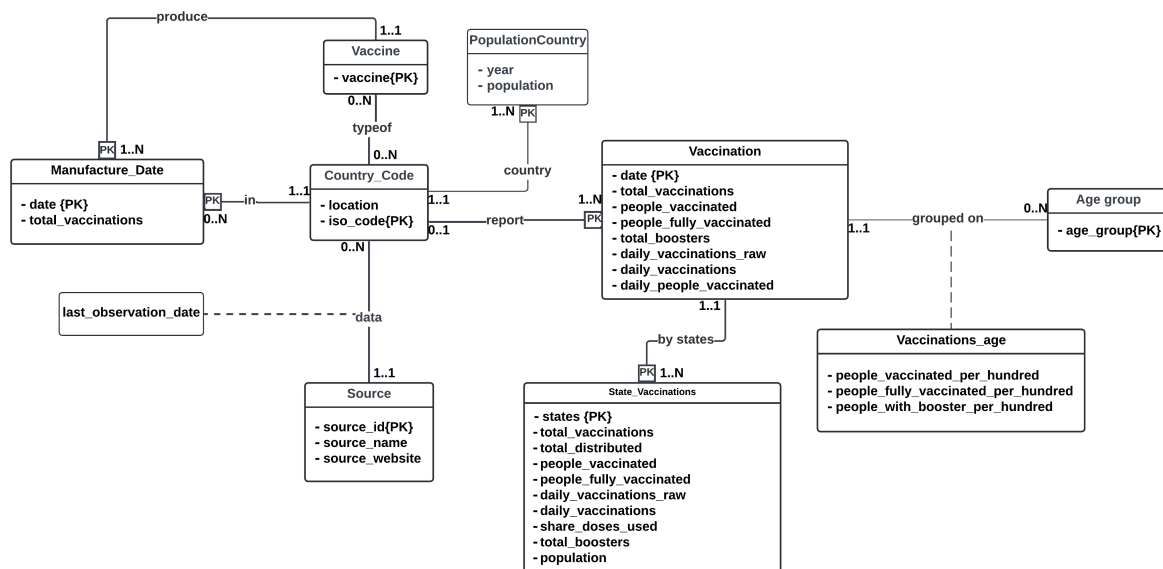# ISYS1055 - Database Design Project

S4068455 - Aswin kumar Sridhar

Part B: Designing the Database

ER diagram:

Assumption:

- *iso_code* is the root to derive all the data, thus all country code exist in Country_code schema and iso code is used as unique key to avoid anomalies

- date of the vaccination gives the detail about the vaccination and the age group vaccination and the states vaccination. Thus the dates which are not in vaccination record is considered as data inconsistency and excluded from this sub relations table.

- Since age_group is separated out as a strong entity, the vaccination_age will occur in attribute relationship as it does have unique key to identify

- Population by country table give the data to calculate all the derived time calculation in terms of per_hundred and per_million population attributes

- Not all country in Vaccination entity is been observed in Observation entity

- Since we don't have the real number on the vaccination_age entity, we are storing the derived values as given

- Last_observation_date is invoked between source and iso_code

- No state information is given except us_state_vaccination and the data of states cannot be derived from vaccination hence new entity with same attribute as vaccination per states

Normalisation Challenge:

- *location → iso_code, source_name → source_website* hence separated it to a strong entity

- *age_group* is redundant and might cause anomalies hence separated as a strong entity

- vaccination by country is a redundant information which can be derived from vaccination entity and removed.

- All the entity are dependent on *location* or *dates* hence made a weak entity.

- Auto increment *source_id* as same source name have different source_website

- Not all country in **Vaccination** entity is been observed in **Observation** entity

- "2023-05-10" is in state_vaccination but not in vaccination table thus new entry is created.

- Vaccine by age group for the dates where the record is not found in the vaccination.csv sheet per country is not considered as there is no data for these days in vaccinations.csv.

Database schema

1. **Strong entity**

    1. **Age_group(**<u>age_group</u>**)**

    2. **Vaccine(**<u>vaccine</u>**)**

    3. **Country(**<u>iso_code</u>, location)

    4. **Source(**<u>source_id</u>**,** source_name**,** source_website**)**

2. **Weak entity**

    1. **Manufacture_Date(**<u>iso_code*, date, vaccine*</u>, total_vaccinations)

    2. **Vaccination(**<u>iso_code*, date</u>, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_booster, daily_vaccination_raw, daily_vaccination, daily_people_vaccinated)

    3. **State_Vaccinations(**<u>iso_code*, date*,states</u>, total_vaccinations, total_distributed, people_vaccinated, people_fully_vaccinated,

daily_vaccination_raw, daily_vaccinations, share_doses_used, total_boosters, population)

    4. **Vaccination_Age**(iso_code*, date*, age_group*, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)

    5. **Population_Country(**iso_code*, year, population**)**

3. **One-One relationship:-** No 1-1 relationship between entity

4. **One-Many relationship:-**

    1. **Country(**iso_code*, location, source_id**,** last_observation_date**)**

5. **Many-Many relationship:-**

    1. **Country_Vaccine**(iso_code*, vaccine*)

6. **Multivalued attribute:-** No multivalued attribute as the vaccine type itself is created as an strong entity

7. **Degree of relationship:-** There is no degree of relationship exist

Final Schema:

- **Country_Vaccine**(iso_code*, vaccine*)

- **Country(**iso_code*, location, source_id**,** last_observation_date**)**

- **Manufacture_date**(iso_code*, date, vaccine*, total_vaccinations)

- **Vaccination**(iso_code, date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_booster, daily_vaccination_raw, daily_vaccination, daily_people_vaccinated)

- **State_Vaccinations**(iso_code*, date*,states, total_vaccinations, total_distributed, people_vaccinated, people_fully_vaccinated, daily_vaccination_raw, daily_vaccinations, share_doses_used, total_boosters, population)

- **Vaccination_age**(iso_code*, date*, age_group*, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)

- **Population_Country**(iso_code*, year, population)

- **Age_group(**age_group**)**

- **Vaccine(**vaccine**)**

- **Source(**source_id**,** source_name**,** source_website**)**