

Tavola rotonda

Verso un sistema unificato per la gestione dei dati di Scienze della Terra?

Premessa

Oreste Signore

CNUCE-CNR, via S. Maria, 36 - Pisa

La gestione integrata dei dati concernenti le Scienze della Terra costituisce indubbiamente un obiettivo estremamente ambizioso, e non pochi esperti nutrono manifesti dubbi sulla fattibilità di un simile approccio. Tuttavia, in molti casi le perplessità appaiono dettate più che altro dalla consapevolezza che le moli di dati da gestire sarebbero di dimensioni probabilmente insostenibili. È evidente, allora, che le obiezioni sono più di carattere tecnico che di principio. Anzi, molti auspicano la realizzazione di un approccio unitario, che permetta di scambiare le informazioni tra le tante iniziative in corso e consenta efficaci economie, evitando la duplicazione di informazioni, dettata spesso solo dalla inaccessibilità o inutilizzabilità di quelle già esistenti.

Quindi la strada percorribile è senz'altro quella di identificare uno schema comune di riferimento, in grado di accogliere i vari aspetti caratteristici delle varie aree specialistiche. La creazione di un'unica base di dati è invece un'operazione per la quale il rapporto costo/benefici è senz'altro troppo elevato.

In figura 1 viene rappresentato in modo molto schematico, e forse anche un po' naïf, il processo che, partendo dalla raccolta dei dati grezzi, porta, attraverso un'opera di filtraggio e di elaborazione in cui sono di importanza fondamentale la competenza scientifica e la sensibilità del geologo, alla produzione della carta geologica, che costituisce da sempre uno strumento di lavoro fondamentale per il geologo.

Il database rappresentato in figura rappresenta una moltitudine di basi di dati, specializzate per tematica e relative a vari livelli di attendibilità dei dati (dati grezzi, dati interpolati, dati validati da prendere come base per la restituzione cartografica, etc.). La rappresentazione come base di dati unica intende enfatizzare il fatto che le

single basi di dati vanno considerate come viste diverse costruite su uno schema di riferimento comune.

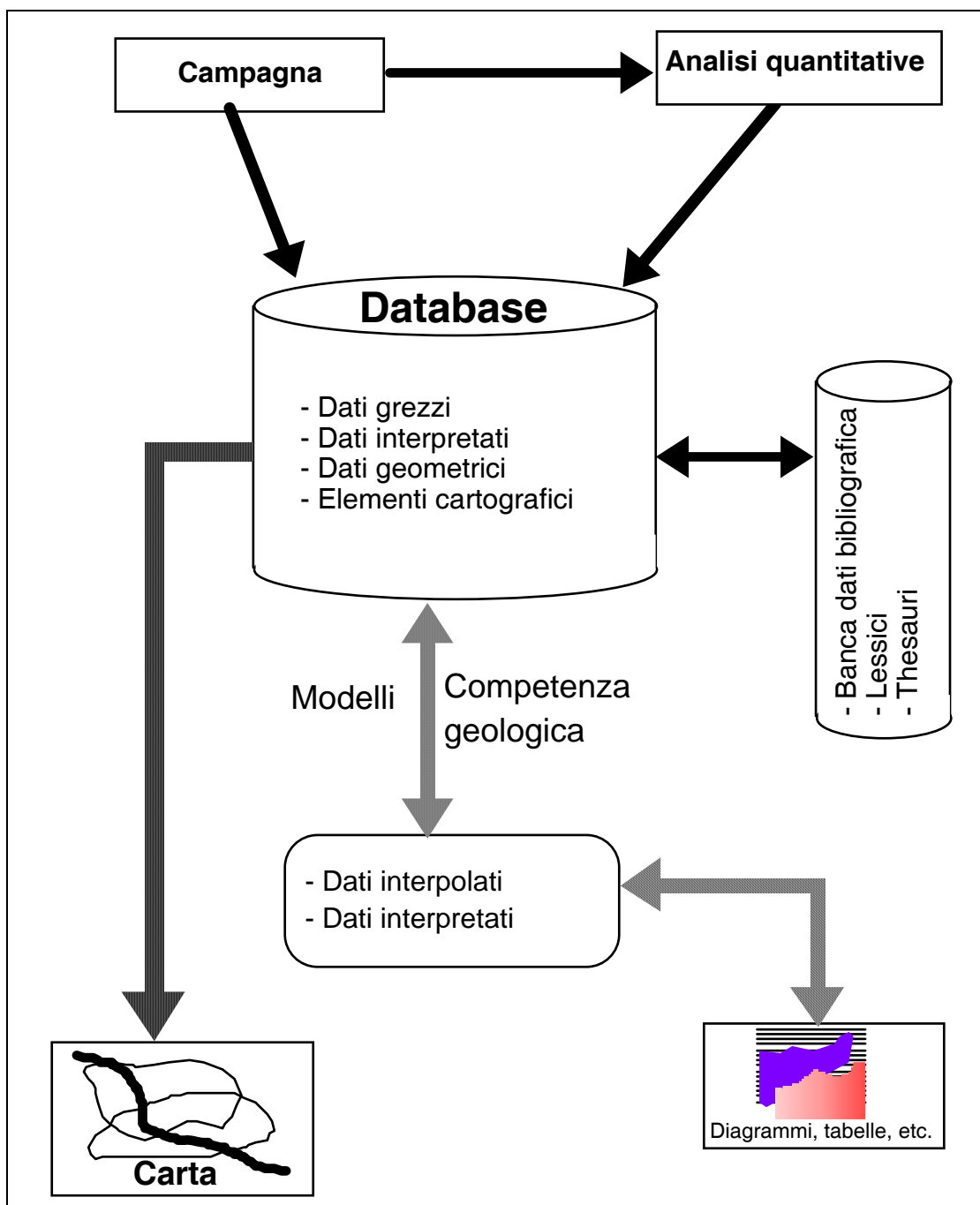


Fig. 1 - Lo schema generale per la raccolta e restituzione di dati di interesse geologico

I problemi fondamentali sono:

- la rappresentazione degli oggetti che posseggono delle ben definite caratteristiche spaziali (elementi puntuali, contorni e linee, elementi areali) con i conseguenti problemi di discretizzazione dell' informazione e della definizione di opportuni operatori spaziali.

- La rappresentazione sia dei dati grezzi e che dei dati interpolati, che vanno ben distinti, in quanto questi ultimi, che sono essenziali perché arricchiti della competenza del geologo, potrebbero in futuro essere modificati alla luce di nuove teorie o conoscenze. In ogni caso, è estremamente importante riuscire a tener traccia dell' apporto di conoscenza fornito in fase di interpretazione.
- La necessità di gestire sia dati di tipo raster, per i quali viene adottato un modello di dati basato su una quadrettatura del piano, che dati di tipo vettoriale, basati su una suddivisione del piano in poligoni irregolari o su una serie di contorni. Inoltre, alcuni autori hanno sollevato da tempo il problema della opportunità di gestire anche una coordinata temporale, per cui ogni dato è identificato dalle quattro coordinate x, y, z, t .
- La normalizzazione dei concetti e la definizione di standard del linguaggio, come componente essenziale in tutte le attività di scambio e integrazione dei dati.

Come si può notare, alcuni di questi problemi sono quelli tipicamente affrontati dai sistemi informativi geografici (GIS), ma è anche evidente che l' adozione sic et simpliciter di un GIS tradizionale non costituisce una soluzione completa del problema, in quanto resterebbero comunque scoperti alcuni dei problemi menzionati. Per ulteriori approfondimenti sull' argomento si rimanda allo special issue " GIS Design Models" di Computer e Geosciences (Vol. 18, N. 4, maggio 1992)

Nel settore della definizione di uno schema comune dei dati, due iniziative italiane di particolare rilievo sono costituite dalla proposta di schema concettuale formulata dal Servizio Geologico Nazionale e dall' esperienza condotta dalla Regione Emilia-Romagna nella produzione della carta 1:10.000. La prima si segnala per l' importanza istituzionale dell' ente proponente, che ha la possibilità di definire degli standard che devono essere rispettati da tutti gli enti che collaboreranno alla compilazione della carta geologica nazionale, la seconda per il rigore scientifico con il quale è stata condotta la progettazione della base di dati e per l' introduzione esplicita di elementi geometrici puntuali, lineari e poligonali.

La normalizzazione dei concetti e della definizione di standard come componente essenziale in tutte le attività di scambio e integrazione dei dati è invece un settore al quale ha lavorato intensamente il Centro Studi Alpi Centrali del CNR, collaborando proficuamente con vari organismi internazionali.

I tre interventi seguenti illustrano, sia pure in modo necessariamente breve, la proposta del Servizio Geologico Nazionale, l' esperienza della Regione Emilia Romagna, il ruolo degli standard geologici.

**Il Disegno Logico della Base Informativa Territoriale del Sistema Informativo
Geologico Nazionale**

Pierluigi Cara

*** intervento di Cara ***

**La “Filosofia” del Modello Concettuale per la Carta Geologica alla Scala
1:10.000 della Regione Emilia-Romagna**

Paolo Artioli

*** intervento di Artioli ***

Integrazione dei Dati: Una Sfida per gli Anni '90

Roberto Potenza

Da molti anni i geologi incontrano problemi di interpretazione quando tentano di arricchire i propri modelli utilizzando dati prodotti da altri; negli ultimi anni, sono sempre più estesamente disponibili fonti di dati, il cui impiego è fondamentale per affiancare il mezzo automatico al lavoro, in passato esclusivamente intellettuale, per la costruzione di modelli geologici e per la produzione di carte. La necessità di disporre di dati omogenei acquista quindi un'importanza crescente sia per il crescere della quantità di fonti e di dati in esse disponibili, sia per la tendenza sempre più diffusa ad inquadrare i problemi locali in un ambito interdisciplinare, regionale o addirittura "globale". Ai tentativi di soluzione proposti da singoli operatori, soprattutto all'interno di gruppi e organizzazioni commerciali e di ricerca, si vanno oggi affiancando approcci coordinati per la soluzione dei problemi di integrazione dei dati di origine e natura differente in modelli complessi.

Questi problemi sono in gran parte intrinseci ai dati e all'ambiente in cui sono stati prodotti, sì che il loro superamento all'origine appare pressoché insormontabile e lo è, in realtà, per i dati già esistenti.

L'inomogeneità dei dati dipende da diversi fattori, a volte isolati, ma più spesso combinati fra loro, tra cui essenzialmente:

- Unità di misura
- Approssimazioni
- Linguaggio
- Codifiche "esterne", adottate dal primo utilizzatore dei dati
- Codifiche interne ai sistemi di archiviazione
- Struttura dei modelli concettuali dei sistemi di archiviazione
- Chiusura dei sistemi di trattamento e rappresentazione dei dati

Conseguenza della situazione oggettiva è la necessità di convertire i dati ricavati da fonti diverse per renderli concettualmente e fisicamente compatibili con il sistema di trattamento che si intende utilizzare. Milioni di righe di programma sono state scritte a questo scopo, ma in molti casi la conversione è onerosa al punto da indurre alla rinuncia all'utilizzo di dati anche potenzialmente importanti.

Dal 1990 l'integrazione dei dati è oggetto di numerose riunioni e convegni, tra cui va ricordata la sessione 22-4 del Congresso Internazionale di Geologia (Kyoto, agosto

1992), a seguito dei quali, fra l'altro, sono state definite le linee guida per la costruzione di basi di dati di uso generale in Scienze della Terra.

Gli elementi principali evidenziati in queste occasioni sono i seguenti:

- architettura delle basi dati orientata all'utilizzatore geologo, piuttosto che all'informatico;
- stabilità nel tempo delle strutture dei dati,
- distinzione tra dati primari e derivati,
- separazione tra dati di interesse locale e dati di uso generale,
- minimizzazione nell'uso di codifiche,
- normalizzazione dei dati archiviati
- standardizzazione dei dati.
- standardizzazione delle procedure

Questi ultimi due punti aprono ovviamente un panorama amplissimo di iniziative per l'adozione e la costituzione di standard, numerici (unità di misura, forma numerica, approssimazioni), terminologici (lessici e thesauri), concettuali (riferimenti cronostatigrafici, paleontologici, petrologici, sedimentologici,...), metodologici (rilevazione dei dati in campagna e in laboratorio), strutturali (modelli concettuali di basi dati), elaborativi (software di gestione e, soprattutto, di interpretazione).

La definizione degli standard geologici è oggetto dell'attività di commissioni dell'IUGS, alcune delle quali operano da alcuni anni e hanno già prodotto i primi risultati praticamente applicabili (nomenclatura delle rocce ignee, scala cronostatigrafica, thesaurus multilingue). Tra i sistemi di riferimento proposti su iniziativa di gruppi accademici o industriali, l'ASTM per le prove sui materiali, è ormai generalmente accettato, mentre il PIDD per la nomenclatura nell'industria petrolifera, è sottoposto dallo scorso anno al collaudo pratico. Un'iniziativa congiunta UNESCO-IUGS ha inoltre gettato le basi per un sistema comune di gestione dei dati geologici dei Paesi africani (PANGIS).

Nonostante queste iniziative, il cammino da percorrere è ancora lungo, e richiede un grande impegno umano e finanziario: ad ogni problema specifico risolto, nuove esigenze si presentano, sia per l'evolversi delle tecniche, sia per l'esigenza di utilizzare dati prima non considerati. Cresce nello stesso tempo l'interesse ad un accesso agevole ai miliardi di dati rilevati ogni giorno, non solo a scopi geologici, in prospezioni, monitoraggi, telerilevamenti, come pure l'esigenza di integrarli in modelli di sempre maggiore complessità e apertura disciplinare. Già ora grandi

organizzazioni pubbliche e private stanno investendo ingenti risorse in quello che si propone come uno degli obiettivi più impegnativi dell'informatica geologica per questo decennio.

Conclusioni: Alcune Esperienze e le Iniziative in Corso

Oreste Signore

A livello italiano, la proposta del Servizio Geologico e l'esperienza della regione Emilia-Romagna individuano chiaramente nel modello entità relazioni lo strumento metodologico più indicato per la descrizione del problema in modo formale, che tuttavia consente un semplice e immediato colloquio tra il geologo e l'informatico.

Da una visita effettuata presso il Geological Survey of Canada (GSC) e lo United States Geological Survey (USGS), è emerso che in entrambi i paesi il processo adottato per passare dai dati raccolti in campagna all'elaborato cartografico è sostanzialmente analogo e riconducibile, settore per settore, a quello descritto in figura 1, anche se la costituzione delle basi di dati è stata fatta in genere senza adottare metodi formali, per cui vi sono casi di incompatibilità tra basi di dati specifiche dei vari settori. In nessuno dei due paesi, però, è stato tentato un approccio globale formalizzato, che viene tuttavia considerato molto interessante e utile, ma anche estremamente impegnativo..

Va però segnalato che presso il GSC la raccolta ed elaborazione dei dati utilizza strumenti informatici fin dalla raccolta dei dati grezzi, per cui già in campagna si ha una prima elaborazione dei dati, effettuata su personal computer portatili, e l'immissione dei dati grezzi può avvenire mediante interazione diretta su una base topografica.

Presso l'USGS è in fase di sviluppo un progetto di grossa portata, il National Water Information System, che ha adottato un metodo più formale per la realizzazione della base di dati, adottando metodologie consolidate (Entity Relationship Diagrams e Data Flow Diagrams) e strumenti di progettazione automatizzati, ampiamente diffusi nel settore dei sistemi informativi. Va sottolineato che il progetto ha una durata prevista di 12 anni (1988-2000) e coinvolge una notevole quantità di risorse umane, sia con competenze scientifiche nello specifico settore che con competenze informatiche. La sola fase di analisi dei requisiti e progettazione della base di dati ha richiesto un impegno di otto gruppi di utenti, che hanno lavorato per circa un anno e mezzo seguendo le direttive di un gruppo di pianificazione strategica composto da dieci top manager. La progettazione della base di dati è stata condotta da gruppi di geologi che, dopo una prima fase di addestramento in cui hanno appreso le basi metodologiche e le modalità di utilizzo degli strumenti, hanno proseguito in piena autonomia, ricorrendo al supporto informatico solo in alcuni momenti di verifica o nei casi in cui era necessario individuare soluzioni informatiche particolarmente sofisticate.

In Italia, da due anni il CNR ha finanziato un progetto che ha come obiettivo la realizzazione di uno schema concettuale integrato per le Scienze della Terra.

Il primo passo effettuato è stato quello di recuperare gli schemi concettuali del SGN e della Regione Emilia-Romagna, adottando strumenti automatici di progettazione (in particolare il tool Bachman) al fine di procedere a una loro integrazione, e successivamente all'arricchimento dello schema risultante con altri aspetti non considerati allo stadio attuale.

Il processo di integrazione dei due schemi è risultato molto più complesso di quanto fosse stato stimato in un primo tempo, perché in effetti i due schemi sono stati costruiti partendo da due approcci diversi. Lo schema proposto dal SGN esplicita molte relazioni di tipo geometrico, e tiene in particolare conto le necessità legate al prodotto finale dell'intero processo, e cioè la produzione della carta geologica. L'approccio adottato dalla regione Emilia-Romagna, invece, evidenzia le caratteristiche geometriche delle varie entità per cui le relazioni spaziali tra di esse devono essere gestite da opportuni operatori.

Di conseguenza, non è possibile procedere a una integrazione dei due schemi adottando le tecniche usuali nel settore della progettazione di sistemi informativi, mentre va operata una scelta metodologica di fondo che non può essere effettuata solo su basi informatiche, ma deve coinvolgere a fondo qualificate e consistenti competenze geologiche.