



**UNIVERSITÀ  
DEGLI STUDI DI BARI  
ALDO MORO**

**DIPARTIMENTO DI INFORMATICA  
CORSO DI LAUREA MAGISTRALE IN DATA SCIENCE**

**GESTIONE DI DATI STRUTTURATI E NON STRUTTURATI**

-

**PROGETTO DATI SPAZIALI**

**Docente:**

**Prof. Corrado Loglisci**

**STUDENTE:**

**Walter Mangione**

**Anno Accademico 2024-2025**

## ABSTRACT

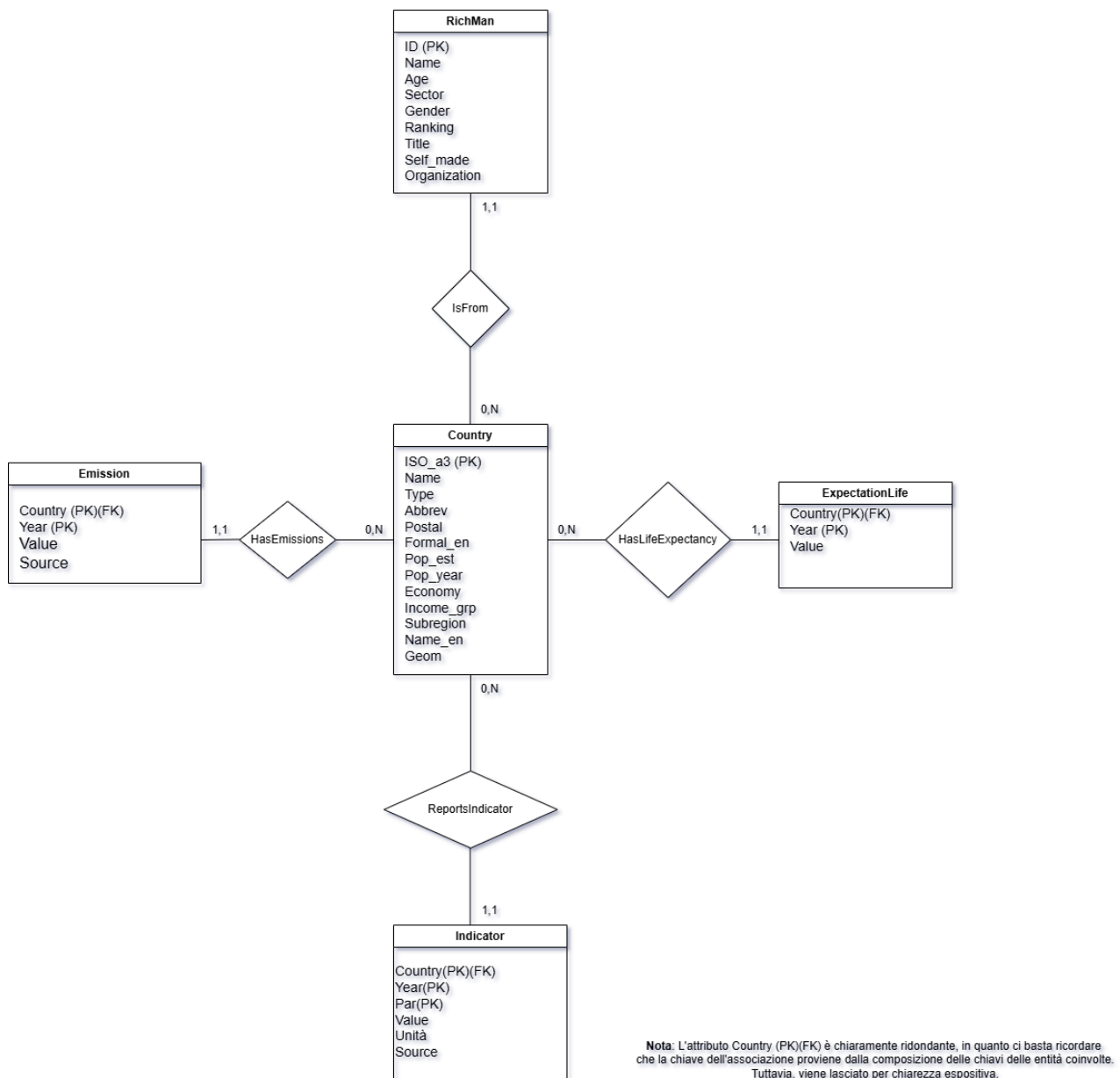
Il presente documento descrive la gestione di dati spaziali mediante l'impiego di due *layer* distinti: uno relativo a geometrie terrestri e l'altro a geometrie marittime. A partire dalla creazione delle relative tabelle in *PostgreSQL*, vengono illustrate le fasi di popolamento e successiva interrogazione dei dati. A supporto delle informazioni spaziali sono stati integrati dati *a-spaziali*, in particolare statistiche ambientali riguardanti l'inquinamento, l'aspettativa di vita ed informazioni circa gli uomini più ricchi del Paese. L'obiettivo del progetto è stato quello di analizzare congiuntamente dati spaziali e *a-spaziali*, mediante interrogazioni mirate, al fine di evidenziare correlazioni e caratteristiche rilevanti del territorio, sfruttando entrambi i *layers*.

## DATI UTILIZZATI

I [dati spaziali](#) in questione contengono due layers principali, estratti dai due shapefiles messi a disposizione. Il primo *layer* è composto da zone terrestri, con ogni riga, nella tabella creata, che corrisponde ad un determinato Paese. Il secondo *layer*, invece fa la stessa cosa, ma questa volta con aree marine (quindi fiumi, oceani, laghi, eccetera). Entrambi i *layers*, tuttavia, contengono poche informazioni *a-spaziali* interessanti.

I dati *a-spaziali*, invece, sono stati estratti da 4 tabelle sorgenti differenti. Il [primo](#) descrive dati relativi agli uomini più ricchi del pianeta. Il [secondo](#), invece, contiene indicatori relativi a diversi aspetti ambientali, quali aree forestali per ogni paese, dal 1992 al 2020. Invece, la [terza sorgente](#) di dati contiene informazioni relative all'aspettativa di vita nei vari Paesi, dal 1990 al 2021. Infine, [l'ultima sorgente](#) dati fa riferimento alle emissioni di *CO2*, all'interno di vari Paesi, provenienti da diverse sorgenti, quali gas, petrolio, cemento. Tutte le 4 tabelle descritte possono essere *collegate*, mediante riferimento esterno, alla tabella descrivente il *layer* di dati

terreni. Di seguito, potremmo immaginare nel seguente modo una rappresentazione concettuale del nostro *minimondo*.



Di conseguenza, il diagramma in questione è stato tradotto e *trasformato* in un insieme di tabelle logiche. Da qui, le tabelle sono state popolate (per i dettagli si rimanda ai file *.sql*). Interessante, anche, la modifica apportata durante la traduzione. Infatti, le entità *expectation\_life*, *emission* e *indicator* condividono i campi *country*, *year*, *value*, implementati nel nostro *DBMS* con una tabella astratta

*country\_stats*, da cui ereditiamo le prime 3 tabelle. Avremmo potuto anche modificare il diagramma, aggiungendo la tabella principale ed una serie di relazioni di generalizzazione. Si è preferito il primo approccio per evitare di appesantire il diagramma e renderlo più complesso da leggere.

## QUERY EFFETTUATE

Tra le query effettuate, iniziamo citando una implementazione di un algoritmo di esplorazione dei percorsi minimi (molto simile a *Dijkstra*, se non fosse che viene troncato prima poiché siamo interessati solo alle coppie) tra paesi confinanti, basandosi sui dati geografici e sulle superfici dei paesi, che vengono usati poi come pesi degli archi che collegano due Paesi confinanti. Si tratta, infatti, di un piccolo navigatore capace di trovare il percorso minimo tra qualsiasi coppia di Paesi, sfruttando, quindi, la superficie di un Paese come peso tra due Paesi collegati. Ad esempio, vi è un arco che va dall'Italia alla Svizzera, con un peso uguale alla superficie della Svizzera. Viceversa, vi è un arco dalla Svizzera all'Italia, con un peso uguale alla superficie dell'Italia.

Lo stesso concetto, poi, è stato esteso anche all'altro layer, collegando i due citati, vale a dire il *layer* terreno ed il *layer* marittimo. In questo ultimo caso, è possibile ottenere il percorso tra qualunque coppia di oggetti geografici, valutando se sia meglio giungervi via mare o via terra. Si noti che questa query sfrutta le funzioni di *vicinanza* messe a disposizione dai metodi stessi geometrici; tuttavia, la definizione di distanza, come abbiamo visto, non sfrutta alcun concetto geometrico in senso stretto (non si fa riferimento alla distanza come distanza tra i punti più vicini, più lontani, o tra i centroidi, per intenderci). Pertanto, le restanti query spaziali sono le seguenti:

1. Paesi confinanti con un Paese fissato che hanno almeno un uomo ricco;
2. Paesi bagnati da mari o laghi che hanno almeno un uomo ricco;

3. Paesi bagnati da mari/laghi con media di emissioni superiore a una soglia tra *start\_year* e *end\_year*;
4. Paesi interni che hanno almeno un uomo ricco;
5. Confronto tra paesi marini e paesi interni sulle emissioni;
6. Confronto tra paesi marini e paesi interni sulla media di un certo indicatore;
7. Paesi che contengono completamente risorse idriche e hanno almeno un uomo ricco;
8. Paesi marini con media di emissioni tra *start\_year* e *end\_year* e superficie;
9. Età media degli uomini ricchi, suddivisa per layer;
10. Paese più vicino a un punto dato che ha una media di emissioni sopra una certa soglia in un intervallo di anni.

Laddove possibile, le query sono state generalizzate con la creazione di funzioni.

## **RISULTATI**

Alcune interrogazioni hanno evidenziato delle caratteristiche interessanti che differenziano i Paesi nell'entroterra dai Paesi costieri. Infatti, è emerso che vi è una maggiore emissione di CO2 nei paesi costieri, senza considerare chiaramente lo sbilanciamento dovuto alla categoria di maggioranza. Non sono emerse differenze sostanziali, sulla differenza d'età, degli uomini ricchi, suddivisa per Paesi costieri e Paesi nell'entroterra. In entrambi i casi, vi è un'età media di 65 anni. Le altre query, invece, forniscono un supporto circa l'estrazione di informazioni nel dominio in questione. Per un approfondimento, si rimanda al file stesso delle query.

## **CONCLUSIONI E SVILUPPI FUTURI**

Potrebbe essere interessante estendere le query relative alla ricerca dei percorsi di distanza minima, cercando delle tecniche di ottimizzazione che permettano di velocizzarle e renderle più scalabili in contesti realistici. Inoltre, si potrebbe

considerare anche l'opportunità di aggiungere ulteriori layer, rendendo più complessi e realistici i collegamenti tra punti.