Exploration of the Relationship Between Food Production & Consumption with Greenhouse Gas Emissions

Datasets

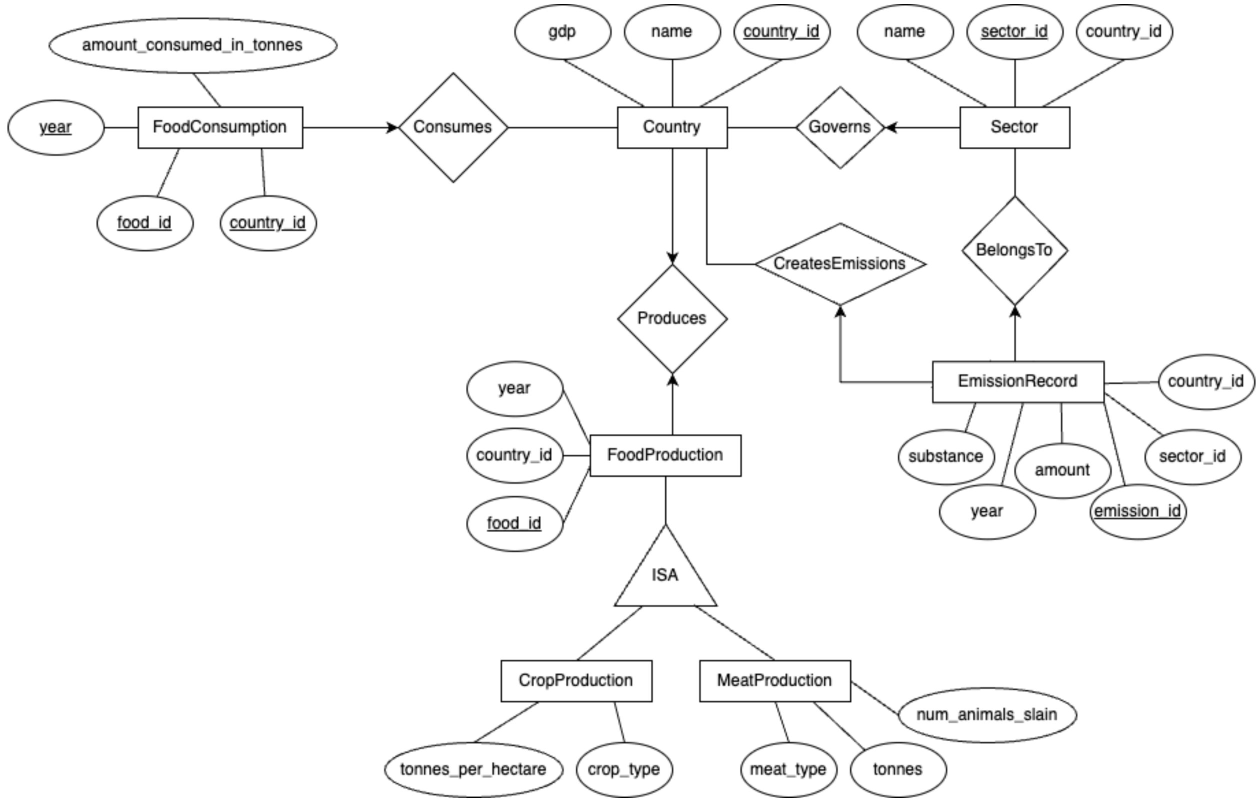
a. Greenhouse Gas Emissions Dataset:
   Link: https://edgar.jrc.ec.europa.eu/report_2023#data_download
b. World Meat Consumption and World Crop Production
   Link: https://data.oecd.org/agroutput/crop-production.htm#indicator-chart
c. Meat Consumption Dataset:
   Link:https://www.kaggle.com/datasets/pragadesh06/meat-production-datasets1961-2018?select=global-meat-production.csv

Entities (**Bold = Primary Key** **,** *italics = Foreign Key)***:**

- Country (**country_id**, name, gdp)
- Sector (**sector_id**, *country_id*, name)
- EmissionRecord (**emission_id**, *country_id*, *sector_id*, substance, year, amount)
- FoodProduction (**food_id**, *country_id*, year)
    - ISA with child:
        - CropProduction (has unique attribute crop_type, tonnes_per_hectare) **[food_id would be the primary and foreign key]**
        - MeatProduction (has unique attributes meat_type, tonnes, and num_animals_slain which describes the number of animals killed per country for that year) **[food_id would be the primary and foreign key]**
- FoodConsumption(country_id, food_id, year, amount_consumed_in_tonnes)

Relationships:

- A country has multiple sectors, a sector belongs to a specific country
- A sector has multiple possible emission records (CO2, GWP_100_AR5_CH4, etc.), an emission record has only one sector and one country
- A country can have multiple food productions, food production can only come from one country



Discussion on Design Choices

1) EmissionRecord as an Entity vs. Attribute of Sector:

- Design Choice: EmissionRecord is represented as an entity rather than an attribute of the Sector entity.
- Rationale: Emission records vary over time and by type of substance. Representing these as attributes would lead to multiple columns for each substance and year, making the Sector table unwieldy.
- Pros: Treating EmissionRecord as an entity allows for scalability, as new substances and years can be added without altering the table structure.
- Cons: More JOIN operations are required to retrieve data that spans sectors and emissions, which could impact performance.

2) FoodProduction and Its Child Entities (CropProduction, MeatProduction):

- Design Choice: FoodProduction is an entity, and CropProduction and MeatProduction are separate entities rather than attributes of FoodProduction.
- Rationale: These productions have distinct attributes that do not overlap well. Combining them into a single table would result in many null values for non-applicable columns.

- Pros: This choice avoids sparse tables and ensures that each entity only contains relevant attributes, simplifying queries.
- Cons: Requires more JOIN operations to collect comprehensive food production data.

3) Sector and Country Relationship:

- Design Choice: Sector is an entity with a relationship to Country, rather than embedding country information within Sector.
- Rationale: Sectors are distinct entities that relate to a country, and a country can have multiple sectors.
- Pros: This maintains normalization and allows sectors to be managed independently of countries.
- Cons: Data retrieval requires a JOIN operation, which can be less performant than denormalized structures.

4) FoodConsumption as an Entity vs. Relationship:

- Design Choice: FoodConsumption is an entity rather than a many-to-many relationship between Country and FoodProduction.
- Rationale: Food consumption data has its attributes, like the year and amount consumed, which are not merely relational but substantive information.
- Pros: It keeps the data normalized and makes it easier to manage and query consumption data.
- Cons: adds complexity to the model and requires an additional JOIN to relate consumption with production.

5) FoodProduction to CropProduction and MeatProduction (ISA Relationship):

- Design Choice: A binary relationship between FoodProduction and each subtype (CropProduction, MeatProduction).
- Rationale: Each food production record will only be one type (crop or meat), making the relationship binary and more straightforward to manage.
- Pros: simplicity in design and querying.
- Cons: It cannot capture complex situations where a single food production event could include both crop and meat aspects, though this is unlikely in the given context.

6) FoodConsumption Primary Key Decision:

- Design Choice: composite primary key of country_id, food_id, and year for the FoodConsumption entity.
- Rationale: Each of these elements is required to uniquely identify a food consumption record.
- Pros: This choice eliminates the need for a separate surrogate key and ensures the uniqueness of each record.
- Cons: Composite keys can make queries and table relationships more complex.