

Global Database of COVID-19 Vaccinations

Entity Relational Diagram

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** where **location** is a state; 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

Forth Assumption:

Age Groupings are the same across all countries and therefore can be recycled

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 adapted 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

Locations

iso_code {PK}
location

LocationVaccines

iso_code {FK}
vaccine_id {FK}
date_available
date_unavailable
data_source_id {FK}

Vaccines

vaccine_id {PK}
name

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

AgeGroupVaccinations

iso_code {FK}
date
age_group_id {FK}
people_vaccinated_per_hundred
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 in which the data stopped showing the vaccine being used in the **date_unavailable** column

Third Assumption:

Countries are using the same names for each vaccine