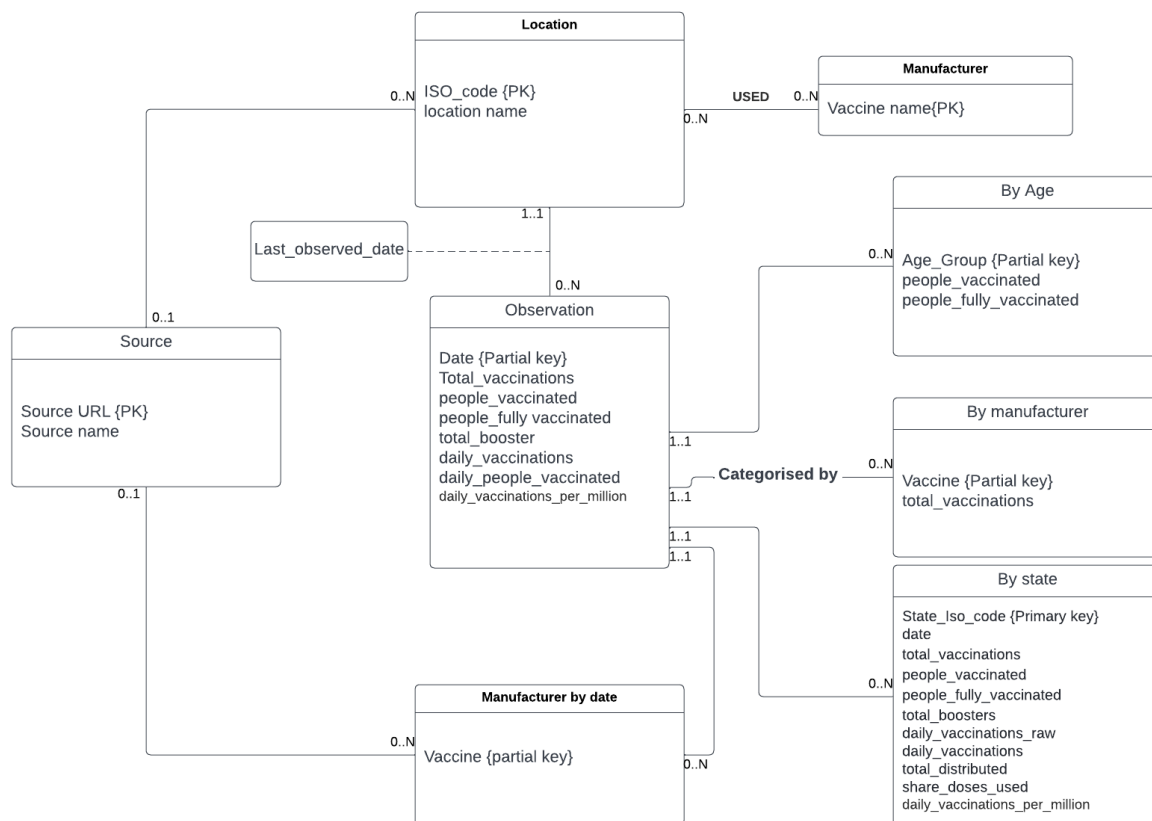# DBC2022S1 Assignment 3

Student name: Matthew John Bentham
Student ID: S3923076

PART B

ER diagram:

Normalisation Process:

<u>Initial design:</u>

```
Locations(Iso_code,location,vaccines[0..N],Last_observation_date,source_name,source_website)
```

```
Vaccinations(Iso_code,location,date,total_vaccinations,people_vacinated,people_fully_vaccinated
,total_booster daily_vaccinations_raw, daliy_vaccinations,
total_vaccinations_per_hundred,people_vaccinated_per_hundred,total_boosters_per_hundred,
daily_vaccinations_per_million,daily_people_vaccinated,daily_people_vaccinated_per_hundred)
```

```
Vaccinations_by_manufacturer(location,date,vaccine,total_vaccinations)
```

```
us_state_vaccinations(date,State,total_vaccinations,total_distributed,people_vacinated,people_f
ully_vaccinated_per_hundred,total_vaccinations_per_hundred,people_fully_vaccinated,people_vacci
nated_per_hundred,total_vaccinations_per_hundred,distributed_per_hundred,daily_vaccinations_raw
,daliy_vaccinations,daily_vaccinations_per_million,share_doses_used
total_booster,total_boosters_per_hundred)
```

```
United_states(location,date,vaccine,source_URL,total_vaccinations,people_vacinated,people_fully
_vaccinated,total_boosters)
China(location,date,vaccine,source_URL,total_vaccinations,people_vacinated,people_fully_vaccina
ted,total_boosters)
England(location,date,vaccine,source_URL,total_vaccinations,people_vacinated,people_fully_vacci
nated,total_boosters)
Australia(location,date,vaccine,source_URL,total_vaccinations,people_vacinated,people_fully_vac
cinated,total_boosters)
vaccinations_by_age_group(Location,date,age_group,people_vaccinated_per_hundred,people_full
ly_vaccinated_per_hundred,people_with_booster_per_hundred)
```

<u>Normalisation:</u>

NOTE: for future relations I have changed the primary key of location to Iso_code.This is purely for a more consistent and interpretable final schema, and to prevent any future mix ups (e.g. if a city is imported that has the same name as a country).

1. Location:
    1NF ? → Location is not in 1st normal form as it contains multivalued attributes in vaccines attribute

    ```
    Locations(Iso_code,location,Last_observation_date,source_name,source_URL)
    Locations2(Iso_code,vaccine_used)
    ```

    FD1: Iso_code → Lased_observed_date , source_name
    FD2: source_website → source_name

    2NF ? → yes
    3NF ?→ NO , source_name (a non_key field) is dependent on another non key field (source_website)

```
Locations(Iso_code,location,Last_observation_date, source_website*)
Locations2(Iso_code,vaccine_used)
Locations3(source_website,source_name)
```

2. Vaccination by manufacturer:
   In 3 NF

3. United_states/Australia/China/England:

```
Australia(Iso_code,date,vaccine_used,source_URL*,total_vaccinations,people_vac
inated,people_fully_vaccinated,total_boosters)

China(Iso_code,date,vaccine_used,source_URL*,total_vaccinations,people_vacinat
ed,people_fully_vaccinated,total_boosters)

United_states(Iso_code,date,vaccine_used,source_URL*,total_vaccinations,people
_vacinated,people_fully_vaccinated,total_boosters)

England(Iso_code,date,vaccine_used,source_URL*,total_vaccinations,people_vacin
ated,people_fully_vaccinated,total_boosters)
```

FD1: Iso_code , date → total_vaccinations , people_vacinated , people_fully_vaccinated, total_boosters

NF1 ? → No , Vaccines used is a multivalued attribute

```
Australia(Iso_code,date,source_URL*,total_vaccinations,people_vacinated,people
_fully_vaccinated,total_boosters)
Australia2(Iso_code,vaccine_used)
China(Iso_code,date,source_URL*,total_vaccinations,people_vacinated,people_ful
ly_vaccinated,total_boosters)
China2(Iso_code,vaccine_used)
United_states(Iso_code,date,source_URL*,total_vaccinations,people_vacinated,pe
ople_fully_vaccinated,total_boosters)
United_states2(Iso_code,vaccine_used)
England(Iso_code,date,source_URL*,total_vaccinations,people_vacinated,people_f
ully_vaccinated,total_boosters)
England2(Iso_code,vaccine_used)
```

4. Vaccinations

```
Vaccinations(Iso_code,date,total_vaccinations,people_vaccinated,people_fully_vacci
nated,total_booster daily_vaccinations_raw, daliy_vaccinations,
total_vaccinations_per_hundred,people_vaccinated_per_hundred,total_boosters_per_hu
ndred,daily_vaccinations_per_million,daily_people_vaccinated,daily_people_vaccinat
ed_per_hundred)
```

In 3NF

5. Us_state_vaccinations:

```
us_state_vaccinations(date,State,total_vaccinations,total_distributed,people_vacin
ated,people_fully_vaccinated_per_hundred,total_vaccinations_per_hundred,people_ful
ly_vaccinated,people_vaccinated_per_hundred,total_vaccinations_per_hundred,distrib
uted_per_hundred,daily_vaccinations_raw,daliy_vaccinations,daily_vaccinations_per_
million,share_doses_used
total_booster,total_boosters_per_hundred)
```

In 3NF

6. Vaccinations_by_age_group

In 3NF, however in order to match the rest of the data I will convert all variable per hundred to per person

Normalised Schemas:

In order remove redundancies and improve readability relations can be renamed and merged when necessary .

- Australia , france , us and israel1 can all be merged with vaccinations_global , and the vaccines used for each can be merged into one relation called `Vaccines_used_bydate`
- Because the USA data in vaccinations and vaccinations_us differ these relations cannot be merged even though they have the same primary keys
- Because the vaccine used data for each studied country does not line up with the data in vaccine_by_manufacturer these tables cannot be merged

```
Locations(Iso_code,Location,Last_observation_date, source_name*)
Vaccines_used Locations2(Iso_code,vaccine_used)
Datasources Locations3(source_website,source_name)

Australia(Iso_code,date,vaccine_used,total_vaccinations,people_vacinated,people_fully_v
accinated,total_boosters)

Australia2(Iso_code,date,vaccine_used)
France(Iso_code,date,vaccine_used,total_vaccinations,people_vacinated,people_fully_vacc
inated,total_boosters)
France2(Iso_code,date,vaccine_used)
United_states(Iso_code,date,vaccine_used,total_vaccinations,people_vacinated,people_ful
ly_vaccinated,total_boosters)
United_states2(Iso_code,date,vaccine_used)
Israel(Iso_code,date,vaccine_used,total_vaccinations,people_vacinated,people_fully_vacc
inated,total_boosters)
Israel2(Iso_code,date,vaccine_used)

Vaccines_used_bydate(Iso_code,date,vaccine_used)

Vaccinations_globalVaccinations(Iso_code,date,vaccines_used,total_vaccinations,people_v
accinated,people_fully_vaccinated,total_booster daily_vaccinations_raw,
daliy_vaccinations,total_vaccinations_per_hundred,people_vaccinated_per_hundred,total_b
oosters_per_hundred,daily_vaccinations_per_million,daily_people_vaccinated,daily_people
_vaccinated_per_hundred)

us_state_vaccinations(date,State,total_vaccinations,total_distributed,people_vacinated,
people_fully_vaccinated_per_hundred,total_vaccinations_per_hundred,people_fully_vaccina
```

```
ted,people_vaccinated_per_hundred,total_vaccinations_per_hundred,distributed_per_hundre
d,daily_vaccinations_raw,daliy_vaccinations,daily_vaccinations_per_million,share_doses_
used
total_booster,total_boosters_per_hundred)
Vaccinations_by_manufacturer(Iso_code,date,vaccine,total_vaccinations)
vaccinations_byage ~~Vaccinations_by_age_group~~ (Iso_code,date,age_group,
people_vaccinated,people_fully_vaccinated,people_with_booster)
```

REMOVING REDUDANT INFORMATION:

As population of each location at each date is not provided for both vaccinations_global and vaccinations_us I will keep the column daily_vaccinations_per_million which can then be used to calculate:

> total_vaccinations_per_hundred , people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred,, total_boosters_per_hundred, daily_people_vaccinated_per_hundred, distributed_per_hundred , last_observed_date

meaning these columns are redundant.

FINAL SCHEMA

```
Locations(Iso_code,last_observed_date,location, source_website*)
Vaccines_used(Iso_code,vaccine_used)
Datasources(source_website,source_name)
Vaccines_used_bydate(Iso_code,date,vaccine_used)
Vaccinations_by_manufacturer(Iso_code,date,vaccine,total_vaccinations)
Vaccinations_global(Iso_code,date,vaccines_used,total_vaccinations,people_vacc
inated,people_fully_vaccinated,total_booster, daily_vaccinations_raw,
daliy_vaccinations ,daily_people_vaccinated, daily_vaccinations_per_million)
Vaccinations_us(Iso_code,date,vaccines_used,total_vaccinations,people_vaccinat
ed,people_fully_vaccinated,total_booster, daily_vaccinations_raw,
daliy_vaccinations ,daily_people_vaccinated, total_distributed,
share_doses_used, daily_vaccinations_per_million)
vaccinations_byage(Iso_code,date,age_group,people_vaccinated,people_fully_vacc
inated,people_with_booster)
```