Phase 1

# Domain

The domain that we have chosen for our project is COVID-19. We are looking at its global impact, and how each country is responding to the virus in terms of health, government policies, political restlessness, and what the future effects on each country’s population might be.

# Dataset

We are getting our data from a variety of sources:

1. <https://www.coronanet-project.org/download> 🡪 select “Download Core Dataset”
   * The information from this dataset allows us to create the government policies relation, where we will be documenting the different types of policies implemented during COVID by different countries, which citizens it impacts, what the policy even is, and when it was put in place.
   * There isn’t much learning required to interpret the data, except perhaps understanding what some of the policies are and how they work.
   * The dataset is very clean (given in an excel sheet), so no cleaning will need to be performed on this dataset.
2. <https://ourworldindata.org/coronavirus-source-data> 🡪 select .csv for “Download the complete *Our World in Data* COVID-19 Dataset”
   * The information from this dataset allows us to compile the amount of confirmed cases, deaths, and general information (population, median age, life expectancy) about each country, which will be used in the country information and covid effects relations.
   * There isn’t any learning required to interpret the data, as we are just looking at number of deaths, COVID cases and general information about each country with this dataset.
   * The dataset is very clean (given in an excel sheet), so no cleaning will need to be performed on this dataset.
3. <https://www.disasterprotection.org/funding-covid-19-response> 🡪 scroll down to “21 September Update, download the data for your own analysis”
   * The information from this dataset allows us to create the financial aid relation, where we will be documenting the amount of funding each country was given, and the reason it was given to that country.
   * There isn’t much learning required to interpret the data, except perhaps understanding some of the purposes for the countries receiving the grants.
   * The dataset is very clean (given in an excel sheet), so no cleaning will need to be performed on this dataset.

# Questions

1. How does government polices influence the spread of COVID? Do more government policies decrease the amount of COVID cases?
2. Is an increased median age of a country associated with an increased amount of deaths due to COVID?
3. Is the date of grants being approved to certain countries associate with a decrease in the amount of cases of COVID?

# Schema

1. GovermentPolicies (**record\_id**, policy\_id, country, description, date\_started, type, compliance, enforcer)
2. CovidEffects (**date, country**, total\_cases, new\_cases, total\_deaths, new\_deaths)
3. CountryInfo (**country**, population, population\_density, median\_age, aged\_65\_older, aged\_70\_older, gdp\_per\_capita, life\_expectancy)
4. FinancialAid (**financial\_id**, country, **approval\_date**, grant\_amount, grant\_purpose, income\_type)

The referential integrity constraints:

1. FinancialAid [country] ⊆ CountryInfo[country]
2. CovidEffects [country] ⊆ CountryInfo[country]
3. GovernmentPolicies [country] ⊆ CountryInfo[country]
4. FinancialAid [approval\_date] ⊆ CovidEffects[country]