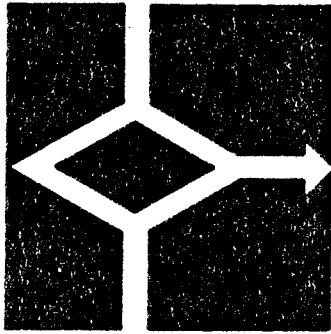


1961



CONGRESSO ANNUALE
ANNUAL CONFERENCE

A.I.C.A. ASSOCIAZIONE
ITALIANA
PER IL CALCOLO
AUTOMATICO

1981

PAVIA 23-25 Settembre 1981

atti
ESTRATTO



LA GESTIONE DEI DATI RAPPRESENTATIVI DI SEZIONI GEOLOGICHE

R. GAGLIARDI * , O. KALIN.** , E. MORREALE *** , E. PATACCA **** , P. SCANDONE ****

R. SPRUGNOLI *** .

(*) c/o Istituto di Geologia e Paleontologia della Università di Pisa

(**) Centro di Studio per la Geologia Strutturale, C.N.R., Pisa

(***) Istituto di Elaborazione della Informazione, C.N.R., Pisa

(****) Istituto di Geologia e Paleontologia della Università di Pisa

*Il modello di Banca Dati ed il relativo Sistema di Gestione descritto in questo lavoro, sviluppato nell'ambito del Progetto Finalizzato "Geodinamica" del C.N.R. ed in corso di avanzata realizzazione presso l'I.E.I. di Pisa, è uno strumento informativo per i geologi e per coloro che operano nel settore delle Scienze della Terra. La possibilità di gestire la totalità dei dati (spesso complessi) provenienti da sorgenti geologiche diverse e l'agevole utilizzazione della Banca Dati Geologici (BDG) da parte di un'utenza che ha poca dimestichezza con il calcolatore elettronico, sono le caratteristiche che distinguono questo sistema da altri funzionanti presso Enti italiani o stranieri. Per questo, una particolare attenzione è stata rivolta alla realizzazione delle due interfacce d'ingresso-dati e d'interrogazione. Per l'organizzazione interna dei dati è stata adottata una struttura "a liste invertite" e la ricerca per chiave avviene attraverso B*Tree. I programmi sono in Assembler 370 e sono implementati su IBM 370/168.*

1. INTRODUZIONE

Nell'ambito di una collaborazione istituitasi tra l'Istituto di Geologia e Paleontologia dell'Università di Pisa e l'Istituto di Elaborazione della Informazione del C.N.R., nel quadro del Progetto Finalizzato "Geodinamica", si è affrontato - a partire dal 1980 - il problema della memorizzazione e gestione di dati rappresentativi di sezioni geologiche di vario tipo (perforazioni profonde in terra ed in mare, dragaggi, sezioni stratigrafiche, gallerie, campionature sparse, etc.).

Tale problema si collega con l'obiettivo di più lunga scadenza tendente ad approntare uno strumento operativo di aiuto nello studio e messa a punto di un modello strutturale tridimensionale del territorio italiano e dei mari adiacenti.

Una volta analizzato il problema ed esaminate le modalità più appropriate per rendere possibile in forma omogenea la raccolta, la rappresentazione e la gestione dei dati relativi ai vari tipi di sezioni geologiche prese in considerazione, il lavoro si è concentrato nella definizione e realizzazione di una Banca Dati Geologici (BDG) e di un Sistema di Gestione (SGBDG) destinato alla più facile accessibilità della BDG. I criteri generali di progetto sono stati principalmente rivolti a garantire:

- la più ampia flessibilità ed articolazione nella descrizione degli oggetti geologici da rappresentare nella BDG,
- la più agevole utilizzazione della BDG da par-

te del geologo, sia nella fase di immissione dei dati, sia nella fase di interrogazione,

ciò al fine di caratterizzare la BDG non solo come struttura di servizio ma anche come strumento di ricerca per il geologo.

In tale senso una particolare attenzione è stata dedicata alla scelta del linguaggio di descrizione, da utilizzare nella fase di immissione dei dati, e del linguaggio di interrogazione, al fine di combinare validamente il rigore necessario per la elaborazione automatica con la utilizzazione delle strutture e dei termini tecnici correntemente impiegati dal geologo nel suo lavoro. La possibilità di accettare dati strutturalmente complessi e la facilità con cui l'utente introduce i dati o interpreta le risposte della BDG, sono gli aspetti che differenziano questo sistema da altri attualmente funzionanti presso enti italiani o stranieri (es. Sistema RIG dell'AGIP, Sistema INGERT, del Servizio Geologico d'Italia, il sistema GEODAS-DCRF dell'Istituto Himeji, Giappone etc.). In questi ultimi l'utente è costretto a codificare i dati e ottiene risposte codificate (e quindi poco comprensibili immediatamente); inoltre l'adozione di codifiche predefinite è spesso inadeguata a descrivere oggetti geologici per i quali è necessario introdurre un vero e proprio "dato descrittivo" articolato, pena la perdita di informazioni geologiche significative. Relativamente alla postelaborazione dei dati risultanti dalle interrogazioni, con particolare riferimento alla presentazione grafica di sezioni di perforazioni geologiche

alla costruzione automatica di vari tipi di carte, si è previsto di affrontare tali problemi successivamente alla realizzazione delle BDG.

Nel seguito del lavoro verrà presentata la soluzione adottata per la BDG, attualmente in corso di avanzata realizzazione e che sarà operativa a Pisa presso il Centro di Studio C.N.R. per la Geologia Dinamica e Strutturale dell'Appennino. La presentazione sarà articolata attorno alla caratterizzazione dei dati (Sez. 2), alla struttura e funzioni del Sistema di Gestione (Sez. 3) ed alla fase di interrogazione (Sez. 4).

2. CARATTERIZZAZIONE DEI DATI

Le attività da svolgere per lo sviluppo di un sistema informativo vengono generalmente raggruppate in tre fasi principali:

- Analisi della realtà per stabilire gli obiettivi,
- Definizione dello schema concettuale del sistema informativo,
- Realizzazione del sistema informativo.

Nell'esposizione di questo lavoro seguiremo l'ordine indicato dai tre punti precedenti, provvedendo a presentare inizialmente la natura delle informazioni geologiche, per poi giustificare la scelta di particolari strutture che hanno reso possibile la loro gestione automatica. Tuttavia ci sembra opportuno ricordare sin dall'inizio, che l'utilità della BDG nell'ambito delle applicazioni geologiche si concentra nella possibilità di risolvere query anche complessi sul suo contenuto. A questo proposito faremo continuo riferimento al seguente esempio informale (non complesso) di "query geologico":

- a) Selezionare la latitudine e la longitudine dei pozzi della Sicilia che hanno incontrato rocce basaltiche appartenenti al Trias.

2.1 Natura dei dati Geologici

Le informazioni geologiche che costituiscono la base informativa provengono da sorgenti diverse: sezioni grafiche di perforazioni a terra o in mare (logs), note descrittive inerenti a campionamenti sparsi, sezioni di gallerie, appunti sulla reinterpretazione di alcune entità geologiche, etc.

Nel seguito indicheremo con il termine "record-geologico", l'insieme delle informazioni fornite da una qualche sorgente sui vari aspetti di un oggetto geologico. Il record geologico contiene, al suo interno, informazioni di natura diversa, che vanno dal nome convenzionale del record all'ubicazione geografica, dalla descrizione dei terreni attraversati alla loro età e appartenenza ad una determinata unità litostigrafica o tettonica, dal numero dei campioni presenti alla descrizione delle singole litolo-

gie, etc.

L'insieme dei record-geologici rappresenta essenzialmente la realtà che la base di dati deve descrivere: il primo passo da compiere è quindi lo studio di questa realtà per individuare l'insieme dei dati che la caratterizzano, tenendo presenti gli obiettivi che si intendono raggiungere. Nel nostro caso, la base di dati geologici (BDG) è stata progettata per descrivere le coperture sedimentarie delle varie unità che costituiscono il territorio italiano e i mari adiacenti. Il contenuto informativo della BDG privilegia pertanto i dati stratigrafici e quelli relativi all'analisi di "facies", trascurando i dati inerenti alle problematiche delle rocce ignee e metamorfiche.

2.2 I moduli di raccolta dati

Il primo frutto della interazione fra geologici ed informatici nella progettazione della BDG, è stata la strutturazione di 19 moduli per la raccolta dei dati, processo che è andato di pari passo con la definizione di un "modello dei dati". I moduli di raccolta dati contengono nel loro insieme la quasi totalità delle informazioni deducibili dalla sorgente informativa che descrive il record geologico in esame.

Ciascun modulo di raccolta dati è identificato dal nome delle entità che descrive tramite l'aggregazione di campi informativi presenti al suo interno: in fig. 1 sono illustrati parzialmente alcuni dei 19 moduli di raccolta.

Il generico modulo può essere sintetizzato nell'associazione di tre campi, nella forma

top; bottom; descrizione .

I campi "top" e "bottom" indicano il limite superiore ed inferiore della profondità in cui è stato incontrato l'oggetto descritto in "descrizione". Ad esempio le informazioni contenute nel modulo AGE di fig. 1 vanno interpretate (se si sta descrivendo il pozzo MODICAL e se l'unità di misura è il metro) come:

"da 780 a 1513 di profondità il pozzo Modical ha attraversato terreni di età Cretaceous, da 1513 a 2820 terreni di età Jurassic etc."

Il tipo di informazioni che compare all'interno del campo "descrizione" può essere un dato elementare o un dato complesso, dipendentemente dall'oggetto che deve essere descritto.

E' importante sottolineare che i moduli sono orientati all'uso immediato da parte dell'utilizzatore della BDG: ciascun modulo, infatti, rappresenta per il geologo che descrive i suoi dati o trascrive dati dalla letteratura, una sequenza logica ed ordinata di campi informativi che possono essere riempiti direttamente senza ricorrere a codifiche. Da ciò deriva che,

LA GESTIONE DEI DATI RAPPRESENTATIVI DI SEZIONI GEOLOGICHE

GENERAL DATA					
RECORD TYPE	<u>well</u>	RECORD NAME	<u>Modica I</u>	RE-EXAM N°	<u>1</u>
OPERATOR	<u>AGIP</u>				
OPERATION DATE	<u>15/4/65</u>	COMPILATION DATE	<u>20/9/63</u>	LAND/SEA	<u>l</u>
CONTINENT	<u>Europe</u>	COUNTRY	<u>Italy</u>	DISTRICT	<u>Sicily</u>
MUNICIPALITY	<u>Modica (RG)</u>	IGM SHEET N°	<u>276</u>	GEOGR SPEC	<u>Ragusa Plateau</u>
OCEAN/FIRST-ORD SEA		SEC-ORD SEA		PHYSIOGR SPECIF	
LAT A	<u>33 53 14</u>	LONG A	<u>19 19 33</u>	LAT B/AZIMUTH	
UNIT OF LENGTH	<u>m</u>	GROUND ELEV/SHA BOT	<u>550</u>	FINAL DEPTH	<u>3060</u>
DEVIATED	<u>n</u>	PRODUCTION	<u>Dry</u>	NUMBER OF SAMPLES/SPACING OF CUTTINGS	<u>43 ; 1/1</u>
TOTAL CORE RECOVERED		% CORE RECOVERY		GEOTECTONIC SETTING	
REFERENCES					

AGE					
Record Name					
TOP	BOT	RELIABILITY	AGE	AGE RELIAB	DATING METHOD
<u>0</u>	<u>180</u>		<u>Langhian</u>		
<u>180</u>	<u>680</u>		<u>Oligocene</u>		
<u>680</u>	<u>780</u>		<u>Eocene</u>		
<u>780</u>	<u>1160</u>		<u>Cretaceous</u>		
<u>"</u>	<u>"</u>		<u>"</u>		
<u>"</u>	<u>"</u>		<u>"</u>		

LITHOLOGIC INTERVAL/SAMPLE DESCRIPTION				
Record Name				
TOP	BOT	LITHOLOGIC INTERVAL	SAMPLE	sample N°
<u>100</u>	<u>1700</u>	<u>Grey marls</u>		
<u>1800</u>	<u>3000</u>	<u>(marls) and (basalts)</u>		
<u>1900</u>	<u>2500</u>	<u>(limestones) intercalated with</u>		
		<u>(shales)</u>		
<u>"</u>	<u>"</u>	<u>"</u>		

LITHOSTRATIGRAPHIC UNITS					
Record Name					
TOP	BOT	RELIABILITY	FORMATION	MEMBER	HORIZON
<u>880</u>	<u>1264</u>		<u>AMERILLO</u>		
<u>1264</u>	<u>1457</u>		<u>HYBLA</u>		
<u>1457</u>	<u>1713</u>		<u>CHIAROMONTE</u>		

Fig.1 - Alcuni moduli di raccolta dati per la BDG.

in alcuni casi, dati costituiti da singoli termini o da liste di termini sono sufficienti a descrivere un oggetto geologico, come per esempio un intervallo litologico, per il quale è necessario specificare i rapporti tra le varie litologie. Un intervallo litologico è ad esempio:

- 1) yellowish dolomites alternating with white calcarenites .

La precisa esigenza avanzata dagli utilizzatori della BDG non solo di mantenere tutte le caratteristiche di informazioni come quelle in 1), ma di individuarle attraverso singoli componenti (nell'esempio specifico: yellowish dolomites; calcarenites; dolomites alternating with calcarenites), ha portato alla utilizzazione di un "linguaggio di descrizione" parentizzato.

Una ulteriore proprietà che caratterizza alcuni dati geologici, è la loro appartenenza ad insiemi di termini legati da una relazione gerarchica. L'esempio più evidente è fornito dalle età geologiche - AGE -: Il Pleistocene fa parte del Quaternary che a sua volta fa parte del Cainozoic e così via. In fase di interrogazione è necessario tenere conto di questa proprietà: infatti per la risoluzione di a) non è sufficiente considerare rocce basaltiche appartenenti al Trias, ma anche rocce appartenenti a Rhaetian, Norian, Carnian etc. In (1) sono elencati i campi che contengono dati appartenenti ad insiemi gerarchici e sono riportati gli alberi di termini che descrivono le gerarchie.

2.3 Modello dei dati

Dopo aver individuato l'insieme di dati che rappresentano i fatti noti della realtà in esame, la seconda operazione da svolgere per lo sviluppo di un sistema informativo è la definizione del "Modello dei dati", cioè lo strumento che rappresenta formalmente il contenuto della Base di dati.

La quantità e la complessità delle informazioni geologiche ricavabili dai "logs" o da altre sorgenti, fa sì che lo schema concettuale della realtà da rappresentare non è riconducibile a nessuno dei tre modelli di dati più popolari (gerarchico, a rete, relazionale). Tuttavia, parlando dei moduli di raccolta, si è già accennato al fatto che i dati geologici si prestano ad essere raggruppati naturalmente in modo tale che ogni gruppo definisce una precisa entità geologica.

La definizione dei moduli di raccolta dati è stata caratterizzata da un processo di astrazione che ha permesso di concentrarsi sugli aspetti essenziali della realtà in esame. In questo modo l'organizzazione dei moduli offre una base valida per la definizione di un Modello dei dati, in termini di entità o associazione e

proprietà: ad esempio l'entità Lithostratigraphic-units, è costituita dalla aggregazione di altre entità: top, bot, formation, member e horizon, che ne diventano i costituenti, e quindi "proprietà" di Lithostratigraphic-units. L'estensione di una entità composta, ovvero l'insieme dei valori che assumono le proprietà che la identificano, può essere pensata come una tabella bidimensionale, caratterizzata dal nome dell'entità: ogni colonna della tabella è associata ad una proprietà e contiene i valori da essa assunti.

Una tabella di questo tipo può essere rappresentata dal generico modulo di raccolta: ad esempio il modulo AGE rappresenta l'entità AGE, costituita dalle proprietà: top, bot, (name-of) age, ed ogni riga del modulo è una n-pia di AGE.

Questo modo di rappresentare il modello dei dati fa pensare all'approccio relazionale descritto da Chen in [8]: tuttavia esistono differenze sostanziali - che verranno evidenziate in seguito - per cui non è sufficiente assumere il modello strettamente relazionale per la realtà in esame. Assumeremo invece, per comodità, parte della terminologia in uso per il modello relazionale, sostituendo i concetti di entità, associazione e proprietà, con quelli di relazioni e attributi.

2.4 Tipologia dei dati

Prima di passare alla descrizione del "Modello dei dati", è opportuno soffermarci sulle caratteristiche dei dati per i quali è stata prevista la gestione dal SGBDG. Oltre ai dati primitivi (STRING , INTEGER , REAL etc.) è stato necessario definire un dato strutturato che permetta la manipolazione di informazioni come la descrizione 1). Questo tipo di dato viene "smistato" su due moduli diversi, poiché il geologo è interessato al suo reperimento attraverso la relazione esistente tra le litologie (dolomites alternating-with calcarenites) e/o attraverso la specifica e la qualificazione dei singoli litotipi (dolomites, calcarenites, white calcarenites, yellowish dolomites). La descrizione dettagliata dei singoli litotipi viene fatta sul modulo Specific Lithology Description (cfr. [1]) dove sono previsti campi per informazioni fino al livello microscopico relativo all'oggetto in esame.

L'associazione tra litologie viene trattato nel modulo Lithologic Interval/Sample Description, ricorrendo all'uso di un linguaggio standardizzato di descrizione che fa uso di parentesi. L'associazione di due litologie può essere espressa nella generica forma:

litologia-1: identificatore di relazione:
Litologia-2 .

In 1) si ha, ad esempio:

- 1a) dolomites alternating with calca-

renites.

Isolando tra parentesi tonde gli oggetti alla destra e alla sinistra dell'identificatore di relazione, 1a) diventa:

1b) (dolomites) alternating-with (calca renites).

In questo modo è possibile riconoscere l'identificatore di relazione degli oggetti che correla. L'estensione a casi più complessi è immediata. Ad esempio:

2) dolomites alternating with calcarenites with intercalation of basalts

viene trascritta nel modulo come:

2a) ((dolomites) alternating-with (calca renites)) with-intercalation of (basalts)

oppure:

2b) (dolomites) alternating-with ((calca renites) with-intercalation of (basalts))

dipendentemente dal fatto che "with intercalation of" associ "dolomites alternating with calcarenites" e "basalts" oppure "calcarenites" e "basalts".

Quest'ultimo esempio spiega il perchè dell'introduzione di un linguaggio standardizzato: accanto alla esigenza di lasciare pressochè inalterata la struttura originale della descrizione e alla possibilità di reperirla attraverso specifiche su componenti che compaiono in essa o attraverso associazioni di litologie esiste il problema dell'ambiguità della descrizione stessa, che avrebbe impedito la gestione automatica del dato geologico senza l'adozione di una regola di standardizzazione.

2.5 La correlazione tra entità geologiche

Descriviamo ora il modello dei dati per l'esempio di fig. 1. Nel seguito faremo riferimento a questo modello per illustrare alcuni concetti essenziali, perciò abbiamo ritenuto opportuno semplificarlo, omettendo alcuni campi che compaiono nei moduli di raccolta. Indichiamo con R $[a_1, a_2, \dots, a_n]$ la relazione R definita sugli attributi a_i per $i = 1, 2, \dots, n$. In questo modo, abbreviando opportunamente i termini, le entità che compaiono in fig. 1 vengono rappresentate da:

Relazioni

GN np,rt, rn, lat, long
AG np, top, bot, age
LU np, top, bot, for, mem, hor
LI np, top, bot, desc

Attributi

np numeric for string
top numeric men string
bot numeric hor string

rt string des string (string)
rn string
lat numeric
long numeric
age string

Il campo "np" non è presente nei moduli di raccolta: tuttavia viene creato e riempito automaticamente dal SGBD al momento dell'ingresso dati con un "numero progressivo di record geologico". Tale numero progressivo serve ad indicare per ogni riga del modulo e quindi, a livello di modello di dati, per ogni n-pla di una qualunque relazione, a quale record geologico la n-pla appartiene.

Questo attributo, insieme a "top" e a "bot", permette un certo tipo di "correlazione" tra relazione diverse: il termine "correlazione" non va inteso nella sua accezione generica, ma con un preciso significato geologico. Ad esempio, dire che: il Pozzo Modica 1 ha attraversato fra 780 e 1160 m terreni "Cretacei" caratterizzati dalla presenza del fossile Globotruncana e dalla litologia "calcarei biancastri marmosi" significa associare implicitamente oggetti contenuti in relazioni diverse (età, fossili, intervalli litologici) attraverso l'identificatore del record geologico (Modica 1) e un intervallo di profondità (780-1160). In definitiva gli attributi "np", "top" e "bot" fungono da coordinate per l'individuazione di un qualsiasi oggetto presente nella base di dati, e permettono l'associazione di oggetti diversi: su queste considerazioni si basa il concetto di query geologico (cfr. cap. 3).

3 DESCRIZIONE DEL SISTEMA DI GESTIONE

In questa parte si intende analizzare la struttura complessiva del Sistema di gestione della base di dati Geologici (SGBDG) da un punto di vista realizzativo, illustrandone le funzioni e descrivendo le strutture dati adottate per la implementazione del sistema. In fig. 2 è rappresentato lo schema complessivo del SGBDG, costituito da:

3.1 Interfaccia d'ingresso dati

Con interfaccia d'ingresso dati si intende un insieme di programmi che gestiscono le operazioni di immissione dati. La caratteristica fondamentale di questi programmi è la loro autonomia dalle routine principali (blocco centrale). L'adozione di un'interfaccia a questo livello è motivata dal fatto che i potenziali utilizzatori della BDG non hanno dimestichezza con gli strumenti di collegamento con il calcolatore (terminali) o con il linguaggio strettamente tecnico di un sistema conversazionale. L'esigenza fondamentale dell'utente si traduce essenzialmente nella richiesta di uno strumento di facile uso e di un linguaggio quanto più possibile vicino al linguaggio naturale.

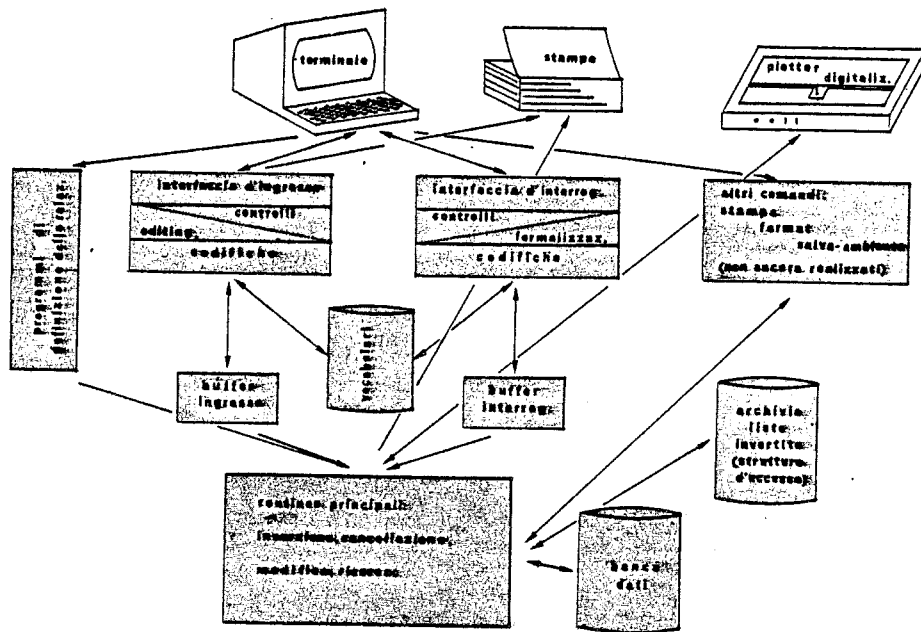


Fig.2-Struttura complessiva del Sistema di Gestione della BASE di DATI GEOLOGICI

Un aspetto ulteriore che interviene a giustificare l'adozione di un'interfaccia di ingresso dati riguarda la possibilità di determinare vincoli di integrità o di correttezza sui dati geologici, ovvero condizioni a cui i dati devono sottostare per essere significativi. La realtà geologica permette di stabilire un buon numero di questi vincoli; è possibile ad esempio prevedere l'estremo superiore o inferiore o l'intervallo dei valori assumibili da un attributo, verificare l'appartenenza di un dato ad un vocabolario prestabilito o esaminare la correttezza sintattica di una descrizione come la 1b) (cfr. par. 1.4). Particolarmente utili, in questo contesto, sono i vocabolari dei termini standard, che costituiscono uno strumento valido per la classificazione degli oggetti geologici. Il linguaggio geologico è infatti per sua natura ambiguo e scarsamente rigoroso, e non di rado uno stesso oggetto viene classificato con nomi diversi.

Le operazioni assolate dall'interfaccia di ingresso sono: a) controllo; b) codifica; c) editing. I programmi operano su un'area di lavoro in cui possono essere memorizzati e trattati circa 30 record geologici diversi. La procedura di input avviene tramite la richiesta all'utente di introdurre il valore di un campo, seguendo l'ordine in cui gli stessi campi compaiono nei moduli di raccolta.

Le verifiche che vengono fatte sui dati, oltre ai controlli elementari su un dato numerico (intero o reale) o alfabetico sono: l'apparte-

nenza ad un determinato intervallo di variabilità, l'assunzione di un valore in conseguenza del valore di un altro dato, il controllo sintattico dei dati strutturati.

L'appartenenza ad un intervallo di variabilità interessa valori numerici o stringhe alfabetiche. In quest'ultimo caso viene controllato l'appartenenza del termine ad un vocabolario. I vocabolari sono mantenuti su dischi magnetici e acceduti sequenzialmente o per ricerca binaria, dipendentemente dalla loro dimensione.

Il controllo sintattico dei dati strutturati del tipo 1b) è effettuato da un programma che realizza un riconoscitore sintattico a due pile (si tratta di un automa a stati finiti, dovendo riconoscere oggetti appartenenti ad un linguaggio libero).

Alcuni dati vengono codificati per limitare l'occupazione di memoria, altri, invece, per permettere l'espansione di una interrogazione quando questa coinvolga termini appartenenti a vocabolari gerarchici. Ad esempio, l'età Quaternary è codificata con A, Olocene e Pleistocene (figli di Quaternary) con AA e AB rispettivamente, Versilian e Tirrenian (figli rispettivamente di Olocene e di Pleistocene) con AAA e ABA etc. La generazione dei codici non presenta grosse difficoltà poiché, dovendo controllare l'esattezza dei termini nei vocabolari, vengono create all'interno di questi le tabelle di corrispondenza: generalmente per la codifica è stata usata la sequenza alfabetica.

Sono stati realizzati comandi di editing che operano sul contenuto dell'area di lavoro, divisi in due gruppi: un primo gruppo è orientato a fornire informazioni generali riguardanti il numero e lo stato di record geologici e dei vocabolari non vuoti, nonché ad eseguire le operazioni di cancellazione, inserzione e stampa dei termini di un vocabolario. Questo perché alcuni vocabolari non sono statici, ma, in seguito alla introduzione di termini nuovi è necessario il loro aggiornamento. Il secondo gruppo di comandi provvede alla cancellazione o alla inserzione di uno o più dati di un record geologico, al reperimento, alla cancellazione e alla stampa di tutte le informazioni relative ad un modulo di raccolta. Ciò permette al Responsabile della Base di dati di operare controlli e modifiche nei dati in ingresso per migliorarne la qualità.

3.2 Routine centrali

Questa parte descrive le routine centrali del Sistema di Gestione. Come struttura di partenza è stato utilizzato il Sistema RESP, ideato presso l'Istituto di Elaborazione della Informazione del C.N.R. di Pisa. Questo sistema, inizialmente progettato per la gestione di un archivio di dati, è stato modificato per la gestione di più archivi. Sono rimasti tuttavia inalterati i comandi di "data definition", con i quali l'utente può definire un archivio dando le specifiche di ogni campo informativo (chiave, non chiave, con descrittori chiave o non chiave, numerico, alfabetico etc.)

L'organizzazione dei dati, del tipo "a liste invertite", fa uso delle seguenti strutture dati:

- 1) archivio principale: è un archivio che può contenere un numero elevato di record (oltre 8 milioni) di lunghezza massima di 800 byte, divisi in campi: ogni campo non può superare 256 byte di lunghezza, e può essere di lunghezza fissa o variabile. Un campo può essere definito campo "chiave" e, in questo caso, può apparire all'interno di un query. Il contenuto di un campo può essere un valore elementare o un insieme di valori: nel secondo caso il campo è definito "campo con descrittori". I descrittori sono considerati come unità singole, che possono essere a loro volta campi chiave.
- 2) Dizionario dei termini: in questo archivio sono contenuti i valori delle chiavi. Il dizionario è strutturato a B*Tree: i nodi terminali dell'albero contengono i puntatori alle liste invertite. Ogni valore nuovo di una chiave viene inserito nel B*Tree e viene creata una lista invertita.
- 3) Archivio delle liste invertite: contiene le liste invertite. Ogni elemento della lista è un puntatore al record contenuto nell'archivio principale nel quale compare il valore

della chiave associato alla lista.

- 4) Archivio ausiliario: serve per l'esecuzione delle operazioni contenute in un query geologico. In questo archivio viene memorizzato il risultato parziale (in termini di indirizzi) di una condizione semplice "attributo-relatore-valore" contenuta in una espressione logica, oppure il risultato parziale di una espressione logica in cui compaiono condizioni sul contenuto di una stessa relazione.

- 5) Buffer pool: area di lavoro permanente, collocata in memoria principale, di 16 pagine di 800 byte. L'inserzione di un dato L'inserzione di un dato provoca l'allocazione del dato stesso nell'archivio principale e, se il dato è chiave, viene creato un riferimento nel dizionario dei termini (se la chiave non è ancora presente) e viene inizializzata (o aggiornata) la relativa lista invertita. Il reperimento delle nple attraverso il valore di una chiave, avviene con le seguenti operazioni: si cerca nel dizionario dei termini (ovvero nel B*Tree associato alla chiave) il valore della chiave e si individua il puntatore alla lista invertita di quel valore, si scandisce la lista invertita estraendo gli indirizzi relativi delle nple nell'archivio principale, si reperiscono le nple. Le routine che implementano le operazioni precedenti interagiscono con le interfacce d'ingresso e d'interrogazione.

4. IL QUERY GEOLOGICO

La definizione di un "query geologico" costituisce la parte centrale di questo lavoro.

L'interrogazione accettata dal SGBDG consiste in una domanda nella quale è individuato l'oggetto della ricerca e sono definite le condizioni al contorno che si desiderano soddisfatte: questo orientamento è stato a lungo discusso con i geologi, futuri utenti del sistema proposto, che lo hanno dichiarato idoneo alle loro esigenze.

Adottando la terminologia introdotta per il modello dei dati in par. 1.5, una domanda geologica si traduce nella selezione di un sottoinsieme dei valori assunti da uno o più attributi di una o più relazioni: tale insieme è individuato da condizioni poste sul valore di altri attributi, espresse esplicitamente nella interrogazione. Occorre tenere presente che, come già accennato nei paragrafi precedenti, gli attributi "np", "top" e "bot", il cui contenuto è analogo per ogni relazione, permettono di associare oggetti provenienti da relazioni diverse, cioè permettono la "navigazione" all'interno della base dei dati.

Tenteremo di dare un procedimento formale che chiarisca e sintetizzi il processo di interpretazione e risoluzione di un "query geologico".

Consideriamo ancora a). Facendo riferimento al

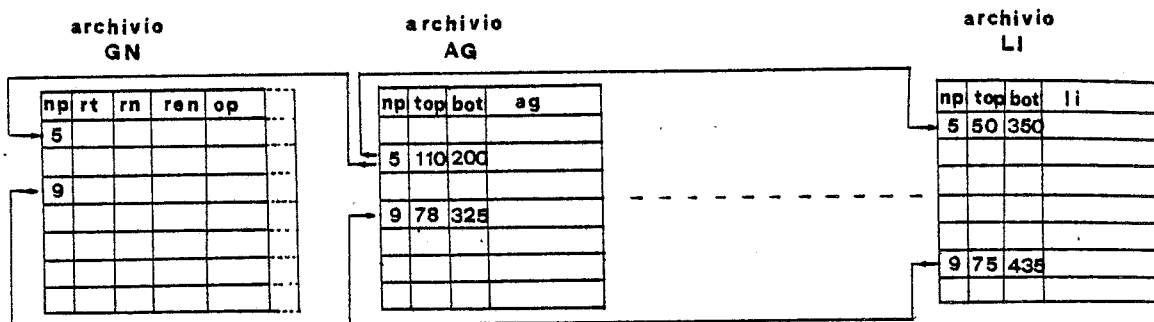


Fig.4 -Correlazione tra i dati geologici attraverso "np-top-bot"

modello dei dati introdotto in par 1.5, le relazioni e gli attributi coinvolti in a) sono:

- GN (general notes) poiché l'oggetto della ricerca sono valori di "lat" e "long" ed una condizione a cui deve soddisfare l'insieme risultato è $rt = \text{POZZO}$ (dove rt è il "tipo di record geologico").

AG (età) per la condizione $age = \text{TRIAS}$

LI (intervalli litologici) per la condizione des (descrizione) = **BASALTO**

Indicheremo con $R[A_1, A_2, \dots, A_n]$ le n-ple della relazione R a_1, a_2, \dots, a_n caratterizzate dai valori A_1, A_2, \dots, A_n di a_1, a_2, \dots, a_n : chiameremo dominio di un attributo l'insieme di valori che questo assume in R . Sostituiranno inoltre gli attributi "top" e "bot" con l'attributo z che rappresenta la "quota di profondità": questa semplificazione rende la trattazione formale del query geologico più compatta, poiché faremo riferimento non più ad intervalli di profondità in termini di estremi superiori e inferiori "top" e "bot", ma come insiemi di quote "z".

L'algoritmo per la risoluzione di a) può essere sintetizzato nei seguenti passi:

a.1) Si determina

$$K [np] = \{ np \mid \exists z (AG [np, z, age=TRIAS] \cap LI [np, z, des=BASALTO]) \}$$

a.2) Si determina

$$R [lat, long] = \{ (lat, long) \mid \forall (rn) GN [np \in K, rt = \text{POZZO}, rn, lat, long] \}$$

In questo caso la profondità non interviene nell'individuazione del risultato, poiché "lat" e "long" sono attributi della relazione GN: questa relazione contiene informazioni anagrafiche di carattere generale, e non ha nessun riferimento alla profondità.

Se l'interrogazione fosse:

b) Seleziona le formazioni incontrate in un qualsiasi record geologico, che contengono rocce basaltiche appartenenti al Trias.

La soluzione di b) è data dalle seguenti azioni:

b.1) Si determina:

$$H [np, z] = \{ (np, z) / AG [np, z, age = \text{TRIAS}] \quad LI [np, z, des = \text{BASALTI}] \}$$

b.2) Si determina

$$E[form] = \{ form \mid \forall (mem, hor) LU [(np, z) \in H, form, mem, hor] \}$$

In questo caso i valori di "form" che soddisfano b), sono reperiti attraverso "np" e "z", poiché la relazione LI contiene il riferimento alla profondità.

I due esempi precedenti illustrano come viene interpretata un'interrogazione geologica: è tuttavia necessario soffermarci sul significato geologico di "inclusione" tra tre intervalli di profondità in termini di "top" e "bot". Supponiamo di dover risolvere b) e di avere memorizzato i seguenti dati relativi allo stesso record geologico 10:

MODULO AGE			MODULO LITHOLOGIC-INTER.
10	80 -90	dogger	10 100-175 basalts
10	100-140	trias	10 75-110 (marls) and (basalts)
10	145-270	permian	10 137-255 shales
10	275-578	carboniferous	10 268-400 basalts

Gli intervalli che soddisfano: "rocce basaltiche appartenenti al "Trias"" sono in questo caso

100-110 e 275-400.

In termini formali, esprimiamo questo fatto, riconducendoci ancora alla profondità "z", come:

$$R[NP, TOP_1, BOT_1] \cap R[NP, TOP_2, BOT_2] =$$

$$= \left\{ R^*[np, z] \mid \begin{matrix} TOP_1 \leq z < BOT_1 \\ TOP_2 \leq z < BOT_2 \end{matrix} \right\} \cap$$

$$= \left\{ R^*[np, z] \mid \begin{matrix} \max(TOP_1, TOP_2) \leq z < \\ \min(BOT_1, BOT_2) \end{matrix} \right\}$$

avendo espresso:

$$R[NP, TOP, BOT] = \{ R^*[np, z] \mid TOP \leq z < BOT \}.$$

Il query geologico, nel modo in cui viene impostato dall'utilizzatore, ha la struttura seguente:

```
Select REL1. ATTR1, ATTR2, ..., ATTRN: REL2.
ATTR1....:
Where REL2. condizione1
REL3. condizione2
" " "
End
```

Select indica l'oggetto della ricerca, espresso da identificatori di relazioni e di attributi; where indica la lista di condizioni che devono essere soddisfatte, sulle singole relazioni.

La condizione è espressa da una combinazione di condizioni elementari legate da operatori " \wedge " e " \vee " e separate da parentesi. La forma sintattica di una condizione elementare è "attributo-relatore-valore", dove attributo è identificatore di un attributo chiave; relatore è uno dei simboli =, #, <, >, >=, <= (inizia per); valore è una costante numerica o una stringa. Gli attributi che compaiono in una condizione devono appartenere alla stessa relazione.

Le condizioni su relazioni diverse sono correlate dagli attributi "np", "top" e "bot" (cfr. fig. 4), come già illustrato nella prima parte di questo paragrafo: ad esempio l'interrogazione b) viene impostata come:

```
Select LU. FORM
where AG. AGE = TRIAS:
LI. DES = BASALTS

end
```

Un esempio più complesso è:

```
C) Select GN. LAT, LONG: LU. FORM
where AG. AGE = TRIAS AGE = DOGGER
GN.RT = WELL
LI.DES = (limestone) with intercalations of (basalt)

end
```

L'interrogazione così formulata viene analizzata da un programma di interfaccia che ne verifica la correttezza ed esegue la codifica dei termini. La correttezza della interrogazione viene testata sia a livello sintattico, sia a livello di significatività dei termini (controllo sul

campo di variabilità, appartenenza ad un vocabolario etc. I termini associati ad un qualche vocabolario (gerarchico o no), sono codificati attraverso le tabelle di corrispondenza poste all'interno dei vocabolari stessi. In fase di codifica i dati strutturati subiscono una ulteriore modifica: vengono eliminate le parentesi e sostituite da operatori " \wedge ". Ad esempio la descrizione di c) viene codificata come: "des = limestone \wedge des = with-intercalation-of \wedge des = basalt". Infatti il dato strutturato viene trattato come valore di un attributo con descrittori chiave, e per ogni termine elementare del dato viene creato un riferimento nelle inversioni relative a quell'attributo.

Ciò che viene recuperato in questo modo è un soprainsieme delle registrazioni che contengono il dato strutturato: per c) si ritroveranno ad esempio, non solo registrazioni contenenti "(limestone) with-intercalations-of (basalt)", ma anche "(basalt) with-intercalation-of (limestone)".

Geologicamente le due descrizioni non rappresentano lo stesso oggetto, perciò un programma di postelaborazione provvederà ad eliminare i dati strutturati che non verificano le condizioni poste nell'interrogazione.

L'estensione dell'interrogazione, quando questa coinvolge termini appartenenti a vocabolari gerarchici, avviene tramite la codifica dei termini in questione e l'utilizzazione dell'operatore ".". (incomincia per). Ad esempio in C) la condizione sulle età geologiche viene modificata in: "AGE.DED \wedge AGE.FEE" essendo DED e FEE le codifiche di "TRIAS" e "DOGGER" rispettivamente.

A tutt'oggi lo stato di avanzamento del progetto può riassumersi nell'aver realizzato l'interfaccia d'ingresso, testata su alcuni pozzi della Sicilia, e nell'aver modificato le routine del sistema RESP, originariamente monorelazionale, per la manipolazione di più relazioni. Sono in via di completamento le routine che gestiscono il "querygeologico". E' nostra intenzione iniziare nel dicembre 1981 la realizzazione dei programmi per le elaborazioni grafiche delle uscite della BDG. Alcune note tecniche: i programmi che costituiscono il SGBDG sono scritti in Assembler 370 e sono implementati su IBM 370/168 del C.N.U.C.E di Pisa.

Riferimenti bibliografici

- [1] Gagliardi, R., Kalin, O., Morreale, E., Patacca, E., Scandone, P., Sprugnoli R., (1980) "Una banca dati geologici del Progetto Finalizzato Geodinamica", (Pubblicazione n. 393 del progetto Finalizzato Geodinamica).
- [2] Bragazzi, F., Grassi, A., Viviani, A., "Implementazione del metodo d'accesso per un sistema di gestione di Basi di dati", (corso di laurea in Scienze della Informazione, Università di Pisa, Anno Accademico 1979-80).

- [3] Maccherani, C., Matteucci, L., Menchetti, M., "Gestione dell'ingresso dati per un sistema di gestione di dati geologici", (corso di laurea in Scienze della Informazione, Università di Pisa, Anno Accademico 1979-80).
- [4] Lafitte, P., "Traitè d'informatique géologique" (1972, Masson, Paris).
- [5] Progetto Finalizzato Geodinamica, Unità Operative 5.3.10, 5.3.3, "Banca Dati Geologici - Guida alla trascrizione delle informazioni da sezioni geologiche (Modello Preliminare)", (Pisa, Aprile 1980, Nota Interna I.E.I. - C.N.R., F80-24).
- [6] Alexander - Marrack P. O., Friend, P. F., Yeats, A. K. (1971) - Mark - Sensing for recording and analysis of sedimentological data - Data processing in Biology and Geology - Edited by Cutbill (Academic Press, London and New York, p. i. - 16).
- [7] Mubaux, A., (1969) - The description of geological object - (Ronéo.CETIS, Euratom, Ispra).
- [8] Chen, P.S., The - entityRelationship - model toward a Unified View of Data, (Tods, 1,1, March 1976, pp.936)
- [9] T.V. Loudon, Computer Methods in geology, - (1979, Academic Press, London and N.York).
- [10] K.Wadatsumi ed al., GEODASDCRF; development of a relational database system, (Computer and Geosciences, 2,3, 1976, pp.351356).