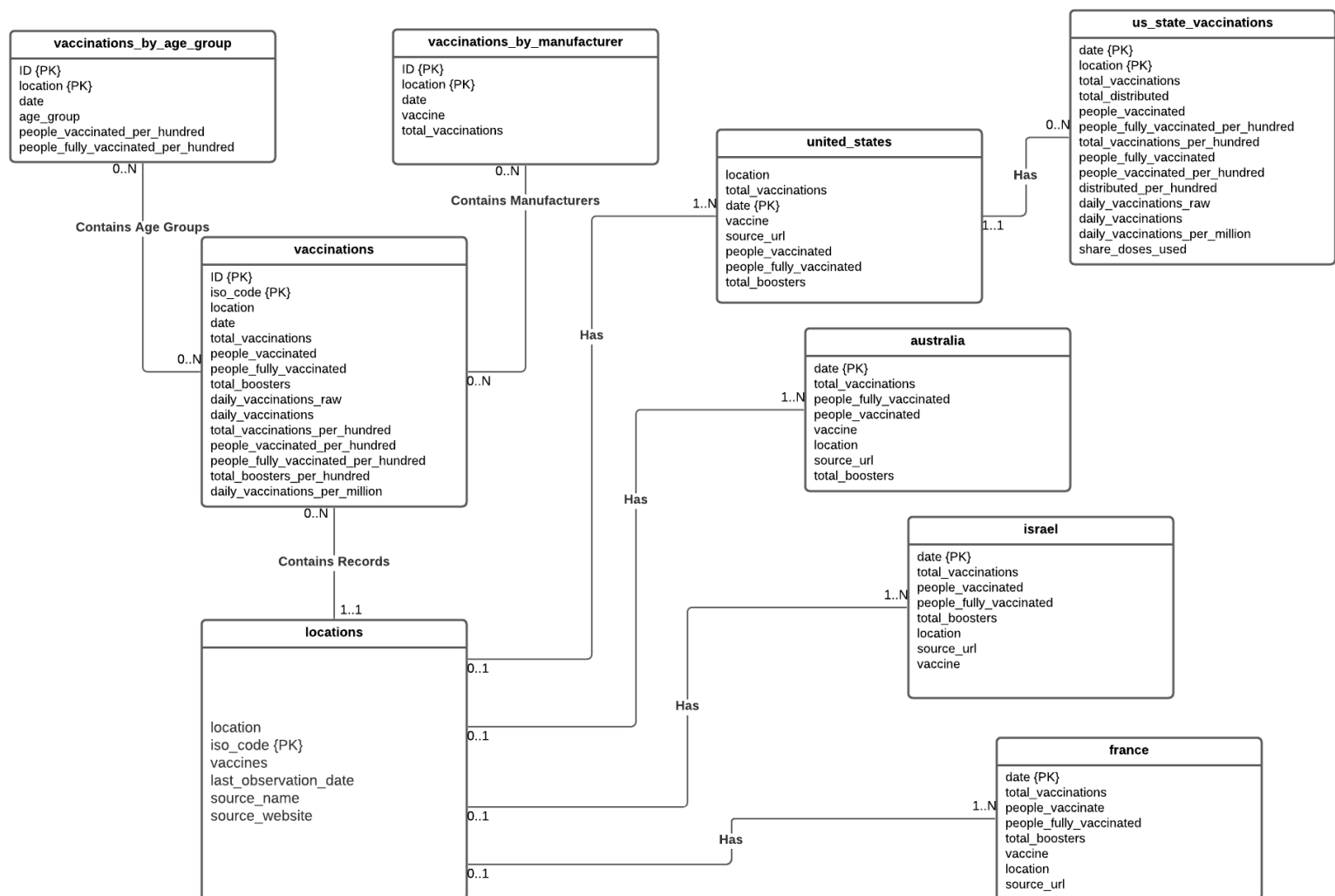


ER Diagram



Assumptions:

1. Not all countries have vaccinations records.
2. IDs were created for some relations to allow for identification.
3. "us_state_vaccinations" relation has "date" and "location" as composite keys because of the dates are only unique across any given state. They are not unique across the board. Therefore, this composite key allows for unique identification.
4. Dates are unique for "united_states", "australia", "Israel" and "france" relations, therefore chosen as primary keys.

Part B: Designing the Database

Initial schema:

vaccinations_by_age_group(ID, location, date, age_group, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred)

vaccinations_by_manufacturer(ID, location, date, vaccine, total_vaccinations)

vaccinations(ID, iso_code, location, date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, daily_vaccinations_raw, daily_vaccinations, total_vaccinations_per_hundred, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, total_boosters_per_hundred, daily_vaccinations_per_million)

locations(location, iso_code, vaccines, last_observation_date, source_name, source_website)

united_states(location, total_vaccinations, date, vaccine, source_url, people_vaccinated, people_fully_vaccinated, total_boosters)

us_state_vaccinations(date, location, 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)

australia(date, total_vaccinations, people_fully_vaccinated, people_vaccinated, vaccine, location, source_url, total_boosters)

israel(date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, location, source_url, vaccine)

france(date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, vaccine, location, source_url)

Step 1: Map Strong Entities

vaccinations_by_age_group(ID, location, date, age_group, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred)

vaccinations_by_manufacturer(ID, location, date, vaccine, total_vaccinations)

vaccinations(ID, iso_code, location, date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, daily_vaccinations_raw, daily_vaccinations, total_vaccinations_per_hundred, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, total_boosters_per_hundred, daily_vaccinations_per_million)

locations(location, iso_code, vaccines, last_observation_date, source_name, source_website)

united_states(location, total_vaccinations, date, vaccine, source_url, people_vaccinated, people_fully_vaccinated, total_boosters)

us_state_vaccinations(date, location, 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)

australia(date, total_vaccinations, people_fully_vaccinated, people_vaccinated, vaccine, location, source_url, total_boosters)

israel(date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, location, source_url, vaccine)

france(date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, vaccine, location, source_url)

Step 2: Map Weak Entities

None

Step 3: Map 1:1 Relationships

None

Step 4: Map 1:N Relationships

united_states(location, total_vaccinations, date, vaccine, source_url, people_vaccinated, people_fully_vaccinated, total_boosters, iso_code*)

us_state_vaccinations(date, location_state, 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, location_country*)

australia(date, total_vaccinations, people_fully_vaccinated, people_vaccinated, vaccine, location, source_url, total_boosters, iso_code*)

israel(date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, location, source_url, vaccine, iso_code*)

france(date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, vaccine, location, source_url, iso_code*)

Step 5: Map M:N Relationships

contains_age_groups(location*, iso_code*)

contains_manufacturers(location*, iso_code*)

Combining into one relation ->

contains_age_groups_manufacturers(location*, iso_code*)

Step 6: Multi-valued Attributes

None

Step 7: Map higher-degree relationships

None

Final Schema:

vaccinations_by_age_group(ID, location, date, age_group, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred)

vaccinations_by_manufacturer(ID, location, date, vaccine, total_vaccinations)

vaccinations(ID, iso_code, location, date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, daily_vaccinations_raw, daily_vaccinations, total_vaccinations_per_hundred, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, total_boosters_per_hundred, daily_vaccinations_per_million)

locations(location, iso_code, vaccines, last_observation_date, source_name, source_website)

united_states(location, total_vaccinations, date, vaccine, source_url, people_vaccinated, people_fully_vaccinated, total_boosters, iso_code*)

us_state_vaccinations(date, location_state, 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, location_country*)

australia(date, total_vaccinations, people_fully_vaccinated, people_vaccinated, vaccine, location, source_url, total_boosters, iso_code*)

israel(date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, location, source_url, vaccine, iso_code*)

france(date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters, vaccine, location, source_url, iso_code*)

contains_age_groups_manufacturers(location*, iso_code*)

Normalisation

First Normal Form (1NF)

All the relations satisfy the 1NF as every value within each tuple component is atomic.

Second Normal Form (2NF)

The relations with only a single primary key that are in 1NF are also in 2NF. This is because all non-primary key attributes are entirely dependent on the single primary key of their respective relations.

The relations that contain a composite key are also in 2NF. This is because all attributes are functionally dependent on the entire composite key.

Third Normal Form (3NF)

All relations are in 3NF as no non-primary key attributes are functionally dependent on any non-primary key attributes.

A relation named “contains_age_groups_manufacturers” is created to store each country’s official names and ISO codes so that other relations can refer to this relation for country names or ISO codes if required.