

会話型データ処理—その10—

GEOCAPSでのデータの入力とソート

吉井 守 正 (鉱床部)
Morimasa YOSHII

1. はじめに

前回にご紹介したGEOCAPS(地球化学データ解析プログラムシステム)でのデータ入力は、その前身のシステムで使っている方式(吉井 1981)を大筋で踏襲している。新しいシステムでは、これにデータのソートと編集の機能が組み込まれた。今回は、データの入力・訂正行程で従来と変った点を中心に要点を述べ、続いてソートの方式についてご紹介する。データの編集については長くなるので後の回にゆずることにしよう。

2. データの入力

a) データの性質とその対応

GEOCAPSで取扱うデータの性質などについて、前置きとして簡単に解説しておこう。データの主要な部分は化学分析値と、それに関連する数値データである。

これに試料番号とコードの文字データが付随する。

このシステムで取扱う数値は、その精度を考慮すると有効数字は6けた(YHP-9845Tでの2分の1精度)あれば十分である。数値範囲も8けた程度で十分と考え、目下のところ8けた(整数の場合)まで受け入れるようにしている。

数値をCRT画面に表示したり印刷するときの、小数点以下のけた数についても考える必要がある。このけた数は、成分およびその他の項目によって異なるので、数値の出力形式を画一化するわけには行かない。このシステムでは、これら数値の表示または印刷をする行程を伴うプログラムでは、そのけた数指定の行程が付いている。数値が欠けて空欄となっている場合の処置も決める必要がある。データの空欄部分は数値0とは異なるので、空欄(ノーデータ)の“数値”を定義する。このシステムでは、これを

$$9 \times 10^{99} \quad (9E63)$$

と定めている。データ処理の際は、この値およびこの値を含む算式について、一切の処理から除外する。数表での表現は……などとしている。

b) データ入力の手順

データの入力・訂正・編集行程の流れ図を第1図に示す。データ入力の行程へ入る経路は2つある。第1は前回に述べたファイル作成の行程から新規のデータ入力をする場合であり、第2は以前入力したデータに新しいデータを追加する場合である。

行程の冒頭で、上に述べた表示けた数の指定を行う。CRT画面には、成分名ファイル中の全項目に、その番号(成分番号)とけた数を付けた表が出される(第2図)。使用者は、画面を見て成分番号の範囲と小数点以下の表示けた数(整数なら0)を入力する。

その上で、プログラムを進めるキーを押す。この操作を1回行うごとに結果が画面に出るので、それを確認しながら、上記の要領で繰り返す。指定作業が終わったら0を入力し、プログラムを進めるキーを2度続けて押す。

つぎに、データの入力(I)、訂正(C)、CRTで表示(D)、印刷(P)、編集(ED)、レコード(R)、および作業の終了(E)を記号で選択する行程に進む。この行程から、各方面へ分岐できるので、以前に“広場”という呼び名を付けた。今回もこれを使うことにしよう。

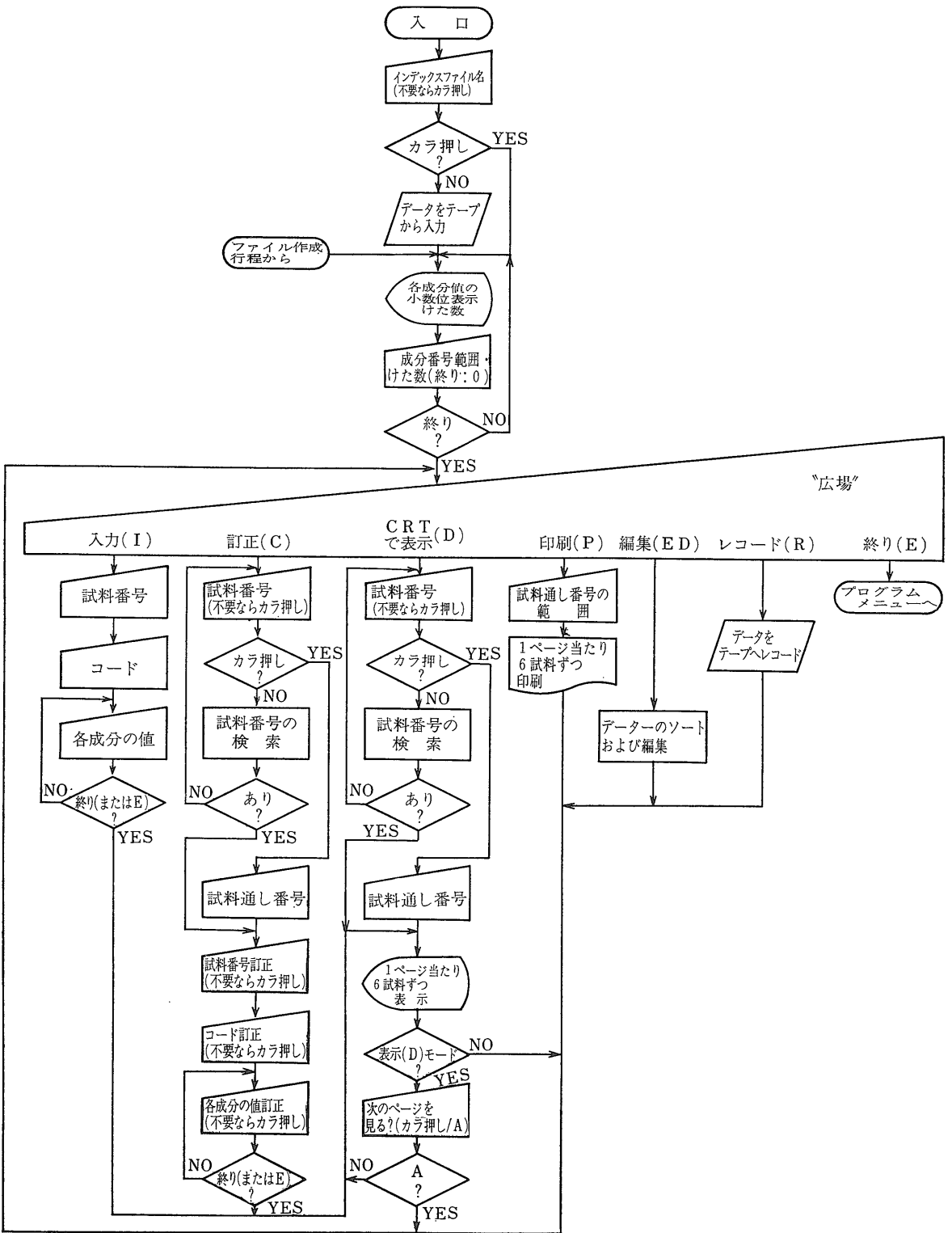
データ入力の行程は、Iの字を入力して選択する。入力作業は、CRT画面に現われる項目名に対応する文字、または数字を順に入力すればよい。1試料ごとに

試料番号
コード
各成分の値

の順に入力する。さきほど述べたノーデータの場合は、数値等を入力しないままプログラムを進めるキーを“カラ押し”する。このシステムでは、このキーのカラ押しによって操作の簡略化と能率化をはかっている。

注) INPUT文実行の際、データを入力しないままプログラムを進めるキーのカラ押しをするとき、パソコンの機種により、その対応が異なる。筆者が実際にテストできた範囲で、参考のため結果を記す。

1. カラ押しの場合以前に入力された数値または文字を保持したまま、プログラムが進む機種——NEC PC-8001 同 PC-8001MkII 同 PC-8201 シャープ X1 同 PC-



第1図 データの入力・訂正・編集行程の流れ図

Current Format:

1 SiO2	2.00	2 TiO2	2.00	3 Al2O3	2.00
4 Fe2O3	2.00	5 FeO	2.00	6 MnO	2.00
7 MgO	2.00	8 CaO	2.00	9 Na2O	2.00
10 K2O	2.00	11 P2O5	2.00	12 H2O+	2.00
13 H2O-	2.00	14 CO2	2.00	15 C	2.00
16 S	2.00	17 Others	2.00	18 Total	2.00
19 Zn	0	20 Pb	0	21 Cu	0
22 Ni	0	23 Co	0	24 Cr	0
25 V	0	26 Li	0	27 Th	1.0
28 U	1.0	29 X	2.00	30 Y	2.00

** Fix the Number of Digits on the Right-Side of the Decimal Point (DDD,ddd)

第2図 数値の出力けた数を指定する行程での表示例

CRT 表示または印刷の際に表示される数値の小数点以下のけた数を 各成分について指定できる。各項の数値と小数位けた数で表現様式を示している

1500 横河ヒューレットパッカーD(YHP)9845T 9836A

2. カラ押しにより 数値は0 文字は空白 (null) が入力されてプログラムが進む機種——NEC PC-9801 同 PC-8801 東芝 PA-7007 同 PA-7010

3. カラ押しするとデータを要求し 先へ進まない機種——カシオ FP-1100 エプソン HC-20 富士通 FM-11 同 FM-8 同 FM-7 日立ベーシックマスター Jr 同 Mk5 シャープ MZ-2000 同 MZ-700 YHP 85.

試料分が CRT に表示され 行程は広場へ戻る。ここでキーのカラ押しをすると 同種の行程へ再び進入する。入力を便利にするために 特殊記号が定義してある。

- / 入力中の試料について 試料番号の行程に戻る。
- = その項目の前回の入力と同値を入力する。
- すでに値の入っている項目をノーデータにする。
- E 数値データの入力を中断して広場へ直行する。

1 試料分の入力が終るごとに その試料を含む最大6

Serial No.	13	14	15	16	17	18
Code	KT203B	KM3	KT143A	KT214A	KT143C	KT216B
	S2TUSC.WTO	S2TUSC.WTO	S2TUSC.WTO	S2TUSC.WTO	S2TUSC.WTO	S2TUSC.WTO
SiO2	61.23	59.18	60.05	43.86
TiO2	.73	.72	.66	1.01
Al2O3	17.60	18.31	18.07	12.57
Fe2O3	1.52	7.59	1.0891
FeO	4.92	0.00	5.24	4.17
MnO	.05	.06	.0760
MgO	2.40	2.27	2.56	1.53
CaO	.76	.65	.92	15.42
Na2O	1.72	1.95	2.33	2.72	2.89
K2O	2.93	2.69	2.76	2.63	1.05	.40
P2O5	.20	.20	.2722
H2O+	4.41	4.86	3.99	2.57
H2O-	.46	.94	.4030
CO2	.2509	11.68
C	.71	.78	.7321
S2233
Others1929
Total	99.89	100.20	99.63	5.35	99.61	.40
Zn	108	134	134	115	73	52
Pb	28	35	18	20	2	2
Cu	34	50	56	45	9	8
Ni	46	46	42	36	15	10
Co	16	19	19	19	15	13
Cr	55	55	.49	49	29
V	176	192	184	155	145	104
Li	51	49	66	69	30	42
Th	12.0	8.5	9.9	1.6
U	2.8	2.2	2.87
X
Y
R-No.	13	14	15	16	17	18

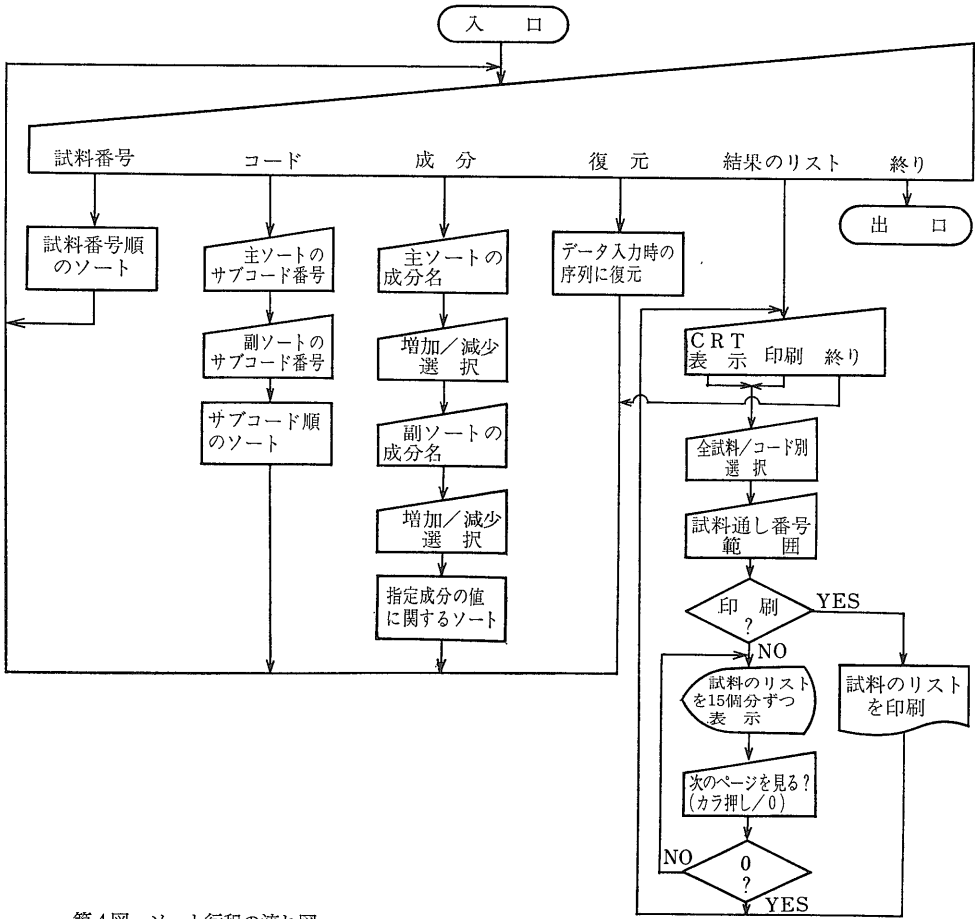
*Data: KITAKAMI SEDIMENTS

*Page 3

第3図 CRT に表示される表の例

点で表わされている欄はノーデータ。 最下段 R-No. はデータの入力時(またはデータテープ内)の序列で最上段の Serial 番号とは最初は一致しているがソート後は食い違ってくる。

入力または訂正された試料を含む最大6 試料までが表示されるので データの横並びによるチェックができる



第4図 ソート行程の流れ図
 データをソートする行程は データ処理用のプログラムにも すべて付けられている。

以上の記号は いわば分岐用のコマンドとして働く。化学分析値など数値データの入力の際にも使用できるようにするために 数値も一度文字列変数で受けて VAL 関数によって数値に変換する仕組みになっている。

1 試料分の入力が終るごとに CRT 画面に表示されるデータの例を第3図に示す。

3. データの訂正

訂正すべきデータは試料番号または (キーのカラ押しにより) 通し番号で 指定できる。データ訂正の行程は入力行程とプログラムの同一部分を共用するので 操作も同じである。ただし 訂正の場合は 各項ごとに順次現データが表示されるので これを確認しながら 訂正箇所では正しい値を入れる。訂正しない項目では キーをカラ押しする。ここでも入力行程の特殊記号が使える。したがって 訂正が終了しだい 数値データ

の項目のところでEを押して脱出すると 作業能率がよい。

4. データのソート

広場の行程で ED を選択するとデータ編集行程へ入る。その中で今回はソートについてご説明しよう。この行程の流れ図を第4図に示す。

a) ソートできる項目

ソート (Sort) とは データをある一定の順序に並べ変えることである。このソートは ASCII (American Standard Code for Information Interchange) 文字コードの値の増加順 または減少順に行われる。

ソートの内容は つぎのとおりである。

ア 試料番号：増加順

Ser.	No.	1	2	3	4	Zn	Ni	R-No.
126	U12B	S	3TL	SC.	SUT	58.000	13.000	53
127	6-608	I	2SE	SD.	E..	58.000	12.000	211
128	KT1009	S	7CA	SC.	CUB	58.000	2.000	119
129	KT133B1	S	US.	CUM	...	59.000	18.000	148
130	KT108	S	2TU	CM.	WTO	59.000	16.000	22
131	71K387X	S	US.	CUM	...	59.000	12.000	144
132	71K18D2	S	US.	CUM	...	59.000	7.000	161
133	KS1005D	S	US.	CFV	...	59.000	1.000	173
134	5-404	I	2SE	HD.	W..	60.000	52.000	255
135	KT635C	S	4KA	SS.	CKE	60.000	30.000	78
136	6-718	I	2SE	SDH	W..	60.000	24.000	233
137	KS0904-2	S	4KA	SI.	SOG	60.000	20.000	40
138	KT513Q	S	US.	CUM	...	60.000	20.000	164
139	KM231	S	3TL	SS.	CKE	60.000	16.000	67
140	11-135B	N	...	SD.	...	60.000	12.000	195
141	KT1001B	S	3TL	SS.	EKA	62.000	19.000	92
142	KT276B	S	3TL	SS.	SUT	62.000	19.000	54
143	71K384E	S	2TU	SS.	WTO	62.000	11.000	21
144	KT1010B	S	US.	CMP	...	63.000	53.000	178
145	71K19A	S	US.	CUM	...	64.000	34.000	166

第5図

ソート結果のリスト例

ここでは Znの値が増加順に並べられており 同値の区間では Niの値の減少順に配列されている。 ソート後の通し番号はSer.で データ入力時のものはR-No.で示される。
この例は使用者が指定する通し番号範囲(126~145)のすべてのデータが印刷されている。 このほかに コードが変化するたびにだけ印刷されるものがあり とくにコードによってソートを行ったときに利用する。

(LIST END)

イ コード：使用者は4つのサブコード中の任意のものを指定でき(主ソート) さらにこのサブコードに同値のものが並ぶとき その区間についてはもう1つ別のサブコードについてのソート(副ソート)ができる。 ソートはいずれも増加順。

ウ 成分(項目)：成分名ファイル中の任意の成分について主ソートの成分と副ソートの成分を指定し それぞれについて 増加順または減少順を選択できる。

h) ソートの方式

このシステムでのソートは 使用計算機 YHP-9845 TのROMを利用して 迅速に行っている。 所要時間は各成分の数値の場合 約1,400試料で16秒程度 約100個では1秒未満である。 この方式では ソートするといっても 実際には 配列内でのデータの移動は行わず いわば“番号札”に記入するというやり方である。

この説明に先立ち このシステムでの数値データの配列について記す。 数値は二次元配列Dに入れられており その行要素が各試料 列要素が各成分に対応している。 したがって 通し番号第I番の第J成分の値は当然

$$D(I, J)$$

にある。 これとは別に整数型一次元配列Nを用意する。 データを配列Dに入力する際に 配列Nに対して

$$N(I) = I$$

の操作を行う。

いま仮に配列Dの第J成分の値についてソートする場

合を考えよう。 このとき配列Dの内容は移動させず ソート後の通し番号第1番となるデータのDの行番号をN(1)に書き込む。 以下第2番以降も同様にして 順次Nに記入する。 配列Dの内容を取り出す際には その行要素に配列Nを用いて

$$D(N(I), J)$$

という形で取り扱うことにより 配列Dそのものには手を付けずに、ソートが実行できたことになる。

この方式は 試料の通し番号について データ入力時とソート後の関係付けを行ったに過ぎない。 その利点として つぎのものがあげられる。

まず メモリー内でのデータの移動を行わないのでソートが短時間で終了する。 つぎに データが複数の配列から構成されていても 連動してソートができる点も見逃せない。 このシステムでは 試料番号とコードが二次元文字列配列No\$に入れられている。 これらについても その内容は それぞれ

$$No\$ (N(I), 1) \dots \dots \text{試料番号}$$

$$No\$ (N(I), 2) \dots \dots \text{コード}$$

として取り扱うことができる。

さらに データ入力時の通し番号N(I)と ソート後の通し番号Iとが つねに簡単に対応できるので便利である。

もうひとつの利点は データをどの項目に着目して何回ソートを繰り返えそうとも その序列をデータ入力時と同じ状態に戻せることである。

すなわち 試料数をSとして
FOR I=1 TO S

N(I)=I
NEXT I

という操作で きわめて簡単に“元のさや”に納めることができるのである。

ソートの行程には 結果を CRT 画面または印刷で見ための行程が付いている。その出力例を第 5 図に示す。

このリストでは ソート後の新しい通し番号 (Ser.) 試料番号 各サブコード 2 成分の数値 および その試料の入力時の序列 (R-No.) とが表示される。これらのうち 2 つの成分については 成分に関するソートを行えば その成分が表示され 初回で成分に関するソートを選択しなかったときは そのデータの第 1 成分と最終成分が表示される。

リストには 2 種類あって 使用者が指定する通し番号範囲の全データを表わすもの (第 5 図) と コード内容が変化する節目のデータだけ出力するものがある。後者は データをコードに関してソートし コード別の試料数を調べるときなどに便利である。

c) ソート結果の利用法

ソートをすれば 使用者が着目した項目に関して データを整理と配列できるわけであるが ただ単にそのような美しさのためだけでなく さまざまな効用がある。

ソートの機能は データ処理の際にも大きな役割を演じるが 今回は データ入力に関係した部分についてだけ述べよう。

まず 入力データの数値範囲 (最小値・最大値) を知ることのほか 有効データ数を知ることができる。すなわち 数値データの場合 増加順にソートすると ノーデータの“値” (9×10^{99}) は リストの末尾に配置されるので その分を差し引けばよい。

コードに関してソートして 結果をコード別リストで見ると コードの種類とそのコードに属する試料数がわかる。あらかじめ定義したコード以外のものが現われる場合は 誤字つまり入力ミスと判定できる。

データが重複して入力されるのを防止することができる。まず適当な成分についてのソートおよび副ソートを行った上で 結果を出力させる。このとき もし同一のデータが多重に入力されていれば 表の中で隣合っ配列されるので そのようなデータを簡単に発見することができる。重複したものは 試料を削除する行程を使って除外する。言い換えれば 追加しようとしている試料が入力済みのものであるかどうかも 事前にチェックできることになる。これらの調査用に あらかじめ任意の成分についてソートした表を常備しておく

追加試料の内容を この表で調べると 簡単に見分けがつく。

このような作業に使う表は ソートの行程で出力されるリストでも通常は間に合う。このリスト中にある 2 成分の数値が異なれば 少なくとも同一試料ではない。もし 2 成分の数値が ともに比較する試料の数値と一致していたら リスト右欄の R-No. の値を頼りに データ入力時などに出力しておいた表と 全項目について 一対一の比較をすればよい。

5. あとがき

このシステムが 使い易さ (大げさに言えば人間工学) を追求していることは これまでも何度か強調して来た。プログラムを進めるキーをカラ押しして 入力の手間を省く方式の多用も そのひとつの表われである。

この操作をただちには受け付けられない機種も多く 使用者側の使い勝手の良さに対するメーカーの姿勢のようなものがうかがわれて 興味深い。

最近こんな話を耳にした。ある個人商店の若い店主が 経理用にいわゆるパソコンとソフトウェアを購入して 商品の在庫管理や売上金の集計に使っているそうだと。ところが そのプログラムでは 入力ミスの訂正がままならず 一ヶ所でも誤りがあると一連のデータを最初から入れ直さねばならないという。この煩わしさにパソコンを操作する女子事務員が悲鳴を上げて 結局 入力の誤りはあとから数値をペンで書き改めることにしたそう。

これでは 高価なパソコンも台無しである。この程度の水準のソフトウェアが 現在でも商品としてまかり通っているとは 筆者にすればむしろ驚きではある。

人間が必ず犯す入力ミスをバックアップする方式は GEOCAPS の前身のそのまた前身のプログラムで 1975 年ごろにはすでに確立していたのである。これはプログラムの技術上少しもむずかしくはないのであって 要はプログラムを作る側が 使用者の心理や生理なども思いやり いかに使え易いプログラムとするかという 言わば哲学の問題である。会話型計算機を使用する人が急増した“パソコン時代”のソフトウェア技術には そのような事柄が当然含まれると信じる。

引用文献

- 吉井守正 (1981) 岩石用化学分析データ入力のプログラム。
地質ニュース no.317, p.51-58.