

# 地質・資源情報の分散型データベース・システム

古宇田 亮一 (鉱床部)  
Ryoichi KOUDA

## はじめに

情報があふれる世の中で どうやったら情報をうまく整理して利用することができるでしょうか。これは誰も頭を悩ますことではないでしょうか。たくさんある情報の中から 今必要なものを探り当てて役に立てるには 人手とファイル箱が使われてきました。あるいは文献カードが。それでも処理しきれない 膨大な情報の処理の為に 電子計算機を使うことができます。電子計算機は使い方次第で便利にも不便にもなり得る道具です。電子計算機を使った情報処理には ファイル処理と呼ばれる ファイル箱の記憶事項を電子計算機の記憶する所に 固有の計算プログラムで記憶させるやり方が使われてきました。しかし ファイル処理では 特定のプログラムに左右されるために 情報の互換性や共用化する点で困難がありました。

この難点を打開してデータを広く共用化できるための 最有力な方法として データベースという概念が確立されてきました。データベースとは 何なのでしょう。データベースを扱い 利用し データベースを開発していくには どんな心構えが必要でしょうか。小論は データベースを地質・資源情報に適用していく上で 必要な概念・述語の使い方を試み 試作中の地質・資源デ

ータベース・システムの基本設計思想を説明して できる限り多くの人々にデータベースの使い方に慣れ親しみ 開発研究に参加していただくことを願って書かれています。用語などには耳新しいものも少なくないと思いますが慣れていただければ有益であろうと思います。

## 分散処理システム

工業技術院の筑波研究センターに このほど大型計算機システムが導入され 9つある研究所の研究情報処理に活躍を期待されています。この大型計算機システム (以後 RIPS と呼びます。研究情報処理システムの略) は 中心に大容量記憶装置として50億バイト以上の利用領域を持ち 中央処理装置 (CPU) として11.8 MIPS以上の演算速度と1600万バイトの記憶空間を有するファコムM 200を2台 密結合しています。将来利用が増大した時は 更に規模を拡大できることになっています。中央サービス室には 昔から使われてきたバッチ処理 (たとえばパンチカードに入れられた情報をプログラムと共に一度に流して処理します) や 10年前から日本でも普通になった TSS 処理 (タイム・シェアリング・システムの略。たとえばテレビ画面を使って 多人数が同時使用していながら あたかも自分しか使用していないかのように意のままに大型計算機

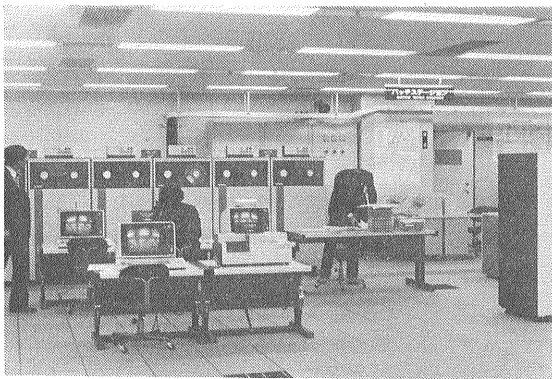


写真1 RIPS サービスセンター内のバッチ処理システム。手前のテレビ画面は入出力情報とその制御 及び一時的なTSS処理に使われる。奥に5台並んでいるのは磁気テープ装置。

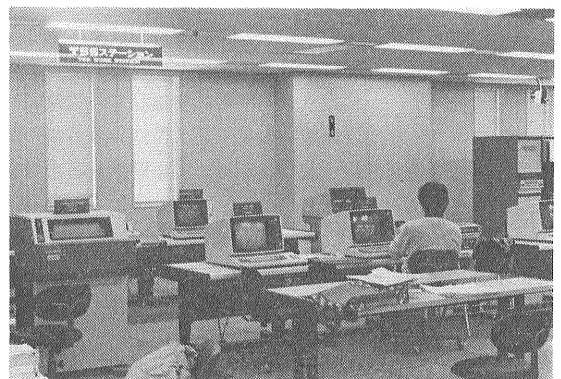
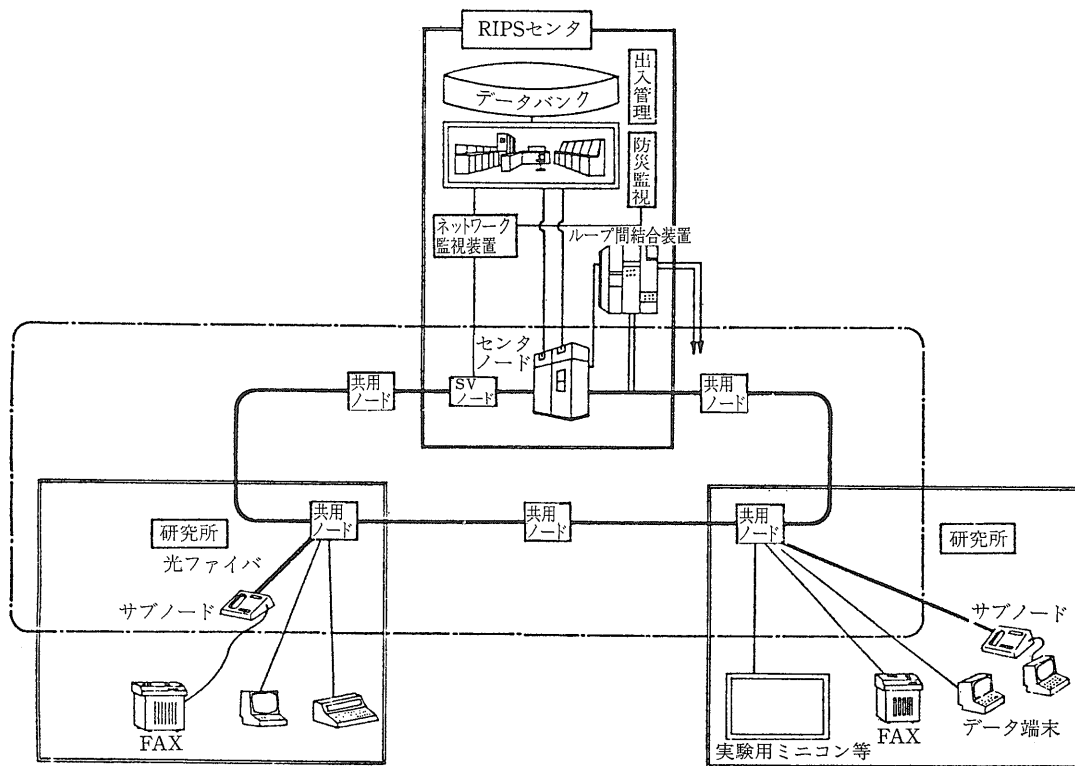


写真2 RIPS サービスセンター内のTSS処理システム。CRT (陰極線管) が多数並び同時に複数の人が処理できる。漢字端末も用意されている。



第1図 RIPSにおける共用回線網の概念図。

RIPSセンターから各研究室に光ファイバが張りめぐらされ 共用ノード (モデム: 本文末参照) によって端末装置  
やミニコンと接続することによりデータ通信が実現される。 (図は RIPS ニュース第1号1980.7.30p.3より転載)。

を操ることができます) 及び 図形・画像処理や日本語処理のほか 多種の周辺機器を揃えています。各研究所にも遠隔サービス・ステーションがあって ほぼ同様の機能が備えられているか 装備可能になっています。従来から各研究所で作動していた電子計算機を集中的に大型化したものですが各研究所にとっては今までと変わりなく運用できます。

しかしこれだけでは在来型の集中処理方式と原理的には大差ないのですが PIPS の実現によって大変便利になったことは 光ファイバを回線に使った共用端末網が手に入ったことにあるでしょう。第1図は 共用ネットワークの概念的あらましです。大規模計算処理や大容量記憶は RIPSセンターの大型計算機が担当します。その運転やデータの入出力は 各研究所の各研究室にある端末と周辺機器装置で全て制御できるようになっています。バッチ処理も可能ですが 普通は TSS 処理を使います。つまり 計算機使用者は わざわざ RIPSセンターや 各研究所のサービス・ステーションまで足を運ぶことなく 各自の居室や実験室に居ながら意のままに

大型計算機を使うことができるわけです。このような汎用回線網アーキテクチャ いいかえると データを処理するためにデータ通信チャンネルで結合される通信手段を使って遠隔地で操作することを一般に分散処理と言います。

分散処理は集中処理に対抗する方法で ここ10年来次第に整備されてきた最新技術です。米国フィリップスのマッ格林博士によれば 「マイクロプロセッサ・衛星伝送システム・回線網アーキテクチャ・通信プロトコル……が駆使される……分散計算・分散データベース・分散制御と監視の意味をもつ……分散処理システムは ごく最近出現したもの」(D.R. Mc-Glynn, 1978) ですが 現在及び将来のデータ処理を根本的に変革し続ける原動力となっています。

### 地質・資源データベース

地質・資源情報の最近の増加傾向は目を見張らせるものがあります。産業活動を支えるエネルギー資源・鉱物資源は使った分だけ再利用しない限り減ってゆきます。

したがって地下に資源が埋蔵されている場所 つまり鉱床は 少なくとも掘り出した分量以上に新しく発見するための探査活動がなければ 安定的な供給が確保できません。たとえば石油危機になったからといってすぐに石油が無くなってしまったわけではありませんでした。掘り出す石油の量にくらべて新しく発見される石油の埋蔵量が追いつかなくなったために 石油問題があれほど騒がれるようになったわけです。確かに石油を含む地下資源は いよいよ発見しにくくなってきています。探査を続けるためには より多くの情報を使って 新しい探査理論を研究することが必要です。

資源探査に用意される情報は極めて膨大なものです。地形図・地質図・鉱床図・変質図を始めとして 地化探データや 重力・磁力・電気探査のデータ 地震・音波探査 リモート・センシングから 各種データを解釈して現実の鉱床探査発見に役立てるための種々の基礎的支援データ 地球物理・地球化学から諸基礎科学のデータを駆使する必要がある このためのデータ処理に大変な時間と労力と資金がかかるカラクリになっています。さらに文献データから適切な情報を得るために 文献検索処理も欠かすことはできません。このような科学的資源探査にとって “兵糧” としてのデータを 利用し易い形に整理して収納するのがデータベース・システムといえるでしょう。私たちの環境としての地質・地震・火山情報や 水資源・海洋資源情報も データベースとして利用し易い形にすることが 公共的・産業的要請として今後高まることが予想されます。私たちはそれに備えておく必要があります。

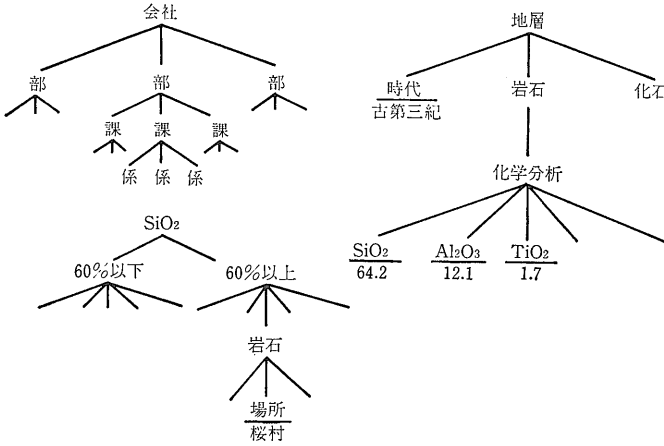
地質・資源情報のデータベースを作成する時 地質情報の性格・その情報構造・形態的特質が考慮されねばなりません。それらを考慮した上で 分散処理ネットワークシステムの上に どんなデータベース管理システムを実現していくのかを見通さねばなりません。地質・資源情報は 現在も各個バラバラに さまざまなやり方で分散されています。分散した研究現場に密着した個々の研究者志向の強い しかも経済的な計算機利用を早急に実現し 将来の変革にも十分耐えうるシステムを求めるとすれば分散処理が最適であるといえるのです。その実用例として 試作段階にあるシステムの設計概念と 地質・資源データベースの望ましい現実的な形について 以下紹介していきます。はじめにデータベース設計の基になるデータモデルの知識を一望しましょう。

## 階層型とネットワーク形のデータモデル

データベースの発想は 使用頻度がかかなり多くて しかも共同で使えるデータを なるべく一つにまとめて反復利用しようというところにありました。そのためにデータを集中制御し かつ共用化することが必要になります。データの呼出しや追加・修正・削除などが なめらかに実行されることも必要です。利用する立場から言えば いろいろな応用形態が考えられるので 各種の応用プログラムに左右されないデータ管理が必要です。データベースの出現以前では 計算機上で ファイル情報処理という形がとられていました。ファイル処理は次の諸点で壁にぶつかります。

- ① 電子計算機出現以前の 人手によるデータ整理・収納・利用形態を そのまま計算機にやらせただけにすぎない。
- ② 利用者の応用プログラムの付属品として扱われているため 他の応用プログラムと共用しにくい。
- ③ 情報の特質がファイル記憶媒体の物理的性質に左右されるため表現されにくい。たとえばパンチカードに打つために 専門家ですら匙を投げてしまうほど用語の無理な省略を重ねたり 複雑で覚えにくく従ってまちがえやすいコーディング・シートが用意されて手作業が大部分になってしまうことなど。
- ④ 本来 計算機がその能力を発揮するはずの 記憶媒体上にデータファイルを実現するプログラムが あまりにも簡素で低い水準でしかなかった。
- ⑤ データの変更や拡張 データ構造の変更などが容易ではなかった。そのため一人の利用者ですら 多数の重複するファイルの管理で制約を受けていた。たとえば 一地域のデータを入力するにあたって 入力時の水準で必要と思われるものしか入力しなかったために 追加データがなめらかに入らず 全て入力し直すのに等しい浪費が ごく普通のこととしてあきらめられていた。

以上のような困難な壁を破るものとして データベース管理システムが整備されるようになります。その源泉は 1955年の W. C. McGee 考案の Report Generator に遡ることができます。その後 商用システムや標準化の発展に合わせて IBM 社の IMS (情報管理システム) データベースが完成され 豊富な経験を持つシステムとして市販され 以後数多くのデータベース管理システムの模範となり なおも改良が重ねられています。同時に 米国防総省の肝入りで発足したデータシステムズ



第2図 階層的な情報構造。

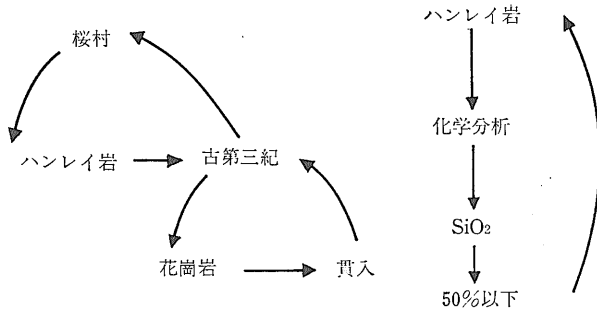
左上図は会社組織の一例。ある係を捜したい時にはまずある会社を指定し特定の部や係をたどることで発見できる(はずである)。実用上はもっと巧みなやり方で検索される。右上図はある地層のある岩石のある化学分析値をたどる場合。左下図のようにある化学分析値からたどる転置ファイルが用意されていれば逆に検索することもできる。あらゆる検索を可能にするためには必要だけ転置ファイルが用意されなければならない。

言語協議会 (CODASYL: The Conference on Data Systems Languages) は独自のデータベース管理システムを提案しこれも商用システムとして IBM 社などの IMS 系列に次いで普及しています。一方情報システムの研究から従来は主に事務計算や科学計算を目的として使われていた各種のコンパイラ言語 (FORTRAN や ALGOL・COBOL・PL/1 など) にデータベース機能を導入する研究も盛んで商業ベースで既に市販されているものも少なくありません。IMS 系列のデータベースは階層型データモデルを採用しています。又 CODASYL 系列のデータベースはネットワーク形データモデルを使っているところが根本的に違います。そこでデータモデルを説明しましょう。

データモデルは現実世界の情報を計算機上に物理的なデータベースとして収納する時に必要になる概念です。これは現実世界の情報が整理されて物理的に記憶される方式を示します。たとえば第2図に図式化しているように私たちの存在する人間社会の情報構造はある一時点をとると階層的な構造を保っていることが比較的多いと考えられています。このことは現実世界の反映と考えられています。江戸時代の身分制度や会社組織上の役職 政党の組織系統などは階層の代表例と言えるでしょう。この階層構造を上からたどっていけばデータの呼出しができますからこの上意下達方式で組織的に必要なデータをまとめたり 不必要なデータを削除したり 修正・追加することもできるでしょう。たとえば軍隊がそうであるように階層構造さえ固定的で不変であれば利用者の命令は必ず迅速に処理されてよいはずで。

しかし世の中はそれほど頑迷では生きていけないことも事実でしょう。途中で割込みをはかったり階層の頂上からではなく途中から呼びだしたい時などや重要な中継がポックリいなくなったりしますとこれだけでは厄介です。たとえばある岩石の SiO<sub>2</sub> 分析値を知ることは簡単ですが逆に SiO<sub>2</sub> 60%以上の岩石を検索することはこのままでは不可能ですべての情報を検索した上であらためて目的に合うような階層をつくり出すことも必要になります。下克上は世のならいと申せましょう。利用者の要求に応じていちいち階層を更新するのは大変ですしそのたびにファイルが複雑化されていき何が何だかわからなくなってしまいかねません。そこで階層型データモデルでは索引データベースを用意し“下克上”に備えて転置表による階層構造を用意します。これが階層型データモデルを実用化するにあたって不可欠な作業ですが冗長さはまぬがれません。全ての情報に対して転置ファイルを作るとは記憶スペースが膨大に必要で処理効率を著しく落とすために大規模データベースには全く不向きです。

IMS 系列では部分転置という概念を採用して不必要と思われる転置を除くことによって商用化に成功しました。ドイツで商用化された ADABAS という部分転置階層型データベース管理システムはこの他に個々の単位データに識別番号を付けることでより緩かで動的な構造を実現しています。これらの階層型データモデルに基づくデータベース管理システム群は実働化の為の妥協を重ねた末手造りの魅力を持ったシステムとなっています。しかしそのためにあとで述べるような地質情報に対しては全転置型ファイルが必要とされるにとどまらない困難さをかかえているため必ずしも有効とはいえないようです。



第3図 ネットワーク形の情報構造。  
 左図は再帰的な情報連関を示す。たとえばハンレイ岩に貫入した花崗岩の年代・産地を知りたい時には連関の輪をたどることで検索できる。第2図で試みた岩石→分析値と分析値→岩石の転置ファイル作りも右図のようにリンクさせれば簡単になる。ネットワーク形データモデル転置ファイルをつくらずに転置が実現できる。

ところで 情報構造は 天地創造があって最後の審判に至る線形又は螺旋形の階層型モデルだけで尽くせるものではありません。永劫回帰の思想が生まれるように現実の情報には 回帰的で始めも終わりもなく人為的区別にたよるだけという構造を持ったものもあります。

これを一般的に拡張しますと 第3図に示した再帰的構造のようなものが考えられるでしょうし この他に部一課一係という階層があってもある係だけは課を飛ばして部に直結するとか 中生代一三疊紀一コノドントという階層の中でもあるコノドントは中生代としか言えないか 又は古生代の可能性もあるというものが出てくるかもしれません。火成岩系列の階層構造に三波川の岩石が位置づけられて 別子型銅鉱床の探査が長く考えられてきましたが 鉱床が発見しにくくなった今頃 もし学説が変わってメランジュ堆積物になり 実は堆積岩系列の階層構造にも含みたくなつた時には 当然探査方針も一新されなければなりません。

黒鉱やキースラーガー鉱床が交代説から堆積説へ180°方向転換したのもわずか20年前のことで その後ガラリと探査指針が変わって 多くの鉱床が発見されました。今また 黒鉱の成因は大きく方向転換しつつあり 現場探査家は 今まで見捨てられていたデータの回収や 同一物のまったく新しい観点からの見直しなどに精力的に取り組んでいます。このような時 もし既存のデータベースの階層構造に新しい考え方が合わないからといって 既存のものを変えるのではなく 新しい考えをひねりつぶしてしまつたらどんなことになるでしょうか。地下資源の開発にとって大きな損失になるのではないのでしょうか。データベース管理システムが固定的で変化に応じられないからといって 研究まで固定されてしまふわけにはいかないでしょう。本来資源探査の役に立つ目的で作られたデータベース管理システムが探査理論の進歩を妨害することがあってはなりません。

階層型データモデルでは 学説が変更されるたびに 新たな転置ファイルを作らねばならず それだけでも不経済ですし 実用上問題が残るでしょう。そこで も

し 火成岩の階層構造に位置しているものでも 堆積岩の階層構造の中にリンクさせることができればはるかに実用的価値が高まるのではないのでしょうか。CODASYLの提案したネットワーク形データモデルは このような目的にも有効です。ネットワーク形データモデルは 単位になる情報に加えて 鍵となる識別情報とその識別情報が他のどの識別情報に連結するかを示すための情報を付属させて 全ての情報を網のように結合する方式です。転置ファイルを作らずに 転置を実現できるわけで 非常に便利なデータベース管理システムでありましょう。ただし データの連結情報をどのように表現するかによって 有効でありうる場合と無意味な場合ができてしまいます。この作業は 地質・資源の専門家でなければ完成しにくいでしょう。しかし 私たちは 素人でも構成できるデータベースを求めているのです。ネットワーク形データモデルは 階層型のシステムに次いで商用化に成功していますが その複雑さのために かえって扱いにくい点も考慮されねばなりません。

### 関係モデル

そこで 階層型構造に基づくシステムにも ネットワーク形構造に基づくシステムにも いずれにも使うことができ しかも独自のデータ構造を持つものとして 関係モデルが登場することになりました。関係という概念は後ほど説明することにして まず 関係モデルを提唱した E. F. Codd (1974) によれば 関係モデルの意義は次のようにまとめられます。

- ① データが応用プログラムに攪乱されない (データ独立)。
- ② 計算機の専門家だけでなく素人にもわかりやすく誰でも共用できる 簡単なデータモデルを作る。
- ③ データ管理の負担を減少し 応用プログラムの開発とその結果の解釈に主力を注げるようにする。
- ④ 人工知能とファイル管理を結合する。

## 恐竜の系統表

時代	系統	肉食小型(属性名)	肉食大型(属性名)	草食中型(属性名)	草食大型(属性名)
組	三疊紀	パレオソールス			テコドントソールス プラテオソールス
組	ジュラ紀	オルニトレステス	アロソールス	カンプトソールス	ケティオソールス プロントソールス
組	白亜紀	オルニトミムス オピラプトル	ティラノソールス	イグアノドン コリトソールス ハドロソールス	
	属性	属性	属性	属性	属性

第4図 恐竜の系統表(一例)

生物進化分類上で使われる系統樹も図のように表形式で表現できる。行は組を示し列は属性を示す。属性名は属性のうちキー又はキー候補(本文参照)を言う。この表形式のまま記憶しても暗黙の了解があるため再現できるかどうかは疑わしい。この表は例示のための仮想的なものである。

⑤ 応用プログラムを集合を対象とするような非手続きの言語の段階に高める。

もともと 計算機を利用するのは その結果をもとにして たとえば探査・開発の実業に役立てたい目的がありました。 共同利用のデータベースを開発するうちにそのデータ管理に多大の負担が加わるようになって本来の目的がかすんでしまったのでは 意味が少なくなります。 単純で保守し易く かつ強力な利用が可能なデータベースを 応用プログラムに迷惑をかけないで築く必要があります。 索引ファイルや転置ファイルを実働化の必要上設けると確かに便利になりますが 利用者は検索性能の向上に労力を使うため 応用プログラムまで修正させられることが多く 整理のための情報管理が逆に混乱と停滞をもたらすことになり このようなデータベースは結局見捨てた方がよいことになりかねません。

この他に 階層型・ネットワーク形いずれにしてもデータ検索のための呼出し経路が設定されていますが 現実のデータが変動してやまないために この呼出し経路によりかかった応用プログラムは 現実の変化についていくことができません。 IMS 系列や CODASYL 系列のデータベース管理システムを使用する最大の難点がここにもあるでしょう。 これらの難点を一掃するために まず情報を計算機の物理的特性から独立させ 検索性能の向上のための手段からも独立させ 呼出し経路からも独立させねばなりません。 こうしてデータベース管理システムに手造りではない 確固たる理論的基礎を提供したのが 70年代に爆発的に盛んになった関係モデル研究です。

第4図はデータを関係形式で示すのに表を利用した方法を示しています。 IMS 系列や CODASYL 系列のデータベースシステムでは 現実の情報をあらわすのに データの内容とデータ相互の関連(呼び出し経路・識別番号)とで別々の表現手法が用意されていました。 しかしデータ相互の関連もまたデータですから 内容も関連も統一的にまとめて表現してしまうことが可能です。 これを関係とよびます。 数値データにしても データの内容であるのと同時に 現象と実数空間との関連を示すわけですから 内容と関連を別の表現にしていたことは 実は無意識的な作為操作又は暗黙の了解にすぎなかったとも考えられるわけです。

第4図で模式化した恐竜系統の表を考えてみましょう。 表形式の最上段「時代・系統・肉食小型・肉食大型・草食中型・草食大型」のように意味のある全体を組(データ処理用語ではレコード)と呼び 各々を属性名又は定義域名(データ処理用語ではデータ項目名・変数名・フィールド名)とよびます。

同様にして たとえばジュラ紀・オルニトレステス・アロソールス・カンプトソールス・ケティオソールス」の全体を組(レコード)とよび 各々を属性又は定義域(データ項目・変数・フィールド)とよびます。 組の全体を関係(データ処理用語ならファイルに相当)とよぶことにします。 この時 第4図の恐竜系統表(関係表)の特徴は次のようになります。

- ① 組の並ぶ順番には意味を持たせない。 従って 順番が必要な時には(たとえば 時代順のように) その情報も表の中に記入させる。
- ② 同じ組は複数個存在しない。

Ⓐ

時代	系統	肉食小型	肉食大型	草食中型	草食大型
三疊紀		パレオソールス			テコドントソールス
三疊紀					ブラテオソールス
ジュラ紀					ケティオソールス
ジュラ紀		オルニトレステス	アロソールス	カンプトソールス	プロントソールス
白亜紀				イグアノドン	
白亜紀		オルニトミムス	ティラノソールス	コリトソールス	
白亜紀		オピラプトル		ハドロソールス	

Ⓑ

時代	順序
三疊紀	1
ジュラ紀	2
白亜紀	3

系統
肉食小型
肉食大型
草食中型
草食大型

Ⓒ

時代	系統	名前
三疊紀	肉食小型 肉食大型	パレオソールス

時代	系統	名前
ジュラ紀	肉食小型	オルニトレステス

第5図 関係正規化.

第4図の仮想的系統表を関係として記憶媒体に収納するためにまずⒶの第一正規形に分割する。この操作で第4図に残されていた組単位のあいまいさはかなり改善され このままでも記憶できるかもしれない。しかし再現するためにはまだ不足している。Ⓒはキー候補としての時代と系統についての第2正規化で 特に時代についての順序という関係を入れてみた。そこでⒶを各々の組と属性で分割するとⒸになる。Ⓒは一部を示しただけで全部を示したわけではない。ⒸとⒸを合わせたものを第3正規形という。第3正規形までで一応Ⓐの表は再現できそうだが まだ個々の属性についていびつなところが残っている。たとえばパレオソールスは肉食型の原型で小型とも大型とも分類されにくい。このような場合にはⒸで作ったようにパレオソールス＝肉食原型という意味の表を作ればよい。又たとえばジュラ紀の大型草食恐竜の場合 ケティオソールスとプロントソールスの前後関係がわかればそれもⒸのように順序として表を作ればよい。これを第4正規形という。ここまでの操作でⒶの表従って第4図は唯一通り再現できる。もし絶対年代がわかればⒸの表に時代とせず年代を記入し Ⓒの表の順序を絶対年代でおきかえておけば第3正規形まででも唯一通りの再現ができる。正規化への道すじは この方法以外にもたとえば組ごとにⒸのようにまとめる属ごとにⒸのようにまとめることもできる。Ⓒの右図ではあえて順序をつけずに暗黙の了解にってしまった。順序づけに意味があるとすればⒸの左図のような表をつけ加えてⒸの右図を削除すればよい。他のモデルのように情報構造の連関の訂正のために いちいち各レコードを呼び出して訂正する必要はない。これが関係モデルの強みの一つである。

③ 属性の集合 (各列) に固有の名前がついていれば列の順番にも意味がない。小型肉食竜の次に大型肉食竜を配せず 大型の次に小型としても同じことになる。もし順番に意味を持たせたい時には その情報も表の中に記入する。

を主キーとすることもできますし 系統を主キーとすることもできます。情報の内容と関連は特に区別しませんから その他のどれでも主キーにすることができます。万物は皆平等というわけです。

しかし 普通は情報の特質やその時点での必要から特定のキーを選びます。キーを検索することにより関係形式の中のある組が抽出できます。このように設定しておきますと 情報の内容や連関がどのように変動しても 関係の組については 唯一通りの識別が与えられますから 関係をデータベースの操作対象とする応用プログラムは 情報の変動や学説による考え方の変化や計算機処理の物理的特性に掻き乱されることなく 安定して使用されることができましょう。データベースは 変動する関係の集まりであると定義することができます。

関係の組を ただ一通りの組合わせで識別するための必要で最小限の属性集合を鍵 (キー) 又はキー候補とよぶことにします。複数のキーについては そのうちの任意の一つを主キーとします。主キーは 情報整理の必要から選ぶこともできますし その日の気分で適当に選ぶこともできます。要するに キーは関係形式の表で表わされるデータを索引するのに便利な属性のことで一つの関係の中に複数個存在します。属性はすべてキーになり得るわけです。第4図では たとえば時代

関係モデルにおける関係に対して 関係代数と関係論理と呼ばれる完成された集合論的演算を施こして情報検索が実行できます。その他 データの追加・更新・削除を容易にし かつ安定した関係の集まりを変動によらずに得るために 正規化という設計指針があります。通常のデータは重複や分岐があり 複雑です。その一部を変更したり削除したり追加すると データベース全体に影響を及ぼすことが普通です。そこで なるべく単純な組に分割してしまっ て その小さな組合わせの集合を関係と考えることにより 全体のデータ系をも表現しようというのが 正規化の目的です。

第4図の恐竜の系統は このままでは使いにくいので すべての属性が それ以上分解できないところまで分割します。これを第1正規形とよびます。第5図は正規化のやり方を示したものです。第1正規形にするとデータ構造が平坦になって 階層構造がかなり消滅してきます。

次に キーについて個々の組が独立できるようにならしたものを 第2正規形といいます。属性についても平坦化を行い 唯一つの概念を表現するところまで関係を小さく分割したものを 第3正規形とよびます。この他に ある属性に対して 一つの属性でなく属性の集合が決まってしまうような第3正規形をさらに分割した第4正規形まで完了すると 情報構造からは 暗黙の了解のようなもの (たとえば ジュラ紀は三畳紀のあとだけれど “常識” だからわざわざ書かない) は一掃されて 平坦化が完成します。素人にとって専門家がわかりにくいのは 暗黙の了解があまりにも多いことでしょう。学説の変動によって 暗黙の了解が変更を余儀なくされてもこれなら十分に対応できるわけです。E. F. Codd(1971)によれば 第3正規形への最適な分割方法が提案されています。

- ① 分割した関係は すべて第3正規形になる。
- ② 分割した関係の全体は もとの関係の情報をすべて含んでいる。
- ③ どの分割された関係も もとの関係群で繰り返し表現する必要のない従属的な属性をそのまま繰り返してしまうことがない。
- ④ ①②③を満たすうちで関係の個数が最小のものをデータベースとする。

正規化された関係形式は組単位で関係記憶システムに収納されます。これをデータベースとして利用するためには データ制御や高水準の最適化システムを通じて

利用システムにつなげます。たとえば関係モデルによる要素操作型 (XRM など)・関係代数型 (PRTV など)・関係論理型 (ALPHA など)・写像型 (SQL など)・例題型 (QBE など)・自然言語型 (RENDEZVOUS・ヤチマタなど)の操作言語系列か 又は IMS 系列・CODASYL 系列の言語系などの利用システムに連結し 応用プログラムに投入できるように組むことが必要です。このような考え方に基づいたものとして IBM 社の実験的な System-R などがありますが 商用として完成されたものにはまだなっていないようです。

ここで データベースマシンについても少し触れておきましょう。今までのデータベースは 既存の計算機システムに ソフトウェアを組むだけで実現するものが大部分でした。つまり “腕力” でソフトウェアのプログラムを組めることのみで決まってしまうものでした。しかし データベースを作る立場から考えますと 専用の計算機があればはるかに処理しやすくなりますし 連想記憶や並列処理 さらにデータフローマシンのものができれば より大規模なデータの処理も容易になるだろうと考えられます。現在の水準では データベースシステムを複数の専用ミニコンに分散処理させる方法がいろいろと試されています。国鉄のみどりの窓口に使われている計算機システムもこの一種と考えられています。最近では ヘッドごとにプロセッサを付加した可動ヘッド型電子ディスクを使って並列処理をおこない大容量記憶の稼働回数を少なくする試みも電子技術総合研究所で開発される (植村1979) など この方面の進歩は著しいものです。本格的なデータベースマシンの登場によって 関係モデルに基づくデータベース管理システムは長足の進歩を遂げると期待されています。

## 地質・資源情報の特質

前節でも少しふれましたが 地質・資源データが地域に密着した情報であることは 地質学や地理学を専攻する人々以外には なかなか理解しにくいことかもしれません。情報そのものが地域性に富むために ある地域・ある国で真理であったことでも 他の地域・他の国に持っていくと たちまち有効性を失うことは しばしば見られることです。また まったく同一場所の地質にも 適用する理論が少し異なるだけで 全然反対の結論が出てしまうことも少なくありません。それどころか 共同で調査し 同一の採集試料について議論しているのに 二人の研究者が同一共著論文の中で 海深50mと海深2000m という互いに矛盾した結論を主張したものすらあります。もちろん一時的に普遍性のありそうな結



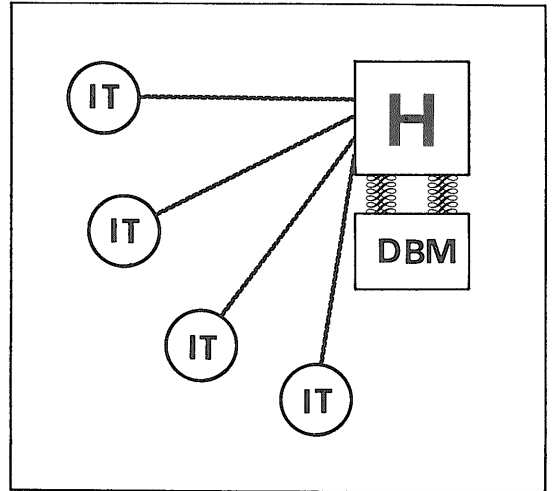
果も あるいはありますが 必ずしも安定したものではなく それだけで地質学をまとめきれものでもありません。 百人の地質屋がいれば 同一地域に百枚以上の地質図がかけるとまで極論されるくらいです。 しかも個々の情報は 事実としては普遍的であり得ても 新しい事実が添加されるたびに 見方が変わり 見方が変わるたびに修正・更新が必要になります。

生物分類の進化系統樹や火成岩・堆積岩の別に見られるように 何となく地質情報にも階層構造があるように錯覚されがちです。 物事を固定的で一面的に考えるなら そう見てもさしつかえないかもしれませんが 現実には 今まで火成岩だと誰もが信じていた岩石が堆積岩になってしまったり 逆に堆積物と思われていたものが砕屑性の貫入岩になっていたり 浮遊性有孔虫と信じられて年代まで決められていたものが底生有孔虫になったり 3億年も昔の古生代と思われていた地層がずっと新しい中生代になり 日本列島の地史が大きく書きあらためられるなど枚挙にいとまがないほどです。 たとえ個々の事実が不変でも階層的に分類・整理しデータとして収納することは永久不変ではありません。 これは資源探査の考え方に密接に関わってきます。 花崗岩の近くに出る鉄床の探査においても 磁鉄鉱系・チタン鉄鉱系の2種類の花崗岩系列があることがわかるまでは 花崗岩中の磁鉄鉱の存在の有無を 階層の中心に置こうとは考えられていなかったようなこともあり得ます。 二次孔隙率の考え直しによって 石油鉄床の探査法が一新されることもあるでしょう。

地質情報は他の物理学的情報や化学的情報に比較してはるかに階層的情報構造が流動的で 分類整理が困難化しやすい特質を持っています。 おまけに 複雑で同一物の別名がやたらと多いなど 今までのデータベースの考え方からすれば まるで手に負えない情報構造に見えてしまうのが地質・資源情報であったのかもしれませんが

なんと錯綜していることかと思うかもしれませんが この流動性こそが 地質学をして柔軟で強力な探査指針にできた原因でもありますから 必ずしも欠点とはいえないでしょう。 逆に 現象記載の蓄積が情報構造の変化ももたらすことを考慮できるデータベースでなければ 実用上役には立たないのです。

情報が変動することは 部分的な修正というにとどまらず 過去の情報システムが変わって 変動した新たな情報システムが 更につけ加えられることであって ますます情報が氾濫することを意味します。 変動しない情報でも新たなシステムの中での位置付けは変化するわ



第6図 分散型データベース管理システム概念図。  
 H (大型ホスト・コンピュータ: M200など)  
 DBM (データベース・マシン: 電総研 EDC など)  
 IT (インテリジェント端末: YHP9845T など)。  
 大規模データベースの管理と高速アクセスは ホスト・コンピュータとそれに後置したデータベース・マシンが担当する。 各種データの入出力やホスト・コンピュータの操作 ソフトウェア作成は インテリジェント端末のターミナル機能と自身のマイクロ・プロセッサで処理する。 小規模データの管理と使用はインテリジェント端末の記憶・計算機能を用いて処理する。

けです。 階層型データモデルやネットワーク形データモデルでは 情報の内容と個々の関連が別物として扱われ 特に個々の関連の変更は数値データの変更ほどに容易ではないことが少なくありませんでした。 氾濫する情報を最小限の労力で整理でき 利用し易くし続けることはできないでしょうか。 常に動的変化に富む複雑データも 単純化して強力に処理できないでしょうか。 それに応えるのが関係モデルであると思われます。

筆者は 関係モデルに基づく分散処理方式の地質・資源データベース管理システムを緩やかに開発中です。 現在 試作段階として 国内の重金属資源探査用小規模分散型データベースシステムを 卓上パーソナル・コンピュータ (パソコン) 端末に実装して RIPS に接続を開始しはじめております。 卓上計算機の記憶容量が小さいため 大規模なものには向きませんが RIPS との結合によって大規模データベース化も可能になるでしょう。 複雑で変動的な地質情報でも 理論的な裏付けによって単純で不変的なデータとして整理・収納できる志向を持ちます。 情報の追加・修正・変更・削除は たとえ大規模化しても極めて容易であり 割り込み機能も単純です。 この方式は既存のデータベース管理システ

ムと結合可能なために 種々の商用・実験用検索システムが利用できます。何よりも 素人が扱える容易さがあるため データベース管理にあまり時間をとられる心配をしなくて済みます。

関係モデルで考えるにあたって たとえば試料採集の場所や日付・採集者、地質現象記載の場所・日付・記載者などの情報は 不変的なキーになります。座標値・日付・人名のデータがそう簡単には変わらないのが地質情報の特長です。関係モデルでは階層構造があれば それだけを他の情報と切離して表化してしまったり 分類がどう変化しようと 分類だけを表化して その内容は別に分けているため 変化した分類表をあらたにつけ加えるだけで 他はいじらなくて済むことになります。表形式のデータをまとめて入力すれば あとは計算機が自動的に正規化し(分割) 関係モデルの集合論的な理論に従って検索を実行してくれます。私たちはフィールドノートなどの記載を一定の表形式にまとめることを念頭に入力を実行していればよいわけです。地形データやルートマップ ボーリング柱状図に検層データ 化学分析値 顕微鏡観察 同位体情報など あらゆる地質及び関連するデータが関係形式に分割されて整理・収納されます。分類方法や記号 考え方のちがいが 歴史的いきさつなども 別に入力されます。これを改めて階層的データやネットワークに作成し直すことも可能で転置ファイルを作るのにくらべて はるかに楽でまちがいが少ないこともわかります。

地質・資源情報には 非数値データ・自然言語型データが大変多く 又地域性に富むために 各種の国語のデータも必要になります。とりわけ 漢字かな混じり文を いちいち英語化して処理することは 誤りの可能性を拡大し 時間的浪費も多いものです。正確さと迅速さと理解の容易さのために 是非とも日本語処理をとられることが重要でしょう。

### 分散型データベース管理システム

地質調査所の地質・資源情報処理システムには いくつかの試みがあります。地質標本のための GEMS システムは 2万個程度の岩石標本情報をパンチカードを通じたバッチ処理で磁気テープにおさめてあり 現在も続行中です(坂巻・小野 1976)。鉱山情報のための MINES システムは 全国455の金属鉱山基礎資料を 階層構造的に編成したもので 同じくパンチカードから磁気テープへというバッチ処理の流れに属するデータベースで すでに終了しています(山田ほか 1980)。地熱資

源のために新たに導入された SIGMA システムは IBM 社の IMS を使用し 階層構造で組立てた 地図・坑井(柱状図・検層図)・地熱地情報をデータファイル化しているもので 最近開始されています(小川ほか 1980)。以上の小へ中規模集中化処理の方向とは別に 小出(1979)は小規模分散型データ処理の柔軟なモデルへの試みを提案しています。筆者のシステムは この提案を関係モデルの立場で改良し 卓上端末に実装化したものにあたります。

大学関係では 大阪市立大学と京都大学を中心に 情報地質研究会が組織され 東京大学の TOOL-IR(山崎・島内 1979)を参考にした COBOLによる GEODAS システムがつくられて精力的なデータベース研究が続けられており 関係モデルに基づくデータベース試案(弘原海他 1976)もされています。

地質調査所には 重力・磁力等の物理探査データや リモートセンシングの画像データ 地下水観測データや 化学分析データ処理システム(吉井・平野 1977)など 大小のデータ及びその処理システムが存在しています。これらを地質学の立場から有効利用していくために 地質学的に有効な分散型データベース管理システムを開発することが 基礎研究・応用研究・探査開発の発展にとって 有力な武器ともなり得るでしょう。

一例として 鉱床部の情報処理系をとりあげてみます。これは RIPS 共用サービス・ステーションを利用した TSS・バッチ処理と 共用回線を利用した TSS インテリジェント端末システムに分かれています。計算機は卓上型YHP9845Tを使用し 周辺機器としてグラフィックディスプレイ・感熱プリンタ・磁気テープ装置・デジタイザ・XYプロッタなどが付いています。その利用

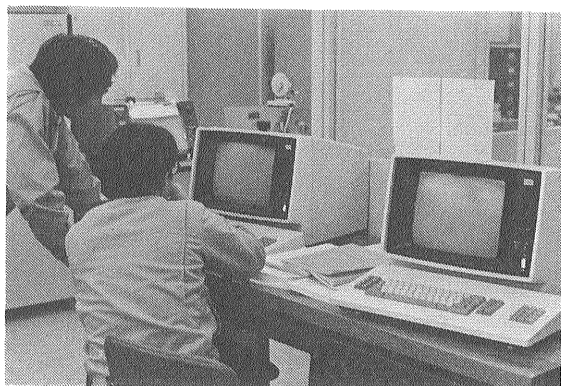


写真3 地質調査所共同利用ステーション内に2台ある TSS 端末。センターと同じく 複数の人が同時にテレビ画面を操作できる。

法については 既に地質ニュース誌上で何回か掲載されています (電卓シリーズ・対話型シリーズ)。RIPSから光ファイバによる端末回線の復変調装置 (モデム) が届けられており YHP9845T の外部インタフェースで接続することにより データ転送ができるほか 普通の TSS 処理もできます。もちろん YHP9845T の今までの計算機能もそのまま使えます。デジタイザのおかげで図形情報の関係モデル化も進みつつあるところです。

試作段階の分散処理システムは拡張 BASIC で記述され自動的に正規化される10万字ほどの微小規模データを一時的に磁気カセットに收容し TSS 処理によって 大型計算機に転送されて普通の磁気テープに収納する方式をとっています。わざわざこうしたのは YHP9845T の方が RIPS より先に稼動していて システムを実装した時点では まだ RIPS に接続されていなかったからなので 将来的には RIPS 上に非手順の言語 PASCAL による構造化プログラミングによって実装し直すことも考慮されています。YHP9845T だけで小規模処理する時のデータの安全確保は パーソナルな問題でしたが RIPS 上にデータベース化する時には ADABAS にあるような 4 ビットの暗号化キーに近いものを導入して不当な使用から防衛することを構想しています。第 6 図に端末からキー入力されたデータ (一行目) と正規化されて分割・整理・収納されるデータを示します。これは黒鉛鉱床の出現する地域の試験地質情報の場合で 座標系を一応のキー候補として第 3 正規形ないしは第 4 正規形まで分割したものです。この方式は原資料に書いてある通りに 何の省略も順序も無視して 5 種の句読点を

使って乱雑にデータを入れられるシステムです。このため まったくの素人や中・高校生でも使用が容易である点が長所です。専門家用にも種々の省略を設けることができ 誤りを自動訂正する機能を付加するなどの工夫も加えつつあります。そして最大の特長は 情報の変動がどんなに進もうと 迅速に対応できてシステムそのものは安定で 応用プログラムに影響しない点にあります。個々の研究室でミニコンやパソコンを使って入力した小規模データを 統一規格にまとめて使用することもできますし 関係正規化されていない在来データも正規化して情報の変動に対して安定させることも可能です。

以上をもとに 今後の地質・資源情報のデータベース化に必要なと思われる概念を要約しておきましょう。

- ① 関係モデル
- ② PASCAL や拡張 BASIC などの会話型処理
- ③ 分散処理
- ④ 並列処理とデータフローマシン
- ⑤ 日本語処理

以上 5 点が現時点で即実現できるわけではなく 特に④のハードウェアがネックともなっています。しかし④以外はすべて現在でも普通になりつつあることなので 遠くならず実現に至るだろうと期待できます。関係モデ

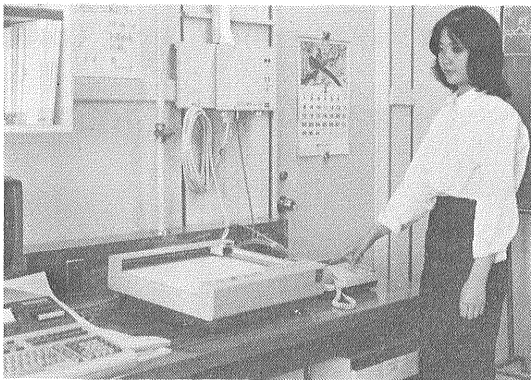


写真 4 操作者の指さしているのが共用回線の復変調装置 (モデム)。光ファイバは RIPS センターからここまで伸びてきて 左の YHP9845T システムに接続される。接続の操作は簡単なボタン操作で開始できる。音響カップラー用の受話機もついている。モデムの左隣りは 4 色ペン付 XY プロッタである。

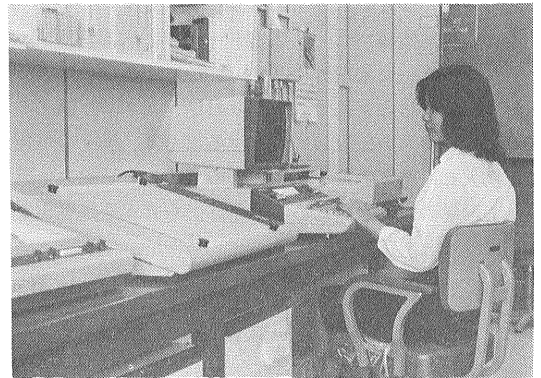


写真 5 実際に YHP9845T インテリジェント端末システムのデジタイザを利用して図形処理をおこなっているところ。簡単で覚えやすい言語で対話型処理ができるため使用者は増大する一方である。FORTRAN を習ったことのある人なら BASIC は 2 時間でマスターでき まったくの素人でも 2 日あれば BASIC の専門家になれると言われているほど容易。PASCAL も 2〜3 日あれば十分使いこなせるようになる。これに反して PL/I や COBOL まして FORTRAN は 長時間の訓練が必要となろう。

CORD, HQ-68, 36.00-37.70m; ROCK, BLACK SHALE; RS, M1-a

シタリ データーヲ STOREシマス。 FILE=HQ-68チュ-リ イチワ Ii= 721 テ-ス。  
/CORD, HQ-68, 36.00-37.70m; T=81.03.10.R.KOUDA( TEST)

シタリ データーヲ STOREシマス。 FILE=HQ-68チュ-リ イチワ Ii= 881 テ-ス。  
/CORD, HQ-68, 36.00-37.70m; ROCK, BLACK SHALE; T=81.03.10.R.KOUDA( TEST)

シタリ データーヲ STOREシマス。 FILE=HQ-68チュ-リ イチワ Ii= 1041 テ-ス。  
/CORD, HQ-68, 36.00-37.70m; RS, M1-a; T=81.03.10.R.KOUDA( TEST)

第7図 黒鉱鉱床地域 試雑データの関係正規化の一例。

一行目はキー入力したデータで 座標位置; 岩石名; 岩石シンボルの順に並べてある。これはあくまで一例にすぎずこの並び方の順序はどうでもよい。操作者が実際にキー入力するのはここまでの破線以下はは計算機上のCRT(陰極線管)に自動的に映し出され記憶される。破線下の2行は座標値検索用の関係で その下は岩石名と岩石シンボル(RSとした)に分割したデータで 各々正規化されている。座標を入力した時の日付・入力者の情報も付加される。他の属性についても同じことである。破線より上は一度のキー入力が約2万字まで可能であり セミコロンの間は 2千字まで可能のため たいていの地質記載データは収容できるであろう。この他に試作中のデータ辞書は 誤りをチェックしたり 略号と総称の変換をしたり 文章化された記載を意味のある熟語(群)に変換する機能を持つ予定となっている。

ルの短所と言われている 専用データベースマシンの未開発も 現在 多方面から開発がすすんでいますから そう遠くない将来に 短所は解消されるでしょう。そして何よりも 関係モデルでデータベースを組んでおけば 現在商用化されているあらゆるシステムにのせることができるだけでなく 地質・資源情報の流動特性にも強靱かつ柔軟に対処できるのは 大変な魅力であると申せましょう。

## 謝 辞

試作段階の重金属資源探査用小規模分散型データベースシステムの作成は 佐藤壯郎課長(鉱床部)の御すめによりました。小論の内容については 坂巻幸雄・吉井守正・佐藤岱生(以上鉱床部)・小出仁(環境地質部)の各氏から有益な御助言を賜りました。特に小出仁博士の強いはげましが何よりの原動力でした。記して以上の方々に厚く感謝する次第です。

## 参 考 書

### ・分散処理

D. R. McGlynn(1978) Distributed processing and data communications. John Wiley & Sons.  
(邦訳: 分散処理とデータ通信. 共立出版).

### ・データモデル

J. D. Ullman(1980) Principles of data base systems. Pitman.  
C. J. Date(1977) An introduction to database systems (2nd ed.). Addison Wesley.  
植村俊亮(1979) データベースシステムの基礎. オーム社.

C. C. Tsichritzis and F. H. Lochovsky(1977) Database management systems. Academic Press.

(邦訳: データベース管理システム. 科学技術出版社)

E. F. Codd(1971) Normalized database structure: a brief tutorial. Proc. 1971 ACM SIGFIDET Workshop p.1~17. ACM.

E. F. Codd(1974) Recent investigations in relational data base systems. IFIP74 p.1017~1021.

### ・データベースシステム

坂巻幸雄・小野晃司(1976) あたらしい標本管理・検索システム~GEMSへの誕生. 地質ニュース. 526 p.30~35.

山田敬一・須藤定久・佐藤壯郎・沢俊明・服部仁・佐藤博之・相川忠之(1980) 鉱物資源予測手法の開発. 地質調査所報告第260号.

小川克郎・花岡尚之・矢野雄策・大久保泰邦・佐藤功(1980) 地熱情報データベースシステム(上) 序報. 日本地熱学会昭和55年度講演要旨 p.51.

小出仁(1979) パーソナル・データ・ファイルとその利用について. 地質調査所月報第30巻 p.580.

山崎昶・島内武彦(1979) 大規模データベースの利用法. 近代科学社.

弘原海清・宮脇富士夫・村山秀次郎・東谷勝(1976) GEODAS マスターファイル(GMF)と全転置型リレーショナルデータベース(IRDB)の作成と更新操作——プロトタイプ GEODAS 利用者マニュアルその3. 情報地質. No.2 p.33~44.

吉井守正・平野英雄(1977) ノルム計算プログラム用テストデータの考案. 地質調査所月報第28巻 p.401~412.

(地質ニュース連載の電卓シリーズ・対話型シリーズも小規模処理の試みとして 一読をおすすめしたい.)