

IL DATABASE TOPOGRAFICO DEL SIT DELLA REGIONE PUGLIA: DAL MODELLO LOGICO AL MODELLO FISICO

Giuseppe CHIELLI, Tina CAROPPO, Michele GRAMEGNA

Tecnopolis CSATA Strada Provinciale Valenzano-Casamassima km 3 70010 Valenzano (Bari)
tel: 0804670513-0804670217 fax: 0804670217 - email: (g.chielli, c.caroppo, m.gramegna)@tno.it

Riassunto

Nell'ambito del Sistema Informativo Territoriale (SIT) Regionale, la Regione Puglia ha in corso la costruzione di un database topografico secondo le specifiche nazionali IntesaGIS; il database topografico rappresenta una componente fondamentale a supporto di svariati servizi applicativi forniti dal sistema e, come per tutte le altre banche dati geografiche del SIT, viene gestita all'interno di un'infrastruttura basata sul DBMS Oracle 10g e il gateway ArcSDE ESRI di tipo Enterprise.

Il database topografico è derivato dalla carta tecnica in costruzione su tutto il territorio regionale alla scala 1:5000 e ne eredita la precisione; essendo il modello dati adottato di tipo relazionale, tutti i contenuti richiedono un'organizzazione in una struttura relazionale per la quale la fase di progettazione è estremamente importante. La metodologia seguita parte da un'analisi concettuale in cui sono individuate tutte le entità e le relazioni più significative, per poi passare alla definizione di un modello logico di tipo E-R in cui sono stabilite le tabelle relazionali, con nomi e attributi, e le loro reciproche relazioni. Risultato finale è il modello fisico, ottimizzato rispetto agli strumenti tecnologici impiegati. Parallelamente, vi è l'individuazione di alcuni vincoli che i contenuti geometrici ed alfanumerici devono rispettare, il cui soddisfacimento è verificato da procedure di validazione a cui sottoporre i dati in ingresso al geodatabase fisico, garantendo in questo modo la qualità dei dati nella fase di popolamento.

Abstract

In the context of the its Territorial Information System, the Apulia Region is involved in the construction of a topographical database according to the IntesaGIS national standard; this database represents an essential component supporting the applicative functionalities provided by the system and it is managed in an infrastructure based on Oracle 10g DBMS and the gateway ESRI ArcSDE Enterprise. The relational model is used as data model, so all the contents of the topographical database, derived from a cartography on the whole regional territory at the scale 1:5000, must be organized in a relational structure. The followed methodology begins from a conceptual analysis in which entities and the most important relationships are found, then it passes to the definition of a E-R logical model in which relational tables, with their names and fields, and mutual relationships are established. The final result is given by the physical model, optimized from the point of view of the chosen technology parallelly some constraints are defined on the geometrical and alphanumeric data: the check of the data is made by validation procedures to run before storing data in the physical geodatabase. In this way data quality is assured even from the first step of database creating; then, it must be maintained in the subsequent phases of database updating.

Premesse

Nodo basilare nel progetto del Sistema Informativo Territoriale della Regione Puglia è rappresentato dalla costruzione della Carta Tecnica Regionale relativa all'intero territorio regionale

e, soprattutto, dall'implementazione di un database topografico (nel seguito DBT) aderente alle specifiche nazionali IntesaGIS. Nella fase di primo impianto, il DBT è derivato direttamente dalla carta tecnica, con contenuti perlopiù orientati ad un approccio fotogrammetrico. Il DBT copre l'intero territorio regionale con un dettaglio tipico della scala 1:5000, mentre aree urbane di un sottoinsieme di Comuni riportano la precisione della scala 1:2000, per inglobare informazioni già rilevate in precedenti lavori regionali; in tali aree è acquisita la toponomastica e la numerazione civica secondo i medesimi standard. Ciò fa della multiprecisione una peculiarità che rende la banca dati estremamente innovativa.

Nucleo fondamentale del patrimonio informativo del SIT e dei servizi applicativi offerti ai vari enti locali, il DBT si traduce in un modello fisico finale implementato in un'infrastruttura basata su sistema di gestione di basi dati relazionali; la tecnologia adottata consta del DBMS Oracle 10g +, ArcGIS Server ESRI con integrato ArcSDE Enterprise, pertanto il modello fisico si rifà ad una struttura dati completa ed "intelligente".

Obiettivo del presente lavoro è illustrare le principali fasi nel processo di progettazione del DBT fino a descrivere i tratti salienti del modello fisico e delle procedure di validazione dei dati da utilizzare in fase di caricamento/aggiornamento nel database fisico. Nelle linee generali, la metodologia seguita si articola nelle tre fasi di progettazione concettuale, logica e fisica, secondo la tradizionale teoria delle basi di dati relazionali.

La modellazione concettuale: corrispondenza dei contenuti CTR con le classi IntesaGIS.

La definizione del modello concettuale consta nell'individuazione delle entità caratterizzanti il database topografico e delle eventuali reciproche relazioni; gli assunti fondamentali di questa fase sono i seguenti:

- il database topografico è derivato dalla cartografia per cui deve almeno includerne tutti i contenuti geometrici ed alfanumerici attinenti alla classificazione degli oggetti;
- i contenuti rappresentati si riferiscono tutti al Catalogo degli Oggetti IntesaGIS (documento 1n1007_1_2) per cui le codifiche presenti in carta tecnica vanno riportate a questo mediante un meccanismo di corrispondenza;
- partendo da una base fotogrammetrica, il database topografico deve essere predisposto, anche solo a livello di schema, per poter inserire altre informazioni derivanti da fonti alternative (dati Acquedotto Pugliese, Autorità di Bacino Puglia, etc.); questo nella logica di estendere la banca dati aldilà di una semplice traduzione della carta tecnica.

I contenuti geometrici del database topografico sono caratterizzati dalla stessa accuratezza posizionale della carta tecnica; la fotorestituzione è legata a specifiche più restrittive delle precisioni imposte da IntesaGIS in quanto aderenti al Capitolato Speciale d'Appalto della Regione Puglia attualmente vigente in materia di produzione di cartografia. Ciò implica che non vi è ammessa incoerenza tra le geometrie della carta tecnica (strutturate in linee, punti e poligoni) e le componenti geometriche delle classi IntesaGIS prodotte. Acquisendo i dati con soglie di acquisizioni più basse, non sempre è possibile restituire come aree oggetti per i quali lo standard prevede componenti geometriche poligonali: l'adattamento delle specifiche nazionali ai requisiti della carta tecnica pugliese ha indotto ad un uso intensivo del *collassamento geometrico*.

Inoltre, il repertorio di classificazione degli oggetti territoriali in carta tecnica si basa su un insieme di codici numerici che rendono conto anche della tipologia geometrica degli oggetti (se linee, punti o poligoni) e ai quali è associata una descrizione; preso ciascun codice, la categoria rappresentata è fatta confluire in una classe IntesaGIS avvalorando la n-pla di attributi prevista per la classe. In tal modo, resta individuato l'insieme base di classi acquisite e, per ciascuna classe, le componenti geometriche (puntuali, lineari o poligonali). In alcuni casi, il mapping è immediato e diretto sia dal punto di vista geometrico sia dal punto di vista dei contenuti alfanumerici descrittivi.

La tabella 1 di seguito riportata esemplifica il mapping diretto nel caso di oggetti puntuali che in carta tecnica individuano dei punti quotati e che nel DBT confluiscono nella classe "Punto quotato" con determinati valori per gli attributi previsti e con una componente geometrica puntuale come da Catalogo degli Oggetti IntesaGIS:

CODICE CARTA TECNICA	DESCRIZIONE CARTA TECNICA	PT_QUO_SEDE	PT_QUO_RIL	PT_QUO_SCA
16110000	quota al suolo	01	02	02 o 03 a seconda della scala
16110001	quota al suolo per scale diverse	01	02	02 o 03 a seconda della scala
61600000	quota piede fabbricato	0301	02	02 o 03 a seconda della scala
11160000	quota terrazzo piano edificio	03	02	02 o 03 a seconda della scala
11170000	quota gronda edificio	0302	02	02 o 03 a seconda della scala
11180000	quota colmo edificio	0302	02	02 o 03 a seconda della scala

Tabella 1 – Esempio di corrispondenza diretta tra contenuti di carta tecnica e classi del DBT applicata ai punti quotati

Sempre nell'ambito dell'altimetria, anche per le curve di livello la corrispondenza è semplice.

Pure in molti esempi relativi alla vegetazione il mapping è diretto, per cui, ad esempio, dalla carta tecnica si derivano aree direttamente riversabili nelle classi "Bosco" o "Coltura agricola" con classificazioni coerenti con i codici di partenza.

Per altre classi, invece, il discorso è più complesso; ad esempio per quanto concerne l'edificato, l'approccio fotogrammetrico non consente di avvalorare molti degli attributi previsti per la classe edificio: la problematica più rilevante è data dall'attributo di "uso" che non può essere avvalorato per corrispondenza diretta con la CTR. Per differenziare i teatri dagli ospedali, ad esempio, viene sfruttata informazione aggiuntiva contenuta in scritte cartografiche o legata a indagini a terra in fase di ricognizione.

In altre situazioni, la corrispondenza diretta non può neanche essere parziale, né geometricamente né sotto il profilo delle classificazioni: nel derivare le classi della viabilità, i codici di carta tecnica sono solo un punto di partenza, poiché le componenti geometriche poligonali sono ricostruite interamente ex-novo e i relativi attributi avvalorati sulla base di operazioni di fotointerpretazione, anche a posteriori, rispetto alla prima fase di fotorestituzione cartografica.

Anche per le classi degli strati 00 e 10 (informazioni aerofotogrammetriche e aree di pertinenza, rispettivamente) la costruzione dei dati va operata ex-novo pur elaborando informazioni di base disponibili in carta tecnica, nelle altre classi o in archivi provenienti da altre fonti.

Il problema del collassamento geometrico è già affrontato nell'analisi concettuale: per ciascuna classe si individuano le componenti geometriche fornite, non sempre coincidenti con quelle previste da IntesaGIS. Per quanto riguarda la classe "Area di viabilità mista secondaria", alle componenti geometriche areali si aggiungono le componenti lineari collassate pur non previste dal Catalogo degli Oggetti, in quanto i sentieri in carta tecnica sono sempre restituiti come linee; analogamente per le scarpate, per cui non sempre è possibile distinguere il ciglio dal piede durante il processo di fotorestituzione cartografica, e per gli argini. Viceversa, molti collassamenti nell'ambito dell'edificato vengono gestiti da classi IntesaGIS: è il caso delle localizzazioni puntuali di manufatti di varia natura, tra cui contenuti puntuali di carta tecnica trovano collocazione senza difficoltà. Il collassamento segue le direttive proposte nell'Allegato A predisposto dal CNIPA per le Regioni aderenti al progetto PR5SIT.

Molte classi sono introdotte nel modello pur senza disporre di contenuti nell'immediato, in prospettiva di integrazione con altre fonti.

Le relazioni tra le varie entità sono mutate essenzialmente dal modello UML delle specifiche IntesaGIS e dall'Allegato A; relazioni associative di cardinalità complessa sono introdotte per definire componenti geometriche di alcune entità alfanumeriche in termini di geometrie elementari:

per esempio, un'estesa amministrativa è messa in relazione con le istanze della classe *Area stradale* per definirne la pertinenza, con gli elementi stradali per definirne il tracciato di dettaglio e con i tratti stradali per individuarne il tracciato di sintesi. Tali relazioni associative possono, talvolta, essere caratterizzate da attributi: nell'esempio dell'estesa amministrativa, una proprietà dell'associazione è data dall'ente gestore. Le relazioni sono previste per rendere più integrate e "navigabili" le informazioni disponibili.

Al termine dell'analisi concettuale, è ben delineato l'insieme delle entità (componenti geometriche delle classi IntesaGIS con relativi attributi) e il set di relazioni eventualmente impostate tra esse.

Definizione del modello logico

Nella fase di definizione del modello logico, tutte le entità individuate nella fase precedente sono tradotte in tabelle in uno schema logico relazionale complessivo non legato alla tecnologia in adozione. Nella predisposizione dello schema, sono stabiliti i nomi delle tabelle e i nomi dei campi di ciascuna tabella, dando per acquisita la differenziazione tra tabelle geometriche e tabelle alfanumeriche. Da un punto di vista esclusivamente logico, le varie tabelle sono organizzate in aree tematiche rispetto ai contenuti memorizzati nel loro interno; ogni area è caratterizzata da un codice mnemonico di due caratteri. La notazione è quella tipica degli schemi ER.

I nomi delle tabelle seguono una convenzione simile a quella scelta nel progetto SigmaTer; in ciascun nome, è presente il codice dell'area tematica, un numero progressivo che contraddistingue univocamente la tabella e un carattere che esplicita la natura della tabella, se geometrica o alfanumerica. Allo scopo di semplificare le procedure di caricamento, i nomi dei campi sono i medesimi degli shapefile di fornitura del database topografico, dati di input al processo di caricamento all'interno del database fisico.

Nella Figura 2 è riportato, a titolo di esempio, lo schema logico relazionale relativo all'area tematica *Copertura relativa a mobilità e trasporti* (CM).

Dall'esempio si astraggono alcune caratteristiche generali dello schema normalizzato:

- presenza delle tabelle dei tratti di contorno **D** associati alle geometrie poligonali: tra le tabelle SITC M0G A CV EI contenente tutte le sottoaree della classe *Area di circolazione veicolare* e SITC M0G T RATTOC ONTA CV EI contenente i relativi tratti di contorno che specificano se il contorno è reale o fittizio sussiste una relazione di cardinalità uno a molti per ricostruire il contorno di ciascun poligono;
- presenza di tabelle codificate che memorizzano i valori codificati previsti per ciascun campo;
- oltre ai campi propri per ciascuna classe, sono presenti dei campi comuni accessori relativi ai metadati operativi previsti dall'Allegato A;
- gestione normalizzata del collassamento mediante l'introduzione di una tabella alfanumerica contenente i campi comuni alle geometrie collassate e non collassate (esempio: viabilità mista secondaria).

Ogni geometria è caratterizzata da un identificativo univoco indipendente dalla tecnologia usata (chiave primaria) che funge da chiave esterna nella relazione. Un buon esempio di tabella associativa dato dalla tabella SITC M3 A RS TRX E SA MM che mette in relazione le estese amministrative con le aree stradali per determinare la pertinenza areale di ciascuna estesa.

Nel modello relazionale sono introdotte tabelle esterne per la gestione degli attributi multivalore; anche gli attributi che per ciascun oggetto specificano rispettivamente il codice ISTAT del comune di appartenenza e il taglio cartografico di riferimento (taglio IGM per la scala 1:5000) sono gestiti come multivalore evitando segmentazioni fisiche sui limiti comunali e sui bordi delle tavole.

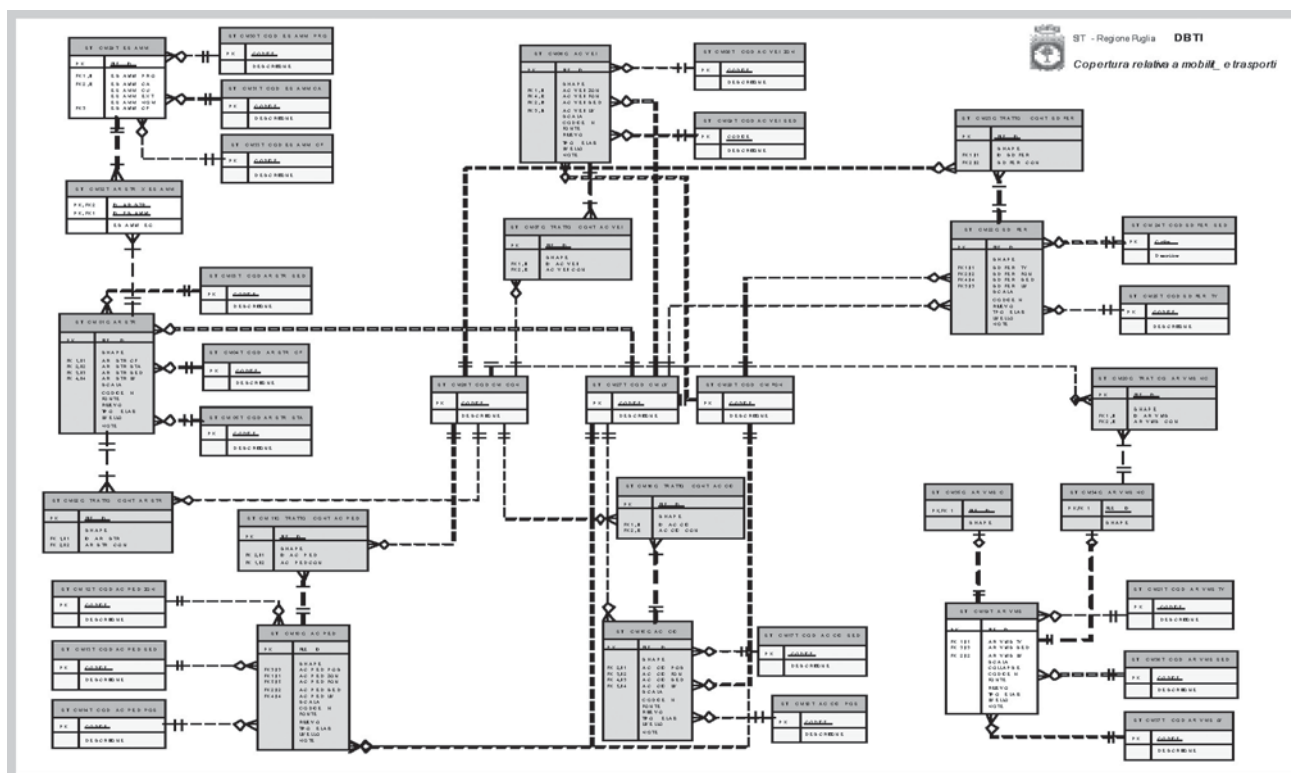


Figura 2 – Schema relazionale dell’area tematica “Copertura relativa a mobilità e trasporti”

Il modello fisico finale: geodb, procedure di validazione e caricamento dei dati

La costruzione del modello fisico tiene conto della tecnologia in adozione per cui traduce il progetto logico in un geodatabase implementato in ambiente ArcSDE + Oracle 10g. Tra le scelte effettuate evidenziamo:

- presenza di unico feature dataset DBTI in cui sono contenute le feature class geometriche; vi sono impostati il sistema di coordinate (UTM WGS84) e l’extent geografico corrispondente all’intero territorio regionale;
- impostazione dei domini per controllare tutti i campi i cui valori prevedono delle codifiche ad hoc; i domini sono tradotti anche in tabelle alfanumeriche per agevolare le procedure di controllo sul rispetto dei domini;
- introduzione di *relationship class* che traducono tutte le relazioni nel modello logico, eccetto quelle di decodifica dei valori dei campi codificati per cui è sufficiente la presenza dei domini; l’uso delle *relationship class* consente di mantenere in fase di editing (aggiornamento) l’integrità e la consistenza dei contenuti geometrici e non;

Il caricamento, in ambiente ArcCatalog, avviene mediante una procedura parametrizzata realizzata come tool di geoprocessing nel linguaggio Python; uno specifico file testuale di configurazione effettua il matching tra file di input (shapefile o file dbf alfanumerico) e tabella di destinazione del modello fisico: un file di log riporta l’esito di ogni operazione evidenziando eventuali anomalie verificatesi in fase di caricamento. Un insieme di tool realizzati con lo stesso approccio consentono di verificare alcuni vincoli importanti previsti sui contenuti del DBT. Le tipologie di controlli effettuati sono le seguenti:

- verifiche generali sui dati dal punto di vista formale (controllo della struttura del tracciato record di ogni tabella, dei valori utilizzati rispetto ai domini codificati, dell’obbligatorietà dei campi, dell’unicità degli identificativi univoci degli oggetti);
- verifiche generali dal punto di vista geometrico (presenza di self-intersection, di geometrie duplicate, di coerenza tra tratti di contorno e frontiera dei poligoni);

- verifiche di carattere specialistico che riguardano le geometrie di singole classi IntesaGIS o di sottoinsiemi di classi (relazioni topologiche).

Anche nel caso delle verifiche particolari, l'implementazione delle procedure di controllo si basa su una forte astrazione per cui i tool effettuano tipologie di controlli attuabili poi su contenuti diversi a seconda della configurazione iniziale; i file di configurazione consentono, nella maggior parte dei casi, di impostare anche dei filtri sui dati. Per esemplificare, riportiamo alcune classi di controlli specialistici che vengono effettuati sui dati:

- verifica della congruenza nella segmentazione fisica tra elementi areali ed elementi lineari dei reticoli rispetto agli attributi comuni;
- verifica che l'unione delle aree di circolazione veicolare sia contenuta nell'unione delle aree stradali (analogamente per le aree di circolazione pedonale e ciclabile);
- verifica di connessione dei reticoli stradali, assumendo che ogni elemento lineare debba essere coperto sulla frontiera da due nodi;
- verifica che gli ingombri al suolo dei cassoni edilizi siano coperti dagli ingombri al suolo degli edifici, che un ingombro al suolo di edificio sia coperto da una o più unità volumetrica per le aree alla scala 1:2000;
- verifica del corretto posizionamento degli accessi rispetto agli edifici.

Di fondamentale importanza è l'insieme di procedure che verificano la corretta copertura topologica del suolo: dato per acquisito l'insieme esaustivo delle classi che partecipano alla copertura areale del territorio (con eventuali filtri), viene testato che non vi siano "buchi" e che non vi siano sovrapposizioni non accettabili tra poligoni della stessa classe e di classi distinte.

Le procedure, eseguite su ciascun lotto di produzione del DBT coincidente con l'area contenuta in un foglio IGM 1:50.000, garantisce in maniera incrementale una buona qualità dei dati in fase di primo impianto del database fisico.

Alcune considerazioni

Rispetto ad una prima ipotesi di derivare il DBT direttamente dalle foto aeree, senza il passaggio intermedio dalla CTR, è stato necessario rivedere il percorso per una serie di motivi, primo tra tutti la difficoltà degli operatori coinvolti di confrontarsi con un nuovo processo e con un nuovo prodotto, la cui formulazione definitiva è stata lunga e laboriosa.

Il modello fisico e le procedure di validazione messe a punto seguono le direttive nazionali seppure con un necessario livello di adattività. Il modello è replicabile e la strutturazione degli shapefile di produzione del DBT compatibile con le linee guida definite dal CNIPA per le regioni aderenti al PR5SIT.

Le procedure di validazione assicurano un buon livello qualitativo dei dati in fase di primo impianto e sono in corso sperimentazioni mirate a valutare modalità e metodologie per un corretto aggiornamento che ne salvaguardi la consistenza.

Riferimenti bibliografici

CNIPA (2008) *Allegato A – Linee guida per la realizzazione dei database topografici nelle Regioni del progetto PR5SIT*

Comitato Tecnico di Coordinamento IntesaGIS (2006) *Il Catalogo degli Oggetti-Revisione delle specifiche di contenuto In1007_1 e In1007_2*

Comitato Tecnico di Coordinamento IntesaGIS (2004) *Specifiche di contenuto-Lo schema del contenuto in GeoUML*