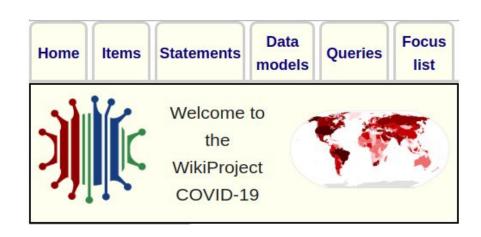
# Wikidata WikiProject COVID-19 : a community effort to curate the pandemic



Tiago Lubiana University of São Paulo, Brazil

ISWC Wikidata Workshop, 24 October 2022



Talking from Gamboa, Panama

# Personal background

- 2014 2017: Bachelor's in Biomedical Sciences
- 2018 03/2020: Master's in bioinformatics
- $08/2020 \rightarrow Starting PhD to study Wikidata$

Building a biological knowledge graph via Wikidata with a focus on the Human Cell Atlas

📵 Tiago Lubiana; 📵 Helder Nakaya

Funded PhD project proposal sent to the São Paulo Research Foundation in November 2019.

# Personal background

- 2014 2017: Bachelor's in Biomedical Sciences
- 2018 03/2020: Master's in bioinformatics
- $08/2020 \rightarrow Starting PhD to study Wikidata$

Building a biological knowledge graph via Wikidata with a focus on the Human Cell Atlas

📵 Tiago Lubiana; 📵 Helder Nakaya

Funded PhD project proposal sent to the São Paulo Research Foundation in November 2019.

# Personal background

- 2014 2017: Bachelor's in Biomedical Sciences
- 2018 03/2020: Master's in bioinformatics
- $08/2020 \rightarrow Starting PhD to study Wikidata$

Building a biological knowledge graph via Wikidata with a focus on the Human Cell Atlas

🕞 Tiago Lubiana; 🕞 Helder Nakaya

Funded PhD project proposal sent to the São Paulo Research Foundation in November 2019.

What to do between M.Sc. and Ph.D? WikiProject COVID-19!

# Mid-March 2020 - reaching out to editors

#### Core set:

- Contributors of COVID-19 related items
- Participants of WikiProject Zika Corpus

#### Would you like to contribute to a WikiProject COVID-19? [edit source [Add topic]

Hello,

I saw you contributed to COVID-19 (Q84263196) and I was wondering if you would be interested in helping to create a Wikidata WikiProject COVID-19.

# 18 people joined in the first day!

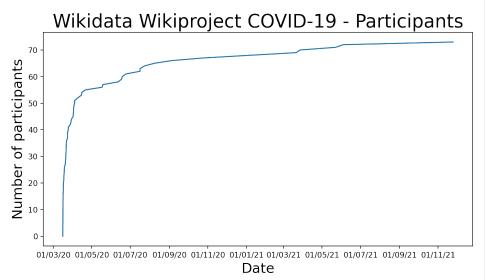
- TiagoLubiana 01:35, 16 March 2020 (UTC) [reply]
- Daniel Mietchen 01:42, 16 March 2020 (UTC) [reply]
- Jodi.a.schneider 02:45, 16 March 2020 (UTC) [reply]
- Chchowmein 02:45, 16 March 2020 (UTC) [reply]
- Dhx1 03:38, 16 March 2020 (UTC) [reply]
- Konrad Foerstner 06:02, 16 March 2020 (UTC)
   Salgo60 09:11, 16 March 2020 (UTC) [reply]

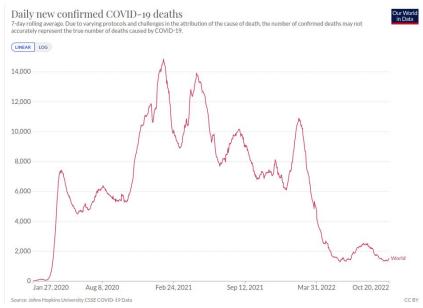
- Netha Hussain 06:19, 16 March 2020 (UTC) [reply]
- Bodhisattwa 06:56, 16 March 2020 (UTC) [reply]
- Neo-Jay 07:04, 16 March 2020 (UTC) [reply]
- John Samuel 07:31, 16 March 2020 (UTC) [reply]
- KlaudiuMihaila 07:53, 16 March 2020 (UTC) reply
- [reply]

- Andrawaag 10:12, 16 March 2020 (UTC) [reply]
- Whidou 10:16, 16 March 2020 (UTC) [reply]
- Blue Rasberry 15:07, 16 March 2020 (UTC) [reply]
- TJMSmith 16:15, 16 March 2020 (UTC) [reply]
- Egon Willighagen 16:49, 16 March 2020 (UTC) reply
- Nehaoua 20:32, 16 March 2020 (UTC) [reply]

# Initial boom of participants

Initial lockdowns, huge uncertainty → Very high engagement

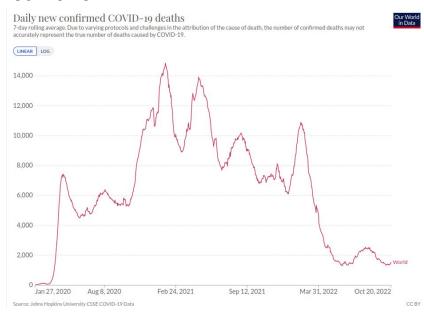




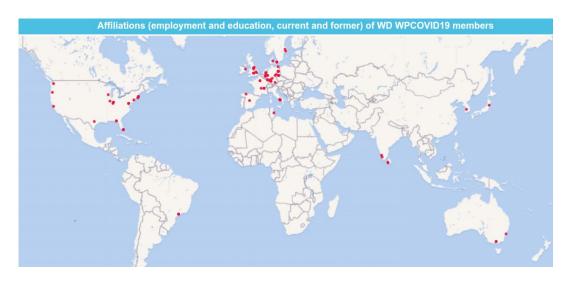
# Documenting: 3-day status updates

- Curated from 24/03 to 06/28
- Proxy for activity correlation with first wave





# Who were the people in WP COVID19?



Jodi Schneider Jasper Koehorst Aleiandro González Hevia Peter Murray-Rust José Emilio Labra Gayo Andra Waagmeester Alexander R. Pico Érica Azzellini Ralf Stephan Houcemeddine Turki Denny Vrandečić

Map: <a href="https://w.wiki/5p9D">https://w.wiki/5p9D</a>

Network:https://scholia.toolforge.or

g/event/Q114647284

# COVID-19-related properties

#### Epidemiology:

- wd:P1120 number of deaths
- wd:P1603 number of cases
- wd:P1846 distribution map
- wd:P8010 number of recoveries (new)
- wd:P8011 number of medical tests (new)
- wd:P8049 number of hospitalized cases (new)
- wd:P8204 tabular case data (new)
- wd:P9107 number of vaccinations (new)

#### Viral biology:

- wd:P3492 basic reproduction number
- wd:P8339 entry receptor (new)
- wd:P9632 PANGO lineage code (new)

#### Bibliography:

• wd:P8150 - COVIDWHO ID (new)

and more!

https://w.wiki/5p\$M

### COVID-19 Data Models

- Data models/COVID-19 apps
- Data models/COVID-19 death
- Data models/Contact tracing apps
- Data models/Curfews and lockdowns
- Data models/Effects and impacts
- Data models/Emergency measures
- Data models/Hospitals
- Data models/New properties
- Data models/Outbreaks
- Data models/Peer-reviewed articles
- Data models/Scientists
- Data models/Template
- Data models/Tests
- Data models/Viral genes
- Data models/Virus strains

#### Properties [edit | edit source]

#### Generic properties [edit | edit source]

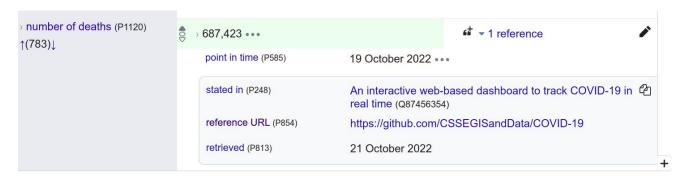
Title \$	ID ¢	Data type \$
publisher	P123	Item
creator	P170	Item

```
PREFIX xsd: <a href="http://www.w3.org/2001/XMLSchema#">http://www.w3.org/2001/XMLSchema#>
PREFIX wdt: <a href="http://www.wikidata.org/prop/direct/">http://www.wikidata.org/prop/direct/</a>
PREFIX wd: <a href="http://www.wikidata.org/entity/">http://www.wikidata.org/entity/>
#Reference: https://www.wikidata.org/wiki/Wikidata:WikiProject COVID-19/Data models/COVID-19 apps
start = @<contacttracingapp>
<contacttracingapp> EXTRA wdt:P31 {
  wdt:P31 [ wd:Q89288125 ] ;# instance of a COVID-19 contact tracing application
  wdt:P1476 LITERAL* :#title
  wdt:P366 .* ;#use
  wdt:P123 .{1} ;#publisher
  wdt:P178 .* ;#developers
  wdt:P495 .* ;#country of origin
  wdt:P306 .* :#operating system
  wdt:P856 .* ;#official website
  wdt:P577 .? ;#publication date
  wdt:P5008 .* ;#on focus list of Wikimedia project
```

Manual update → Bot

# COVID-19 pandemic in Brazil (Q86597695)

ongoing coronavirus pandemic in Brazil



- Manually update items x use a bot
- Bloated items x delete historical data x tabular case data
  - o Still open!



Minh Nguyễn User:Mxn tabular case data

Data:COVID-19 (STCenter)/IR/Q794.tab

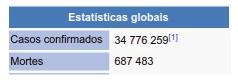
Fonte: Wikimedia Commons

COVID-19 cases in Iran

P585_date	P1603_confirmed	P1120_deaths	P8010_recovered
texto	número	número	número
Date •	Confirmed +	Deaths +	Recovered \$
id	89182	89182	89182
2020-01-22	0	0	0

Pandemic in Brazil 21 October 2022
 Using Wikidata

ptwiki, frwiki, eswiki, itwiki...



(overwritten in the Brazil page)



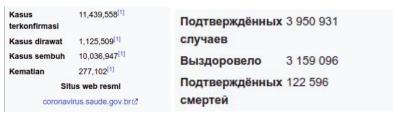


Dati :	statistici <sup>[1]</sup>
Numero di casi	34 771 320 (19 ottobre 2022)
Numero di guariti	14 964 631 (1° giugno 2021)
Numero di morti	687 423 (19 ottobre 2022)
Sito is:	tituzionale ௴
	Modifica dati su Wikidata · Manuale

#### Not using Wikidata

newiki, ruwiki, idwiki...





(viwiki and enwiki have local bots)

• Sometimes overwriting Wikidata leads to mistakes

Dati	statistici <sup>[1]</sup>
Numero di casi	16 547 674 <sup>[2]</sup> (1° giugno 2021)
Numero di guariti	14 964 631 (1° giugno 2021)
Numero di morti	462 966 (1° giugno 2021)
Sito is:	tituzionale ௴
	Modifica dati su Wikidata · Manuale



Dati	statistici <sup>[1]</sup>
Numero di casi	34 771 320 (19 ottobre 2022)
Numero di guariti	14 964 631 (1° giugno 2021)
Numero di morti	687 423 (19 ottobre 2022)
Sito is	tituzionale 🗗
	Modifica dati su Wikidata • Manuale

- CovidDatahubBot running from my personal computer (!!)
- If my internet connection was down, the bot stopped working

Now on toolforge, thanks Chico Venancio



Donatas Dabravolskas, CC-BY-SA

# Beyond Wikidata:

- Hackathons (Wikidata Lab XXII, BioHackathon Europe 2020)
- 6 group calls on Google Meet
- Active communication on Twitter
- Many scientific publications

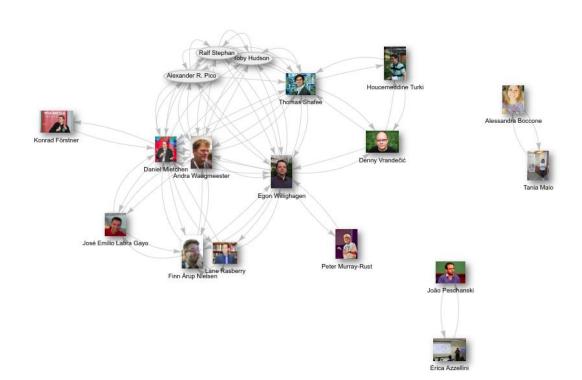


https://www.wikidata.org/wiki/Wikidata:WikiProject\_COVID-19/Project\_meeting/Archive

https://etherpad.wikimedia.org/p/open meeting COVID-19 20%2F04

# Scientific collaborations before the WikiProject

 WikiProject members were publishing together before the event!

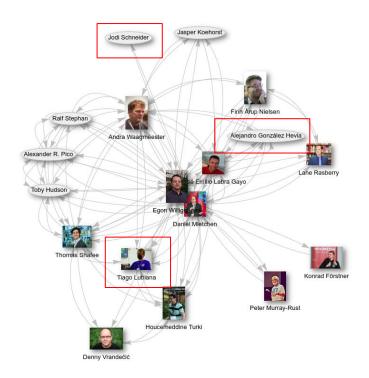


# Scientific collaborations after the Wikiproject

- New Wikidatians added to the mix
- New collaborations inside the network







# But that shows only a tiny bit! Tales of colabs:

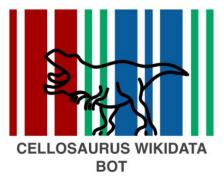
WP COVID-19: Reviving the Cellosaurus Bot



Cellosaurus - SARS-CoV-2 relevant information

#### Table of contents

- (A) Cell lines that can be used to grow SARS-CoV-2
- (B) Engineered cell lines useful for SARS-CoV-2 research
- (C) Cell lines shown to be not suitable to grow SARS-CoV-2
- (D) Cell lines used in COVID-19 vaccine production
- (E) Hybridomas producing mAb against SARS-CoV-2 proteins



~100,000 cell lines!



Amos Bairoch, creator of SwissProt and co-founder of UniProt

https://www.wikidata.org/wiki/Topic:Vky3mdxgv0ye0j6p https://github.com/calipho-sib/cellosaurus-wikidata-bot

# But that shows only a tiny bit! Tales of colabs:

#### Brazilian connections in Wikidata Lab COVID-19





João Vitor Ferreira Cavalcante

Wikidata for 5-star Linked Open Databases: a case study of PanglaoDB

Format: Pre-recorded with live Q&A Moderator(s): Robert Hoehndorf

<u>Tiago Lubiana</u>, University of São Paulo, Brazil João Vitor Ferreira Cavalcante, Federal University of Rio Grande do Norte. Brazil

Presentation Overview: Show

https://www.iscb.org/cms\_addon/conferences/ismbeccb2021/tracks/bio-ontologies

https://pt.wikipedia.org/wiki/Wikip%C3%A9dia:Edit-a-thon/Atividades\_em\_portugu%C3%AAs/Wikidata\_Lab\_XXII

# But that shows only a tiny bit! Tales of colabs:

**ELOHACKATION** COVID-19 Biohackathon



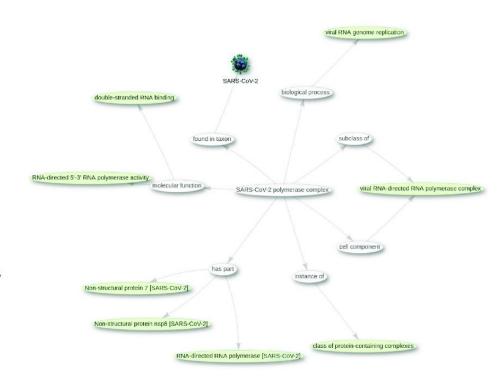


Birgit Meldal

#### Complex Portal 2022: new curation frontiers 3

Birgit H M Meldal, Livia Perfetto, Colin Combe, Tiago Lubiana,
João Vitor Ferreira Cavalcante, Hema Bye-A-Jee, Andra Waagmeester, Noemi del-Toro,
Anjali Shrivastava, Elisabeth Barrera, Edith Wong, Bernhard Mlecnik, Gabriela Bindea,
Kalpana Panneerselvam, Egon Willighagen, Juri Rappsilber, Pablo Porras,
Henning Hermjakob ▼, Sandra Orchard ▼

Nucleic Acids Research, Volume 50, Issue D1, 7 January 2022, Pages D578–D586,



https://academic.oup.com/nar/article/50/D1/D578/6414048

# Publications related to the WikiProject COVID-19



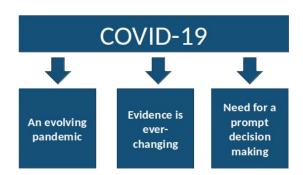
Houcemeddine Turki
User:Csisc

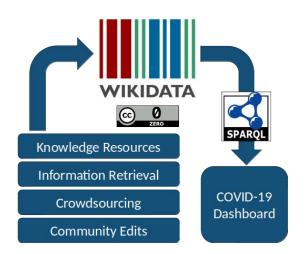
# Representing COVID-19 information in collaborative knowledge graphs: The case of Wikidata

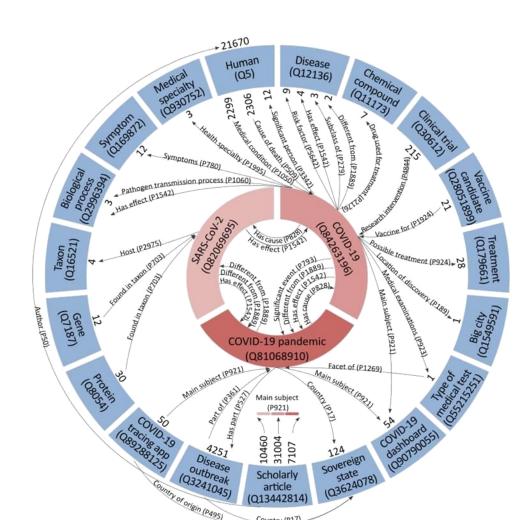


```
Article type: Research Article
```

```
Authors: Turki, Houcemeddine<sup>a; b</sup> | Hadj Taieb, Mohamed Ali<sup>b</sup> | Shafee, Thomas<sup>c</sup> | Lubiana, Tiago<sup>d</sup> | Jemielniak, Dariusz<sup>e</sup> | Aouicha, Mohamed Ben<sup>f</sup> | Labra Gayo, Jose Emilio<sup>g</sup> | Youngstrom, Eric A.<sup>h</sup> | Banat, Mus'ab<sup>i</sup> | Das, Diptanshu<sup>j; k</sup> | Mietchen, Daniel<sup>l; m; *</sup> | on behalf of WikiProject COVID-19, **
```

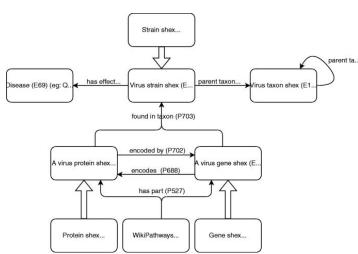






# Publications related to the WikiProject COVID-19



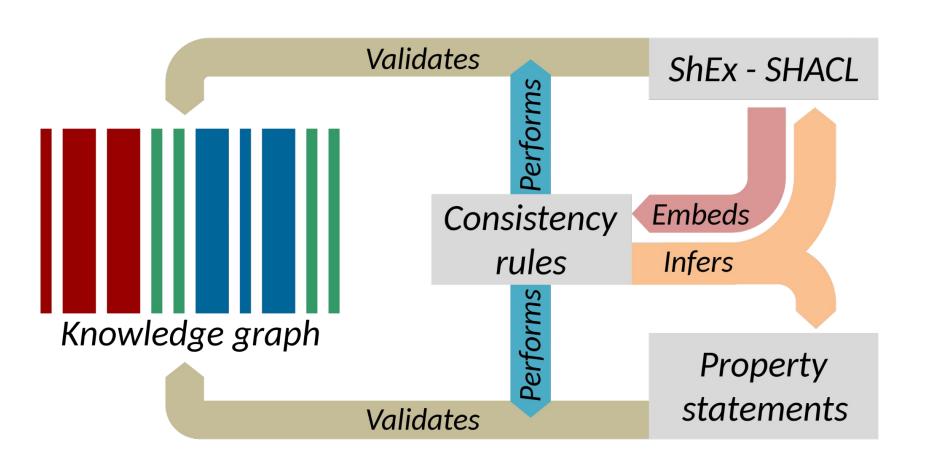


# Publications related to the WikiProject COVID-19



# Using logical constraints to validate statistical information about disease outbreaks in collaborative knowledge graphs: the case of COVID-19 epidemiology in Wikidata

Houcemeddine Turki<sup>1</sup>, Dariusz Jemielniak<sup>2</sup>, Mohamed A. Hadj Taieb<sup>1</sup>, Jose E. Labra Gayo<sup>3</sup>, Mohamed Ben Aouicha<sup>1</sup>, Mus'ab Banat<sup>4</sup>, Thomas Shafee<sup>5,6</sup>, Eric Prud'hommeaux<sup>7</sup>, Tiago Lubiana<sup>8</sup>, Diptanshu Das<sup>9,10</sup> and Daniel Mietchen<sup>11,12,13,14</sup>



# Uso de Wikidata y Wikipedia para la generación asistida de un vocabulario estructurado multilingüe sobre la pandemia de Covid-19

Using Wikidata and Wikipedia for assisted generation of a structured multilingual vocabulary about the Covid-19 pandemic



Tomás Saorín https://orcid.org/0000-0001-9448-0866 Universidad de Murcia Facultad de Información y Documentación Depto. de Información y Documentación Campus Universitario 30100 Espinardo (Murcia), España

tsp@um.es



https://orcid.org/0000-0002-1677-1059
Universidad de Murcia
Facultad de Información y Documentación
Depto. de Información y Documentación
Campus Universitario
30100 Espinardo (Murcia), España

Juan-Antonio Pastor-Sánchez 🖂



María-José Baños-Moreno
https://orcid.org/0000-0001-9137-1330
Universidad de Murcia
Facultad de Información y Documentación
Depto. de Información y Documentación
Campus Universitario
30100 Espinardo (Murcia), España
mbm41963@um.es

Tomás Saorín; Juan-Antonio Pastor-Sánchez; María-José Baños-Moreno

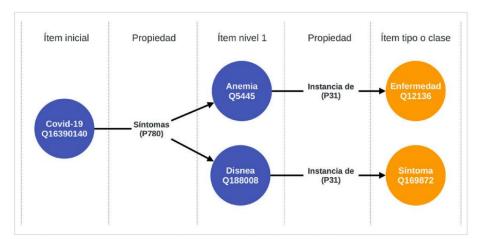


Figura 5. Esquematización de la tipología de entidades relacionadas.

# Wikidata Queries around the SARS-CoV-2 virus and pandemic

© 2020-2022 Addshore, Daniel Mietchen, Egon Willighagen

DOI 10.5281/zenodo.3977414



Egon Willighagen, Addshore, Daniel Mietchen,

### [ en ja nl es pt ]

#### **Authors and translators**

- Addshore
- Carolina Prado
- Daniel Mietchen
- Egon Willighagen
- Marvin Martens
- Sharon Prado
- Tiago Lubiana
- Yayamamo

# Wikidata Queries around the SARS-CoV-2 virus and pandemic

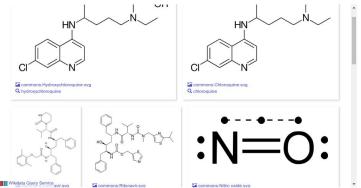
© 2020-2022 Addshore, Daniel Mietchen, Egon Willighagen



#### **Snippets:**

#### Drug repurposing

The Section 8.3 already listed an overview of clinical trials. It gives an idea of what symptoms people are worried about. It also gives insight in what drugs are studied for repurposing. Some of these have received a lot of attention, others less so. They look like this:



The following query lists the number of clinical trials by intervention:

SPARQL sparql/clinicalTrialsByIntervention.rq (run, edit)

#### **Variants**

Multiple variants of the virus genome made it into the international news. Originally thes were known as a Danish variant, a South-African variant, and a South-England variant. But the variants were only first discovered there, and the variant is not caused by anything related to the region. The following variants are listed in Wikidata, and includes the PANGO lineage code:

variant	pango
SARS-CoV-2 Lineage BA.1.1 (edit)	BA.1.1
SARS-CoV-2 Lineage BA.2.75 (edit)	BA.2.75
SARS-CoV-2 Lineage BA.2.75.2 (edit)	BA.2.75.2
Cluster 5 (edit)	
SARS-CoV-2 Lineage BF.7 (edit)	BF.7
Lineage B.1.427 (edit)	B.1.427
Lineage B.1.616 (edit)	B.1.616

# Wikidata Queries around the SARS-CoV-2 virus and pandemic

© 2020-2022 Addshore, Daniel Mietchen, Egon Willighagen



#### Snippets:

#### Genes

The RNA of SARS-CoV-2 has been sequenced. Therefore, the open reading frames are known and identified. We can query for the gene information in Wikidata with thie query:

SPARQL sparql/virusGenes.rq (run, edit)

```
SELECT ?gene ?geneLabel ?ncbigene WHERE {
    ?gene wdt:P703 wd:Q82069695 ; wdt:P31 wd:Q7187 .
    OPTIONAL { ?gene wdt:P351 ?ncbigene }
    SERVICE wikibase:label { bd:serviceParam wikibase:language "en,en". }
}
```

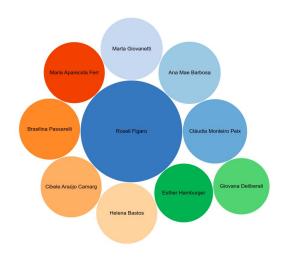
Which gives us these genes:

gene	ncbigene
surface glycoprotein (edit)	43740568
ORF1a polyprotein;ORF1ab polyprotein (edit)	43740578
ORF3a protein-encoding gene (edit)	43740569
envelope protein (edit)	43740570

### And much more! Some resources:

- Scholia-related and bibliographic catalogs
- E.g. works on COVID-19 by latin-american women:
   <a href="https://searchy.toolforge.org/search/Q84263196?gender=Q6581072&region=Q12585">https://searchy.toolforge.org/search/Q84263196?gender=Q6581072&region=Q12585</a>





# And much more! Some resources:

# Wikidata-based Dashboards[edit | edit source]

- External use of Wikidata data and queries Tunisia
- External use of Wikidata data and queries India
- San Francisco Bay Area medical cases by county
- San Francisco Bay Area case rate map
- COVIWD: COVID19 Wikidata Dashboard
- PAWS dashboard of cases and people affected, including demography

- Take the time to manage the crowd
  - Bring the people together

- Take the time to manage the crowd
  - Bring the people together
- Document and publicize the result
  - We are doing great stuff on Wikidata, let's share it with the world.
  - Bake sure you partner up with people with stamina!

- Take the time to manage the crowd
  - Bring the people together
- Document and publicize the result
  - We are doing great stuff on Wikidata, let's share it with the world.
  - Bake sure you partner up with people with stamina!
- Use the hype to push Wikidata forward
  - Interest might fade quickly, so move fast to contribute the most

- Take the time to manage the crowd
  - Bring the people together
- Document and publicize the result
  - We are doing great stuff on Wikidata, let's share it with the world.
  - Bake sure you partner up with people with stamina!
- Use the hype to push Wikidata forward
  - Interest might fade quickly, so move fast to contribute the most
- Have fun, the WikiProjects never end

# Thank you!















ACRISO SESSECTIONED, 1911

- Sj 23:15, 3 April 2020 (UTC)

- Doniel Metchen 01:42, 16 March 2020 (UTC)
- Jod a schneider 02:45, 16 March 2020 (UTC)

 813gan 18:25, 11 June 2020 (UTC) Tania 06:57, 17 Ame 2020 (UTC)

- Salgo60 09:11, 16 March 2020 (UTC)
- Whidou 10:16, 16 March 2020 (UTC)
- Blue Rasberry 15:07, 16 March 2020 (UTC)

Kasyap 02:45, 17 March 2020 (UTC)

. Denny 16:21, 17 March 2020 (UTC)

Joalpe 22:47, 17 March 2020 (UTC)

. Finn Arup Nielsen frielsen) 10:59, 18 March 2020 (UTC

- Sāsha 13.56, 24 October 2020 (UTC) . B bradley26 14:54, 16 July 2020 (UTC) • Fa\_suisse
- Rattentains 12:35, 23 July 2020 (UTC)

Yupik 20:23, 20 March 2020 (UTC)

Csiso 23:05, 20 March 2020 (UTC)

Gnoeee (A) 12:28, 21 March 2020 (UTC)

Jikoehorst 14:27, 21 March 2020 (UTC)

Scen 08:58, 22 March 2020 (UTC)

- - . Erfurth 19:20, 24 Nov 2021 (UTC)

Mathieu Kappler (talk) 01:17, 27 March 2021 (UTC)

- AlexanderPico 23:34, 27 March 2020 (UTC)
- Higa4 02:51, 29 March 2020 (UTC)
- Aleigh 11:04, 1 April 2020 (UTC)
- WII (WKI Edil) 17:36, 1 April 2020 (UTC)
- Ranithsii 04:47, 2 April 2020 (UTC)
- Hannolans 17:22, 2 April 2020 (UTC)
  - Familyouth 21:15: 3 April 2020 (UTC)

https://www.wikidata.org/wiki/Wikidata:WikiProject COVID-19

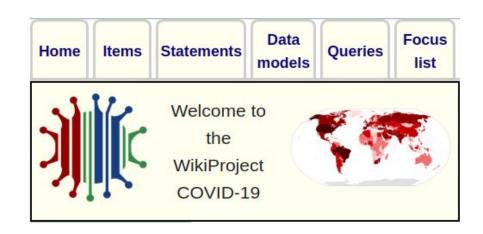


# Wikidata WikiProject COVID-19 : a community effort to curate the pandemic



Tiago Lubiana University of São Paulo, Brazil

ISWC Wikidata Workshop, 24 October 2022



I'm on Gamboa, Panama