

# Global Database of COVID-19 Vaccinations

## Database Relational Model

By Adam Mutimer (S3875753)

### First Assumption:

**iso\_code** is the strongest link in the data provided, csv files not using **iso\_code** have a **location** which is a country name and can be converted using a query. EXCEPT: for **us\_state\_vaccinations.csv** however knowing from the file name its US data when importing into **CountryStateDailyTotals** we can add the **iso\_code** column with the value USA.

iso\_code = ISO 3166-1 alpha-3 - Three letter country code

### CountryStateDailyTotals

iso\_code {FK}  
location  
date  
total\_vaccinations  
total\_distributed  
people\_vaccinated  
people\_fully\_vaccinated\_per\_hundred  
total\_vaccinations\_per\_hundred  
people\_fully\_vaccinated  
people\_vaccinated\_per\_hundred  
distributed\_per\_hundred  
daily\_vaccinations\_raw  
daily\_vaccinations  
daily\_vaccinations\_per\_million  
share\_doses\_used

### Fifth Assumption:

**us\_state\_vaccinations.csv** could be apapted in the future to track states in other countries so the table for this data should be universal. So I created a table **CountryStateDailyTotals** to handle this type of data with the additional column **iso\_code**

### CountryVaccByManufacturer

iso\_code {FK}  
date  
vaccine\_id {FK}  
total\_vaccinations

### Vaccines

vaccine\_id {PK}  
name

### Third Assumption:

Countries are using the same names for each vaccine

### Locations

iso\_code {PK}  
location

### LocationVaccines

iso\_code {FK}  
vaccine\_id {FK}  
date\_available  
date\_unavailable  
data\_source\_id {FK}

### CountryVaccinations

iso\_code {FK}  
date  
total\_vaccinations  
people\_vaccinated  
people\_fully\_vaccinated  
daily\_vaccinations\_raw  
daily\_vaccinations  
total\_vaccinations\_per\_hundred  
people\_vaccinated\_per\_hundred  
people\_fully\_vaccinated\_per\_hundred  
daily\_vaccinations\_per\_million

### DataSource

data\_source\_id {PK}  
iso\_code {FK}  
source\_name  
source\_website  
last\_observation

### AgeGroups

age\_group\_id {PK}  
group

### Forth Assumption:

Age Groupings are the same accross all countries and therefor can be recycled

### AgeGroupVaccinations

iso\_code {FK}  
date  
age\_group\_id {FK}  
people\_vaccinated\_perhundred  
people\_fully\_vaccinated\_per\_hundred

### CountryDailyTotals

iso\_code {FK}  
date  
data\_source\_id {FK}  
total\_vaccinations  
people\_vaccinated  
people\_fully\_vaccinated

### Second Assumption:

based off all data in csv files i will be tracking the date a country started using a vaccine in the **LocationVaccines** table using **date\_available** as well as tracking the date inwhich the data stopped showing the vaccine being used in the **date\_unavailable** column