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

SPECIAL CASE: OWID_XXX - not all are listed ion locations.csv but exist in vaccinations.csv - USE Query to add these to "DataSources"

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

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**

Sixth Assumption:
All "date" values are not NULL or EMPTY and in CSV are in format "DD/MM/YYYY" this is a horrible date format to sort with in SQLLITE and will need to be converted to YYYY/MM/DD

Second Assumption:
based off all data in csv files i will be tracking the date a country started using a vaccine in the ***LocationVaccinces*** table using **date_avaliable** as well as tracking the date inwhich the data stopped showing the vaccine being used in the **date_unavaliable** column

LocationVaccinces:
Added "date_avaliable" to track when it was first observed being used in a country
Added "date_unvaliable" to track when it was stopped being used in a country

Third Assumption:
Countries are using the same names for each vaccine

Seventh Assumption:
All "last_observation" values are not NULL or EMPTY and in CSV are in format "DD/MM/YYYY" this is a horrible date format to sort with in SQLLITE and will need to be converted to YYYY/MM/DD

AgeGroups
- age_group_id {PK}
- group

Locations
- iso_code {PK}
- location

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

AgeGroupVaccinations
- iso_code {FK}
- date
- age_group_id {FK}
- people_vaccinated_perhundred
- people_fully_vaccinated_per_hundred

LocationVaccinces
- iso_code {FK}
- vaccine_id {FK}
- date_avaliable
- date_unavaliable
- 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

CountryDailyTotals
- iso_code {FK}
- date
- data_source_id {FK}
- vaccine_id {FK}
- total_vaccinations
- people_vaccinated
- people_fully_vaccinated

CountryVaccByManufacturer
- iso_code {FK}
- date
- vaccine_id {FK}
- total_vaccinations

Vaccines
- vaccine_id {PK}
- name

DataSource
- data_source_id {PK}
- iso_code {FK}
- source_name
- source_website
- last_observation