# Wikidata WikiProject COVID-19

Tiago Lubiana - University of São Paulo

June 25th, 2020

# Quick intro to Wikidata

- Wikidata is a Wikimedia project that wants to *structure* infos on Wikipedia



Item: *Earth (Q2)*    Property: *highest point (P610)*    custom value: *Mount Everest (Q513)*

# Quick intro to Wikidata

- Basic data structure (not sure what Karen meant)

# Quick intro to Wikidata

- Relationship to Wikipedia
  - Move structured data from Wikipedia templates

# Wikiprojects

- WikiProjects are groups of contributors who want to work together as a team to improve Wikidata.
- Focus on specific tasks or areas
- Know more about projects and quality metrics, check Wikidata Lab XVI

```
https://pt.wikipedia.org/wiki/Wikip%C3%A9dia:
Edit-a-thon/Atividades_em_portugu%C3%AAs/Wikidata_Lab_XVI
```

# Wikiprojects

- There are Wikiprojects about a great variety of topics:
  - Wikidata:WikiProject_Books
  - Wikidata:WikiProject_Ancient_Rome
  - Wikidata:WikiProject_Biology
  - Wikidata:WikiProject_Schemas

And anyone can get envolved!

# Wikiproject COVID-19 - Overview

- A group of users working to structure information about the COVID-19 pandemic at Wikidata.
- Project was created on 16 of March.
- +50 participants from all over the world.

# Wikiproject COVID-19 - Overview

- Motivations behind the WikiProject:
    - Data quality monitoring and data modeling for a quickly evolving situation
    - "Oiling the machine" for future crises
    - Standardization of data accross Wikipedias.
    - Providing structured data for secondary visualizations

```
wikidata.org/wiki/Wikidata:WikiProject_COVID-19
```

# Wikiproject COVID-19 - Quality monitoring and modelling

- Collaborative approach
- Bottom-up, iterative process

# Wikiproject COVID-19 - "Oiling the machine"

- Create new data models
- Improve the process itself of creating the model
- Finding consensus
- Expressing constraints
- Monitoring item quality

# Wikiproject COVID-19 - Standardization of Wikipedias

- Open problem: standardization.
- Examples of different approaches for number of cases :
  - Q83873593 (France)
  - Q84081576 (Sweden)

# Wikiproject COVID-19 - Providing structured data for secondary visualizations

- Wikidata Queries using the SPARQL system
- External databases that use Wikidata COVID-19 data.

# Wikiproject COVID-19 - Data Models (schemas)

-
   Example:https://www.wikidata.org/wiki/Wikidata:WikiProject_COVID-19/Data_models/Outbreaks
- Properties and human-readable models
- ShEx and computer readable models

# Wikiproject COVID-19 - Data Models (sources)

- Reuse of Wikidata properties
- Inspirations in external modelling approaches

# Wikiproject COVID-19 - Data Models (community)

- Who creates a data model?
- Where the discussions happen?
- How is consensus reached?
  - Data models are innocent until proven otherwise.
- How are these models enforced?

# Wikiproject COVID-19 - Data Models (community)

- If something is not modelled for other use cases, how to insert this on Wikidata?
- Property Proposals and Wikidata Property Admins.

# Diving into ShEx)

- Language for data structures
- ShEx core, optional properties and so on
- Tools to generate ShEx and validate Wikidata items

# Wikidata WikiProject COVID-19

Tiago Lubiana - University of São Paulo

June 25th, 2020