lady with an Ermine, displayed as cards with their titles and dates.

Timeline navigation: My timeline | Create Timeline

Lady with an Ermine

Lady with an Ermine



Lady with an Ermine

Add to Timeline

W Wikipedia
Twitter
Facebook
TV & Film
Music
YouTube

Multiple timelines – Prado Museum

MUSEO DEL PRADO | **TIMELINE**

Look for an artist, a work, an event ...

Music

Painting

PRADO MUSEUM

This image shows a digital timeline from 1823 to 1828, featuring a grid of historical figures and their corresponding artworks or events. The timeline is divided into three main sections: Music, Painting, and PRADO MUSEUM.

Music: This section includes figures like Francisco Asenjo Barbieri (1823-1894), Carl Reinecke (1824-1884), Bedrich Smetana (1824-1884), Anton Bruckner (1824-1896), Francis James Child (1825), Eduard Hanslick (1825-1899), Johann Strauss II (1825-1899), Ludwig Minkus (1826-1864), Stephen Foster (1826-1864), Jovan Dordevic (1826-1900), Prince Gustaf, Duke of Uppland (1827-1871), and Josef Strauss (1827-1870). It also features the premiere of Beethoven's Symphony No. 9 in 1824.

Painting: This section includes figures like Raff (1822-1890), Édouard Lalo (1822-1892), César Franck (1822-1890), Prilidiano Olga Nikolaevna of Russia (1823), Alfred Stevens (painter) (1823-1906), Alexandre Cabanel (1823-1889), Gustave Boulanger (1823-1888), Jean-Léon Gérôme (1824-1904), Pierre Puvis de Chavannes (1824-1898), Hans Gude (1825), George Inness (1825-1894), William-Adolphe Bouguereau (1825-1905), Gustave Moreau (1826-1898), Ramon Martí Alsina (1826-1886), Karl von Piloty (1826-1886), Jean-Baptiste Carpeaux (1827-1875), and Arnold Böcklin (1827-1901).

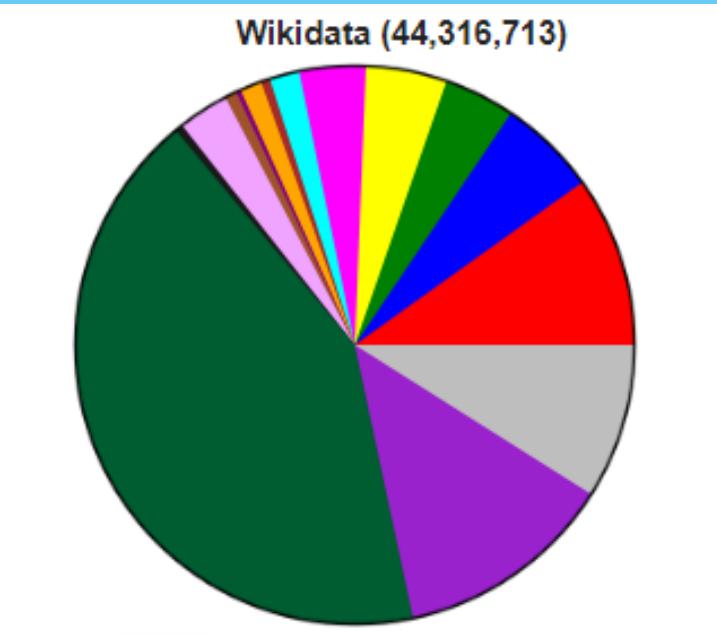
PRADO MUSEUM: This section displays five specific artworks: "Maria Josepha of... López Portaña,..." (1823), "The flying dog Goya y Lucientes..." (1824), "Real Establecimiento Litográfico..." (1825), "Haes, Carlos de..." (1826), and "Sketch for Allegory of... López Portaña,..." (1827).

The timeline is marked with years 1823, 1824, 1825, 1826, 1827, and 1828 at the bottom.

<https://tinyurl.com/yb49xguq>

WikiCite

- Ingest source metadata into Wikidata
- One pilot - all citations in PubMed,
~ 20M academic articles
- 40% of WD items -
citations in WikiCite



human	4,378,742	(9.9%)
taxon	2,509,963	(5.7%)
administrative territorial entity	1,812,457	(4.1%)
architectural structure	2,119,774	(4.8%)
occurrence	1,651,237	(3.7%)
chemical compound	759,000	(1.7%)
film	268,609	(0.6%)
thoroughfare	522,260	(1.2%)
astronomical body	155,105	(0.3%)
Wikimedia list article	299,810	(0.7%)
Wikimedia disambiguation page	1,280,367	(2.9%)
Wikinews article	211,930	(0.5%)
scholarly article	18,771,018	(42.4%)
other P31/P279	5,602,972	(12.6%)
no P31/P279	3,973,469	(9.0%)

Scholia

Generates academic profiles
(like a Google Scholar or Elsevier SciVal profile)
based on WikiCite and other data.

Example:

<https://tools.wmflabs.org/scholia/author/Q7183255>

author

 topic

Tim Berners-Lee (Q80)

Sir Timothy John Berners-Lee (born 8 June 1955), also known as TimBL, is an English engineer and computer scientist best known as the inventor of the Worldwide Web. He is a Professorial Fellow of Computer Science at the University of Oxford and a professor at the Massachusetts Institute of Technology (MIT). Berners-Lee proposed an information management system on 12 March 1989, then implemented the first successful communication between a Hypertext Transfer Protocol (HTTP) client and server via the internet in mid-November. Berners-Lee is the director of the World Wide Web Consortium (W3C) which oversees the continued development of the Web. ... (from the English Wikipedia)

Related: Rob Lewis · James E. Gunn · John Edward Hopcroft · Maurice Wilkes · Bernard Lovell · John Bardeen · David Mumford · Glenn Seaborg · Eugene Merl Shoemaker · Robert Giffen

<https://orcid.org/0000-0003-1279-3709> | [@timberners_lee](#)

List of publications [RSS](#)

Show 10 entries

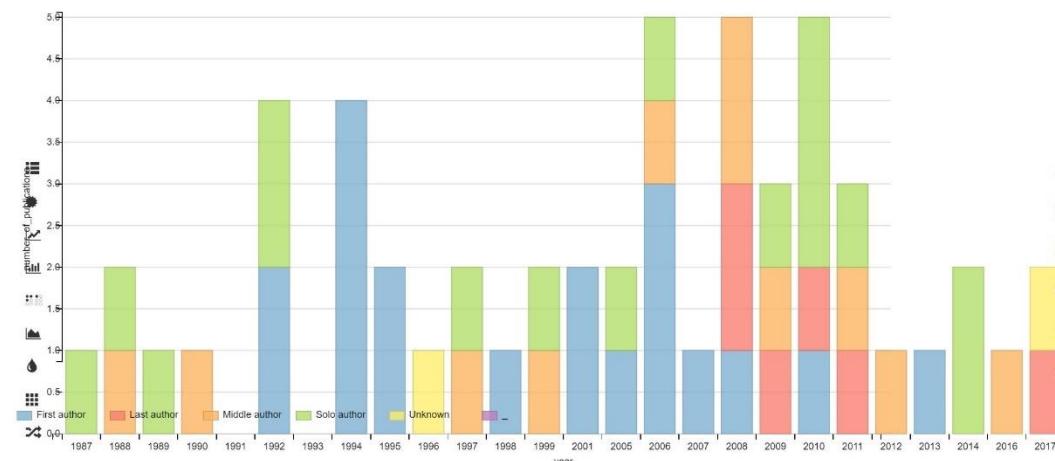
Search:

Date	Work	Type	Pages	Venue	Authors
2017-01-01	Linked Data Notifications: A Resource-Centric Communication Protocol	scholarly article		The Semantic Web: 14th International Conference, ESWC 2017, Portorož, Slovenia, May 28 – June 1, 2017, Proceedings, Part I	Christoph Lange, Amy Guy, Sarven Capadisi, Sören Auer, Tim Berners-Lee
2017-01-01	Decentralised Authoring, Annotations and Notifications for a Read-Write Web with dokielic	scholarly article		Web Engineering: 17th International Conference, ICWE 2017, Rome, Italy, June	Christoph Lange, Amy Guy, Ruben Verborgh, Sarven Capadisi, Sören Auer

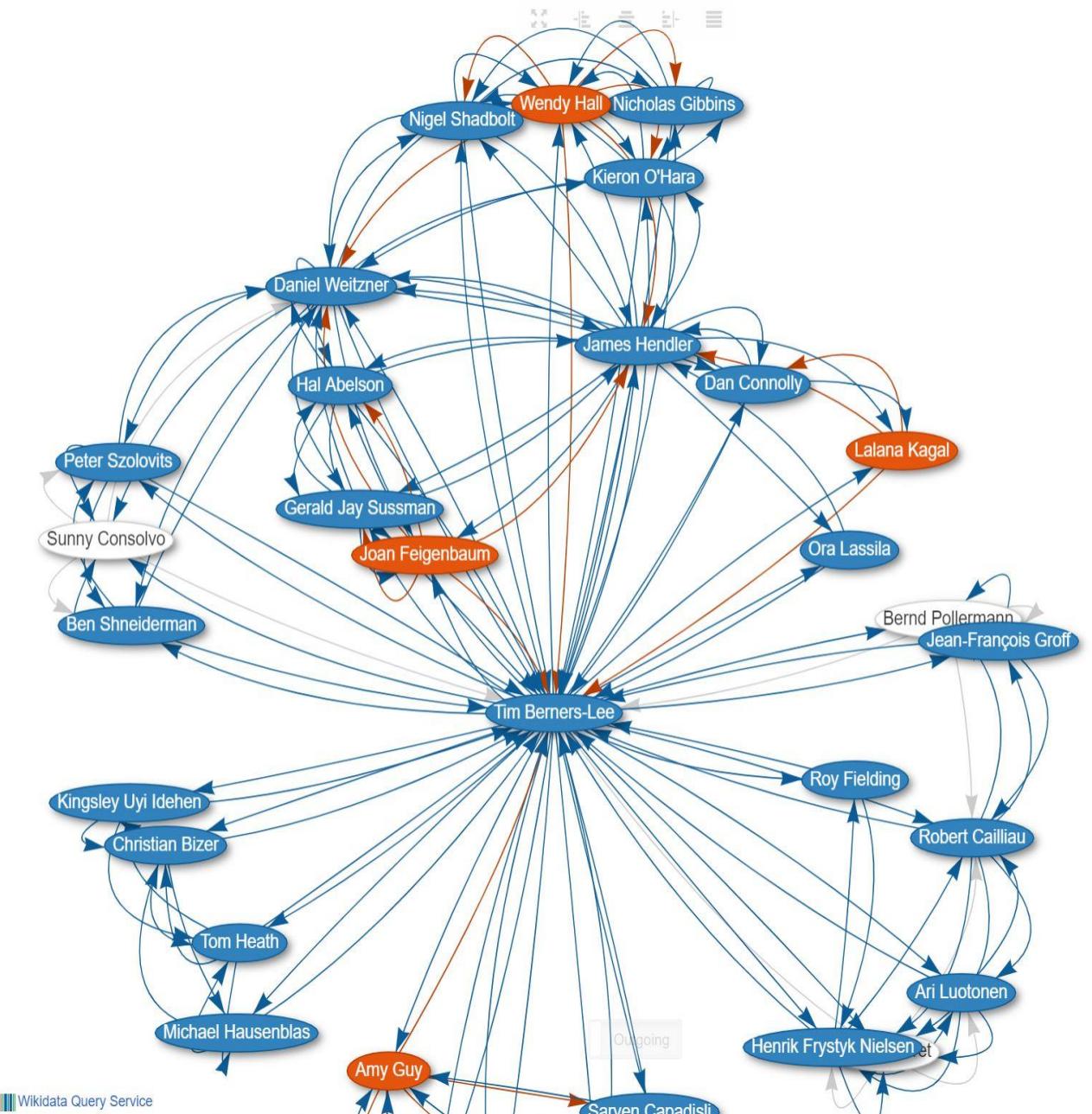
Showing 1 to 10 of 56 entries

Previous 1 2 3 4 5 6 Next

Number of publications per year



Co-author graph



Ask Platypus

Ask me anything Alpha version



who is luke skywalker's dad?



Result

Show internal results

Anakin Skywalker [G](#) [Q](#)

character of Star Wars

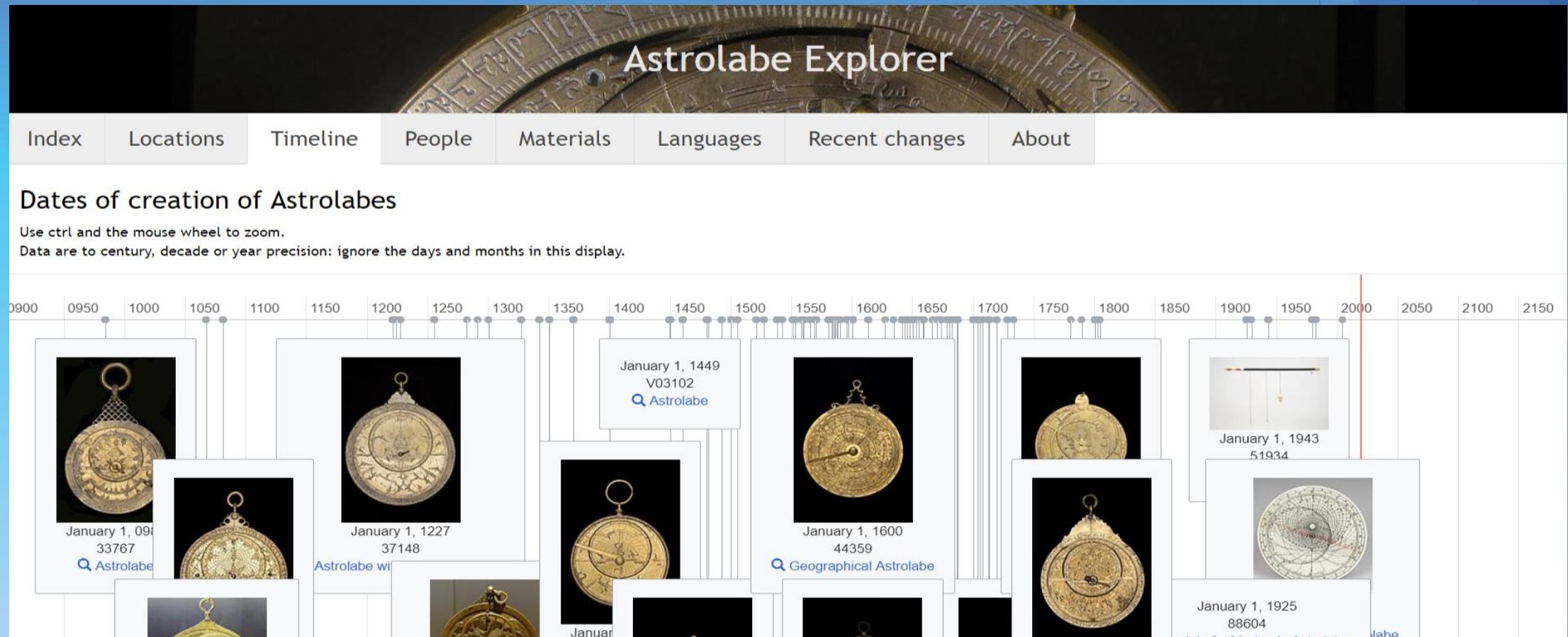
[Ask Platypus](#)

Part 3: Exploring Uses

Some inspiration from around the world
mainly in Education, Research, GLAM



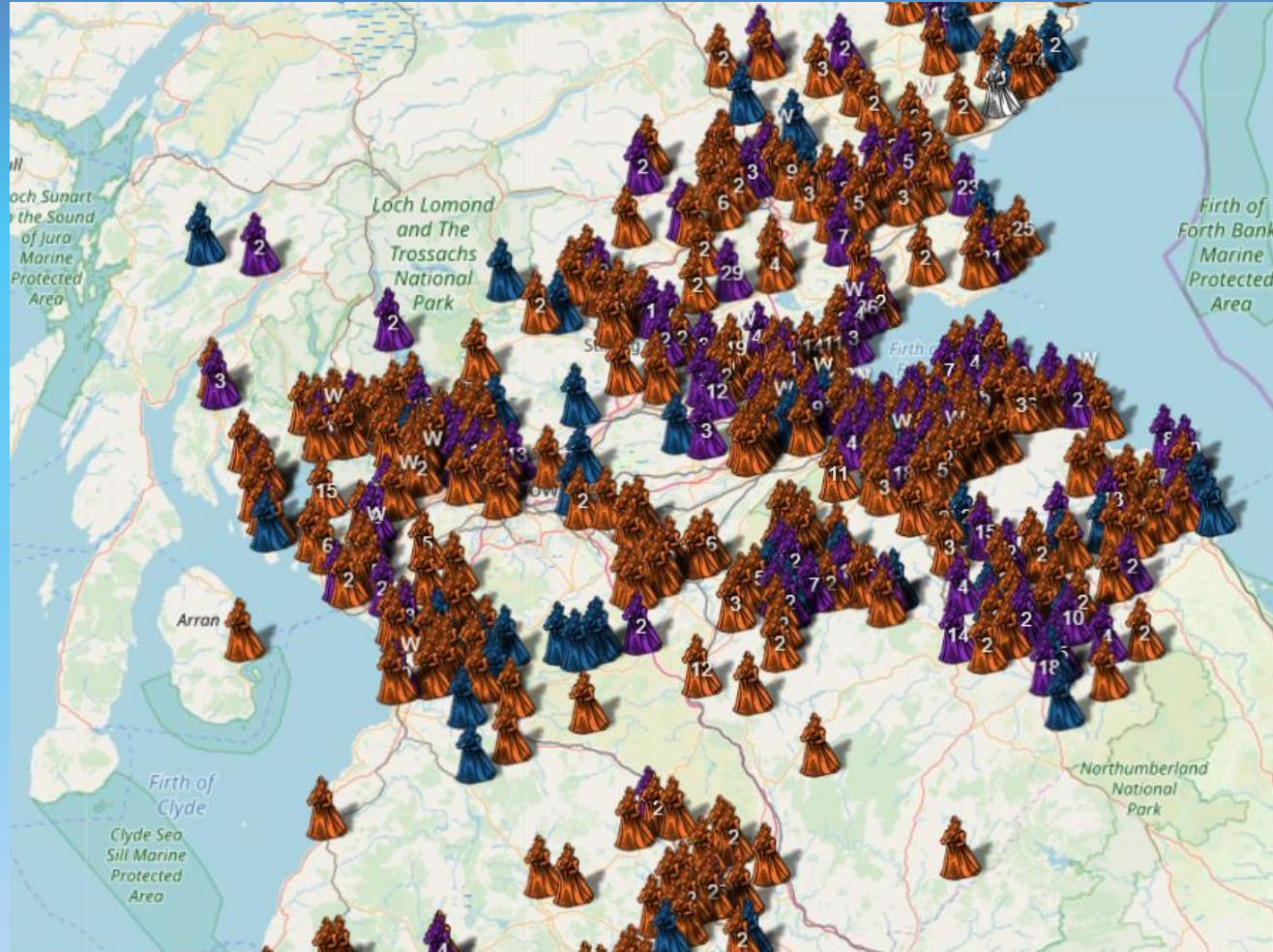
The Bodleian Libraries, University of Oxford, UK: The Astrolabe Explorer



Use: Telling the story of your collection

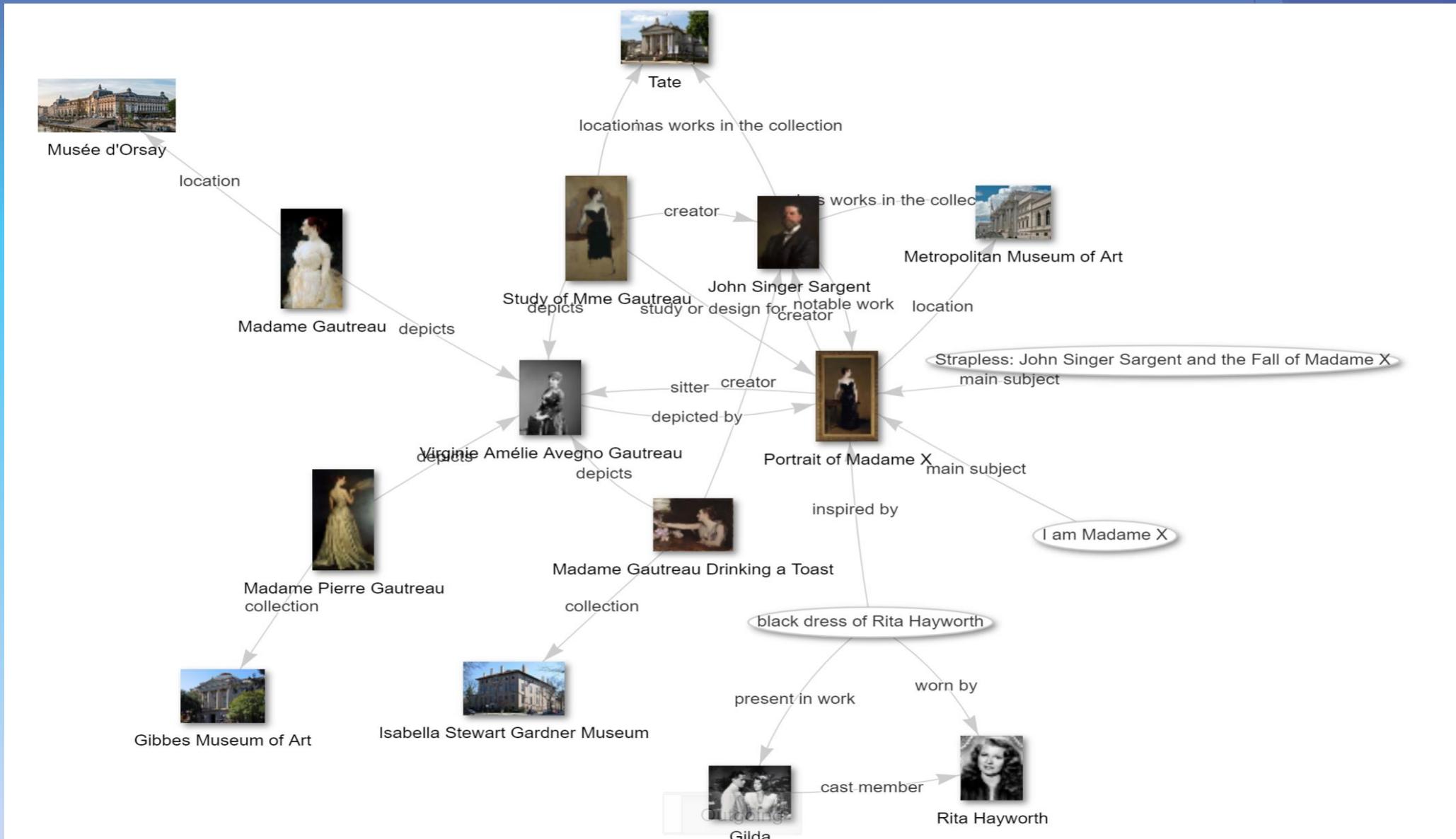
University of Edinburgh 16th- and 17th-Century Scottish Witch Hunts

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



[Link to article about the project](#)

The Met - Unraveling links, connections & relationships



The Met Dashboard

Wikidata:GLAM/Metropolitan Museum of Art/All objects

< Wikidata:GLAM | Metropolitan Museum of Art

Dashboards for The Metropolitan Museum of Art Open Access project



This table describes the completeness of properties in Wikidata for all Met objects (10,000+). A good reference for item structure for visual works: [Wikidata:WikiProject_Visual_arts/Item_structure](#).

[all entities](#) | [count](#) | [all groupings](#) | [without grouping property](#) | [below threshold \(1\)](#)

Manually update

Top groupings (Minimum 1 items)		Top Properties (used at least 0 times per grouping)								
Name	Count	image (P18)	inception (P571)	height (P2048)	width (P2049)	creator (P170)	The Met object ID (P3634)	inventory number (P217)	described at URL (P973)	
painting (Q3305213)	11024	67.77% (7471)	98.71% (10882)	71.77% (7912)	71.62% (7895)	95.5% (10528)	100.0% (11024)	100.0% (11024)	2.5% (276)	
sculpture (Q860861)	419	67.78% (284)	82.1% (344)	11.46% (48)	8.83% (37)	50.84% (213)	100.0% (419)	100.0% (419)	1.91% (8)	
drawing (Q93184)	356	65.17% (232)	82.87% (295)	42.42% (151)	42.13% (150)	79.78% (284)	100.0% (356)	100.0% (356)	0.28% (1)	
woodcut print (Q18219090)	172	99.42% (171)	84.3% (145)	20.35% (35)	20.35% (35)	100.0% (172)	100.0% (172)	98.84% (170)	17.44% (30)	
watercolor painting (Q18761202)	138	93.48% (129)	90.58% (125)	84.06% (116)	84.06% (116)	89.13% (123)	100.0% (138)	100.0% (138)	0% (0)	
literary work (Q7725634)	119	86.55% (103)	99.16% (118)	13.45% (16)	3.36% (4)	14.29% (17)	100.0% (119)	100.0% (119)	0.84% (1)	
figure (Q29527347)	104	8.65% (9)	31.73% (33)	0% (0)	0% (0)	0% (0)	100.0% (104)	100.0% (104)	0% (0)	
photograph (Q125191)	89	92.13% (82)	93.26% (83)	13.48% (12)	13.48% (12)	55.06% (49)	100.0% (89)	98.88% (88)	3.37% (3)	
vase (Q191851)	88	57.95% (51)	64.77% (57)	7.95% (7)	3.41% (3)	22.73% (20)	100.0% (88)	100.0% (88)	1.14% (1)	
print (Q11060274)	63	55.56% (35)	84.13% (53)	9.52% (6)	9.52% (6)	68.25% (43)	100.0% (63)	100.0% (63)	3.17% (2)	
furniture (Q14745)	53	79.25% (42)	96.23% (51)	5.66% (3)	5.66% (3)	43.4% (23)	100.0% (53)	100.0% (53)	0% (0)	
statuette (Q16738862)	49	93.88% (46)	95.92% (47)	4.08% (2)	2.04% (1)	6.12% (3)	100.0% (49)	100.0% (49)	0% (0)	
textile artwork (Q22075301)	44	9.09% (4)	90.91% (40)	0% (0)	0% (0)	81.82% (36)	100.0% (44)	100.0% (44)	0% (0)	
evening wear (Q1908472)	44	0% (0)	47.73% (21)	0% (0)	0% (0)	22.73% (10)	100.0% (44)	100.0% (44)	0% (0)	
tapestry (Q184296)	40	92.5% (37)	95.0% (38)	55.0% (22)	55.0% (22)	70.0% (28)	100.0% (40)	100.0% (40)	0% (0)	
codex (Q213924)	38	100.0% (38)	100.0% (38)	23.68% (9)	15.79% (6)	13.16% (5)	100.0% (38)	100.0% (38)	0% (0)	
evening dress (Q2144456)	36	11.11% (4)	55.56% (20)	0% (0)	0% (0)	30.56% (11)	100.0% (36)	100.0% (36)	0% (0)	
bowl (Q153988)	35	57.14% (20)	71.43% (25)	11.43% (4)	5.71% (2)	8.57% (3)	100.0% (35)	100.0% (35)	0% (0)	
metal artwork (Q29382731)	29	100.0% (29)	100.0% (29)	17.24% (5)	17.24% (5)	10.34% (3)	100.0% (29)	100.0% (29)	0% (0)	
illuminated manuscript (Q48498)	27	100.0% (27)	100.0% (27)	74.07% (20)	74.07% (20)	77.78% (21)	100.0% (27)	100.0% (27)	7.41% (2)	
bronze sculpture (Q928357)	27	100.0% (27)	100.0% (27)	11.11% (3)	7.41% (2)	29.63% (8)	100.0% (27)	100.0% (27)	0% (0)	
stained glass (Q1473346)	26	100.0% (26)	100.0% (26)	19.23% (5)	19.23% (5)	76.92% (20)	100.0% (26)	100.0% (26)	0% (0)	
hanging scroll (Q2188827)	26	65.38% (17)	96.15% (25)	26.92% (7)	26.92% (7)	65.38% (17)	100.0% (26)	100.0% (26)	0% (0)	

<https://w.wiki/Q9J>

“Depiction” Wikidata Game

Depiction

Decide whether the image depicts the named item.

Entry type : Horse Soldier Dresses House Flower Boat Tree Leaf

Javanese Temple in Ruins [Q20503317]

Reruntuhan candi Jawa Tempelruinen auf Java

Auto

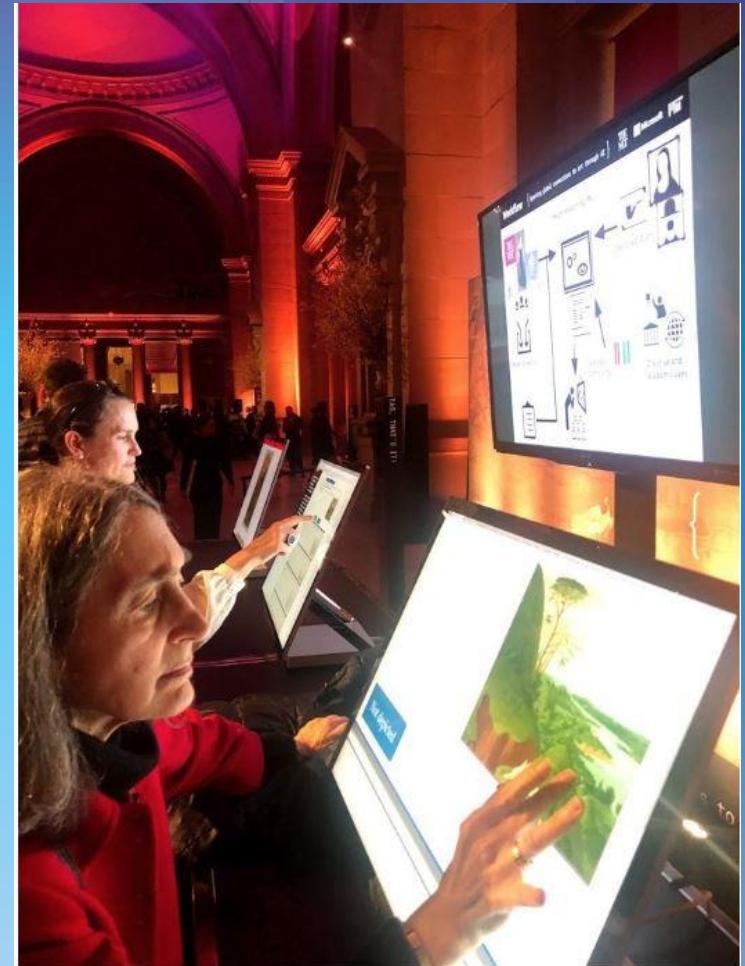
Painting by Raden Saleh, from 1860



Depicts a tree

Skip

Not depicted



Fuzheado, CC-BY-SA 4.0

Auto-generating content in other Wikis: Using WD to create a Wikibook from a GLAM Collection

Screenshot of a Wikibooks page titled "As fotografias de Guilherme Gaensly no acervo do Museu Paulista". The page features a large historical black and white photograph of a busy street scene in São Paulo, showing several multi-story buildings and people walking. Below the image, there are two sections of text with numbered lists:

Introdução

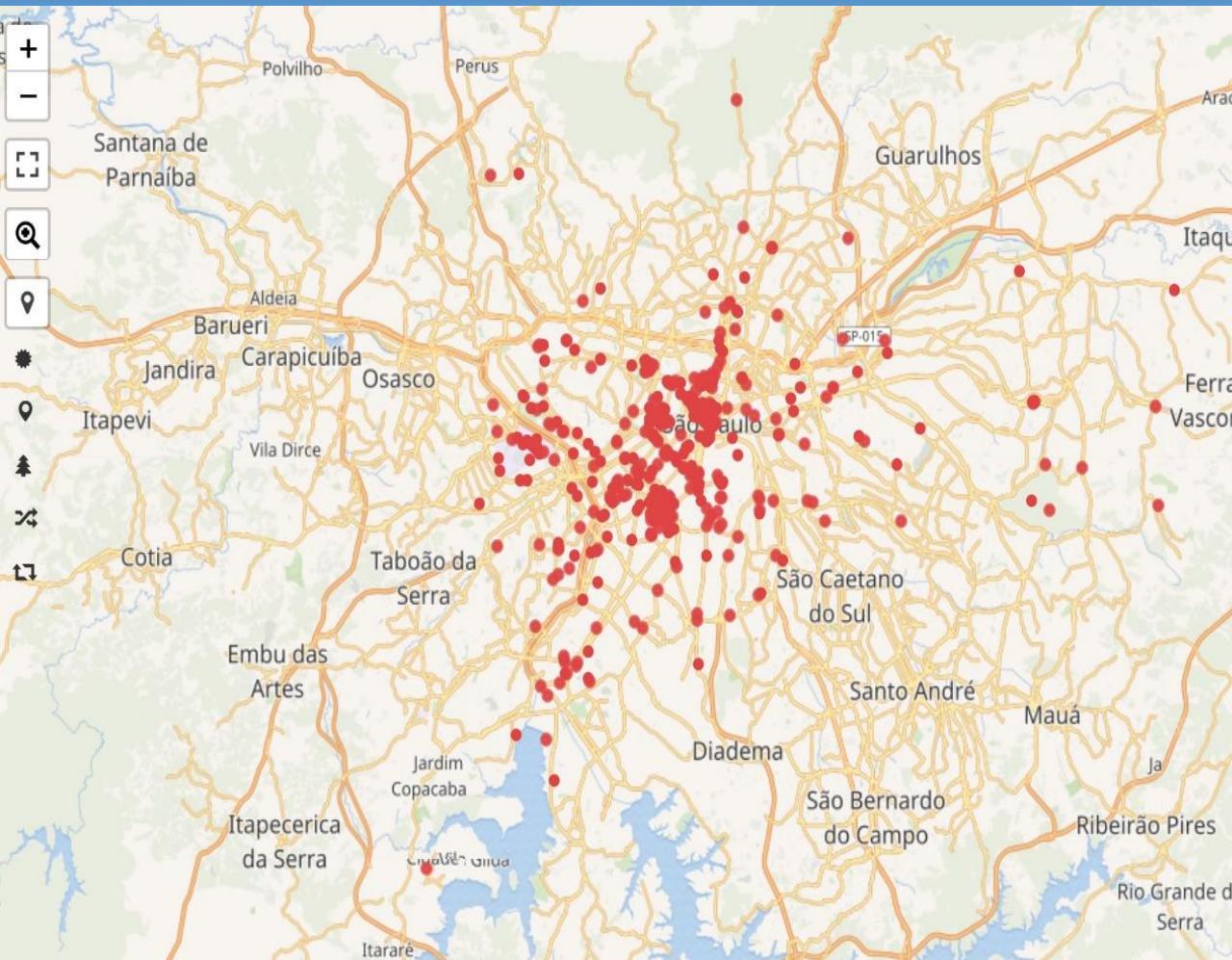
1. Capa
2. A Coleção Guilherme Gaensly do Museu Paulista e a Iniciativa Wikipédia

O Museu Paulista

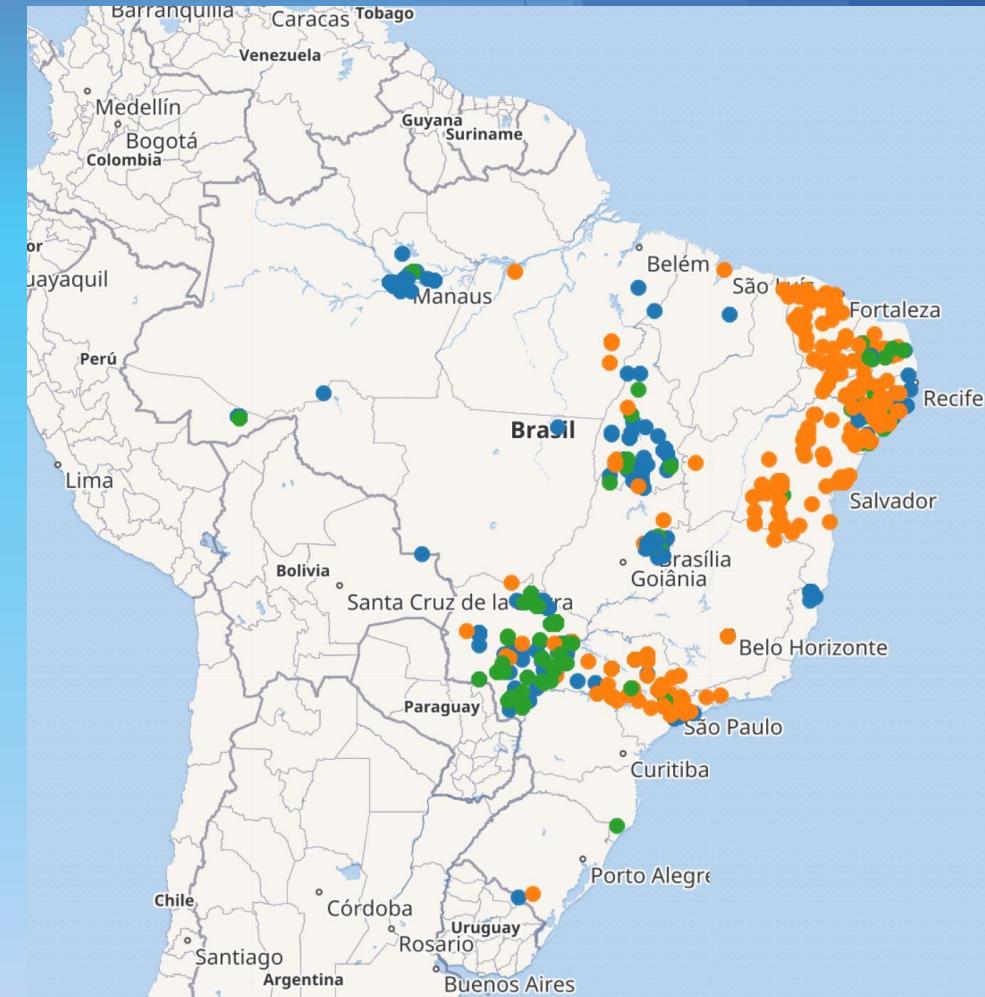
1. Edifício-monumento do Ipiranga - Fachada e lateral direita
2. Edifício-monumento do Ipiranga - Fachada, galeria, andar superior visto da escadaria central
3. Festejo à Frente do edifício-monumento do Ipiranga
4. Interior do Edifício-monumento do Ipiranga - Andar superior visto da escadaria central
5. Interior do Edifício-monumento do Ipiranga - Bezzil na escadaria central e andar superior visto a partir da escadaria central
6. Interior do Edifício-monumento do Ipiranga - Escadaria central
7. Interior do Edifício-monumento do Ipiranga - Vista a partir da escadaria central
8. Interior do Edifício-monumento do Ipiranga - Vista do andar superior a partir da escadaria central
9. Interior do Ipiranga
10. Interior do Monumento do Ipiranga - Capital
11. Monumento do Ipiranga

Creating new digital objects:

Students Mapping monuments in São Paulo & Dams in Brazil



<https://w.wiki/3EC>



<https://w.wiki/Pjn>

Digitally recreating lost Museum artifacts



Emperor Pedro II stands in front of his burning palace—the National Museum of Brazil..
Felipe Milanez, CC-BY-SA 4.0

“Data Archaeology”

#	Item	Label	Description	of	collection	part of	image
1	Ring bracelets Q56810425	en Ring bracelets pt Pulseiras de argola pt-br Pulseiras de argola	en object in the Polícia da Corte Collection at the National Museum (Rio de Janeiro) pt objeto da Coleção Polícia da Corte do Museu Nacional (Rio de Janeiro) pt-br objeto da Coleção Polícia da Corte do Museu Nacional (Rio de Janeiro) fr objet dans la Collection Polícia da Corte du Museu Nacional (Brésil)	item of collection or exhibition bracelet	Coleção Polícia da Corte	Africans in Brazil	Argolas de metal (pulseiras) MN 01.jpg 
2	Xangô figure Q56810426	en Xangô figure pt Representação de Xangô pt-br Representação de Xangô	en object in the Polícia da Corte Collection at the National Museum (Rio de Janeiro) pt objeto da Coleção Polícia da Corte do Museu Nacional (Rio de Janeiro) pt-br objeto da Coleção Polícia da Corte do Museu Nacional (Rio de Janeiro) fr objet dans la Collection Polícia da Corte du Museu Nacional (Brésil)	item of collection or exhibition oshe	Coleção Polícia da Corte	Africans in Brazil	Representação de Xangô MN 03.jpg 
3	Arrow Q56810427	en Arrow pt Flecha pt-br Flecha	en object in the Polícia da Corte Collection at the National Museum (Rio de Janeiro) pt objeto da Coleção Polícia da Corte do Museu Nacional (Rio de Janeiro) pt-br objeto da Coleção Polícia da Corte do Museu Nacional (Rio de Janeiro)	item of collection or exhibition Ofá	Coleção Polícia da Corte	Africans in Brazil	Flecha de Oxossi MN 03.jpg 

COVID-19

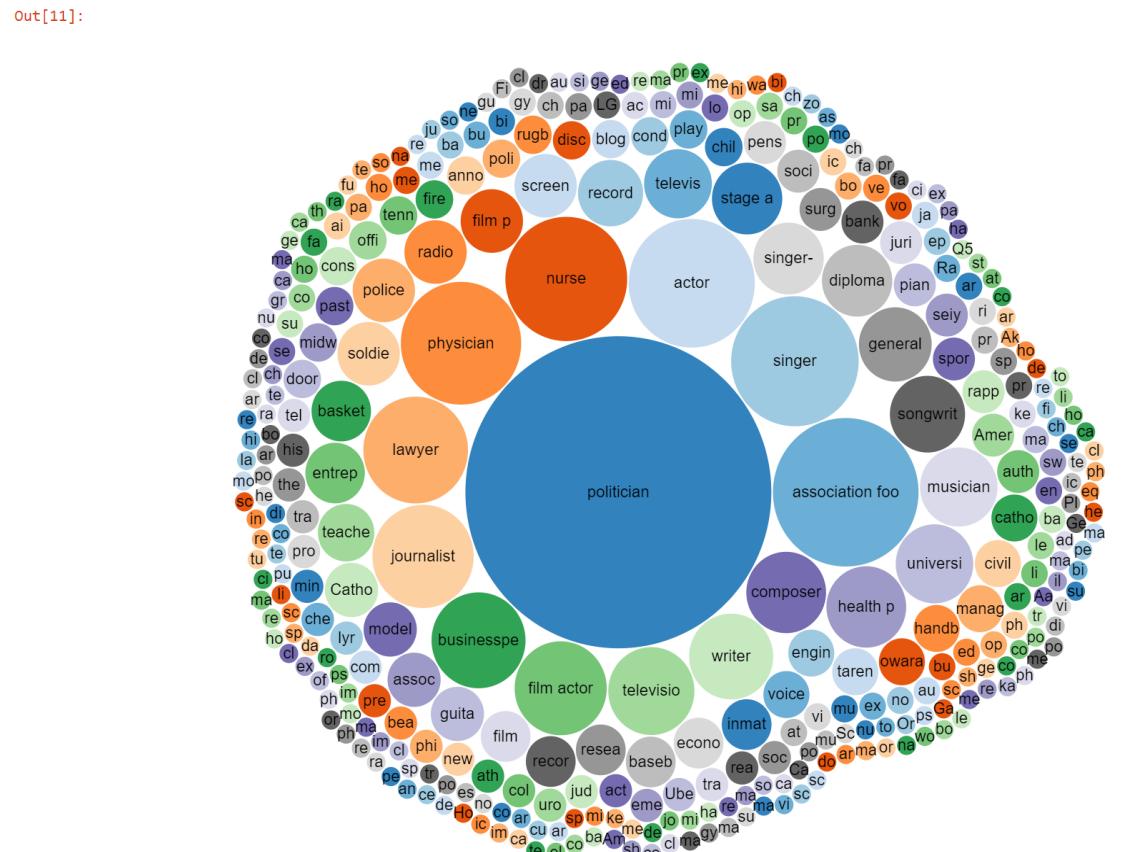
Screenshot of the table that tracks the pandemic's progress, as of May 9rd, 2020

COVID-19 pandemic by country and territory					
Locations[b]	Cases[a]	Deaths[c]	Recov.[d]	Ref.	
229	3,991,010	276,815	1,346,319	[4]	
United States ^[e]	1,328,419	78,787	182,496	[13]	
Spain ^[f]	223,578	26,478	133,952	[14]	
Italy	218,268	30,395	103,031	[15][16]	
United Kingdom ^[g]	215,260	31,587	No data	[18][19]	
Russia ^[h]	198,676	1,827	31,916	[20]	
Germany ^[i]	170,979	7,510	138,330	[22][21]	
Brazil	148,670	10,100	59,297	[23][24]	
France ^[j]	138,421	26,230	55,782	[25]	
Turkey	137,115	3,739	89,480	[26]	
Iran	106,220	6,589	86,064	[27]	
China ^[k]	82,887	4,633	78,046	[28]	
Canada	67,619	4,690	30,980	[29]	
Peru	61,847	1,714	19,012	[30][31]	
India	59,662	1,981	17,847	[32]	
Belgium ^[l]	52,596	8,581	13,411	[34]	
Netherlands ^[m]	42,382	5,422	No data	[36]	
Saudi Arabia	35,432	229	9,120	[37]	
Mexico	31,522	3,160	17,781	[38][39]	
Switzerland	30,207	1,526	25,900	[40][41]	
Ecuador	28,818	1,704	3,433	[42][43]	
Pakistan	27,474	618	7,756	[44]	
Portugal	27,268	1,114	2,422	[45]	
Chile ^[n]	25,972	294	12,160	[52]	
Sweden	25,921	3,220	4,971	[53][54]	
Ireland	22,541	1,429	17,110	[55]	
Singapore	22,460	20	2,296	[56][57]	
Belarus	22,052	126	6,050	[58]	
Qatar	20,201	12	2,370	[59]	
United Arab Emirates	16,793	174	3,837	[60]	
Israel ^[o]	16,409	245	11,007	[61]	
Austria	15,736	614	13,836	[62][63]	
Japan ^[p]	15,663	606	5,906	[64][65]	
Poland	15,366	776	5,184	[53][66]	

A query detailing the occupation of COVID-19 cases in a bubble chart

Occupations of COVID-19 cases

```
In [11]: wdq(width=1000,height=800,query="""
#defaultView:BubbleChart
SELECT ?occupation ?occupationLabel (COUNT(*) as ?count) WHERE {
  ?p wdt:P1050 wd:Q84263196 .
  ?p wdt:P106 ?occupation .
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en" } .
}
GROUP BY ?occupation ?occupationLabel
ORDER BY DESC(?count)
LIMIT 1000
""")
```



Implementing WD in Israeli Academia since 2015

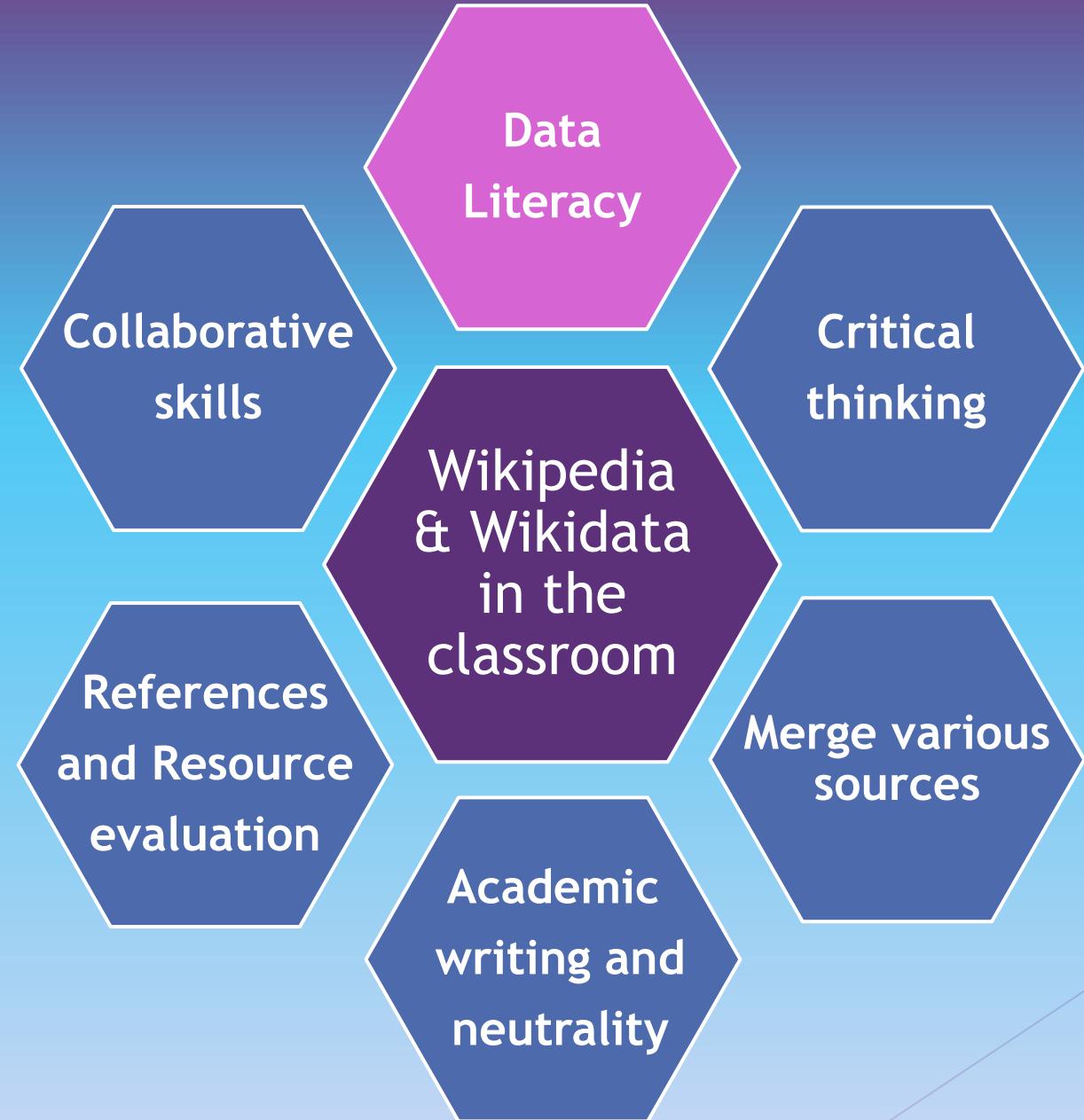
- **Alternative assessment:** help incorporate into courses:
 - University of Haifa – Dept. of Information Systems
 - University of Be'er Sheba – Digital Humanities
 - Bar Ilan University: Computer Sciences Department, Semantic Web Course;
 - In collaboration with the [Israel Antiquities Authority](#)
- **Main assessment:** main topic of my academic course:
 - Elective course at TAU: “From Web 2.0 to Web 3.0, from Wikipedia to Wikidata”

From Web 2.0 to Web 3.0, from Wikipedia to Wikidata

- Initiated in 2018
- Approved by the Rector's office
- Open to all undergraduates
- First in the world to feature Wikidata
- Based on a course model designed and implemented at TAU since 2013



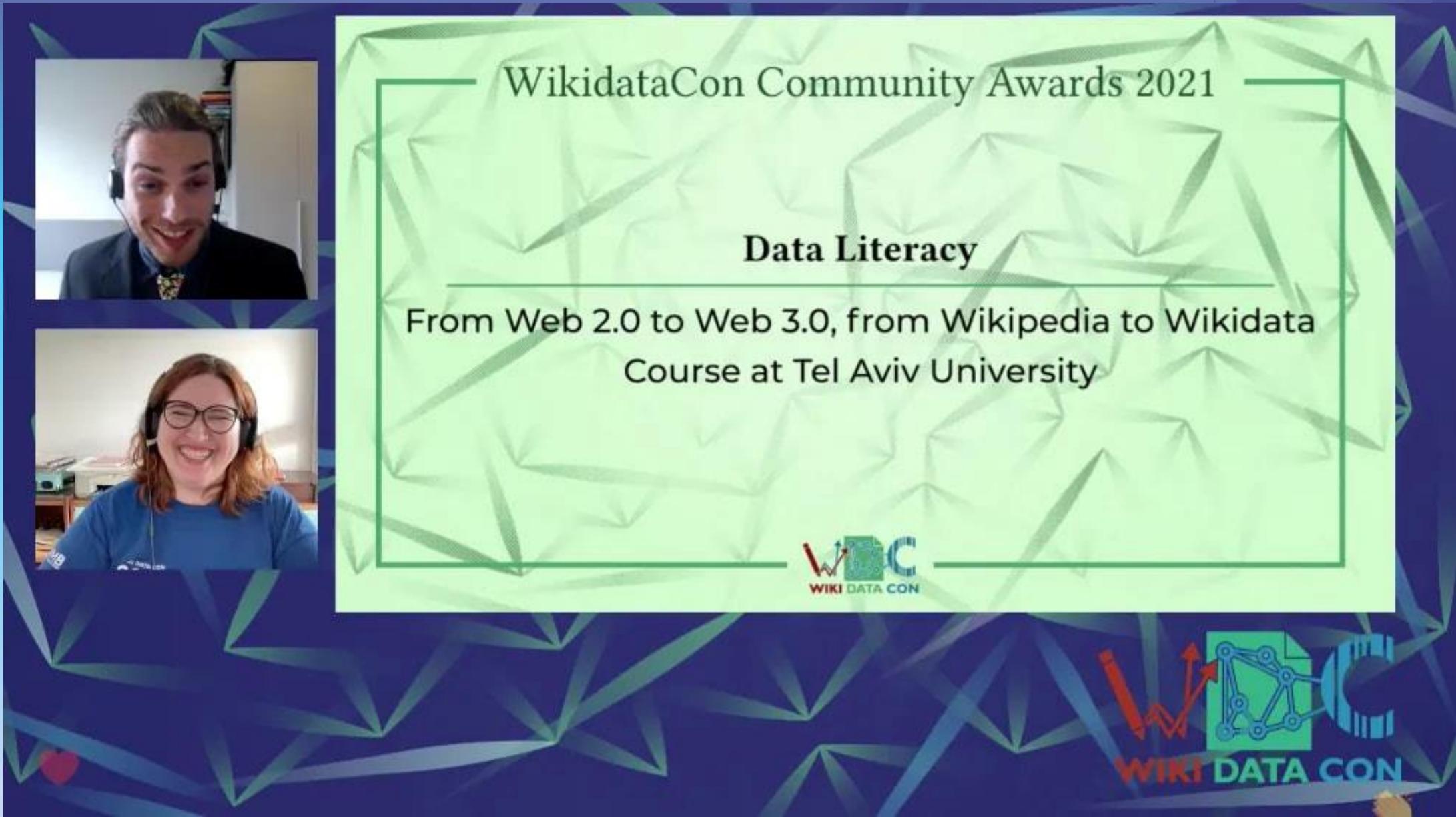
"From Web 2.0 to Web 3.0, from Wikipedia to Wikidata" Elective course at TAU, Class of 2018-2019



What do students say?

- * Active & engaged learning
- * Positive learning experience
- * Work that matters → giving back
- * Improved skills
- * Improved awareness of:
 - Gender gap, bias, knowledge equity
 - Neutrality, verifiability, completeness, convergence of sources
 - Copyrights, accessibility, openness

Thank you, WikidataCon 21!



Part 4: Summing up



Highlights: Potential & Benefits of Wikidata

- ❖ Connect separate databases with different metadata
- ❖ Queries: Getting specific answers to questions
- ❖ Data visualization!
- ❖ Tell stories with Data
- ❖ Help “curate” work – follow completeness, find & correct gaps
- ❖ Basis for applications: histropedia, crotos, genewiki, scholia, platypus
- ❖ Basis for AI: siri, alexa etc
- ❖ Critical thinking and analysis – for both curation and extraction
- ❖ Better evaluation of data – especially in “fake news” era → **Data Literacy**

Challenges

- Data reliability & verifiability, data manipulations
- Data completeness – mind the gap!
- Data modeling → query results might not include everything

Requires skills and critical thinking!

Suggested reading: “[INSIDE THE ALEXA-FRIENDLY WORLD OF WIKIDATA](#)” - WIRED, 2019

Suggested listening: [Questioning Wikidata](#) – keynote at Wikidata Conference, 2019

We are just at the beginning

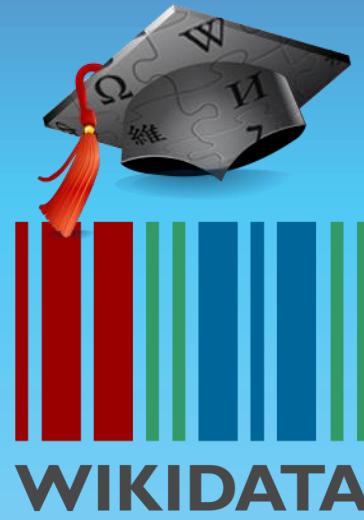
Challenges are opportunities

- Wikibase
- WikiLambda & WikiFunctions

Q & A



Thank you for participating!



Shani Evenstein Sigalov
shani.even@gmail.com

