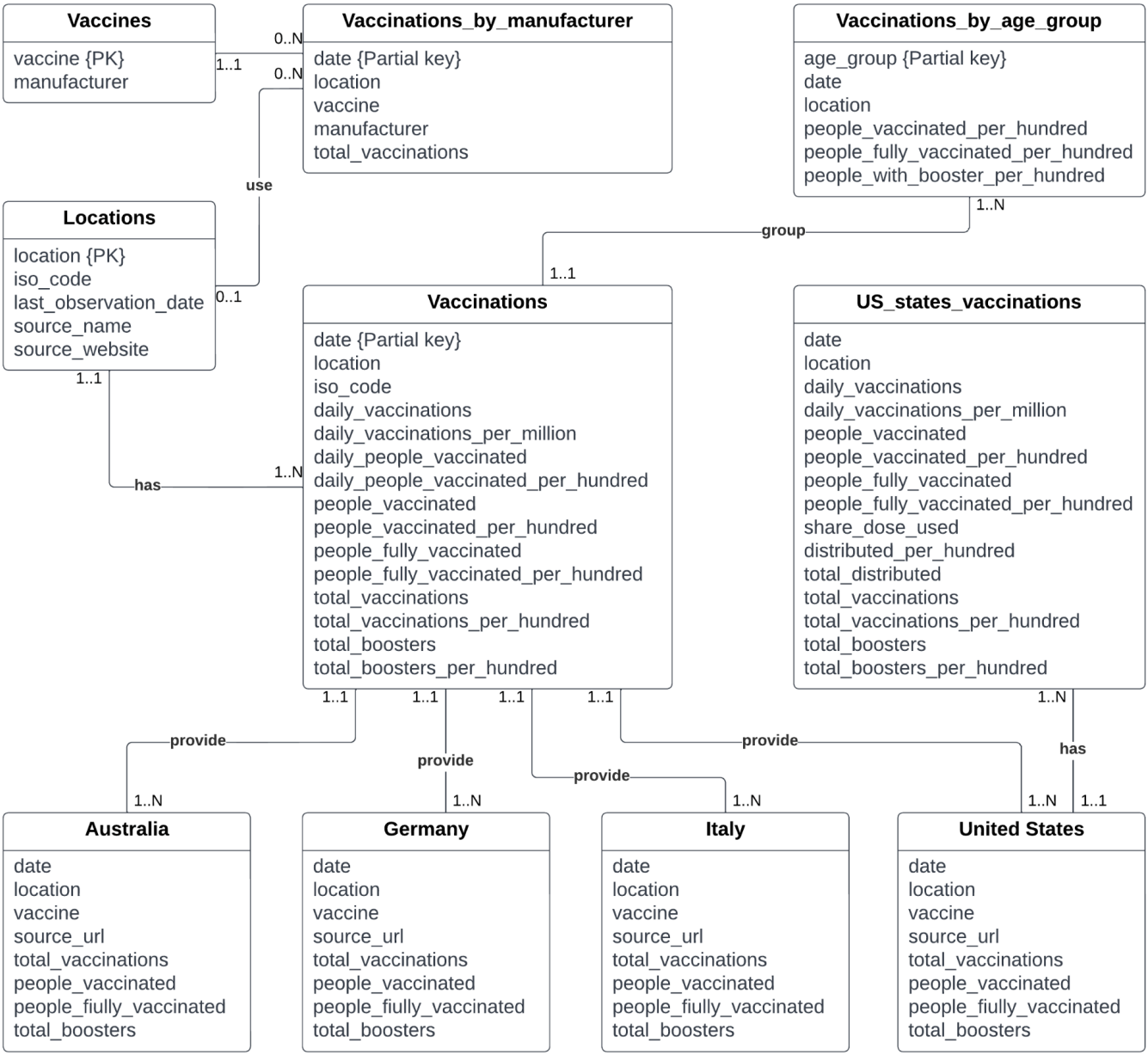
SUKHUM BOONDECHARAK - S3940976
Database Concept (ISYS1055): Final Project

Part B: Designing the Database

ER Diagram



Assumptions:

1. vaccines column in locations.csv is a list of vaccines in vaccinations_by_manufacturer.csv
2. location column in vaccinations_by_manufacturer can be found in location column in locations.csv
3. source_url column in each country file can be found in source_website column in locations.csv
4. daily_vaccinations_raw column in both vaccinations.csv and US_states_vaccinations can be ignored.
5. Only the United States has divided vaccination records into states.
6. location column in US_states_vaccinations is the list of states in the US, and not the country

FDs:

Locations (location, iso_code, last_observation_date, source_name, source_website)

location → iso_code, last_observation_date, source_name

source_name → source_website

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

date, location → vaccine, daily_vaccinations, daily_vaccinations_per_million, daily_people_vaccinated, daily_people_vaccinated_per_hundred, people_vaccinated, people_vaccinated_per_hundred, people_fully_vaccinated, people_fully_vaccinated_per_hundred, total_vaccinations, total_vaccinations_per_hundred, total_boosters, total_boosters_per_hundred

location → iso_code

Vaccines (vaccine, manufacturer)

Vaccine → manufacturer

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

date, location, vaccine → total_vaccinations

Vaccinations_by_age_group (age_group, date*, location*, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)

age_group, date, location → people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred

Australia (date*, location*, vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters)

date, location → vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters

Germany (date*, location*, vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters)

date, location → vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters

Italy (date*, location*, vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters)

date, location → vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters

United States (date*, location*, vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters)

date, location → vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters

US_states_vaccinations (date*, location*, daily_vaccinations, daily_vaccinations_per_million, people_vaccinated, people_vaccinated_per_hundred, people_fully_vaccinated, people_fully_vaccinated_per_hundred, share_dose_used, distributed_per_hundred, total_distributed, total_vaccinations, total_vaccinations_per_hundred, total_boosters, total_boosters_per_hundred)

date, location → daily_vaccinations, daily_vaccinations_per_million, people_vaccinated, people_vaccinated_per_hundred, people_fully_vaccinated, people_fully_vaccinated_per_hundred, share_dose_used, distributed_per_hundred, total_distributed, total_vaccinations, total_vaccinations_per_hundred, total_boosters, total_boosters_per_hundred

Normalisations:

Locations

Locations (location, iso_code, last_observation_date, source_name, source_website)

location → iso_code, last_observation_date

We simply eliminate the vaccine column since the values are not atomic and we assume the list of vaccine in the column are from vaccine column in vaccinations_by_manufacturer.csv

No partial dependency because Locations' PK is single-valued

No transitive dependency

Already in 3NF

Location1 (location, iso_code, last_observation_date, source_name, source_website)

Vaccinations

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

date, location → vaccine, daily_vaccinations, daily_vaccinations_per_million, daily_people_vaccinated, daily_people_vaccinated_per_hundred, people_vaccinated, people_vaccinated_per_hundred, people_fully_vaccinated, people_fully_vaccinated_per_hundred, total_vaccinations, total_vaccinations_per_hundred, total_boosters, total_boosters_per_hundred

location → iso_code

Partial dependency occurs, therefore;

Decompositions:

Vaccinations1 (date, location, vaccine, daily_vaccinations, daily_vaccinations_per_million, daily_people_vaccinated, daily_people_vaccinated_per_hundred, people_vaccinated, people_vaccinated_per_hundred, people_fully_vaccinated, people_fully_vaccinated_per_hundred, total_vaccinations, total_vaccinations_per_hundred, total_boosters, total_boosters_per_hundred

Vaccinations2 (location, iso_code)

Vaccinations1 no longer has a partial dependency → 2NF

Vaccinations2'PK is single-valued → 2NF

Both relations have no transitive dependencies → 3NF

Vaccine

Vaccines (vaccine, manufacturer)

Vaccine → manufacturer

No partial dependency

No transitive dependency

Already in 3NF

Vaccines (vaccine, manufacturer)

Vaccinations_by_manufacturer

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

date, location, vaccine → total_vaccinations

No partial dependency

No transitive dependency

Already in 3NF

Vaccinations_by_manufacturer (date, location*, vaccine, total_vaccinations)

Vaccinations_by_age_group

Vaccinations_by_age_group (age_group, date*, location*, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)

age_group, date, location → people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred

No partial dependency

No transitive dependency

Already in 3NF

Vaccinations_by_age_group (age_group, date*, location*, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)

Australia

Australia (date*, location*, vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters)

date, location → vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters

Values in vaccine column is not atomic → 0NF

After separate the values into different rows with the rest of observation's value remain the same, the relation becomes 3NF

Final relation:

Australia (date*, location*, vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters)

Germany

Germany (date*, location*, vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters)

date, location → vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters

Values in vaccine column is not atomic → 0NF

After separate the values into different rows with the rest of observation's value remain the same, the relation becomes 3NF

Final relation:

Germany (date*, location*, vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters)

Italy

Italy (date*, location*, vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters)

date, location → vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters

Values in vaccine column is not atomic → 0NF

After separate the values into different rows with the rest of observation's value remain the same, the relation becomes 3NF

Final relation:

Italy (date*, location*, vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters)

United States

United States (date*, location*, vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters)

date, location → vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters

Values in vaccine column is not atomic → 0NF

After separate the values into different rows with the rest of observation's value remain the same, the relation becomes 3NF

Final relation:

United States (<u>date*, location*</u>, vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters)

US_states_vaccinations

US_states_vaccinations (<u>date*, location*</u>, daily_vaccinations, daily_vaccinations_per_million, people_vaccinated, people_vaccinated_per_hundred, people_fully_vaccinated, people_fully_vaccinated_per_hundred, share_dose_used, distributed_per_hundred, total_distributed, total_vaccinations, total_vaccinations_per_hundred, total_boosters, total_boosters_per_hundred)

date, location → daily_vaccinations, daily_vaccinations_per_million, people_vaccinated, people_vaccinated_per_hundred, people_fully_vaccinated, people_fully_vaccinated_per_hundred, share_dose_used, distributed_per_hundred, total_distributed, total_vaccinations, total_vaccinations_per_hundred, total_boosters, total_boosters_per_hundred

No partial dependency

No transitive dependency

Already in 3NF

US_states_vaccinations (<u>date*, location*</u>, daily_vaccinations, daily_vaccinations_per_million, people_vaccinated, people_vaccinated_per_hundred, people_fully_vaccinated, people_fully_vaccinated_per_hundred, share_dose_used, distributed_per_hundred, total_distributed, total_vaccinations, total_vaccinations_per_hundred, total_boosters, total_boosters_per_hundred)

Draft Schema

R1: Location1 (location, iso_code, last_observation_date, source_name, source_website)
R2: Vaccinations1 (date, location, daily_vaccinations, daily_vaccinations_per_million, daily_people_vaccinated, daily_people_vaccinated_per_hundred, people_vaccinated, people_vaccinated_per_hundred, people_fully_vaccinated, people_fully_vaccinated_per_hundred, total_vaccinations, total_vaccinations_per_hundred, total_boosters, total_boosters_per_hundred
R3: Vaccinations2 (location, iso_code)
R4: Vaccines (vaccine, manufacturer)
R5: Vaccinations_by_manufacturer (date, location, vaccine, total_vaccinations)
R6: Vaccinations_by_age_group (date, location, age_group, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)
R7: Australia (date, location, vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters)
R8: Germany (date, location, vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters)
R9: Italy (date, location, vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters)
R10: United States (date, location, vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters)
R11: US_states_vaccinations (date, location, daily_vaccinations, daily_vaccinations_per_million, people_vaccinated, people_vaccinated_per_hundred, people_fully_vaccinated, people_fully_vaccinated_per_hundred, share_dose_used, distributed_per_hundred, total_distributed, total_vaccinations, total_vaccinations_per_hundred, total_boosters, total_boosters_per_hundred)

Among these relations:

R1 and R3 have the same primary key and can be combined
R2, R7, R8, R9, and R10 have the same primary key and can be combined

R1: Location1 (location, iso_code, last_observation_date, source_name, source_website)
R2: Vaccinations1 (date, location, daily_vaccinations, daily_vaccinations_per_million, daily_people_vaccinated, daily_people_vaccinated_per_hundred, people_vaccinated, people_vaccinated_per_hundred, people_fully_vaccinated, people_fully_vaccinated_per_hundred, total_vaccinations, total_vaccinations_per_hundred, total_boosters, total_boosters_per_hundred
~~**R3: Vaccinations2** (location, iso_code)~~
R4: Vaccines (vaccine, manufacturer)
R5: Vaccinations_by_manufacturer (date, location, vaccine, total_vaccinations)

R6: Vaccinations_by_age_group (date, location, age_group, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)
~~**R7: Australia** (date, location, vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters)~~
~~**R8: Germany** (date, location, vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters)~~
~~**R9: Italy** (date, location, vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters)~~
~~**R10: United States** (date, location, vaccine, source_url, total_vaccinations, people_vaccinated, people_fiully_vaccinated, total_boosters)~~
R11: US_states_vaccinations (date, location, daily_vaccinations, daily_vaccinations_per_million, people_vaccinated, people_vaccinated_per_hundred, people_fully_vaccinated, people_fully_vaccinated_per_hundred, share_dose_used, distributed_per_hundred, total_distributed, total_vaccinations, total_vaccinations_per_hundred, total_boosters, total_boosters_per_hundred)

Lastly, they can also be renamed to reflect what they stand for, as following:

Final Schema

R1: Countries (location, iso_code, last_observation_date, source_name)
R2: Vaccinations (date, location, daily_vaccinations, daily_vaccinations_per_million, daily_people_vaccinated, daily_people_vaccinated_per_hundred, people_vaccinated, people_vaccinated_per_hundred, people_fully_vaccinated, people_fully_vaccinated_per_hundred, total_vaccinations, total_vaccinations_per_hundred, total_boosters, total_boosters_per_hundred
R4: Vaccines (vaccine, manufacturer)
R5: Vaccine_types (date, location, vaccine, total_vaccinations)
R6: Age_Groups (date, location, age_group, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, people_with_booster_per_hundred)
R11: US_States (date, location, daily_vaccinations, daily_vaccinations_per_million, people_vaccinated, people_vaccinated_per_hundred, people_fully_vaccinated, people_fully_vaccinated_per_hundred, share_dose_used, distributed_per_hundred, total_distributed, total_vaccinations, total_vaccinations_per_hundred, total_boosters, total_boosters_per_hundred)