

## Vaccinations

location, date, total\_vaccination, people\_vaccinated, people\_fully\_vaccinated, total\_booster, daily\_vaccinations, daily\_people\_vaccinated, vaccine, source\_url\*

## VaccinationsManufacturer

location, date , vaccine, total\_vaccinations

## VaccinationsUSstate

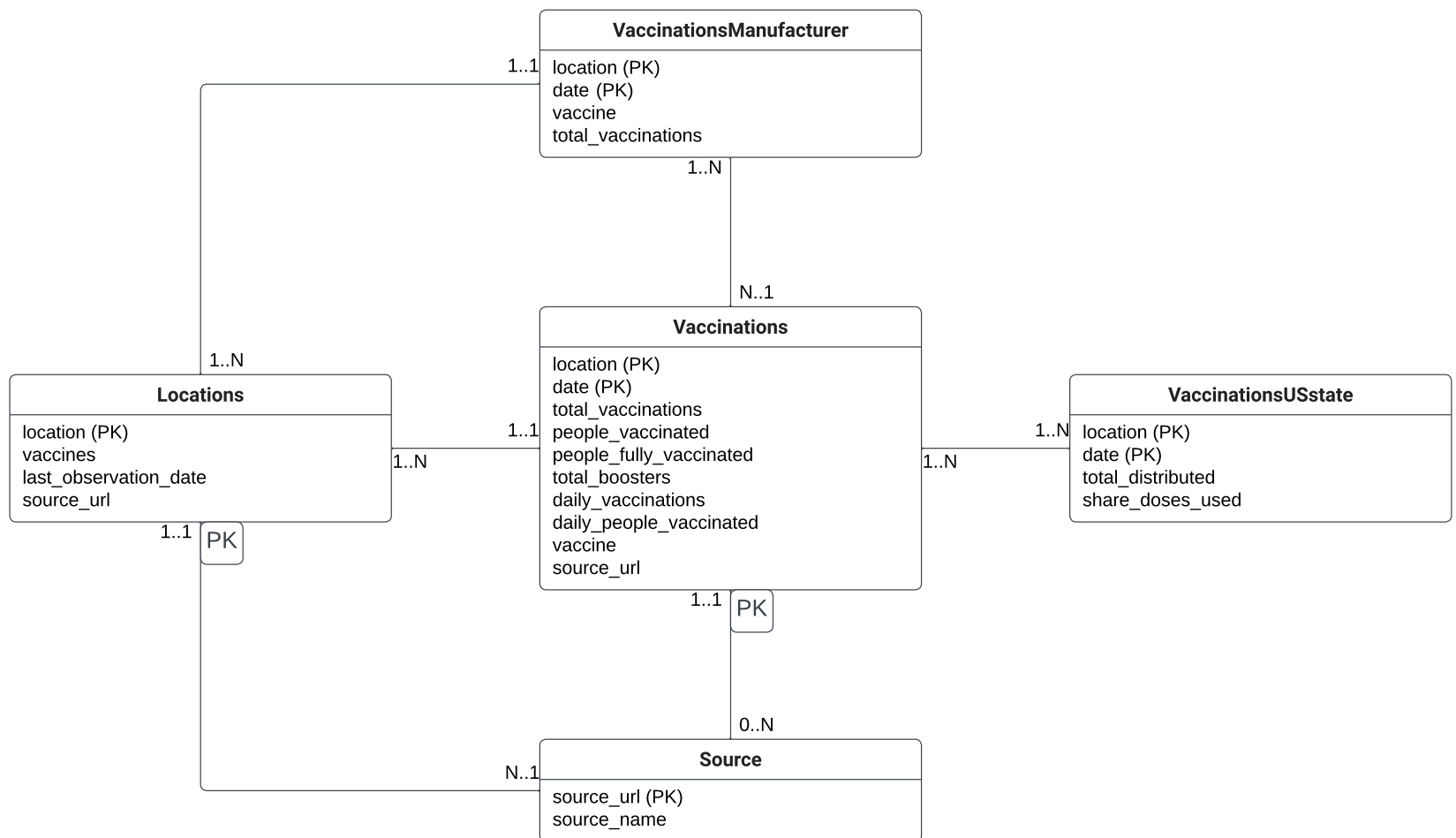
location, date, total\_distributed, share\_doses\_used

## Locations

location, vaccines , last\_observation\_date, source\_url\*

## Source

source\_url, source\_name



1. Data removal was performed from the 'Vaccinations' table, involving 99,187 rows, with the conditions that 'location' matches Asia, low income, upper middle income, lower middle income, high income, Europe, North America, South America, European Union, Africa, South Africa, and Central African Republic.
2. Data was imported into the 'Vaccinations' table, amounting to 54,629 rows, from the 'VaccinationsUSstate' table due to the presence of identical columns, namely 'total\_vaccinations,' 'people\_vaccinated,' 'people\_fully\_vaccinated,' 'daily\_vaccinations,' and 'total\_boosters.'
3. Data was incorporated into the 'Vaccinations' table through a merge operation from the 'Country' dataset, specifically for Australia, England, France, and Germany, as they shared similar columns, including 'total\_vaccinations,' 'people\_vaccinated,' 'people\_fully\_vaccinated,' and 'total\_boosters.' An example of the code used for merging is `IFERROR(FILTER(England!$C$2:$D$829,England!$B$2:$B$829 = B49481),"")`, in Google Sheets.
4. The 'iso\_code' was removed from all tables, as it was deemed unnecessary, and tables that had it already displayed the 'name' value.
5. Data duplicates were removed from the 'Source' table.
6. Data from 'vaccinations-by-age-group' was not incorporated because all its measurements were in 'per\_hundred'.
7. The 'total\_vaccinations' column in the 'VaccinationsManufacturer' table could not be removed or merged into the 'Vaccinations' table, as the 'VaccinationsManufacturer' table had incomplete 'location' data compared to the 'Vaccinations' table.
8. Normalization was not achievable for some tables that contained overlapping data with the 'VaccinationsManufacturer' table, as they also stored data from 'VaccinationsUS' within the 'Vaccinations' table.
9. In some tables, a composite key or primary key with more than one column was necessary, as there were no unique values in a single column, and the designer deemed it unnecessary to implement the concept of a surrogate key.