Wikidata enabled on-the-fly ontological modelling of COVID-19 related concepts

Wikidata WikiProject COVID-19:
modelling the pandemic in real time

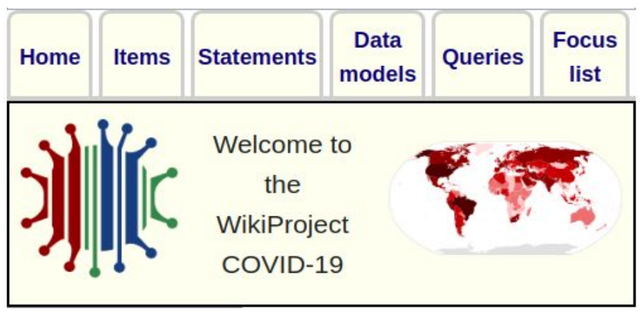
Tiago Lubiana (https://lubianat.github.io/)
Computational Systems Biology Laboratory - INOVA USP ,
University of São Paulo, Brazil

Introduction:

- Wikidata is a knowledge graph with more than 80 million items
- It provides a single hub for knowledge across domains
- We created the WikiProject COVID-19 on Wikidata, to catalyse efforts on modelling COVID-19 knowledge on an open format
- The life sciences community is already benefiting from Wikidata (see related resources)

Participation:

- The project was (and still is) open for volunteers
- Data is seamlessly integrated in real time



Banner of the WikiProject at Wikidata
(wikidata.org/wiki/Wikidata:WikiProject_COVID-19)

Results:

The collaboration resulted in models for Wikidata items related to the pandemic. These models relied both on reuse of established properties and on creation (an approval by the Wikidata community) of new ontological relations

outbreak (E173)

```
<outbreak> {
  wdt:P31 [ wd:Q3241045  ] ;# instance of an outbreak
# Core properties
  wdt:P1603 .{1} ;# number of cases
  wdt:P580 .{1} ;# start time

# Important, optional properties
  wdt:P582 .{1} ;# end time
  wdt:P1120 .* ;# number of deaths
  wdt:P8011 .* ;# number of clinical tests
  wdt:P8010 .* ;# number of recoveries
  wdt:P8045 .* ;# organized responses related to this event
  wdt:P276 .* ;# location
```

disease outbreak (Q3241045)

wd:Q86458103	COVID-19 pandemic in Europe
wd:Q86521237	COVID-19 pandemic in Asia
wd:Q86597685	COVID-19 pandemic in Denmark
wd:Q86597695	COVID-19 pandemic in Brazil
wd:Q834456	2002–2004 SARS outbreak
wd:Q917048	Great Plague of London
wd:Q934777	2009 flu pandemic in Mexico
wd:Q1099260	HIV/AIDS in the United States

Figure 1: Snippet of an Entity Schema (E173) for outbreaks written in Shex (left) and Wikidata items that can be modelled by this schema (right). SPARQL query for result on right: w.wiki/bvj

Entity Schemas for modelling of concepts related to COVID-19

E69	disease	E174	strain
E75	gene	E184	pandemic
E165	virus gene	E185	preprint
E167	protein	E186	macromolecular complex
E169	virus protein	E192	virus taxon
E170	virus strain	E194	Complex Portal entity
E173	outbreak	E189	clinical trial

New Wikidata properties for COVID-19 modelling

P8010	number of recoveries
P8045	organized response related to outbreak
P8049	number of hospitalized cases
P8064	Chinese Clinical Trial Registry ID
P8150	COVIDWHO ID
P8204	tabular case data
P8339	entry receptor

Related resources:

- SARS-CoV-2-Queries: Wikidata Queries around the SARS-CoV-2 virus and pandemic (Addshore, Mietchen, Willighagen; 2020)
- A protocol for adding knowledge to Wikidata, a case report (Waagmeester et al; 2020)
- Using logical constraints to validate information in collaborative knowledge graphs: a study of COVID-19 on Wikidata (Turki et al; 2020)

Acknowledgements:

Wikidata is built on the sum of all knowledge contributed by thousands of editors and the work of the Wikimedia Foundation.. The volunteer list for those that contributed to this project can be seen at:
wikidata.org/wiki/Wikidata:WikiProject_COVID-19#Participants