# Constructing a Question Answering system on the Covid-19 dataset.

By Jonathan Mallet

# RDF Modeling

We are getting the data from a repository from the **Center for Systems Science and Engineering (CSSE) at Johns Hopkins University**

The data are in a CSV format with a file per day and contains information we want like:
- Number of Confirmed cases
- Number of Deaths
- Number of Recoveries
- Place

| FIPS | Admin2 | Province_State | Country_Region | Last_Update | Lat | Long_ | Confirmed | Deaths | Recovered | Active | Combined_Key | Incidence_Rate | Case-Fatality_Ratio |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 45001 | Abbeville | South Carolina | US | 2020-06-02 02:33:08 | 34.22333378 | -82.46170658 | 44 | 0 | 0 | 44 | "Abbeville, South Carolina, US" | 179.39413707342928 | 0.0 |
| 22001 | Acadia | Louisiana | US | 2020-06-02 02:33:08 | 30.2950649 | -92.41419698 | 422 | 23 | 0 | 399 | "Acadia, Louisiana, US" | 680.1515029414135 | 5.450236966824645 |

# RDF Modeling

From this Data we want to create a Knowledge graph that make a relation between a place and the differents data about the Covid-19
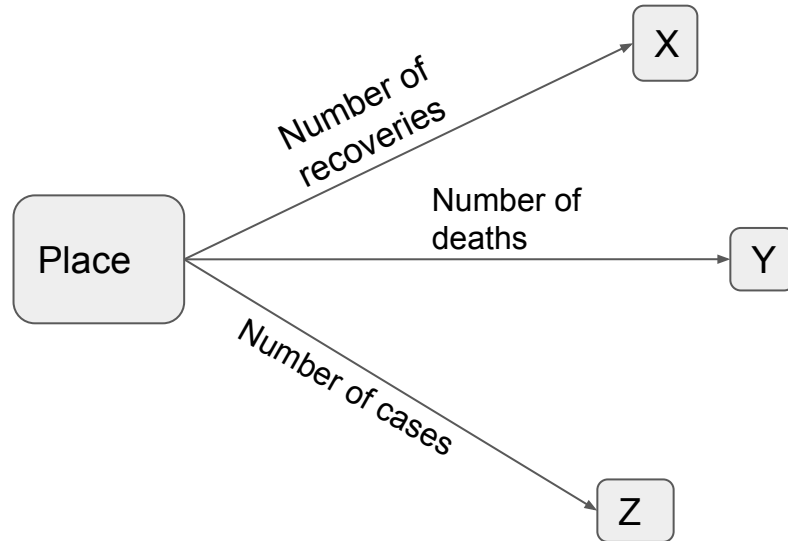
# Entity linking

Expand easily our knowledge about the entity we are using (like the labels in different language etc)

Like that we will be able to aggregate data following some other knowledge.
And add some specific knowledge



Transform the name of a place into the corresponding wikidata entity.
For that we are using nominatim which is a service from OpenStreetMap.

# Entity linking

Adding extra information to our knowledge base

what are the symptoms of the covid19 ?    Go

Confidence : ▬▬▬▬▬▬▬▬▬ 95 %    SPARQL LIST    DID YOU MEAN    DIRECT ANSWER

Is this the right answer?    ○ Yes    ○ No

*COVID-19 / symptoms*

**acute respiratory distress syndrome**    +

**respiratory failure**    +

**WIKIDATA**

Item    Discussion

COVID-19    (Q84263196)

zoonotic respiratory syndrome and infectious disease in humans, caused by SARS coronavirus 2

# RDF Modeling

what is the number of cases in USA ?    Go

Confidence : ████████████ 60 %

SPARQL LIST   DIRECT ANSWER

Is this the right answer? ○ Yes ○ No

*how many (located in the administrative territorial entity / United States of America / number of cases)*

**889,366**

For example "*what is the number of cases in USA ?*" which will give as an answer the sum of the number of case in the place located in USA.

```
SELECT (SUM(?o2) as ?c)  WHERE {
        ?s1  wdt:P131 wd:Q30  .
        ?s1  wdt:P1603  ?o2  .
}    LIMIT 1000
```

# Aggregation

what is the number of cases in USA ?　　　Go

Confidence : ████████ 60 %

SPARQL LIST　　DIRECT ANSWER

Is this the right answer?　○ Yes　○ No

*how many (located in the administrative territorial entity / United States of America / number of cases)*

**889,366**

SELECT (SUM(?o2) as ?c)  WHERE {
        ?s1  wdt:P131 wd:Q30  .
        ?s1  wdt:P1603  ?o2  .
}    LIMIT 1000

Place —— Located in ——> USA

Place —— Number of cases ——> Z

7
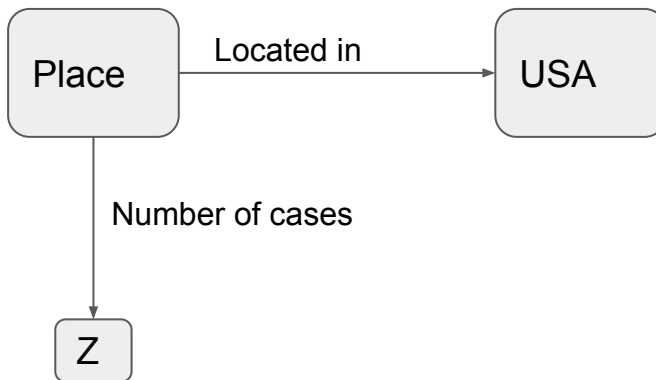
# Future work: Reification

But like that we only can build the knowledge about one day.
So now we want to create something like that :
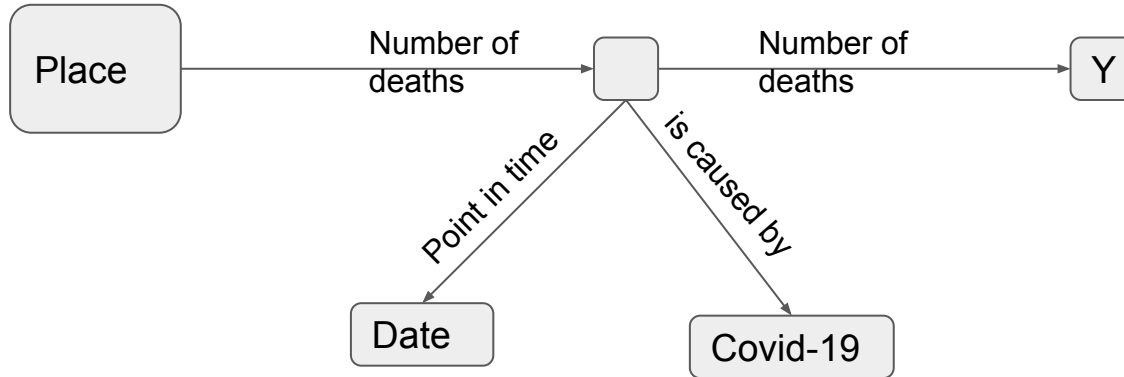


With something like that we will be able to ask the question:
"*What is the number of case at the 3 may in USA?* "
And get a satisfactory answer.

# Thank you