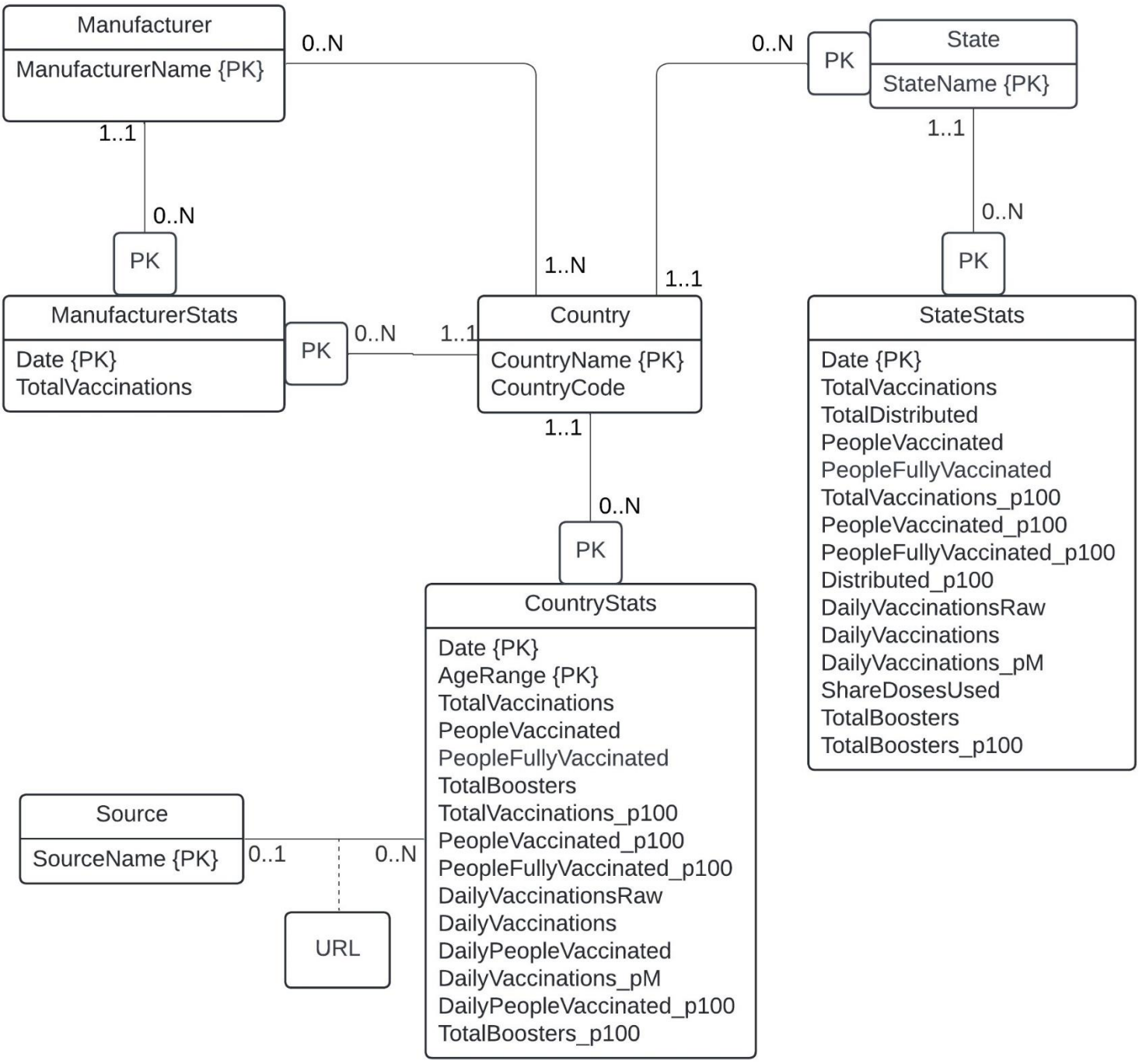
Peter Ljubisic – S4081442

ER Diagram



Extra Assumptions

The **Source** entity does not necessarily have a direct relationship to the **Country** entity. For example, the World Health Organisation is a global institution that shares vaccine statistics but is not confined to any specific country.

Vaccine manufacturers must supply stock to at least one country, but a country does not need to have a vaccine.

Not every **Manufacturer** & **Country** combination contains statistics for the **ManufacturerStats** entity.

Each record in **CountryStats** can only relate to one country, and does not necessarily need a source.

Each record in **StateStats** can only relate to one state, and never has any source.

For **CountryStats** records related to the whole population of the region, the corresponding **AgeRange** value is simply "0+".

Each **Date** only appears once per **StateName** in the **StateStats** entity, and once per (**CountryName** and **AgeRange**) combo in the **CountryStats** entity.

Mapping to Relational Schema in 7 Steps

- Step 1: Map Strong Entities

  Country (CountryName, CountryCode)
  Manufacturer (ManufacturerName)
  Source (SourceName)

- Step 2: Map Weak Entities

  ManufacturerStats (Weak) & Manufacturer (Strong)

  ManufacturerStats (ManufacturerName*, Date, TotalVaccinations)

  ManufacturerStats (Weak) & Country (Strong)

  ManufacturerStats (ManufacturerName*, CountryName*, Date, TotalVaccinations)

  StateStats (Weak) & State (Strong)

  StateStats (StateName*, Date, TotalVaccinations, TotalDistributed, PeopleVaccinated, PeopleFullyVaccinated, TotalVaccinations_p100, PeopleVaccinated_p100, PeopleFullyVaccinated_p100, Distributed_p100, DailyVaccinationsRaw, DailyVaccinations, DailyVaccinations_pM, ShareDosesUsed, TotalBoosters, TotalBoosters_p100)

  State (Weak) & Country (Strong)

  State (CountryName*, StateName)

  CountryStats (Weak) & Country (Strong)

  CountryStats (CountryName*, Date, AgeRange, TotalVaccinations, PeopleVaccinated, PeopleFullyVaccinated, TotalBoosters, TotalVaccinations_p100, PeopleVaccinated_p100, PeopleFullyVaccinated_p100, DailyVaccinationsRaw, DailyVaccinations, DailyPeopleVaccinated, DailyVaccinations_pM, DailyPeopleVaccinated_p100, TotalBoosters_p100)

- Step 3: Map 1:1 Relationships

  No 1:1 Relationships

- Step 4: Map 1:N Relationships

  This step changes only one entity:
  CountryStats (N-Side) & Source (1-Side)

  CountryStats (CountryName*, Date, AgeRange, TotalVaccinations, PeopleVaccinated, PeopleFullyVaccinated, TotalBoosters, TotalVaccinations_p100, PeopleVaccinated_p100, PeopleFullyVaccinated_p100, DailyVaccinationsRaw, DailyVaccinations, DailyPeopleVaccinated, DailyVaccinations_pM, DailyPeopleVaccinated_p100, TotalBoosters_p100, SourceName*, URL)

- Step 5: Map M:N Relationships

  Country & Manufacturer

  CountryManufacturer (CountryName*, ManufacturerName*)

- Step 6: Map Multi-Valued Attributes

  There are no multi-valued attributes

- Step 7: Map Higher-Degree Relationships

  There are no higher-degree relationships

Prototype Schema

Country (CountryName, CountryCode)

Manufacturer (ManufacturerName)

Source (SourceName)

ManufacturerStats (ManufacturerName*, CountryName*, Date, TotalVaccinations)

State (CountryName*, StateName)

StateStats (StateName*, Date, TotalVaccinations, TotalDistributed, PeopleVaccinated, PeopleFullyVaccinated, TotalVaccinations_p100, PeopleVaccinated_p100, PeopleFullyVaccinated_p100, Distributed_p100, DailyVaccinationsRaw, DailyVaccinations, DailyVaccinations_pM, ShareDosesUsed, TotalBoosters, TotalBoosters_p100)

CountryStats (CountryName*, Date, AgeRange, TotalVaccinations, PeopleVaccinated, PeopleFullyVaccinated, TotalBoosters, TotalVaccinations_p100, PeopleVaccinated_p100, PeopleFullyVaccinated_p100, DailyVaccinationsRaw, DailyVaccinations,

DailyPeopleVaccinated, DailyVaccinations_pM, DailyPeopleVaccinated_p100, TotalBoosters_p100, SourceName*, URL)

CountryManufacturer (CountryName*, ManufacturerName*)

Normalisation

All relations outlined in the prototype are in 1NF by virtue of no multi-valued or repeating attributes appearing in the ER diagram and relational schemas.

For a relation to be in 2NF, the non-primary key attributes must be functionally dependent on the entire primary key. Before checking if each relation is in 2NF, it is best to sketch out the functional dependencies for each relation so far.

Country Entity

CountryName -> CountryCode

CountryCode -> CountryName

Manufacturer Entity

N/A (Only One Attribute)

Source Entity

N/A (Only One Attribute)

State Entity

StateName -> CountryName

StateStats Entity

StateName, Date -> TotalVaccinations, TotalDistributed, PeopleVaccinated, PeopleFullyVaccinated, TotalVaccinations_p100, PeopleVaccinated_p100, PeopleFullyVaccinated_p100, Distributed_p100, DailyVaccinationsRaw, DailyVaccinations, DailyVaccinations_pM, ShareDosesUsed, TotalBoosters, TotalBoosters_p100

Note that these attributes are not functionally dependent on **CountryName**, as they only care about one state of a country, and are independent of the statistics associated with the other states of the same country. **StateName** is not functionally dependent on **CountryName** as you cannot single out a state in a country with multiple states using only the name of the country.

It may be mistakenly assumed that more functional dependencies exist in **StateStats**, such as **TotalVaccinations** -> **TotalVaccinations_p100**, however this is not true because it is impossible to determine **TotalVaccinations_p100** from **TotalVaccinations** with the data available. A population attribute would be needed, but there isn't one available. So these are the only functional dependencies of **StateStats**.

CountryStats Entity

CountryName, Date, AgeRange -> TotalVaccinations, PeopleVaccinated, PeopleFullyVaccinated, TotalBoosters, TotalVaccinations_p100, PeopleVaccinated_p100, PeopleFullyVaccinated_p100, DailyVaccinationsRaw, DailyVaccinations, DailyPeopleVaccinated, DailyVaccinations_pM, DailyPeopleVaccinated_p100, TotalBoosters_p100, SourceName, URL)

CountryName, Date, AgeRange, SourceName -> URL

CountryName, Date, AgeRange, URL -> SourceName

Not that there is a relationship between **SourceName** and **URL**, but it is not direct. This is because one **SourceName** can be attributed with many **URLs** (e.g. different URLs for statistics of the United States sourced by the Centers for Disease Control and Prevention), and one **URL** can be associated with many **SourceNames** (e.g. https://data.who.int/dashboards/covid19/ is primarily associated with the World Health Orgaisation, but also lists Africa Centres for Disease Control and Prevention as a source for Mauritania). Thus, these two awkward functional dependencies are defined and a choice has to be made between one or the other for normalisation procedures. The first is selected.

CountryManufacturer Entity

N/A (Attributes are primary keys and independent)

Based on these functional dependencies, we can conclude that the **Country**, **Manufacturer**, **Source**, **State**, **StateStats**, **CountryStats**, and **CountryManufacturer** relations are all in 2NF.

Finally, for the relations above to be in 3NF, all of their non-primary key attributes must only be functionally dependent on their primary keys. This remains true for all relations except **CountryStats**, because of the **SourceName** and **URL** attributes. **CountryStats** needs to be split into two relations.

CountryStats (<u>CountryName*, Date, AgeRange</u>, TotalVaccinations, PeopleVaccinated, PeopleFullyVaccinated, TotalBoosters, TotalVaccinations_p100, PeopleVaccinated_p100, PeopleFullyVaccinated_p100, DailyVaccinationsRaw, DailyVaccinations, DailyPeopleVaccinated, DailyVaccinations_pM, DailyPeopleVaccinated_p100, TotalBoosters_p100, SourceName*)

URLSource (<u>CountryName*, Date*, AgeRange*, SourceName</u>*, URL)

Final Relational Schema

Country (CountryName, CountryCode)

Manufacturer (ManufacturerName)

Source (SourceName)

ManufacturerStats (ManufacturerName*, CountryName*, Date, TotalVaccinations)

State (CountryName*, StateName)

StateStats (StateName*, Date, TotalVaccinations, TotalDistributed, PeopleVaccinated, PeopleFullyVaccinated, TotalVaccinations_p100, PeopleVaccinated_p100, PeopleFullyVaccinated_p100, Distributed_p100, DailyVaccinationsRaw, DailyVaccinations, DailyVaccinations_pM, ShareDosesUsed, TotalBoosters, TotalBoosters_p100)

CountryStats (CountryName*, Date, AgeRange, TotalVaccinations, PeopleVaccinated, PeopleFullyVaccinated, TotalBoosters, TotalVaccinations_p100, PeopleVaccinated_p100, PeopleFullyVaccinated_p100, DailyVaccinationsRaw, DailyVaccinations, DailyPeopleVaccinated, DailyVaccinations_pM, DailyPeopleVaccinated_p100, TotalBoosters_p100, SourceName*)

CountryManufacturer (CountryName*, ManufacturerName*)

URLSource (CountryName*, Date*, AgeRange*, SourceName*, URL)