

ROBERTO POTENZA

METODO PER LA CONSERVAZIONE, IL REPERIMENTO
E L'ELABORAZIONE DEI DATI GEOLOGICI (*)

(Nota preliminare)

SUMMARY. — A method for storage and retrieval of geological data is proposed. It requires no previous format conditions for data to be stored, except a ten digit specification must precede every data set. Although use of this method can be made in every branch of Geological Sciences, at present retrieval programs for use in Geochemistry and Petrology only have been written. As FORTRAN IV language is always been employed, modifications and enlargements of the retrieval system are readily feasible.

Introduzione.

In questi ultimi anni anche in Italia è andato diffondendosi sempre più l'impiego dei calcolatori elettronici per la soluzione di problemi geologici, in concomitanza con l'aumentata necessità di trattare le grandi masse di informazioni acquisite con le tecniche di studio più aggiornate e con la opportunità di confrontare i propri dati con quelli ricavati da pubblicazioni o trasmessi da colleghi italiani o stranieri. Le precise esigenze degli elaboratori elettronici impongono però che i dati siano formalmente omogenei, il che comporta trascrizioni successive con conseguenti possibilità di errori e soprattutto con notevoli perdite di tempo. Il ricorso all'elaboratore si rende d'altra parte indispensabile anche nella fase di archiviazione e ricerca quando sia necessario trattare rapidamente grandi quantità di informazioni.

In sede internazionale l'I.U.G.S. prevede l'organizzazione di archivi di informazioni geologiche predisposti in modo da facilitarne la utilizzazione su vasta scala, ed ha promosso a tale scopo la costituzione

(*) Lavoro eseguito nell'ambito dei programmi del «Centro Nazionale per lo studio geologico e petrografico delle Alpi», sez. IV, del Consiglio Nazionale delle Ricerche.

di un comitato internazionale per lo studio dei problemi relativi. In Italia queste ricerche sono ancora nella fase iniziale: se si eccettuano archivi interni di industrie minerarie e una proposta recente dell'Università di Perugia (CORADOSSI & FARAONE, 1968), praticamente nulla è stato fatto in questo campo.

Gli indirizzi geologici nei quali in Italia si è cominciato a mostrare interesse al trattamento automatico dei dati, e quindi ad una corretta impostazione dei problemi relativi all'informazione, sono la Geochimica e la Petrologia e, successivamente, la Paleontologia e la Sedimentologia. Entro breve tempo però anche le altre discipline afferenti alla Geologia dovranno inserirsi in questa problematica.

Incoraggiato dai colloqui avuti col Dott. Hubaux del C.E.T.I.S. (C.C.R. - EURATOM), mi sono quindi preoccupato di studiare la possibilità di organizzare i dati petrochimici, petrografici e geochimici, rientranti nel mio campo di interessi specifici, e di studiare a livello delle tecniche dell'informazione i metodi migliori per l'acquisizione e la registrazione dei dati, sia in ordine ad una loro utilizzazione il più possibile aperta, sia in funzione di un loro eventuale inserimento in archivi internazionali. In questo studio mi sono valso dei risultati del Committee on Storage and Retrieval of Geological Data, canadese, di cui, tra l'altro, ho adottato con qualche aggiunta la lista dei nomi abbreviati dei minerali (tab. 1).

Sono così giunto a puntualizzare alcuni criteri generali relativi all'organizzazione di un sistema di archiviazione e reperimento dati la cui realizzazione, a livello dell'Istituto di Mineralogia di Milano, è già giunta alla fase operativa. L'allargamento del sistema al trattamento delle informazioni di diversa provenienza rende ora necessario studiare le differenti esigenze degli eventuali utilizzatori. Quanti ritengono di poter essere interessati ai problemi proposti nella presente nota sono quindi invitati a prendere contatto direttamente con l'autore, sì che lo sviluppo del sistema possa procedere in aderenza alle necessità pratiche di ognuno.

Scelta dei linguaggi e delle codificazioni.

Tra i principali requisiti di un repertorio di dati mi è sembrata fondamentale la possibilità che un ricercatore se ne serva senza dover ricorrere a tecniche estranee al suo campo di conoscenze; l'accesso alle informazioni deve quindi essere previsto anche per chi non abbia espe-

TABELLA 1.

Elenco dei nomi abbreviati dei minerali.

ACNL	Actinolite	JSPR	Diaspro
ADLS	Andalusite	KLNT	Caolinite
AHPL	Antofillite	KYNT	Cianite
ALBT	Albite	LMNT	Limonite
ALTR	Alterazione	LPDL	Lepidolite
AMPH	Anfibolo	LUCT	Leucite
APTT	Apatite	MGCL	Microclino
ASBS	Asbesto - Crisotilo	MGNS	Magnesite
AUGT	Augite	MGNT	Magnetite
BOTT	Biotite	MICA	Mica
BRCT	Bruceite	MLBD	Molibdenite
CHMS	Chamosite	MRCs	Marcasite
CHRT	Chert	MSCV	Muscovite
CLAY	Argilla	NPLN	Nefelina
CLCT	Calcite	OLVN	Olivina
CLCP	Calcopirite	OPQS	Opachi
CLPN	Collofane	ORCL	Ortoclasio
CLRD	Cloritoide	PFDP	Feldspato potassico
CLRT	Clorite	PLGC	Plagioclasio
CMNG	Cummingtonite	PLGP	Flogopite
CNCR	Cancrinite	PNLD	Pentlandite
CRBC	Sostanze carboniose	PRNT	Prennite
CRBN	Carbonati	PRTH	Pertite
CRDT	Cordierite	PRTT	Pirrotina
CRMT	Cromite	PRXN	Pirosseno
CRND	Corindone	PYRT	Pirite
CRNT	Coronite	QRTZ	Quarzo
CTRT	Cassiterite	RBCK	Riebeckite
DLMT	Dolomite	RUTL	Rutilo
DPSD	Diopside	SCPL	Scapolite
EPDT	Epidoto	SDLT	Sodalite
EVPR	Evaporiti	SLMN	Sillimanite
EXNT	Euxenite	SLPD	Solfuri
FLDR	Feldspato	SPDM	Spodumene
FLRT	Fluorite	SNDN	Sanidino
GLCN	Glauconite	SPHN	Titanite
GLCP	Glaucofane	SPLR	Blenda (Sfalerite)
GLEN	Galena	SPNL	Spinello
GLSS	Vetro	SRCT	Sericite
GPSM	Gesso	SRPN	Serpentino
GRNT	Granato	SSRT	Saussurite
GRPT	Grafite	STRL	Staurolite
GTHT	Goetite	TALC	Talco
HBLD	Orneblenda	TMLM	Tormalina
HMTT	Ematite	TNNT	Tennantite
HPRS	Iperstene	TRDR	Tetraedrite
HPRX	Ortopirosseno	TRML	Tremolite
ILMN	Ilmenite	WLSN	Wollastonite
IRCB	Carbonato di ferro	ZLTS	Zeoliti
IRXD	Ossidi di ferro	ZRCN	Zircone

rienza di calcolo automatico, purchè segua materialmente poche e semplici istruzioni. Se però si rende necessario modificare le modalità di impiego dei dati, è desiderabile che non siano indispensabili conoscenze troppo approfondite nel campo della programmazione.

```

99999 999 ISTITUTO DI MINERALOGIA, PETROGRAFIA E GEOCHIMICA
99999 998 GABBRO DI SONDALO DISPONIB 50
99999 997 CAMPIONATURA GDS SIGLA RPVT DATA 070063 COP 50 DIST MED M 300
99999 997 CAMPIONI SCELTI MEDIANTE ESTRAZIONE STOCASTICA FRA CAMPIONATURE PRECE
99999 997 IDENTI
99999 2997 CAMPIGLIO POTENZA-FACIES DIORITICHE-ATTISOC SC NAT 103 336
99999 3997 CAMPIGLIO POTENZA-LE FACIES OLIVINICHE-ATTISOC SC NAT 105 102
99999 4997 CAMPIGLIO POTENZA-FACIES A PIROSSENO ROMBICO-ATTI SOC SC NAT 106 193

402 132TP S03603870 QUOTA 1200 DATA 070063 POTENZA
402 1P TAZZALE AEM PRESSO FUSINACCIA
402 1-DIORITE- IN STUDIO COME STANDARD PER ANALISI DI ROCCE IN FLUORESCENZA
402 1 IX OTT 1966
402 2SIO2 AL2O3FE2O3FED MNO MGO CAO NA2O K2O TIO2 P2O5 H2O
402 352.8 16.7 1. 8.38 0.13 4.46 7.07 3.17 1.36 2.98 0.13 1.42
402 1ANALISTA CAMPIGLIO 1964
402 4PLGC AMPH QRTZ BOTT ILMN ALTR
402 552.2 20.5 11.5 10.3 3.2 2.3
402 4ALTR = PER LO PIU A PRODOTTI URALITICI

400 132TP S04843461 QUOTA 1090 DATA 1965 POTENZA
400 1200 M A SE E 70 M PIU IN BASSO DELLE CASE DI MENEGAI
400 10DLVN 2VX AMPH Z/C AMPH 2VX PYRX Z/C HPRX 2VX
400 11 83. 13. 86. 42. 74.

```

Fig. 1. — Organizzazione dei dati da archiviare. Le linee verticali evidenziano le divisioni delle prime 10 colonne in 4 campi: 1° numero d'ordine, 2° codice per i commenti, 3° codice di continuazione, 4° codice delle indicazioni cui si riferiscono i dati.

E' questa la ragione per cui, dopo aver esaminato la possibilità di impiegare linguaggi di programmazione più adatti alla manipolazione di dati misti (COBOL, PL1), o più razionali per la soluzione dei problemi di macchina (linguaggi « assemblatori » tipo MAP), ho preferito ancora ricorrere al FORTRAN IV, la cui conoscenza ha larghissima diffusione tra studiosi e ricercatori nel campo scientifico.

L'uso di calcolatori di grande potenza e la possibilità di registrazione su dischi possono infatti contenere in tempi assai ridotti le elaborazioni, anche le più complesse, richieste per la soluzione di taluni problemi. Se quindi è vero che con l'uso di linguaggi appropriati è possibile una maggiore efficienza, è forse preferibile accontentarsi di un livello di efficienza minore, ma ancora ragionevolmente elevato, a vantaggio di una più larga possibilità di intervento nei programmi di utilizzazione.

Sebbene dalle considerazioni ora formulate emerga l'opportunità dell'impiego del FORTRAN, esso pone il problema della codificazione dei dati. La sua rigidità imporrebbe infatti, per ottenere la massima efficienza di calcolo, una codificazione completa, possibilmente nume-

rica; d'altra parte esigenze di chiarezza e di rispetto per i criteri personali del raccoglitore delle informazioni, vorrebbero che la registrazione avvenisse nella forma originaria. Poichè ciò non è sempre possibile in quanto spesso in contrasto con la elaborabilità delle informazioni registrate, si è cercato un compromesso che permetta un razionale immagazzinamento del materiale con alterazioni minime o nulle delle informazioni originarie. Si è così adottato a livello del nostro Istituto un sistema di organizzazione delle informazioni articolato nel modo indicato nella tabella 2. Le condizioni indicate, va subito precisato, non sono tassative, tuttavia, con il sistema attualmente predisposto per il trattamento, modificare alcune di esse implica alterazioni complesse del sistema di elaborazione, con aumento dei tempi e dei costi. In ordine di difficoltà possono essere apportate le modifiche seguenti:

1) Eliminazione del vincolo sulle prime dieci posizioni di ogni « scheda » d'ingresso. Richiede rimaneggiamenti sostanziali nei programmi di elaborazione.

2) Soppressione del codice. Richiede un ordine costante nell'introduzione dei dati e rende necessaria una elaborazione preliminare per il loro riconoscimento. I numeri di codice possono invece essere variati senza inconvenienti.

3) Soppressione del numero d'ordine. Richiede un programma preliminare per l'assegnazione di un numero di riconoscimento.

Altre condizioni possono essere variate senza difficoltà, pur comportando talvolta un leggero aumento dei tempi di elaborazione, così per esempio l'adozione di formati di lettura particolari o la lettura a formato libero sono possibili ma meno economici delle condizioni qui indicate. Possono essere invece eliminate con vantaggio le indicazioni ripetute delle determinazioni eseguite in serie omogenee di dati.

Si può notare come, nella tabella, il nome del tipo litologico appare nel « commento »: tale posizione non deve dare l'impressione che l'informazione vada perduta perchè, come si è detto prima, tutti i dati registrati sono utilizzabili nell'elaborazione. Si è però ritenuto poco prudente fare del nome un elemento primario di selezione dei dati, in quanto esso è sempre legato ad un particolare criterio di classificazione petrografica, cioè ad elementi assai meno obiettivi dei parametri

misurati sul campione. D'altra parte, per le categorie di rocce che ancora attendono una classificazione quantitativa, nelle quali il nome è tuttora elemento imprescindibile di designazione, l'elaborazione del nome è prevista senza particolari difficoltà.

Organizzazione dei dati.

Il primo passo nell'archiviazione dei dati è la loro suddivisione in insiemi costituiti da tutti i dati relativi ad un « campione », termine con il quale indicheremo una entità geologica indivisibile, sia essa un campione materiale di roccia o minerale, oppure una stazione di misure strutturali o geofisiche, o qualsiasi altro ente definito in una ricerca. I dati omogenei verranno riuniti in una o più « schede », cioè in unità di ingresso di 80 caratteri, che non saranno necessariamente schede perforate. All'interno di un blocco di dati i campioni si possono disporre in un ordine qualsiasi, purchè sia rispettata la continuità delle informazioni relative ad un unico campione; entro il blocco di dati relativi al campione, d'altra parte, l'unica esigenza di ordine è quella di rispettare l'ordine: dati descrittivi-dati numerici, in sequenza biunivoca, quando i primi esistano.

Nemmeno queste esigenze sono tassative in assoluto ed è prevedibile la loro decadenza, ma per ora si ritiene consigliabile soddisfarle, sia per ragioni di relativa semplicità delle operazioni di analisi dei blocchi di dati, sia per una elementare esigenza di ordine.

Nel sistema studiato le schede descrittive dei caratteri relativi al campione sono contrassegnate da un numero di codice di 3 cifre, per cui sono distinguibili fino a 999 categorie di informazioni. Nel nostro sistema il codice è 1 per le informazioni generali relative al campione, 2 e 3 per le analisi chimiche di rocce o minerali, 4 e 5 per le analisi modali di rocce, 6 e 7 per le determinazioni chimiche parziali e per le analisi di tracce, 8 e 9 per gli stessi con indicazioni sulla determinazione (p. es. terne di valori indicanti: dato, deviazione standard, numero di determinazioni eseguite), ecc. Il primo dei due codici indica di regola che la scheda contiene informazioni alfabetiche relative ai dati numerici riportati nella scheda contraddistinta dal codice successivo. Come si è detto queste regole sono del tutto contingenti e i codici possono assumere qualsiasi valore si desidera, purchè sia omogeneo in uno stesso blocco di dati sufficientemente grande (possibilmente entro lo stesso ente che opera la ricerca il codice sia unico).

Quando i dati occupano più di una scheda vanno ripetuti il numero d'ordine ed il codice, a cui si premette un numero progressivo da 1 a 9; ogni serie di informazioni può quindi essere distribuita su 10 schede, con una disponibilità totale di 700 caratteri alfanumerici. Quando poi si vogliono aggiungere dati descrittivi per esteso, commenti, bibliografie, ecc., un numero di una cifra premesso alle quattro suddette permetterà l'introduzione di informazioni sia dopo una serie di dati di un determinato tipo, sia in riferimento al campione nel suo insieme; possono così essere aggiunti, ad ogni gruppo di dati designati da un codice, 6300 caratteri alfanumerici di commenti, anch'essi interamente elaborabili, più che sufficienti per qualsiasi esigenza.

TABELLA 2.

*Programmi per il trasferimento, l'archiviazione, il reperimento
e la trascrizione dei dati.*

Nome	Funzioni svolte dal programma
TRASFE	Trasferisce i dati, scritti in condizioni standard, da schede perforate a nastro magnetico.
LIBER	Interpreta e ricolloca i dati scritti su schede senza formato definito.
ARCHIV	Programma principale del sistema di reperimento e trascrizione dati. Accetta istruzioni su schede perforate in ordine ai criteri di selezione ed al tipo di dati di cui è richiesta l'uscita. I dati da selezionare sono su nastro magnetico.
STRECK	Seleziona il materiale petrografico in base alla classificazione modale di Streckeisen (1967).
NIGGLI	Seleziona i campioni di rocce in base ai « valori Niggli ».
PARG1	Seleziona i campioni in base ad un parametro prefissato entro limiti definiti.
PARG2	Seleziona i campioni in base a più parametri prefissati entro limiti definiti.
SEL	Raccoglie i dati di selezione elaborati dai programmi precedentemente utilizzati e predispone per l'uscita i dati richiesti nella forma voluta (schede perforate, nastro, disco magnetico).

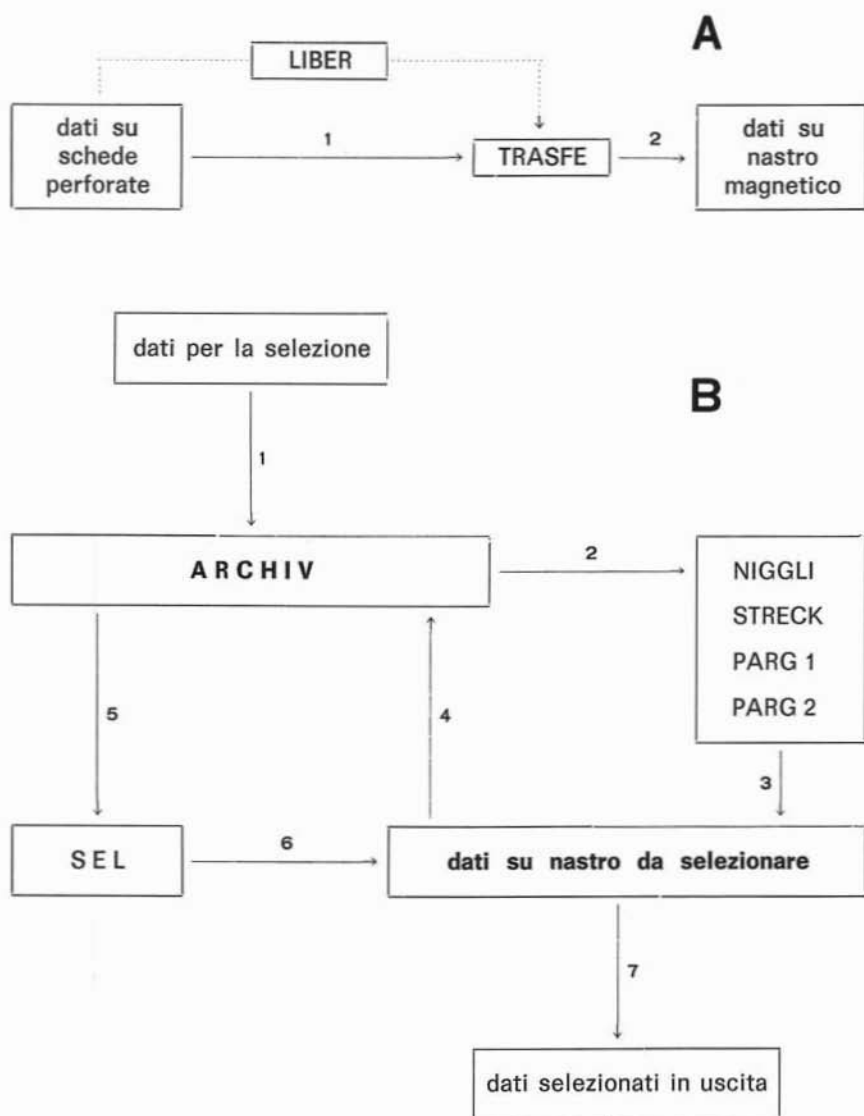


Fig. 2. — Schema delle operazioni che vengono eseguite per l'archiviazione (A) e per il reperimento dei dati (B):

In A i dati perforati su schede vengono, se necessario, interpretati e riordinati dal programma LIBER, quindi trasferiti su nastro magnetico. In B i dati di selezione su schede perforate vengono letti dal programma ARCHIV (1) il quale fa intervenire uno dei programmi asserviti (2). Questo legge i dati su nastro, li elabora se necessario (3), e li confronta con quelli indicati nelle richieste (4). Il controllo passa allora al programma SEL (5) il quale rilegge i dati (6) e pre-dispone quelli richiesti per l'uscita (7).

I programmi di elaborazione dei dati.

L'elaborazione dei dati avviene, come si è detto, a mezzo di programmi in FORTRAN IV a struttura il più possibile semplice. Quelli scritti finora prevedono l'uso di nastri magnetici, ma la conversione per i dischi è immediata. I programmi attualmente predisposti sono quelli elencati in tabella 2. I primi due sono programmi preliminari per l'organizzazione di dati scritti a formato libero e per la loro trascrizione su nastro magnetico, il quale sarà l'archivio in cui verranno conservati i dati. La ricerca, selezione ed estrazione avvengono mediante una serie di programmi, anch'essi trascritti su nastro, i quali formano una struttura complessa che opera sotto il controllo di un programma principale su schede perforate. E' quest'ultimo che, secondo le esigenze, fa intervenire i singoli programmi asserviti per governare l'uscita dei dati desiderati.

Attualmente è possibile ottenere la scrittura di qualunque dato inserito nell'archivio scegliendolo in base al tipo petrografico (classificato sec. Streckeisen), ai parametri magmatici di Niggli, o a qualsiasi altro parametro o gruppo di parametri prefissati in un dato intervallo.

Questi programmi coprono le principali esigenze della ricerca petrolologica e geochimica nel nostro Istituto. Altri criteri di reperimento ed utilizzazione possono però venire agevolmente introdotti nel sistema in vista di esigenze di diversa natura.

Questo tipo di organizzazione presenta perciò notevoli vantaggi rispetto ad un grande programma unico: innanzitutto qui vengono richiamate le sole sezioni utili, garantendo quindi una rapida esecuzione, inoltre la segmentazione dei programmi permette, evolvendosi i criteri classificativi ed interpretativi, di adeguarsi con poco sforzo alle nuove esigenze, mentre la semplicità dei singoli programmi permette questi adeguamenti anche da parte di personale con una piccola base di esperienza nella programmazione dei problemi scientifici (ed è qui che si dimostra l'utilità di un linguaggio di largo impiego come il FORTRAN), infine, con il presentarsi di necessità imprevedute, nuovi programmi di ricerca-dati possono essere predisposti in poco tempo dall'utilizzatore stesso ed aggiunti al sistema. Altro vantaggio da sottolineare è la possibilità di servirsi del materiale archiviato nella sua completezza, analizzando anche i concetti registrati in forma discorsiva; tali analisi sono

già possibili in parte, ma diverranno assai più redditizie in un prossimo futuro quando l'impiego di calcolatori di grande potenza renderà economiche le complesse elaborazioni necessarie a questo scopo.

Conclusioni.

Da quanto detto finora si può concludere che l'utilità del metodo proposto appare attualmente valida e destinata a valere anche in seguito all'evoluzione di nuovi criteri per il trattamento dei dati, nonchè all'introduzione di sistemi internazionali di archiviazione delle informazioni geologiche. Allargando le possibilità di trattamento, pur senza imporre rigide codificazioni comuni, il sistema proposto permette di superare le difficoltà di organizzazione dei dati che sorgevano finora ogni volta che si proponeva un nuovo problema, riducendole a problemi materiali di programmazione e di scrittura su schede. La possibilità di comparare dati di più fonti ripropone però urgentemente il problema della soggettività delle informazioni, in quanto misure fisiche di parametri eseguite su una entità scelta con criteri soggettivi costituiscono a loro volta un insieme di valori correlato in modo soggettivo ai caratteri dell'insieme entro cui è stata scelta l'entità misurata. Questo grande problema da cui spesso si crede di poter prescindere nel comunicare i risultati delle proprie ricerche, infirmandone così buona parte del valore, non è stato risolto nel sistema proposto, in quanto uno studio in tal senso sarebbe esulato dal tema affrontato; non è stata però nemmeno trascurata la sua esistenza e sono stati cercati dei criteri preliminari atti a migliorare la significatività del campione. Va comunque tenuto presente che, rispetto ai sistemi finora proposti, questo pone il minor numero possibile di trasposizioni concettuali fra il ricercatore e i dati archiviati, avvicinando così notevolmente la soluzione di tale problema.

*Istituto di Mineralogia, Petrografia e Geochimica dell'Università di Milano.
Marzo 1969.*

BIBLIOGRAFIA

- CORADOSSI N. & FARAONE D. - *Archiviazione e trattamento di dati bibliografici petrografico-geochimici*. Assisi 1968.
- NATIONAL ADVISORY COMMITTEE ON RESEARCH IN THE GEOLOGICAL SCIENCES (CANADA) - *A national System for storage and retrieval of geological data in Canada*. Geological Survey of Canada, 1967.