

米国地質調査所のデータベース

花岡尚之(物理探査部)
Naoyuki HANAOKA

高度経済成長の過程で日本の技術力は著しく向上した。分野によっては欧米の進んだ技術を学んでくるなどと云えば無知をさらけ出すことにもなる。コンピュータの利用技術については基本的な部分では差があってもそれは小さいものと思われる。しかし我々に身近な資源とか地球科学の分野で総合的な情報システムと限定すると実績の面で大きな差があるようである。

地質調査所では地殻熱部と物理探査部が中心となって地熱情報データベース・システムの研究開発を行っている。これについては別に報告する機会があると思う。この研究に関連して米国において地熱資源探査及び評価にオンライン情報システムをどのように使用しているかさらに科学技術データ一般の管理をどのようにシステム化しているかを見聞する機会を得た。ここにその概要を報告したい。

データベースは当所にとっても馴染みがないわけではない。標本館が所蔵する岩石 鉱物 及び化石の地質標本について各種の情報を蓄積し必要に応じて検索し様々なレポートを出力するためのシステム GEMS (Geological Survey Museum Samples) がある¹⁾。これは既に実用に供されている。また 鉱物資源予測手法の開発研究で作成された MINES (Mineral Resources Inventory and Evaluation System) もある²⁾。これは鉱物資源に関し 鉱床ごとに位置 地質 鉱化作用 生産量などのデータをファイルしている。システムの一部としてデータの解析 鉱床の予測を行うソフトウェアが開発されている。

これらの例があるにもかかわらず あえて米国のものを紹介するのは 資源を始め 地質調査所の関係する科学技術分野で多くのデータベースが構築され 実用的な情報システムとして継続的に利用者へ供する体制がとられているからである。このようなシステムが我々にとって必要であるのか否か 必要であるとすればだれがそれを構築し だれが運営し維持するのがよいかを考えてみる時期に来ていると思われる。

データベースといっても大規模なものから小さなものまで様々である。その規模によっても性格を異にする。

ちょうど文献ファイルがそうであるように 小さなものに研究者の個人的なデータ・ファイルがあり プロジェクトごとのファイルがあり 大きな中央図書館のようなデータベースもある。また 100人の研究者の文献ファイルを寄せ集めても統一のある大きな文献ファイルになることがないように 大きなデータベースは小さなデータファイルの集合ではあり得ない。データベースの基本設計から始まって 構築 維持 利用に至るまで システム的な発想がなければ大きなデータベースは成長できないであろう。

データベースの話の背景として 米国地質調査所 USGS のコンピュータの話を少し。USGS の主なコンピュータ・システムには パッチ処理用とタイム・シェアリング用がある。パッチ・システムはバージニア州レストンにある AMDAHL470V/7 で 1980年度に IBM 370/155 の2台からなるシステムを置き換えたばかりである。タイム・シェアリング・システムとしては レストン デンバー 及びメンローパークの各本部に Honeywell Multics システムが置かれている。このほか ミニ・コンピュータが多数導入されており プロセス・コントロール ファイル管理 ローカルな計算処理に使われている。

コンピュータとともに USGS が重視しているのはデータ通信網の整備である。ARPANET という各種のコンピュータを接続してコンピュータ資源を分け合うネットワーク・システムがある。機能的にいえばいくつかのコンピュータが連結され どこにいても各所に分散しているデータとかプログラムとかのファイルを瞬時に転送して利用することが可能である。研究情報とかメッセージの交換もファイル転送と同じであるから速く しかも正確に行える。また コンピュータの間で負荷が異なるとき ジョブを転送して平準化することも容易である。効率的なデータ通信網はデータベースにとっても重要なことである。東海岸のレストンのコンピュータに置かれているデータベースが レストンの科学者だけにしか使えないとしたら その利用価値はかなり小さいと言わなければならない。それが通信網を整備することにより 全米にちらばる USGS の科学

者がレストンにあるデータベースの恩恵にあずかることができる。

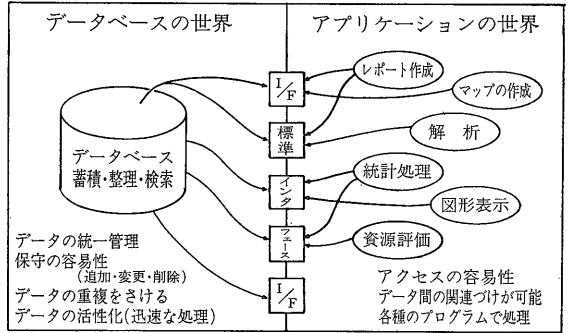
データベースとは

USGSのデータベースについて報告する前に データベースとは何か を紹介しておきたい。

我々は 研究に必要なデータを地図や表として あるいはカードの形式でファイリング・キャビネットに整理している。 データベースは データの詰まったファイリング・キャビネットをオンライン化したものと理解してよい。 キーボードから適当な命令を入力すると 必要な情報が自動的に目の前に現われてくる。

データベースの基本的な機能は データの蓄積 整理及び検索である。 膨大な量のデータは ランダム・アクセスが可能な高密度・大容量記憶装置 たとえば磁気ディスク装置 に整理して蓄えられる。 これを高速演算で検索することにより 希望する条件のデータを迅速に利用者のファイルに書き出すことができる。 このような基本的な機能をコンピュータ上に実現しているソフトウェアは データベース管理システム (DBMS) と呼ばれている。 データベースのファイル様式の定義 データベースへのデータの書き込み 及びデータベースの検索は全て DBMS の言語でプログラムを書くことにより行われる。

データベースの最も重要な概念は データベースのアプリケーション・プログラムからの独立である (第1図参照)。 データベースに納められたデータは 多数の利用者によってそれぞれ特殊な目的 (アプリケーション) に使われる。 たとえば温泉のデータベースがあったとする。 ある人は特定の地域の分析データを取り出して地球化学温度を計算し その分布図を描くであろう。 また別な人は 全国の温泉の湧出量と温度 及び湧出母岩の統計的研究から温泉の成因を考察するかもしれない。このほか あらかじめ決められていないデータの利用に対しても データベースのファイルの内容を変更せずに対処するためには データベースのアプリケーションからの独立を図らなければならない。 これは標準インターフェースを確立することにより実現する。 データベースの側では標準インターフェースに合うことだけを考慮してデータベースの管理を行えばよいことになる。 標準インターフェースがなく アプリケーションが増えるたびにデータベースを手直しするのは データの管理が繁雑になるばかりでなく そのデータベースを使う全てのアプリケーション・プログラムが影響を受けることになる。 したがって データベースの独立は 安定な情報システムの基本的な要請である。



第1図 データベースの独立。 データ・ファイルとアプリケーション・プログラムを分けると データの統一的な管理が可能になる。

USGS のデータベース管理システム

USGSの地質部局では 各種のデータベースを管理するソフトウェア (DBMS Database Management System) を使用している。 このうち 大規模なデータベースに使われているものは オクラホマ大学が開発した GIPSY (General Information Processing System) と USGSが開発した GRASP (Geologic Retrieval and Synopsis Program) である。 このほか Honeywell が提供している GFMS (General File Management System) や ミニ・コンピュータ・メーカーが提供している DBMS が使われている。

GIPSYはファイル管理システムで データ・ファイルの構築 保守管理 更新 検索 及びプリント出力を簡単ではあるが柔軟な命令で実行するプログラムである³⁾。 これは高速ディスク装置を使用している。 特徴は可変長のレコードを処理するように設計されていることである。 蓄積したデータに対する検索機能は非常に高い。 反復機能によって 使用者は段階的に細かなレベルまで検索できる。 また 検索の過程で復帰の機能を使い いつでも任意の以前のステップまで戻することも可能である。 検索したレコードは 各種の配列でプリントすることもできるし 他の後処理にまわすこともできる。

GIPSYのファイルの定義は レコード (ひと組の記録) 内の各フィールド (記入欄) を識別する一連のラベル (記号) からなる。 ラベルは 制御情報と関連した記述とともに デイクシオナリー (辞書) と呼ぶ別ファイルに納められている。 全てのラベルは同じレベルにあるから 階層構造はない。 レコードはディスクに順不同で蓄積される。

可変長フィールド 可変長レコード・フォーマットのため ファイルの設計が非常に自由であり ディスク装置を有効に使える。プログラムはフィールドをラベルだけで識別するから フィールドはレコード内のどこにあってもよい。また可変長であってもよい。この特徴のため 1つのマスターファイルの内に各種のセクション(分野別の記録)を入れ込むことが可能である。また同時に個々のセクションの識別もなされる。また同じフィールド(ラベル)を使用して 必要であれば様々なレポート様式を作成できる。

固定長のやり方とは対照的に 空白フィールドを蓄積しないので ディスク容量の効率的な使用ができる。データを含むフィールドだけを実際にディスクに書くので レコードが圧縮される。さらに レコードの一部が1つのトラック(磁気ディスク装置の記録単位)の終りになり 他の部分を次のトラックへオーバーフローさせる スパンド・レコード機能により 蓄積効率が向上している。

可変長のフィールド 可変長のレコードの良さは 磁気ディスクの容量を節約できることだけではない。野外における観察や測定の結果を一定の様式に整理することを考えてほしい。観察地点によっては あらかじめ決められた様式のなかに書き表わせず 備考にメモを記述したくなることがある。可変長のフィールドのよさは このようなメモを任意の長さだけ書けることである。

GIPSY の検索は次のようなステップで進む。

- (1)使用者が検索するデータベースを決める。
- (2)使用者が探している情報を決める。
- (3)検索過程は使用者がコントロールする。普通は全体のデータベースから使用者はまず大ワクで情報の小集合を選ぶ。さらにそこから次の小集合を選ぶことを反復して 必要な情報を抽出する。また検索の過程でいつでも復帰の機能を使って前の段階へ戻ることもできる。
- (4)全体のデータベースあるいは選び出した小集合に対して検索を行った都度 GIPSY は検索結果の統計的データを自動的に出力する。
- (5)統計が満足のゆくものであれば使用者は出力コマンドを使って検索情報の全部又は一部をディスプレイし 内容を確かめる。場合によっては さらに小集合に対する検索を反復し あるいは検索をやり直す。

GIPSY を使い易くしているものの1つは QUESTRAN (QUESTion TRANslator) と呼ばれる命令言語である。これは検索 様々な中間処理 及びプリント出

力の機構を提供している。使用者はこれによりプログラマーの手助けなしにコンピュータと対話できる。各種の QUESTRAN 命令と指定パラメータを与えるると使用者の指令が一連のコマンドと処理に翻訳されて コンピュータに理解される。

GRASP は 地球科学データベースへの会話的なアクセス(読み出し)のために USGS で開発された⁴⁾。各種のコンピュータに乗り 使い易く またデータベース一般に向くものである。

できる限り多くのコンピュータで動かすために GRASP はフォートランIVで書かれている。ほとんどのミニ・コンピュータはフォートランIVを扱えるから 事実上どのタイムシェアリング・コンピュータでも使用できる。このため データ・ファイルがどこにあっても GRASP でのアクセスが可能である。

GRASP は データベースが普通の表の形式(マトリックス)に書ければ 何でも扱える。表形式では 行が記録の単位で 列が記録の属性である。たとえば地球化学データベースは表形式をとり 列は各元素の化学分析値を表わし 各行は1つのサンプルを表わす。GRASP はどのようなデータ・マトリックスでも扱えるが ただ1つの前提は 変数(列)はあらかじめ型(文字か数字か)が定義されており 必要ときには文字変数に対して文字入力の手書が準備されていることである。つまり GRASP は表を志向した型であるので どのような固定長の表でも扱うことができる。

GRASP を使うためには フォートランなどコンピュータ言語を知らなくてよい。GRASP の命令言語で 使用者はデータベースに質問することができ 質問に合うデータだけを取り出すことができる。

ところで USGS のデータベース管理部門の要請でコンピュータ部局が現在出回っている主要なデータベース管理システムについて 最適であるものを調べた。その結果 モデル 204 がデータベース管理部門の要求に合うものとして選定され AMDAHL 用に導入された。モデル 204 は データベースの生成 検索 及び報告書作成が1つの使用者言語でできる唯一のものである。このユニークな特徴のため 使用者は容易にあらゆる面のデータの取り扱いができる。このような DBMS の使い易さが 科学的な部門の場合に極めて重要であるとされている。

USGS のデータベース

USGS のデータベースは 1979年の調査で223ファイ

GEOOTHERM

Sample File

A10 _____
A20 _____

SAMPLE SOURCE IDENTIFICATION

* Name of source B20 _____
KGRA B11 _____
Well/Spring no. B116 _____
 township range section letters
API no. B12 _____
Warning no. B15 _____

COMPILATION INFORMATION

* Compiled by A50 _____
* Date (YY/MM) A60 _____ last _____ first _____ m. i.
* Affiliation A70 _____
Compiler index A75 _____

GEOGRAPHIC DESCRIPTION

* Country B50 _____ Latitude, Longitude
* State B60 _____ * B70 _____ degrees _____ minutes _____ N/S
* County B65 _____ * B80 _____ degrees _____ minutes _____ E/W
Geologic Province B128 _____ * Township B95 _____ (ex.07N)
Map reference B82 _____ * Range B105 _____ (ex. 007E)
* Section(s) B115 _____
Quarter(s) B117 _____ of _____
Meridian B125 _____
Other locality information B83 _____

SAMPLE DESCRIPTION AND CONDITIONS

* Source type (circle one) A34 SPRING WELL FUMAROLE Other: _____
Point of collection P55 _____
Sample type (circle one) S10 WATER STEAM
Analysis(s) in this report A31 WATER CONDENSATE GAS
Sample number M190 _____
Collection date (YYYY/MM/DD) M200 _____/_____/_____
Collector(s) (last, first, m.i.) S20 _____

Sample temperature (°C) M210 _____ At depth M210A _____
Ambient temperature (°C) M26 _____
Discharge (well or spring) M220 _____ Flow units M223 _____
Deposits or alteration S30 _____
Water treatment data M234 _____
Pertinent lithology M235 _____

Other sample information S50 _____

* Reference M790 _____

FOR WELLS:
Depth (m) M25 _____
Gradient (°C/km) M27 _____
Wellhead pressure (kg/cm²) N30 _____
1st separation pressure (kg/cm²) P60 _____
2nd separation pressure (kg/cm²) P70 _____
3rd separation pressure (kg/cm²) P80 _____
Mass flow of steam (kg/s) N50 _____
Total flow enthalpy (j/g) N60 _____

*REQUIRED

第2図 GEOOTHERM のサンプル・ファイルの記入フォーマットの一部分。 岩相とか備考が自由に書けるところに特徴がある。

ルが登録されている⁵⁾。 このうち 地質部局が管理しているものは62ファイルである。 これを分野別に見ると 半数は資源関係で 残りは分析データとか基礎科学データ 研究の運営管理などに関するものである。 資源では 鉱物鉱床 鉱産物についてのデータベース CR-IB (Computerized Resources Information Bank) の関係が14ファイル 石炭資源の情報システム NCRDS(National Coal Resources Data System) が6ファイル 石油の情報システム PDS (Petroleum Data System) が5ファイルを擁している。 地熱資源のデータベースはGEOOTHERM が1つである。 資源以外でファイル数の多いのは海洋調査で得たサンプルの分析結果を分析項目ごとにデータベース化したもので10ファイルある。

そのほか 参考までになるほどと思われるデータベースの名前をあげてみる。

岩石分析値収納システム RASS

地質学的名称 GEONAMES

地質図インデックス GEOINDEX

放射年代データ・バンク RADB

鉱物等の熱力学的性質 THERMO PROP

コア・ライブラリー・データ・ファイル CLDF

クルーズ管理データ ADMIN/CRUIS

これらのうち いくつかの例からデータベースが果し得る役割について 簡単に紹介する。

GEOOTHERM—地熱のデータベース

GEOOTHERM はコンピュータ化した地熱資源情報のファイルである。 ファイルは地熱資源評価に使用するように設計されており 地熱情報を迅速に 効率的にまた経済的に取り出すことができる。 1978年の USGS の地熱資源評価はこのファイルを使用して行われた。 現在も 全米の低温地熱資源の評価作業に使われている。 このファイルは全て一般に公開されており 既に ジェネラル・エレクトリックの Mark III 情報サービス・ネットワークに乗っているという (J.Bliss による)。

GEOOTHERM は IGIEP (国際地熱情報交換プログラム) の1つとして開始された⁶⁾。 その目的は 新しい地熱

情報の迅速な交換と配布にある。初期の GEOTHERM は USGS のほか イタリアのピサにある CNR (International Institute of Geothermal Research) とカリフォルニアのローレンス・バークレー研究所にも置かれていた。

GEOTHERM は 現在 バージニア州レストンのコンピュータ部局にあるバッチ処理を主とするコンピュータ AMDAHL に置かれている。使用している DBMS は GIPSY である。GEOTHERM から見て GIPSY を使う主な理由は 32000 字までの可変長のレコードが使える点と 検索機能が非常に高いことである。

GEOTHERM は 3つのセクション 地熱地の地質と生産記録 分析データ 及び掘削記録からなる。各セクションの内容は次のとおりである。

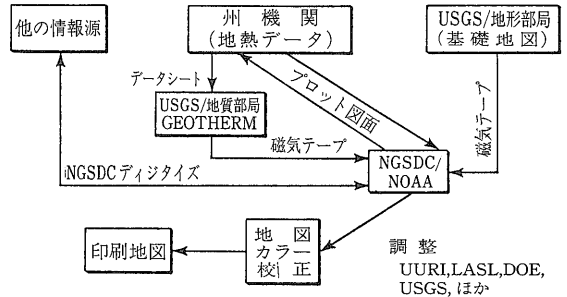
セクションA 地熱地域： 地熱地域の場合 地表地熱徴候 工業的開発 地下温度と規模 基礎化学 熱エネルギー 一般地球物理 地質 及び関連情報である。このセクションには 510 レコードが記録されている。

セクションB サンプル・データ/化学分析： 温泉 坑井からの化学分析データである。3種類の分析値 水 凝縮水 及びガス を入れるスペースがとってある。データ項目は 報告者 場所 サンプルの記述 産状 及び流体の物理化学的データである。このセクションには アメリカ ニューゼーランド 及びメキシコの坑井と温泉の 4600 レコードが納められている。

セクションC 地熱坑井： 地熱の生産及び開発のために掘削された坑井の深度 ケーシング 流量などの物理的データである。USGS の保護局から提供された 436 レコードが記録されている。

データベース GEOTHERM へのアクセスのうち 90% 以上は セクションBのサンプル・ファイルに対するものであると言う。このファイルのフォーマットの例を第2図に示す。これはサンプルの属性を記入する部分である。各記入欄(フィールド)の前に B20 のように書かれているのがラベルである。このラベルは1つのレコードの内ではユニークである。検索はこのラベルを目当てに行われる。フィールドは 様式が決まっています。数字 コードを記入するもの 記述式のもの その後に自由に備考を記述するもの がある。このような自由さが 可変長のフィールドを許容するデータベース管理システム GIPSY の良さである。

データベースの応用面について言えば GEOTHERM が USGS の地熱資源評価に果たした役割は次のようなも



第3図 州別の地熱資源賦存図の出版に至るデータの流れ。GEOTHERM は地熱情報のファイリングを担当している。州政府 USGS NOAA 大学 エネルギー省が協力して 1つのプロジェクトが成立している。

のである。(1)データの収集とファイリング。(2)データの編集と保守：レコードの変更と追加は GIPSY のアップデート・プログラムを使用して容易にできる。(3)地図作成：様々な縮尺と投影法の地図を60葉以上作成した。(4)データの処理と解析：GEOTHERM のデータは地球化学温度の計算 熱勾配の計算 統計解析 容積と熱量の決定 グラフの表示 回帰解析などに使われた。さらに 回収可能な熱量も計算した。(5)データの表示：データの迅速な検索 分類 及び表示のために GEOTHERM は不可欠であった。

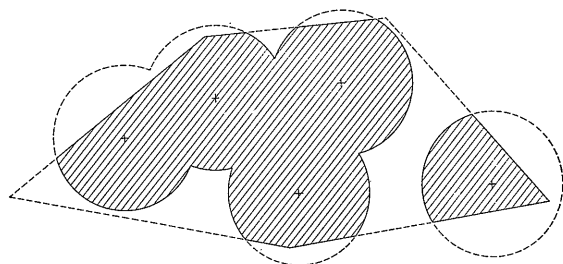
現在アメリカでは エネルギー省が State Coupled Project と称して 各州の中～低温地熱資源のデータを精力的にコンパイルしている。その一方で 米国地質調査所が全米の低温地熱資源の評価作業を進めている。この中で GEOTHERM がどのように位置づけられているかを見ることは興味深い。このプロジェクトは中～低温の地熱資源の分布とその資源量予測を公表し 地熱資源の多目的利用を促進することをねらっている。このプロジェクトの1つの支柱は 州別の地熱資源賦存図を出版することである。第3図は地熱資源賦存図の出版に至る過程を示している。この図は同時にこのプロジェクトに参加している機関の関係も示している。地熱情報の収集は主として州の機関が担当している。図には表だっって書かれていないが 各州の作業に対する技術的なサポートは ユタ大学研究機構(UURI)の地球科学研究所(ESL)ほかが与えている。各州で集めた情報はデータ・シートに記入して GEOTHERM に提出される。ここでは提出されたデータについて 論理テスト 品質検査を行い データベースに記入する。NOAA の NGSDC (The National Geophysical & Solar-Terrestrial Data Center) は このプロジェクトのなか

で地熱資源賦存図の出版を担当している。GEO-THERM から地熱情報を受け取るとともに USGS の地形部局から資源図のベースになる地図情報を受け入れる。これらの情報をもとにプロッタで地熱資源の素図を作成する。これを各州の機関と調整して最終的な図面を編集する。NGSDC ではフラット・ベットのプロッタにカッターをつけて製版に使う密着焼原図までを作成している。これはプロッタ出力をさらに製図工程にまわすことによる時間と費用と精度の無駄を省くためである。

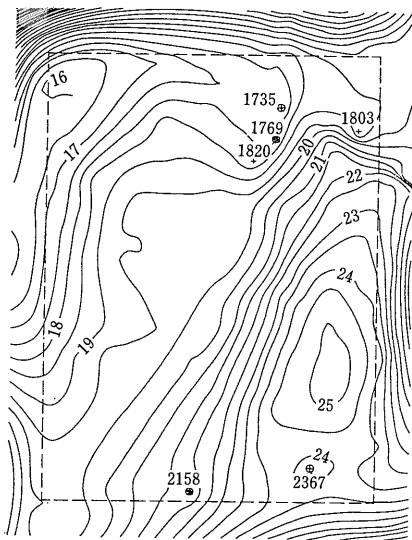
GARNET—図形による資源評価

GARNET (Graphic Analysis of Resources by Numerical Evaluation Techniques 数値評価法による資源の図形解析) は USGS が1973年の石油危機ののち国内の石炭資源を評価するために作成した情報システム NCRDS (National Coal Resources Data System) の一部をなす。NCRDS には 2つの資源データベース WCOAL (ミシシッピ川から西の鉱床の16,000レコード) と ECOAL (同じく東の鉱床の15,000レコード) 及び石炭分析のデータベース USALYT がある。資源データベースを管理するDBMS は PACER と呼ばれ GRASP の機能を拡張したものである。データベースの説明は割愛して検索したデータをグラフィックディスプレイに図形として表示しこれを会話的に解析するシステム GARNET について述べる。

GARNET は PACER から検索した石炭資源の点情報を使用して 2次元の図形による資源量の評価を行うソフトウェアである⁹⁾。点情報を迅速に処理する構造図を描く一定の地域の資源量を推定するなどの処理が容易にできるように設計している。開発のポイントは5つある。コンピュータの使用時間が短いこと、使い易いこと、オンラインで会話的な処理ができること、コアメモリが最小限で済むこと、ポータブルであること



第5図 資源を評価するとき 実測点のまわりにある大きさの実測値の有効範囲を想定する。評価する領域は多角形で示す (GARNET)。



第4図 GARNET で 2次元の曲面を会話的に求めた結果。詳しくは本文を参照されたい。

である。つまり いつでも だれでも どこでも 気楽に使えるソフトウェアを目指している。

点情報は 座標 X Yとその点の値 Zからなっている。Zは標高 硫黄含有量 あるいは炭層の厚さなどである。GARNET の最も基本的な機能の1つは 点情報から最適な 2次元の曲面を会話的に定義することである。まず 2次元多項式 又は 2次元三角級数で最小二乗的に傾向面を求め 格子点上の値を計算する。格子点の値からコンター図をCRT (ブラウン管) 上に描く。各観測点のデータについて傾向面からの残差を求め 二乗平均平方根の2倍より残差が大きい点をコンター図上に表示し Z値も書き出す。このCRT上の図を見て 疑わしい点を除く 値を変更する そのままにする などの編集を行う。データ点を取り除くときは カーサーをその点の位置に合わせ Dとキーインする。変更は Cとキーインすると入力指示がCRT上に表示されこれに従って数字を入力する。編集を終えた点は丸印等で表示される。あまりに込み入った部分は 適当にウインドをかけて表示を拡大して処理する。編集ののちまず得られるのは修正した傾向面のコンター図である。しかし これは全体の傾向を示しはするが 局所的な特徴を表わし得ない。そこで コントロール点 (実測点) からの距離で指数的に減少する重みをつけて 格子点の値を評価するようにしている。このようにして得られる重み付き曲面は 地質の専門家が手描きする図面によく似たものとなる (第4図)。

この曲面解析の目的は資源情報の計算 図化である。上述のようにして会話的に求めた炭層の上面と下面の差から石炭の容積が 炭層上面と地表面の差し引きからは

ぎ取らなければならない被ふく層の容積が求まる。等層厚線図などは容易に CRT 上に表示できる。

実測した資源量は3つのおもなカテゴリーに分けている。確定 推定 及び予想である。地図上の多角形で定義される範囲で資源量を評価するとき 各実測点を中心にある半径の円を描き 実測値の有効範囲とする(第5図)。実測点が隣接してあるときは影響圏が重なることもある。削はくされた部分 採掘を完了した部分は 多角形で示すことにより取り除かれる。資源量の計算は各カテゴリーごとに円に含まれる範囲を格子単位で評価し 加え合わせて行う。

CRIB に見る情報の流れ

(Computerized Resources Information Bank)

データの保守管理を容易にし 統一的な情報システムを必要とするということで データベースを作成し 維持体制を整えたものとする。では 利用面はどうか。コンピュータは高い演算処理能力を持っているから データベースと高速データ通信システムを組み合わせれば 地理的に広い範囲にちらばった多数の利用者にサービスを提供できる。費用の面から見ると データをコンピュータが読める型に変換し データベースを更新する データに含まれる虫(誤り)を発見してデータベースの質を高く保つ などの保守管理に必要な固定費用が相当の額にのぼる。したがって 利用者が受け取る情報のコストは データベースが広く開放され 利用

者が多ければ多いほど低くなる。逆に 利用者が相当多くないと データベース・システムを作成する意義がないとも言える。

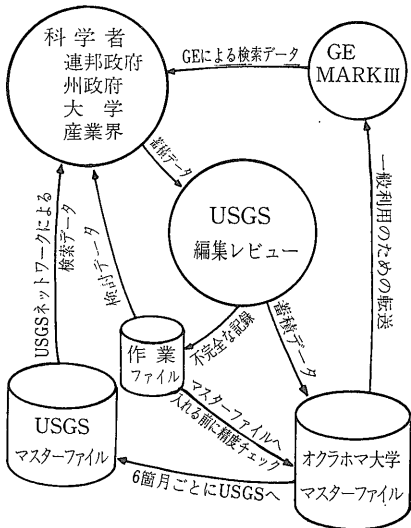
CRIB は鉱物資源情報のデータベースである。そのデータ・ソースと利用者は 連邦政府や州政府の科学者ばかりではなく 大学や産業界の科学者まで広がっている。CRIB の情報の流れを第6図に示す。USGS や各州の地質調査所を中心に各界の科学者がデータ・ソースとなって CRIB の成長に寄与すると同時に 利用者としてサービスを受けている。データベースの情報の質は USGS が行う入力データの編集レビューの作業によって保たれている。レビューをパスした情報は オクラホマ大学が管理するマスター・ファイルに書き込まれる。情報の一般への提供はジェネラル・エレクトリックの情報システムにファイルを移して行われている。もちろん 利用者は GE のシステムの使用料程度は支払わなければならない。

このように CRIB は USGS が自らの必要のために開発した専用の情報システムというよりも 社会的な役割を与えられた1つの事業と理解した方がよい情報の流れになっている。CRIB に限らず USGS が管理しているデータベースは無制限に一般に公開されているものが数多くある。これは 連邦政府の予算で得られた成果は納税者に還元されなければならないとする理念に裏うちされた結果である。ついでに述べれば コンピュータ・プログラムも連邦予算で開発されたものは 実費程度でコピーを頒布する機関が存在し サービスを提供している。

RASS—化学分析値のデータベース

USGS の分析ラボラトリーが分析したサンプルのデータベース RASS (Rock Analysis Storage System) は 研究の基礎データを組織化して成功した1つの例である。1965年から今日までに 500,000 レコードのデータが蓄積されている。研究者が現地で採取したサンプルの分析値は その研究者が関心を持つ領域で研究成果を公表すれば価値が無くなるものではない。分析値のほかわずかな情報 採取場所や地質概要などを付け加えさえすれば 永い寿命を持ち得るものである。

RASS の概要は第7図に示されている⁹⁾。研究者は分析を依頼するとき コーディング・シートになった第8図のような分析依頼書をサンプルとともに提出する。これには依頼者 所属 日付 研究テーマなどが記入され 各サンプルごとに ID 採取場所 地層 岩石名のほか コード化した記号で産状 地質年代 及び組成等の



第6図 CRIB の情報の流れ。情報の質は USGS の編集レビューに依存している。一般の利用はジェネラル・エレクトリックの情報システム Mark III をとおして行われている。

観察項目を記入するようになっている。サンプルが提出されると依頼書はカードパンチされて RASS のファイルに入れられる。これとは別に分析データは自動分析装置のものは磁気テープからその他のものはカードや最近では TSS ターミナルから RASS に入力される。RASS のファイルから検索したデータは統計解析のためのプログラムの入力様式で STATDAC ファイルに書き出すことも RASS のプログラムで更に検索するためサブファイルに入れることもラインプリンタにそのまま出力することもある。

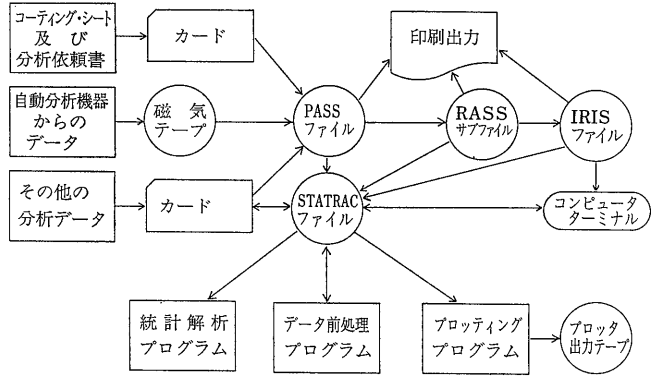
STATPAC はデータの入出力処理統計解析データの前処理及び図形出力などを行う地球化学データ処理パッケージである。サブファイルは IRIS ファイルに変換して GRASP で会話的に処理することもできる。

RASS の利益には2つの面がある。サンプルを提出したプロジェクトに対するものと、いつか後のプロジェクトに対するものである。サンプルを提出したプロジェクトにとっては、データがデータベースに入れられる結果、各種の分類・検索が容易にできること、データ処理のプログラムで解析するためデータが自動的に転送されることなど時間を節約できる効果がある。将来のプロジェクトに対する利益は、たとえば、関心ある地域にどのようなデータがあるか調べたり、特定の条件に合う物質のデータを検索するなどして、研究及び研究計画に役立てられることなどである。STATPAC のプロット用のプログラムを使えば、研究者がある地域の調査に出かけるとき、それまで RASS に蓄積された関連情報を全て地図上にプロットして、綿密な調査計画を立てることができる。

おわりに

今回の調査旅行で、データベースがお話とか研究テーマとかではなく、USGS という巨大な行政機関の欠くことのできない情報システムとして日々の活動に組み込まれていることがわかった。といっても、訪問者のなんとはない印象に過ぎないのであるが、専らデータベースの管理を担当する少なからぬ人々に会って話を聞いてうちに、研究者が片手間で作るのは全く異なる体制であることに気付いた。

データベースは博物館の業務に似ている。岩石を受け入れ、登録するに値するか判断し、登録するものは一



第7図 RASS-STATPAC における地球化学データの流れ。データ処理が容易なソフトウェア体系になっている。

定の規準に従って分類整理し、保管する。研究者から問い合わせがあれば、条件に合う標本を検索し、もし存在すれば貸出しの手続きをとる。データベースもこれと同じである。博物館で、岩石、化石、生物と分野ごとに Curator (管理者) がいるように、データベースではデータ・ファイルごとにデータベース管理者が必要である。データベースがプロジェクト限りのものでないならば、あるいはデータベースが1つの機関のワクを越えて一般にサービスするものならば、データベース管理者の責任は大変に重いものである。データベースの情報にある割合以上の誤りがあったり、最新の情報がファイル化されるのが遅かったりすれば、データベースはたちまち信用を失って巨大な情報の墓場となるであろう。

データベースはいやがおうでもデータの取り扱いを統一化する。そこでデータベースを本格的に導入しようとするとき、データの扱いについてコンセンサスが前提となる。たとえば RASS では分析したデータは直ちにデータベース化されるから、データに対する研究者の優先権は認められていないことになる。その辺の扱いは考え方しだいであるが、研究所の運営の基本問題にふれざるを得ない一面をデータベースは持っている。いずれにしても、コンピュータはファイルの管理、データ処理、図形の入出力に高い処理能力を持っており、使い易いシステムを組むことによって、研究環境は大幅に改善されるであろう。またデータ通信が自由化するにしたがい、研究所間で役割を分担して、総合的な情報ネットワークが検討される日も近いであろう。

謝辞： この報告の基になった米国の地熱探査技術調査 (56.2.8—56.3.29) は、サンシャイン本部の支援で実現した。また、訪問の日程について R. I. ティリング氏

Submitted by: JANE DOE
 Address: DENVER
 Collected by: (if other than submitter)
 Send carbon of report to:

UNITED STATES DEPARTMENT OF THE INTERIOR GEOLOGICAL SURVEY CODING SHEET AND REQUEST FOR ANALYSIS		Work Requested: (List elements for partial anal.) SiO ₂ Al ₂ O ₃ Fe ₂ O ₃ FeO MgO CaO Na ₂ O K ₂ O H ₂ O ⁺ H ₂ O ⁻ TiO ₂ P ₂ O ₅ MnO CO ₂
Page 1 of pages: 1	Total No. of Samples: 4	LABORATORY USE ONLY
Date Submitted: SEPT 1, 1976	Date Results Needed: OCT 15, 1976	Report No.
Subdivision & Branch: MINERAL RESOURCES WESTERN	Project Title: MISC. INVESTIGATIONS	Report Date:
Project No. 1234-56789	Lot No.	Job No.
Approved by: (Branch Chief or Representative)	Lab. Comment:	Report No. Type of analysis

Assigned Lab. No.	FIELD NO.	S.E. Corner - 7 1/2' Quad.		STATE	COUNTY	FORMATION	SAMPLE NAME & DESCRIPTION																											
		LAT.	LONG.																															
X065163	74L-11	38°01'	107°15'	COLO	HINSDALE	HINSDALE	TYPICAL BASALTIC ANDESITE																											
COMMENT:		11	12	13	14	15	16	17	19	20	21	22	23	24	25	26	27	28	29	30	35	36	37	38	39	70	71	72	73	74	75			
		T	A	A	U				A	A	D	F	T																					
X065164	74L-12	38°01'	107°15'	COLO	HINSDALE	HINSDALE	RHYOLITIC PHASE																											
COMMENT:		11	12	13	14	15	16	17	19	20	21	22	23	24	25	26	27	28	29	30	35	36	37	38	39	70	71	72	73	74	75			
		T	B	A	U				A	A	D	C	1																					
X065165	74L-13	37°59'	107°46'	COLO	HINSDALE	STONEY MTN STOCK	RHYOLITIC DIKE																											
COMMENT:		11	12	13	14	15	16	17	19	20	21	22	23	24	25	26	27	28	29	30	35	36	37	38	39	70	71	72	73	74	75			
		T	A	A	T	U			A	J	D	A	1																					
X065166	74L-14	38°01'	107°46'	COLO	HINSDALE	STONEY MTN STOCK	GRANODIORITE																											
COMMENT:		11	12	13	14	15	16	17	19	20	21	22	23	24	25	26	27	28	29	30	35	36	37	38	39	70	71	72	73	74	75			
		T	A	A	T	U			A	L	G	F	P																					
COMMENT:		11	12	13	14	15	16	17	19	20	21	22	23	24	25	26	27	28	29	30	35	36	37	38	39	70	71	72	73	74	75			

INSTRUCTIONS: Prepare original plus 5 copies of this form: keep 1 copy; enclose 1 copy with samples; send original and 3 copies with memo to Laboratory. Do not mark in unnumbered boxes. Receipt of samples will be acknowledged.

Date Received _____ Estimated date of Completion _____ Signed _____
 Liason Officer

第8図 分析ラボラトリーの分析依頼書。 RASS のコーディング・シートになっている。 そのまま RASS のファイルに入れられる。

(USGS) の協力がなければ 豊かな成果は得られなかった。 記して謝意を表す。

文 献

- 1) 地質調査所(1978):地質標本管理検索システム (GEMS システム). ユーザーズ・プログラム 使用説明書第3集 98p.
- 2) 山田敬一・須藤定久・佐藤壮郎・藤井紀之・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.
- 3) Calkins, J. A., Keefer, E. K., Ofsharick, R. A., Mason, G. T., Tracy, P., and Atkins, M. (1978): Description of CRIB, the GIPSY Retrieval Mechanism, and the Interface to the General Electric MARK III Service. USGS Circ, 755-A.
- 4) Bowen, R. W. and Botbol, J. M. (1975): The Geologic Retrieval and Synopsis Program (GRASP). USGS Prof. Paper 966.
- 5) Scientific and Technical, Spatial, and Bibliogra-

phic Data Bases of the U.S. Geological Survey, 1979. USGS Circ. 817.

- 6) Clark, A. L., Calkins, J. A., Tongiorgi, E., and Stefanell, E. (1975): A Report on the International Geothermal Information Exchange Program, 1974-1975. Proceedings 2nd U. N. Symp. on the Development and Use of Geothermal Resources, p. 67-99.
- 7) Teshin, V. N., Swanson, J. R., and Orris, G. J. (1979): GEOTHERM-Geotherm Resources File. Geothermal Resources Council Transactions, V. 3, p. 721-724.
- 8) Olson, A. C. (1977): Graphic Analysis of Resources by Numerical Evaluation Techniques (GARNET). Computers & Geosciences, V. 3, p. 539-545.
- 9) Van Trump Jr., G., and Miesch, A. T. (1977): The U.S. Geological Survey RASS-STATPAC System for Management and Statistical Reduction of Geochemical Data. Computers & Geosciences, V. 3, p. 475-488.