

GIS e Database Geo-spaziali. Un contributo per evitare di confonderli

Andrea Favretto (a)

(a) Università di Trieste, via Lazzaretto 8, 34123 Trieste, 040 5583641, afavretto@units.it

Introduzione

Come è ben noto, i dati in un sistema informativo geografico sono tradizionalmente suddivisi in: "grafici" e "attributo". I primi sono organizzati in strutture vettoriali o raster mentre i dati attributo sono costituiti da tabelle collegate ai dati grafici e collegabili fra loro attraverso le regole dei database relazionali (mediante l'applicazione di comandi SQL sulle strutture tabellari opportunamente preparate - relazioni attraverso l'uguaglianza di campi chiave)¹.

Le strutture grafiche vettoriali sono quelle che utilizzano maggiormente i comandi SQL per il collegamento fra le tavole attributo (comandi "select" e "join"). Generalmente, in questo modo viene prodotta la cartografia tematica, spesso parte integrante di ampi e approfonditi studi territoriali.

In tali studi talvolta viene utilizzato il termine "database" in modo improprio, nel senso che si definisce database l'insieme delle tabelle collegate a dei layer grafici, molto spesso non collegate fra loro in alcun modo.

Un database relazionale è invece tutt'altra cosa ovvero un archivio di dati digitali integrati e condivisi (cfr.: Date, 1994), attraverso le relazioni poste in essere dalle varie chiavi presenti nei campi delle tabelle.

L'errore è, a mio avviso, giustificato dal fatto che, in effetti, la differenza fra un GIS e un Database relazionale con componente spaziale (anche detto "Geodatabase"), è molto sfumata. Le similitudini sono molte e, considerando la sola parte dati di un GIS, i due si equivalgono. Un GIS può infatti essere visto sia da un punto di vista funzionale che da quello dei dati in esso presenti. Se lo si considera in questa ultima veste, sono dell'opinione che GIS e Geodatabase siano sinonimi.

Il nocciolo della questione è relativo all'integrazione dei dati²: si può parlare di database se e solo se questa condizione sia vera e ciò si ottiene creando le relazioni fra le tabelle. Memorizzare nel proprio disco rigido una moltitudine di tabelle collegate a dei layer grafici non significa costruire un database relazionale.

Questa breve nota si pone come un piccolo contributo per chiarire e ricordare la questione ovvero la differenza fra un Geodatabase e una serie di layer GIS

¹ Per chi volesse approfondire GIS e Database relazionali, la bibliografia disponibile è molto vasta. Fra i testi italiani recenti dedicati ai GIS si possono segnalare: Pesaresi, 2017; Migliaccio e Carrion, 2016. Per quanto riguarda i Database relazionali, un testo di riferimento sulla metodologia non collegato a nessun software può essere: Date, 1994.

² In un database relazionale i dati sono integrati quando non ci siano duplicazioni degli stessi nell'archivio, la qual cosa si ottiene costruendo delle relazioni fra le tabelle.

collegati a tavole tra di loro scollegate. Dopo un breve paragrafo che richiama alcune principali caratteristiche proprie di un database, viene mostrato come introdurre la componente spaziale in una struttura relazionale, utilizzando interamente software liberi, facilmente reperibili in rete.

Database relazionale, Geodatabase

Un database può essere definito come un archivio di dati in formato digitale. Generalmente un database è gestito da un software, che viene detto DBMS (DataBase Management System). La gestione di un database consiste di tutta una serie di operazioni per il suo funzionamento, tra cui: la memorizzazione dei dati, la cancellazione, la visualizzazione (accesso al database), l'aggiornamento, l'elaborazione.

Per quanto riguarda i dati, questi sono memorizzati solo attraverso tavole (matrici generalmente non quadrate). Le tavole sono formate da colonne (campi) e righe (record).

I dati di un database devono essere integrati e condivisi. Tale caratteristica è di fondamentale importanza per il corretto ed efficiente funzionamento del database. Come si è detto, integrazione vuol dire evitare la duplicazione di un dato nell'archivio. Se un dato è memorizzato una sola volta, è naturalmente necessaria la sua condivisione. L'integrazione garantisce che le informazioni siano gestite in modo efficiente (ad esempio: gli aggiornamenti sono più veloci) ed è una buona premessa per evitare errori all'interno dell'archivio (presenza di informazioni discordanti riferite allo stesso soggetto – disallineamento della base dei dati).

L'integrazione dei dati si ottiene mettendo in relazione le diverse tavole che popolano il database. Le relazioni possono essere di vari tipi, le più diffuse sono le cosiddette "uno sta a molti" e "molti stanno a molti". Le relazioni si realizzano attraverso l'uguaglianza di campi comuni. Tali campi comuni sono generalmente detti campi "chiave". Le chiavi possono essere primarie (Primary Key – PK) e secondarie (Secondary Key – SK). Una PK è generalmente un codice che identifica univocamente il singolo record mentre una SK è un codice, generalmente presente in più record di un'altra tabella (nell'ipotesi che si voglia costruire una relazione uno sta a molti fra le due). Quando il valore dei due codici (PK e SK) è uguale, i record delle tavole in oggetto sono messi in relazione.

Se nelle tavole del database sono memorizzati anche componenti geometriche, si può parlare di database spaziali o geodatabase. In tal modo la struttura database acquisisce delle potenzialità aggiuntive, legate alle analisi spaziali (spesso anche molto evolute) messe a disposizione. Il software di gestione diviene uno "Spatial DBMS", in grado di porre in essere query spaziali ed elaborazioni geografiche dei dati da esse ricavati.

Un riferimento open source

Di seguito si propone una soluzione a basso costo per la realizzazione e gestione di un Geodatabase. Non si desidera presentare un esempio concreto per non essere riduttivi. L'obiettivo è quello di fornire una sequenza di operazioni logiche per impostare la struttura di un database spaziale e predisporre una serie di strumenti applicativi, cercando di prediligere le

interfacce grafiche (se disponibili), rispetto ai controlli a tastiera. In tal modo si spera di facilitare l'uso degli strumenti proposti in diversi ambiti tematici.

Si è deciso di non mostrare i singoli comandi che l'utente deve imporre (fornendo lunghe e noiose sequenze di schermate software). La rapida obsolescenza del software (che viene spesso aggiornato con versioni nuove), è il principale motivo di tale decisione. Inoltre si pensa che capire una sequenza logica di operazioni aiuti a consolidare maggiormente i concetti rispetto alla memorizzazione di una serie di singoli passaggi. In tal modo, sperabilmente, l'utente è messo in grado di lavorare con più tipi di software (magari su piattaforme diverse).

Si è deciso di far riferimento ai seguenti software, liberamente scaricabili in rete:

- PostgreSQL come database relazionale (<https://www.postgresql.org/>);
- pgAdmin come piattaforma grafica per gestire e sviluppare strutture database relazionali PostgreSQL (<https://www.pgadmin.org/>);
- PostGIS come estensione spaziale per un database relazionale PostgreSQL (<https://postgis.net/>);
- QGIS come sistema informativo geografico che si interfaccia al Geodatabase (<https://www.qgis.org/it/site/index.html>);
- DB Manager Plugin come interfaccia grafica per la gestione dei formati database spaziali che QGIS supporta (PostGIS nel nostro caso), (https://docs.qgis.org/2.14/en/docs/user_manual/plugins/plugins_db_manager.html),

Dopo aver installato questi software sul proprio pc ma soprattutto dopo aver pensato la struttura del Geodatabase, si devono eseguire le seguenti operazioni.

- Su pgAdmin: si deve creare e denominare il database che si intende costruire; poi si deve inserire PostGIS come estensione spaziale nel database che si è creato.
- Su QGIS: si deve costruire una connessione al database creato in pgAdmin (il nome della connessione è libero mentre il nome del database è quello scelto in pgAdmin (attenzione a dichiarare "localhost" come server; il nome utente, se non lo si è cambiato in fase di installazione di PostgreSQL, è "postgres"; la password è quella che si è scelta in fase di installazione di PostgreSQL sul proprio pc).
- Su QGIS: si apre il plugin "DB manager", si sceglie PostGIS fra le estensioni spaziali disponibili e quindi la connessione che si è appena fatta al database; a questo punto si possono creare le tabelle del proprio database che hanno la componente spaziale (attenzione a scegliere un sistema di riferimento cartografico per le geometrie che si disegneranno, dichiarando un codice EPSG noto – ad esempio 4326, corrispondente a WGS84, coordinate geografiche); una volta costruita la tabella, la si può aprire ed editare, disegnando le geometrie del tipo prescelto in fase di realizzazione della tabella stessa (punti, linee o poligoni).

- Su pgAdmin: si possono vedere tutte le tavole con componente spaziale che si sono create con QGIS. Si possono costruire anche le altre tavole (quelle senza componente spaziale) previste dalla struttura del database che si è pensata e, naturalmente, realizzare tutte le relazioni, query e viste che si desidera.
- Su QGIS: le tavole senza componente spaziale, costruite in pgAdmin, sono visibili; le si può visualizzare insieme a quelle con componente spaziale e si possono creare delle query spaziali e tabellari, utilizzando la finestra SQL di DB Manager.

Conclusioni

In questa breve nota si è cercato di contribuire nel fornire chiarezza su di un concetto a volte poco chiaro in ambiente GIS, ovvero la differenza fra le tabelle degli attributi dei layer grafici (memorizzate nei file system), e le tabelle di un Geodatabase (un database con componente spaziale). Come si è accennato, non sempre le due entità sono precisamente delineate nella mente (e negli interventi), dei neofiti in materia di GIS, con il rischio di confondere le une con le altre.

Poiché si pensa che il miglior metodo per capire e consolidare dei concetti sia quello di esercitarsi, di "toccare con mano", si è pensato di fornire alcune indicazioni di massima per la costruzione di un Geodatabase, utilizzando degli applicativi liberamente disponibili in rete. Dopo aver costruito un semplice Geodatabase, l'utente è a nostro avviso in grado di apprezzare pienamente la differenza fra una serie di tavole attributo di layer grafici e delle vere e proprie tavole di un Geodatabase, collegate alle altre tavole della struttura relazionale mediante delle relazioni.

La chiave di volta per capire il concetto potrebbe essere proprio l'esistenza delle relazioni in un caso e l'assenza delle stesse nell'altro. Le relazioni, come si è visto, sono infatti la condizione dell'integrità dei dati in una struttura database.

Bibliografia

Date C. J. (1994), *An Introduction to Database Systems*, Addison Wesley, New York.

Migliaccio F., Carrion D. (2016), *Sistemi Informativi Territoriali. Principi e applicazioni*, UTET.

Pesaresi C. (2017), *Applicazioni GIS. Principi, metodologie e linee di ricerca. Esercitazioni ed esempi*, UTET.

poter