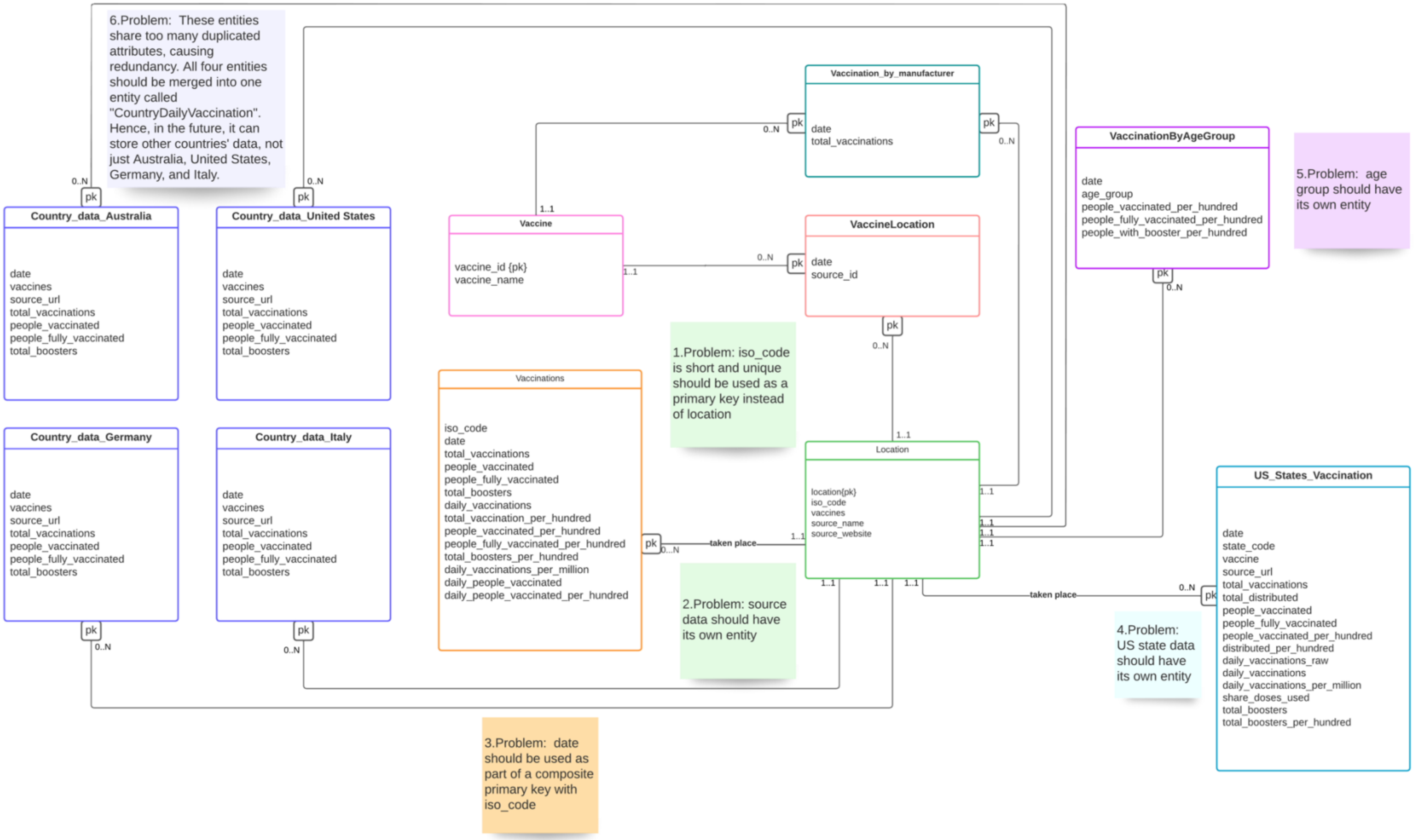
# A ER diagram for a Global Database of COVID-19 Vaccinations
# (Initial Diagram & Problem)



6.Problem: These entities share too many duplicated attributes, causing redundancy. All four entities should be merged into one entity called "CountryDailyVaccination". Hence, in the future, it can store other countries' data, not just Australia, United States, Germany, and Italy.

**Country_data_Australia**
date
vaccines
source_url
total_vaccinations
people_vaccinated
people_fully_vaccinated
total_boosters

**Country_data_United States**
date
vaccines
source_url
total_vaccinations
people_vaccinated
people_fully_vaccinated
total_boosters

**Country_data_Germany**
date
vaccines
source_url
total_vaccinations
people_vaccinated
people_fully_vaccinated
total_boosters

**Country_data_Italy**
date
vaccines
source_url
total_vaccinations
people_vaccinated
people_fully_vaccinated
total_boosters

**Vaccination_by_manufacturer**
date
total_vaccinations

**Vaccine**
vaccine_id {pk}
vaccine_name

**VaccineLocation**
date
source_id

**VaccinationByAgeGroup**
date
age_group
people_vaccinated_per_hundred
people_fully_vaccinated_per_hundred
people_with_booster_per_hundred

5.Problem: age group should have its own entity

1.Problem: iso_code is short and unique should be used as a primary key instead of location

Vaccinations
iso_code
date
total_vaccinations
people_vaccinated
people_fully_vaccinated
total_boosters
daily_vaccinations
total_vaccination_per_hundred
people_vaccinated_per_hundred
people_fully_vaccinated_per_hundred
total_boosters_per_hundred
daily_vaccinations_per_million
daily_people_vaccinated
daily_people_vaccinated_per_hundred

Location
location{pk}
iso_code
vaccines
source_name
source_website

2.Problem: source data should have its own entity

**US_States_Vaccination**
date
state_code
vaccine
source_url
total_vaccinations
total_distributed
people_vaccinated
people_fully_vaccinated
people_vaccinated_per_hundred
distributed_per_hundred
daily_vaccinations_raw
daily_vaccinations
daily_vaccinations_per_million
share_doses_used
total_boosters
total_boosters_per_hundred

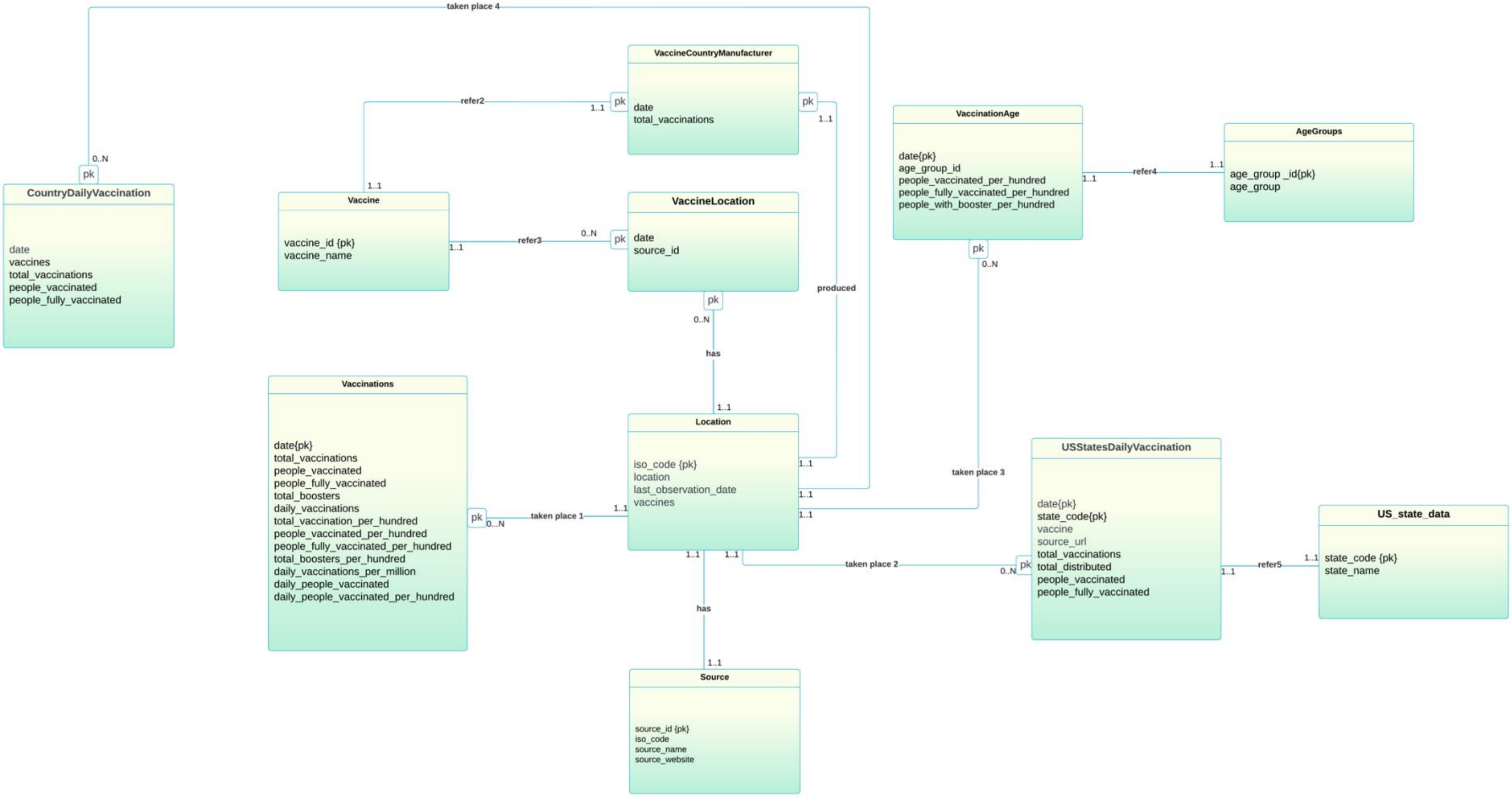
4.Problem: US state data should have its own entity

3.Problem: date should be used as part of a composite primary key with iso_code

### Problems found in initial design and changes made to the diagram

1. iso_code is short and unique should be used as a primary key instead of location
2. Source data should have its own entity
3. Date should be used as part of a composite primary key with iso_code
4. US state data should have its own entity
5. Age group should have its own entity
6. These entities share too many duplicated attributes, causing redundancy. All four attributes should be merged into one entity called "CountryDailyVaccination". Hence, in the future, it can store other countries' data, not just Australia, United States, Germany, and Italy.

# A ER diagram for a Global Database of COVID-19 Vaccinations (Final Diagram)



## Assumptions

1. Each source can come from only one location
2. Zero or many vaccinations can be taken place in each location. However, each vaccination can only be taken place in one location
3. Each daily vaccination of each country can only be taken place in many locations, whereas each location can have only 1 country daily vaccination
4. Each vaccination with age record can only refer to only one age_group_id
5. Each US States Daily Vaccination can only refer to only one state_code
6. Each vaccine can come from only one country manufacturer or one location

## Mapping ER Model to Relational Model

Step 1: Strong Entities

Location (iso_code, location, last_observation_date, vaccines)

Vaccine (vaccine_id, vaccine_name)

AgeGroup (age_group_id, age_group)

US_state_data (state_code, state_name)

Source (iso_code*, source_id, source_name, source_website)

Step 2: Weak Entities

CountryDailyVaccination (iso_code*, date, vaccines, total_vaccinations, people_vaccinated, people_fully_vaccinated)

VaccineCountryManufacturer (iso_code*, vaccine_id*, date, total_vaccinations)

VaccineLocation (iso_code*, vaccine_id*, date, source_id)

VaccinationAge (iso_code*, date, age_group, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred , people_with_booster_per_hundred)

Vaccinations (iso_code*, date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters , daily_vaccinations, total_vaccination_per_hundred, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred , total_boosters_per_hundred, daily_vaccinations_per_million, daily_people_vaccinated, daily_people_vaccinated_per_hundred)

USStatesDailyVaccination (iso_code*, date, state_code, vaccine, source_url, total_vaccinations, total_distributed , people_vaccinated, people_fully_vaccinated).

Step 3: One-to-one Relationships

Source (source_id, source_name, source_website, iso_code*, location, last_observation_date, vaccines)
US_state_data (state_code, state_name, iso_code*, date, state_code, vaccine, source_url, total_vaccinations, total_distributed , people_vaccinated, people_fully_vaccinated)
AgeGroup (age_group_id, age_group, iso_code*, date, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)

Step 4: One-to-many Relationships
Nothing to do

Step 5: Many-to-many Relationships

Nothing to do

Step 6: Multi-valued Attributes

Nothing to do

Step 7: Higher-degree Relationships
Nothing to do

## Relational Database Schema before normalisation

| | |
|---|---|
| Location | (iso_code, location, last_observation_date, vaccines) |
| Vaccine | (vaccine_id, vaccine_name) |
| AgeGroup | (age_group_id, age_group) |

US_state_data (state_code, state_name)
Source (source_id, iso_code*, source_name, source_website)
CountryDailyVaccination (iso_code*, date, vaccines, total_vaccinations, people_vaccinated, people_fully_vaccinated)
VaccineCountryManufacturer (iso_code*, vaccine_id*, date, total_vaccinations)
VaccineLocation (iso_code*, vaccine_id*, date, source_id)
VaccinationAge (iso_code*, date, age_group, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred , people_with_booster_per_hundred)
Vaccinations (iso_code*, date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters , daily_vaccinations, total_vaccination_per_hundred, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred , total_boosters_per_hundred, daily_vaccinations_per_million, daily_people_vaccinated, daily_people_vaccinated_per_hundred)
USStatesDailyVaccination (iso_code*, date, state_code, vaccine, source_url, total_vaccinations, total_distributed , people_vaccinated, people_fully_vaccinated)
Source (source_id, source_name, source_website, iso_code*, location, last_observation_date, vaccines)
US_state_data (state_code, state_name, iso_code*, date, state_code, vaccine, source_url, total_vaccinations, total_distributed , people_vaccinated, people_fully_vaccinated)
AgeGroup (age_group_id, age_group, iso_code*, date, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred , people_with_booster_per_hundred)

## Normalisation challenges

- ### Functional Dependencies

**1.Location**
Location (iso_code, location, last_observation_date, vaccines)

**FDs:**
FD1: iso_code → location, last_observation_date, vaccines

Only iso_code are the attributes that has not been determined by any other attributes.

The correct primary key is <iso_code>

**Normal Form:**

This relation is in 1NF because there are no multi-valued attributes.

This relation is in 2NF because the primary key is simple primary key, as such there is no way it is not in 2NF. To clarify, there is no partial functional dependency here.

This relation is in 3NF because there is no transitional dependency here.

**Decomposition:**

Location: iso_code → location, last_observation_date, vaccines

**Final Schema:**

(iso_code, location, last_observation_date, vaccines)

**The highest normal form for this relation is 3NF.**

**2.Vaccine**

Vaccine (vaccine_id, vaccine_name)

**FDs:**

FD1: vaccine_id→ vaccine_name

Only vaccine_id are the attributes that has not been determined by any other attributes.

The correct primary key is < vaccine_id >

**Normal Form:**

This relation is in 1NF because there are no multi-valued attributes.

This relation is in 2NF because the primary key is simple primary key, as such there is no way it is not in 2NF. To clarify, there is no partial functional dependency here.

This relation is in 3NF because there is no transitional dependency here.

**The highest normal form for this relation is 3NF.**

**Decomposition:**
Vaccine: vaccine_id→ vaccine_name
**Final Schema:**
Vaccine (vaccine_id, vaccine_name)

**3.AgeGroup**
AgeGroup (age_group_id, age_group)
**FDs:**
FD1: age_group_id → age_group

Only age_group_id are the attributes that has not been determined by any other attributes.
The correct primary key is < age_group_id >

**Normal Form:**
This relation is in 1NF because there are no multi-valued attributes.
This relation is in 2NF because the primary key is simple primary key, as such there is no way it is not in 2NF. To clarify, there is no partial functional dependency here.
This relation is in 3NF because there is no transitional dependency here.

**The highest normal form for this relation is 3NF.**

**Decomposition:**
AgeGroup: age_group_id → age_group
**Final Schema:**
AgeGroup (age_group_id, age_group)

**4. US_state_data**

US_state_data (state_code, state_name)

**FDs:**

FD1: state_code → state_name

Only state_code are the attributes that has not been determined by any other attributes.

The correct primary key is < state_code >

**Normal Form:**

This relation is in 1NF because there are no multi-valued attributes.

This relation is in 2NF because the primary key is simple primary key, as such there is no way it is not in 2NF. To clarify, there is no partial functional dependency here.

This relation is in 3NF because there is no transitional dependency here.

**The highest normal form for this relation is 3NF.**

**Decomposition:**

US_state_data: state_code → state_name

**Final Schema:**

US_state_data (state_code, state_name)

## 5. Source

Source (source_id, iso_code*,_source_name, source_website)

**FDs:**

FD1: source_id → iso_code, source_name, source_website

Only source_id are the attributes that has not been determined by any other attributes.

The correct primary key is < source_id >

**Normal Form:**

This relation is in 1NF because there are no multi-valued attributes.

This relation is in 2NF because all non-primary key attributes are dependent on the composite primary key (iso_code, source_id)
This relation is in 3NF because there is no transitional dependency here.

**The highest normal form for this relation is 3NF.**

**Decomposition:**
Source: source_id → iso_code, source_name, source_website
**Final Schema:**
(source_id, iso_code*, source_name, source_website)

### 6. CountryDailyVaccination

CountryDailyVaccination (iso_code*, date, vaccines, total_vaccinations, people_vaccinated, people_fully_vaccinated)

**FDs:**
FD1: iso_code → date, vaccines, total_vaccinations, people_vaccinated, people_fully_vaccinated
Only iso_code is the attributes that has not been determined by any other attributes.
The correct primary key is < iso_code >

**Normal Form:**
This relation is in 1NF because there are no multi-valued attributes.
This relation is in 2NF because all non-primary key attributes are dependent on the composite primary key (iso_code, vaccine_id)
This relation is in 3NF because there is no transitional dependency here.

**The highest normal form for this relation is 3NF.**

**Decomposition:**

CountryDailyVaccination: iso_code → date, vaccines, total_vaccinations, people_vaccinated, people_fully_vaccinated

**Final Schema:**

CountryDailyVaccination (iso_code*, date, vaccines, total_vaccinations, people_vaccinated, people_fully_vaccinated)

## 7. VaccineCountryManufacture

VaccineCountryManufacturer (iso_code*, vaccine_id*, date, total_vaccinations)

**FDs:**

FD1: iso_code, vaccine_id, date → total_vaccinations

Only iso_code, vaccine_id, date are the attributes that has not been determined by any other attributes.

The correct primary key is < iso_code, vaccine_id, date >

**Normal Form:**

This relation is in 1NF because there are no multi-valued attributes.

This relation is in 2NF because all non-primary key attributes are dependent on the composite primary key (iso_code, vaccine_id, date)

This relation is in 3NF because there is no transitional dependency here.

**The highest normal form for this relation is 3NF.**

**Decomposition:**

VaccineCountryManufacturer: iso_code, vaccine_id, date → total_vaccinations

**Final Schema:**

VaccineCountryManufacturer (iso_code*, vaccine_id*, date, total_vaccinations)

## 8. VaccineLocation

VaccineLocation (iso_code*, vaccine_id*, date, source_id)

**FDs:**

FD1: iso_code, vaccine_id, date → source_id

Only iso_code, vaccine_id, date are the attributes that has not been determined by any other attributes.

The correct primary key is < iso_code, vaccine_id, date >

**Normal Form:**

This relation is in 1NF because there are no multi-valued attributes.

This relation is in 2NF because all non-primary key attributes are dependent on the composite primary key (iso_code, vaccine_id, date)

This relation is in 3NF because there is no transitional dependency here.

**The highest normal form for this relation is 3NF.**

**Decomposition:**

VaccineLocation: iso_code, vaccine_id, date → source_id

**Final Schema:**

VaccineLocation (iso_code*, vaccine_id*, date, source_id)

### 9. VaccinationAge

VaccinationAge (iso_code*, date, age_group, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred , people_with_booster_per_hundred)

**FDs:**

FD1: iso_code, date → age_group, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred , people_with_booster_per_hundred

Only iso_code and date are the attributes that has not been determined by any other attributes.

The correct primary key is < iso_code, date >

**Normal Form:**
This relation is in 1NF because there are no multi-valued attributes.
This relation is in 2NF because all non-primary key attributes are dependent on the composite primary key (iso_code, date)
This relation is in 3NF because there is no transitional dependency here.

**The highest normal form for this relation is 3NF.**

**Decomposition:**
VaccinationAge: iso_code, date → age_group, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred
, people_with_booster_per_hundred
**Final Schema:**
VaccinationAge (iso_code*, date, age_group, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred
, people_with_booster_per_hundred)

**10. Vaccinations**
Vaccinations (iso_code*, date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters
, daily_vaccinations, total_vaccination_per_hundred, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred
, total_boosters_per_hundred, daily_vaccinations_per_million, daily_people_vaccinated,
daily_people_vaccinated_per_hundred)

**FDs:**
FD1: iso_code, date → total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters
, daily_vaccinations, total_vaccination_per_hundred, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred
, total_boosters_per_hundred, daily_vaccinations_per_million, daily_people_vaccinated,
daily_people_vaccinated_per_hundred
Only iso_code and date are the attributes that has not been determined by any other attributes.
The correct primary key is < iso_code, date >

**Normal Form:**
This relation is in 1NF because there are no multi-valued attributes.
This relation is in 2NF because all non-primary key attributes are dependent on the composite primary key (iso_code, date)
This relation is in 3NF because there is no transitional dependency here.

**The highest normal form for this relation is 3NF.**

**Decomposition:**
Vaccinations: iso_code, date → total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters , daily_vaccinations, total_vaccination_per_hundred, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred , total_boosters_per_hundred, daily_vaccinations_per_million, daily_people_vaccinated, daily_people_vaccinated_per_hundred
**Final Schema:**
Vaccinations (iso_code*, date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters , daily_vaccinations, total_vaccination_per_hundred, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred , total_boosters_per_hundred, daily_vaccinations_per_million, daily_people_vaccinated, daily_people_vaccinated_per_hundred)

**11. USStatesDailyVaccination**
USStatesDailyVaccination (iso_code*, date, state_code, vaccine, source_url, total_vaccinations, total_distributed , people_vaccinated, people_fully_vaccinated)

**FDs:**
FD1: iso_code, source_id, state_code → vaccine, source_url, total_vaccinations, total_distributed , people_vaccinated, people_fully_vaccinated.
Only iso_code and source_id are the attributes that has not been determined by any other attributes.
The correct primary key is < iso_code, source_id, state_code >

**Normal Form:**

This relation is in 1NF because there are no multi-valued attributes.
This relation is in 2NF because all non-primary key attributes are dependent on the composite primary key (iso_code, source_id)
This relation is in 3NF because there is no transitional dependency here.

**The highest normal form for this relation is 3NF.**

**Decomposition:**
USStatesDailyVaccination: iso_code, source_id, state_code → vaccine, source_url, total_vaccinations, total_distributed , people_vaccinated, people_fully_vaccinated.
**Final Schema:**
USStatesDailyVaccination (iso_code*, date, state_code, vaccine, source_url, total_vaccinations, total_distributed , people_vaccinated, people_fully_vaccinated)

**12. Source**
Source (source_id, source_name, source_website, iso_code*, location, last_observation_date, vaccines)

**FDs:**
FD1: iso_code, source_id → source_name, source_website, location, last_observation_date, vaccines
Only iso_code and source_id are the attributes that has not been determined by any other attributes.
The correct primary key is < iso_code, source_id >

**Normal Form:**
This relation is in 1NF because there are no multi-valued attributes.
This relation is in 2NF because all non-primary key attributes are dependent on the composite primary key (iso_code, source_id)
This relation is in 3NF because there is no transitional dependency here.

**The highest normal form for this relation is 3NF.**

**Decomposition:**

Source: iso_code, source_id → source_name, source_website, location, last_observation_date, vaccines

**Final Schema:**

Source (source_id, source_name, source_website, iso_code*, location, last_observation_date, vaccines)

**13. US_state_data**

US_state_data (state_code, state_name, iso_code*, date, state_code, vaccine, source_url, total_vaccinations, total_distributed, people_vaccinated, people_fully_vaccinated)

**FDs:**

FD1: iso_code, state_code → state_code, date, state_code, vaccine, source_url, total_vaccinations, total_distributed , people_vaccinated, people_fully_vaccinated

Only iso_code and state_code are the attributes that has not been determined by any other attributes.

The correct primary key is < iso_code, state_code >

**Normal Form:**

This relation is in 1NF because there are no multi-valued attributes.

This relation is in 2NF because all non-primary key attributes are dependent on the composite primary key (iso_code, state_code)

This relation is in 3NF because there is no transitional dependency here.

**The highest normal form for this relation is 3NF.**

**Decomposition:**

US_state_data: iso_code, state_code → state_code, date, state_code, vaccine, source_url, total_vaccinations, total_distributed , people_vaccinated, people_fully_vaccinated

**Final Schema:**

US_state_data (state_code, state_name, iso_code*, date, state_code, vaccine, source_url, total_vaccinations, total_distributed, people_vaccinated, people_fully_vaccinated)

**14. AgeGroup**

AgeGroup (age_group_id, age_group, iso_code*, date, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)

**FDs:**

FD1: iso_code, age_group _id → age_group, date, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred

Only iso_code and age_group _id are the attributes that has not been determined by any other attributes.

The correct primary key is < iso_code, age_group _id >

**Normal Form:**

This relation is in 1NF because there are no multi-valued attributes.

This relation is in 2NF because all non-primary key attributes are dependent on the composite primary key (iso_code, age_group _id)

This relation is in 3NF because there is no transitional dependency here.

**The highest normal form for this relation is 3NF.**

**Decomposition:**

AgeGroup: iso_code, age_group _id → age_group, date, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred

**Final Schema:**

AgeGroup (iso_code*,age_group_id, age_group, date, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)

## Relational Database Schema after normalisation

Location (iso_code, location, last_observation_date, vaccines)
Vaccine (vaccine_id, vaccine_name)
AgeGroup (age_group_id, age_group)
US_state_data (state_code, state_name)
Source (source_id, iso_code*, source_name, source_website)
CountryDailyVaccination (iso_code*,date, vaccines, total_vaccinations, people_vaccinated, people_fully_vaccinated)
VaccineCountryManufacturer (iso_code*, vaccine_id*, date, total_vaccinations)
VaccineLocation (iso_code*, vaccine_id*, date, source_id)
VaccinationAge (iso_code*, date, age_group, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred , people_with_booster_per_hundred)
Vaccinations (iso_code*, date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters , daily_vaccinations, total_vaccination_per_hundred, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred , total_boosters_per_hundred, daily_vaccinations_per_million, daily_people_vaccinated, daily_people_vaccinated_per_hundred)
USStatesDailyVaccination (iso_code*, date, state_code, vaccine, source_url, total_vaccinations, total_distributed , people_vaccinated, people_fully_vaccinated)
~~Source (source_id, source_name, source_website, iso_code*, location, last_observation_date, vaccines)~~
~~US_state_data (state_code, state_name, iso_code*, date, state_code, vaccine, source_url, total_vaccinations, total_distributed, people_vaccinated, people_fully_vaccinated)~~
~~AgeGroup (iso_code*, age_group_id, age_group, date, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)~~

Some schemas, which are redundant or can be derived from other schemas, will deleted to reduce redundancy.

## Final Relational Database Schema

**Location** (iso_code, location, last_observation_date, vaccines)
**Vaccine** (vaccine_id, vaccine_name)
**AgeGroup** (age_group_id, age_group)
**US_state_data** (state_code, state_name)
**Source** (source_id, iso_code*, source_name, source_website)
**CountryDailyVaccination** (iso_code*, date, vaccines, total_vaccinations, people_vaccinated, people_fully_vaccinated)
**VaccineCountryManufacturer** (iso_code*, vaccine_id*, date, total_vaccinations)
**VaccineLocation** (iso_code*, vaccine_id*, date, source_id)
**VaccinationAge** (iso_code*, date, age_group, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred
, people_with_booster_per_hundred)
**Vaccinations** (iso_code*, date, total_vaccinations, people_vaccinated, people_fully_vaccinated, total_boosters
, daily_vaccinations, total_vaccination_per_hundred, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred
, total_boosters_per_hundred, daily_vaccinations_per_million, daily_people_vaccinated,
daily_people_vaccinated_per_hundred)
**USStatesDailyVaccination** (iso_code*, date, state_code, vaccine, source_url, total_vaccinations, total_distributed ,
people_vaccinated, people_fully_vaccinated)

## Appendix

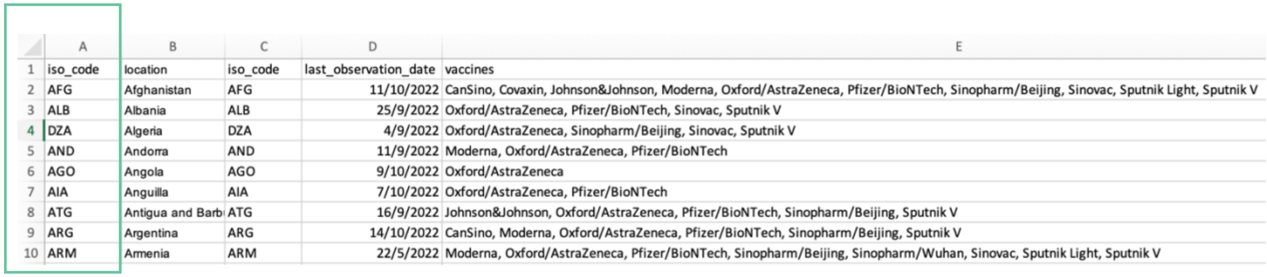
## Other challenge found in creating database in SQLiteStudio

- **Prepare .csv file before importing to the newly created database**

In this section, we need to clean and rearrange the file to reflect the ER diagram and prepare data for other purposes in the future. Since iso_code is short and unique, this primary key will be shared by many weak entities. In some case, we will use VLOOKUP function to build new column in Excel file by looking up values from other tables. Please be noted that INDEX function can return the value after matching as well although I am more familiar with VLOOKUP function.

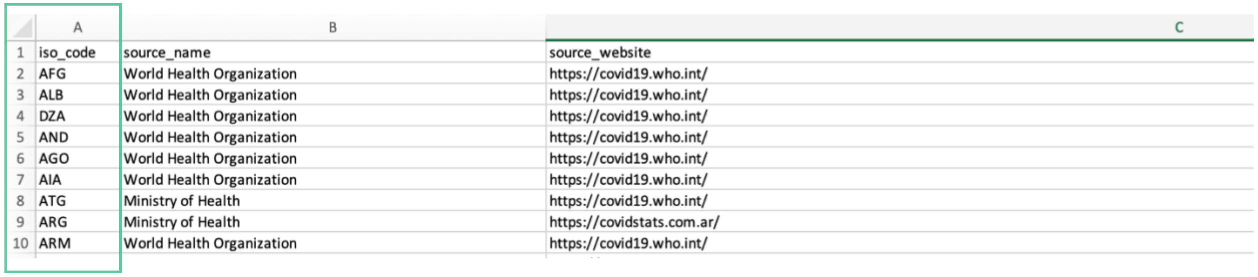
Below are examples of file cleaning and handling the template

1.Location.csv – move the iso_code column to the first column since we use iso_code as a primary key for location table



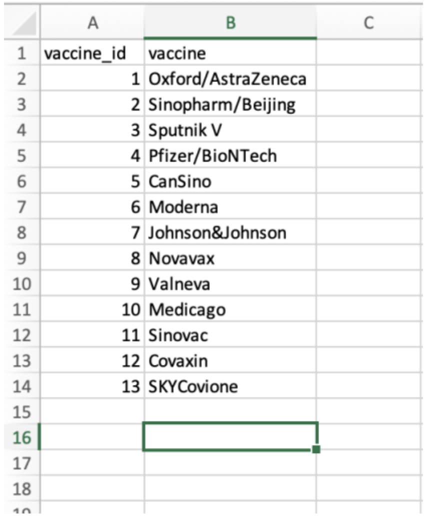
| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | iso_code | location | iso_code | last_observation_date | vaccines |
| 2 | AFG | Afghanistan | AFG | 11/10/2022 | CanSino, Covaxin, Johnson&Johnson, Moderna, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing, Sinovac, Sputnik Light, Sputnik V |
| 3 | ALB | Albania | ALB | 25/9/2022 | Oxford/AstraZeneca, Pfizer/BioNTech, Sinovac, Sputnik V |
| 4 | DZA | Algeria | DZA | 4/9/2022 | Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac, Sputnik V |
| 5 | AND | Andorra | AND | 11/9/2022 | Moderna, Oxford/AstraZeneca, Pfizer/BioNTech |
| 6 | AGO | Angola | AGO | 9/10/2022 | Oxford/AstraZeneca |
| 7 | AIA | Anguilla | AIA | 7/10/2022 | Oxford/AstraZeneca, Pfizer/BioNTech |
| 8 | ATG | Antigua and Barb | ATG | 16/9/2022 | Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing, Sputnik V |
| 9 | ARG | Argentina | ARG | 14/10/2022 | CanSino, Moderna, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing, Sputnik V |
| 10 | ARM | Armenia | ARM | 22/5/2022 | Moderna, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing, Sinopharm/Wuhan, Sinovac, Sputnik Light, Sputnik V |

2.Source.csv – add iso_code to the first column of excel file since we will use this column as a composite primary key (iso_code, source_name) for source table. Also, iso_code will be used as a foreign key for other tables



| | A | B | C |
|---|---|---|---|
| 1 | iso_code | source_name | source_website |
| 2 | AFG | World Health Organization | https://covid19.who.int/ |
| 3 | ALB | World Health Organization | https://covid19.who.int/ |
| 4 | DZA | World Health Organization | https://covid19.who.int/ |
| 5 | AND | World Health Organization | https://covid19.who.int/ |
| 6 | AGO | World Health Organization | https://covid19.who.int/ |
| 7 | AIA | World Health Organization | https://covid19.who.int/ |
| 8 | ATG | Ministry of Health | https://covid19.who.int/ |
| 9 | ARG | Ministry of Health | https://covidstats.com.ar/ |
| 10 | ARM | World Health Organization | https://covid19.who.int/ |

3.Vaccine.csv – remove duplicate values for vaccine name and assign vaccine_id to each vaccine_name. We we use this table for the future reference of vaccine_id.



| vaccine_id | vaccine |
|---|---|
| 1 | Oxford/AstraZeneca |
| 2 | Sinopharm/Beijing |
| 3 | Sputnik V |
| 4 | Pfizer/BioNTech |
| 5 | CanSino |
| 6 | Moderna |
| 7 | Johnson&Johnson |
| 8 | Novavax |
| 9 | Valneva |
| 10 | Medicago |
| 11 | Sinovac |
| 12 | Covaxin |
| 13 | SKYCovione |