DATA STORAGE AND RETRIEVAL FOR THE PALAEONTOLOGICAL COLLECTIONS, UNIVERSITY OF ALBERTA

by BRIAN JONES

ABSTRACT. The Palaeontological Collection at the University of Alberta, comprising one to two million fossils, was formed by combined donations from Shell Canada Ltd., J. C. Sproule and Associates Ltd., Imperial Oil Ltd., Gulf Oil Canada Ltd., Panarctic Oils Ltd., and the University of Alberta type collection. In order to accurately curate and permit easy access to information a computer-based retrieval system was designed using the SPIRES (Stanford Public Information Retrieval System) language. For each collection of fossils, twenty-five categories of information are input. From these, sixteen elements are stored in indices and are available to search procedures. One of the main advantages of the system is that all required catalogues, index cards, dictionaries, and cards for specimen trays can be generated from a single input of information. The SPIRES program is relatively easy to use even by people not familiar with computer terminology.

THE Department of Geology at the University of Alberta has maintained an exhibit of vertebrate and invertebrate fossils since 1912, in addition to building up substantial research and teaching collections. Until recently the collection, although large, did not warrant an indexing system other than standard card indexes. The situation changed dramatically in 1977 when Shell Canada Ltd., Gulf Oil Canada Ltd., J. C. Sproule and Associates Ltd., Imperial Oil Ltd., and Panarctic Oils Ltd. donated their invertebrate fossil collections to the University. The total collection contains an estimated one to two million specimens collected during the donor companies' exploration of northern Canada (primarily in the Yukon and Northwest Territories, including Arctic Canada).

CHOICE OF DATA RETRIEVAL SYSTEM

The acquisition of such an immense collection of fossils meant that considerable thought had to be given to establishing good, strict curating methods and procedures. Of primary concern was that of choosing a suitable data storage and retrieval system. Individual companies had maintained card indexes, each designed to their own specific needs and in their own specific way. If the collections were to be united into a single collection a master indexing system was needed. To this end, the following points were considered:

1. Each company stipulated that they must have easy access to the material they had donated. Since each company had its own unique method of cataloguing, any master system would have to be designed so that each company could access records using their specified parameters.
2. The master system had to be easy to use and provide the maximum return for its cost.
3. The system had to be such that most of the currently available information could be input within the shortest possible time.

The above points indicated that a multi-indexed, computer-based retrieval system would have to be considered. Alternatively numerous card indexes (the actual

[Special Papers in Palaeontology No. 22, pp. 175–187.]

number depending upon the number of indexes needed) or some form of microfilm-based system would have to be used. Card indexes can be keyed on no more than two or three elements; if additional key elements for searching are required, additional card indexes are needed. Obviously such a system could easily become cumbersome and inefficient. The index-card systems provided by the donor companies are keyed on elements such as geologist (i.e. collector), field season, or locality. However, if a request for all Devonian brachiopods were received it would be impossible to obtain the information without manually searching all the existing indexes. This consideration pointed to the fact that it would take a considerable amount of time to establish any additional keyed index systems that might be needed.

A computerized system is probably easier to use than a card index and, more important, is infinitely more accurate.

A computer-based curating system would mean that each record and its pertinent information would have to be typed (and commonly in an abbreviated form), and proof-read only once. All catalogues, tables, labels, tray cards, and any index cards needed could be produced from the single typed input. Alternative systems would require that the same information be typed many times. Thus, the computerized system offers an immediate saving of time.

The above discussion has pointed out some of the practical advantages of using a computerized curating system. An important consideration in this context is that of cost. The following example outlines some of the cost considerations that can be made. Costs that are approximately the same for both systems (e.g. cost of typewriter, paper, etc.) have been omitted from these calculations. The comparison is based on a total of 1 000 000 records and no allowance has been made for repetition, correction of mistakes, or proof-reading.

Card-index system

Cost estimates assume that the information is typed on to forms which have a standard layout and titles printed.

1 copy of each record of average length 400 characters	400 000 000
1 copy of a card suitable for specimen trays—assumes 1 species per card and average of 5 species per record	550 000 000
Sets of index cards, assuming 6 keyed sets and high degree of cross-referencing and abbreviations	200 000 000
TOTAL	1 150 000 000 characters

A good typist typing 15 000 characters per hour would require approximately 50 years of continuous typing to complete the task. Proof-reading, correcting mistakes, updating records, or adding records would add to this.

Computerized index system

Costs for a computerized system are incurred by:

1. Typing data into computer. This aspect of cost is greatly reduced because:

(*a*) each record is input only once; all reports, index cards, tray cards, and listings are generated from the single input;

(*b*) the size of each record is reduced using encoding procedures outlined later in this paper. Thus, the average record size of 400 characters is reduced to an average of 250 characters.

Thus, to input all the data for 1 000 000 records, a total of 250 000 000 characters would be input. Assuming a similar typing rate of 15 000 characters per hour, approximately 10 years of continuous typing would be required.

Computer costs for writing the program, compiling data, and storing data are difficult to estimate accurately since it would be a function of the number of records input. For example, according to the estimates above, approximately 100 000 records would be entered every year. After one year storage costs would be about \$6000–9000; after 10 years the cost would be about \$50 000–80 000, and the total cost over the 10 years about \$300 000–400 000. (These costs take into account estimated costs of storing in the indexes.) However, such costs can be drastically reduced by transferring all the information to computer tapes and thereby making it an off-line system.

Comparison of systems

Assuming that a typist is paid about \$10 000/year, total funds of at least \$500 000 over 50 years would be required. By comparison \$100 000 over the next 10 years for the computerized system would achieve the same. This saving in time and money is of course offset by the computing costs.

Over the next 10 years the absolute cost of the computerized system would be far greater than that of a card-index system. However, in that period all one million records could be input and all necessary listings and cards produced.

In the final analysis a computerized curating system was chosen because:

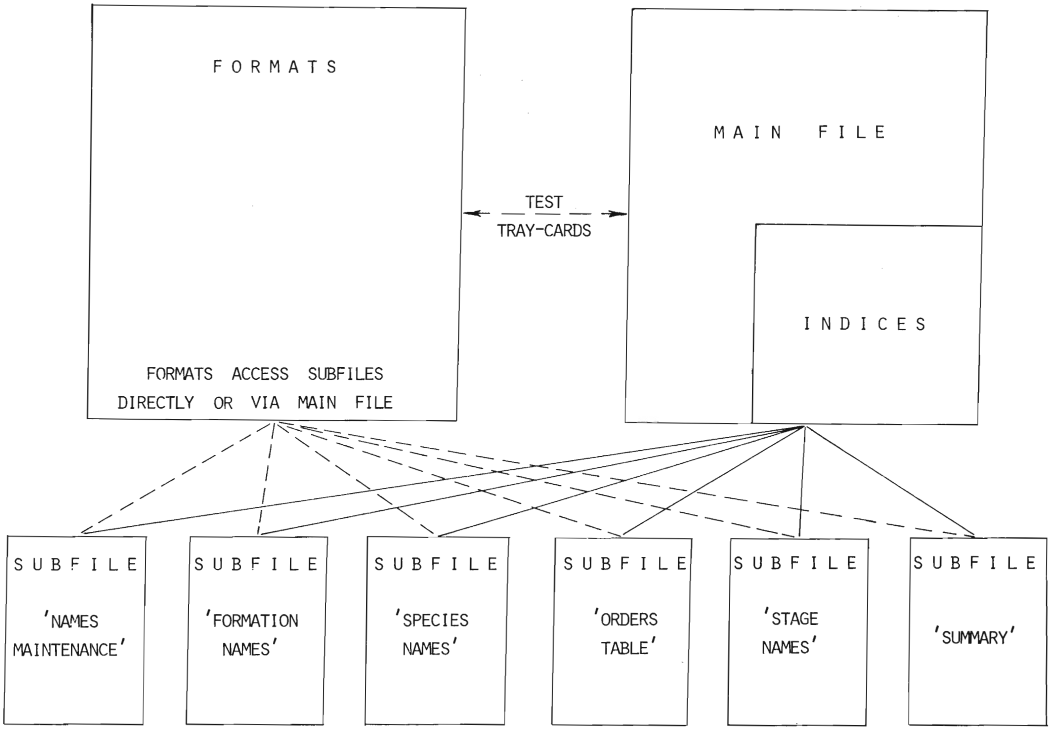
1. All the records could be input within 10 years rather than the 50 years needed for a card-index system.
2. It permits multi-indexed records keyed on far more elements than allowed for in the card-index system previously used as an example.
3. It permits easier and more accurate search procedures.
4. From a single input all required catalogues, listings, tray cards, labels, and index cards can be produced accurately. If the original input is accurate there is no further need for proof-reading.
5. It permits easier updating of records; using the EDITOR facility (Hogg and Tenisci 1976), only updating of original input is needed.

DESIGN OF A COMPUTER-BASED CURATING SYSTEM

The language chosen for designing the curating system was SPIRES (Stanford Public Information Retrieval System) (Jackson 1977; Jackson and Davies 1976) since:

1. Although costs for writing the program, compiling the program, and data input are high, search costs are relatively low.
2. SPIRES is used extensively at the University of Alberta and consultants were readily available in the Computing Science Department. In particular I am grateful to Ron Senda of that department who solved many of the problems encountered during the course of programming.
3. The Department of Computer Sciences at the University of Alberta runs short courses for people who need to use SPIRES. Such courses cover all aspects of the language and provide the basis for programming and use of the system. The system described in this paper was written by the author following attendance at such courses.

Twenty-five categories (composed of forty-five elements) are sufficient to cover all the pertinent information of one record. These categories are all contained in the *main file* (text-fig. 1). Linked to the main file are seven *subfiles* (text-fig. 1).



TEXT-FIG. 1. Diagrammatic illustration of the relationship between main file, indexes, subfiles, and formats in the program used for data storage and retrieval.

Function of subfiles

The subfiles linked to the main file are of two types, namely:

1. A subfile containing all the input records.
2. Subfiles containing encoded data.

These subfiles have three advantages:

1. They permit coded input of elements that:

(*a*) are long (i.e. 10 characters or more); use of subfile permits storage of element using only 2 bytes (1 byte is amount of computer space needed to store 1 character) since integer code of up to 10 characters is converted to a binary length of 4 bytes. This saves on storage costs;

(*b*) in situations where the same name is used time and time again, only the code has to be used. This saves on input time.

2. Coded input is less complicated.

3. Each subfile can be treated as a main file and additional elements pertaining to the main theme of the subfile can be used.

To demonstrate these points, consider a hypothetical example of 100 records containing fossils from 100 different levels in the same formation. If, for example, the formation name was the Read Bay Formation then a total of 1800 characters (blank spaces between words included as 1 character) would have to be input for this element by the time all 100 records had been input. By using a subfile, a code number such as 8, 75, or 1045 is assigned to the formation name. Thus, for the same 100 records only 100, 200, 300, or 400 characters need be input and only 200 characters (since code number is converted to a binary length of 2) are stored. This causes a saving in two ways since:

1. Storage costs are 0·04 cents/month/4096 characters. Thus, by using a subfile for the 100 records a saving of approximately 1·6 cents/month (approximately 20 cents/year) is achieved. This in itself is not very high; however, consider a system with 1 000 000 records and 6 elements coded using subfiles; then assuming an average length of 15 characters per original name, a saving of about $9140/year is achieved (increased to $18 000–27 000 if savings on indexes are included).

2. Typed input is reduced from 18 characters to 1, 2, 3, or 4 characters (depending on length of code). Thus, for the first example of 100 records the input is reduced by 1400–1700 characters. For the second example of 1 000 000 records the input is reduced by 66 000 000–84 000 000 characters. Assuming an average input cost of $8.50/15 000 characters per hour a saving of 4000–5600 hours or $37 400–47 600 is achieved by the time all records have been input.

The use of subfiles and additional encoding in the main file results in a considerable saving both in money and time. For example, a set of 1 000 000 records each having an original average length of 400 characters can be encoded to produce about 40% saving on storage costs per year. The additional saving because of shorter typed input is realized either by reduced time to input a given number of records or by an increase in the number of records input per year.

The disadvantages of using subfiles are twofold:

1. The codes must be determined prior to input.
2. The codes must be used during search procedures.

The codes can easily be obtained from a simple card index or dictionary (which can be generated by the program once information has been input), without any great time involved. These disadvantages are minimal compared with the cost advantages.

The code that is input and stored is translated to the full name during output procedures either in the main file or during formatting of output. The latter approach is used in this system since access via the main file (without formats) can be used for editing procedures.

Example of a subfile

The subfile 'Formation Names' can be used to illustrate the use of any of the six subfiles in the system:

Input to main file would be: FORM = 1;

Main input to subfile would be: CODE = 1; NAME = READ BAY FORMATION;

One advantage of a subfile is that accessory elements pertaining to the main theme of that subfile can be added. For example, in the case of the subfile 'Formation Names' the following information can be input:

1. Location of type section
   - (*a*) Geographical location
   - (*b*) Latitude–longitude
   - (*c*) NTS-location
2. Reference defining type section
   - (*a*) Author(s)
   - (*b*) Year
   - (*c*) Title
   - (*d*) Publication

FORMATION NAME:- READ BAY FORMATION CODE FOR MAIN PROGRAM:- 1 LOCATION OF TYPE SECTION:- READ BAY, CORNWALLIS ISLAND LAT-LONG:- 96W 75N NTS LOC:- 58F15 REFERENCE DESCRIBING TYPE SECTION:- THORSTEINSSON, R. 1958. CORNWALLIS AND LITTLE CORNWALLIS ISLANDS, DISTRICT OF FRANKLIN, NORTHWEST TERRITORIES. GEOLOGICAL SURVEY OF CANADA MEMOIR 294, 134 P.	FORMATION NAME:- LEOPOLD FORMATION CODE FOR MAIN PROGRAM:- 3 LOCATION OF TYPE SECTION:- PORT LEOPOLD, NE SOMERSET ISLAND LAT-LONG:- NTS LOC:- 58D14 REFERENCE DESCRIBING TYPE SECTION:- JONES, B. AND O.A. DIXON, 1975, THE LEOPOLD FORMATION; AN UPPER SILURIAN INTERTIDAL/SUPRATIDAL CARBONATE SUCCESSION ON NORTHEASTERN SOMERSET ISLAND, ARCTIC CANADA. CAN. JOUR. EARTH SCI. V.12, P. 395-411

TEXT-FIG. 2. Example of 5 in. × 3 in. index cards that can be produced from the subfile 'Formation Names'. Note that the pertinent information has been manipulated to suit the specific requirements of the index cards. The solid lines indicate the position of the perforated lines on the computer output card.

Thus, on the index cards used for determining the code for formation, useful information pertaining to the location of the type section and the reference in which it is described is readily available (text-fig. 2).

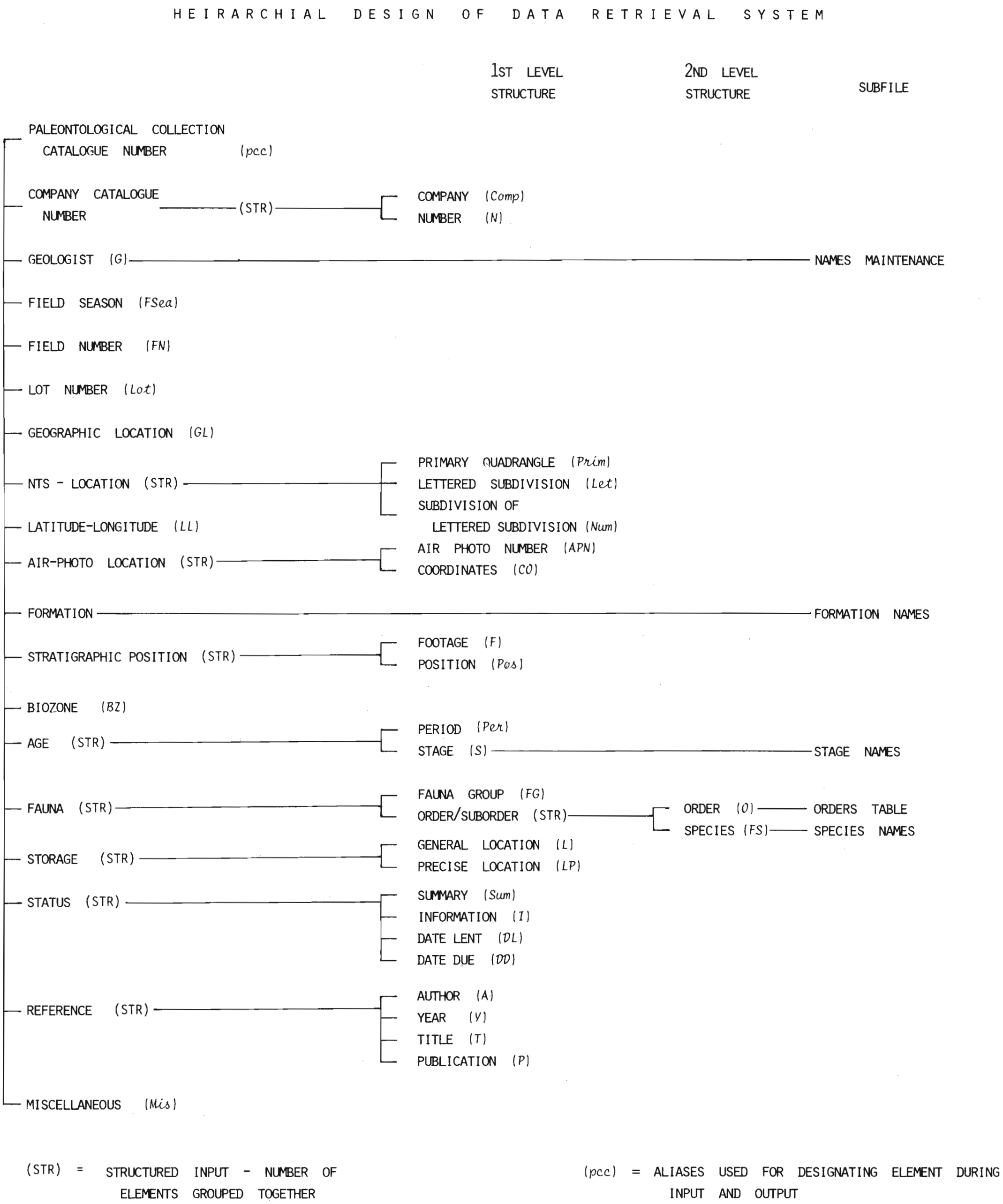
Function of main file

The main file defines the elements that are to be input for each record (text-fig. 3). It is programmed such that only the Paleontological Collection Catalogue Number (alias pcc) and the Company Catalogue Number (aliases comp. or CN) have to be input; the remaining elements are optional. Each element has a specific name with a maximum length of sixteen characters; but definition of an alias permits the shortening of that name for input and search procedures.

Assuming that all the information is available for a particular record then a typical input is given below (element aliases in lower-case letters, input data in small capitals):

pcc = 3000; c = 1; cn = 1678; fsea = 1977; gl = READ BAY, CORNWALLIS ISLAND; ll = 76N 97W; fn = BJ77160; form = 1; bz = NOT SPECIFIED; nts = 15 × WN5513; mis = COLLECTION CONTAINS WELL-PRESERVED BRACHIOPODS; a = BRIAN JONES; y = 1977; t = VARIATION IN THE UPPER SILURIAN BRACHIOPOD ATRYPELLA PHOCA (SALTER) FROM SOMERSET AND PRINCE OF WALES ISLANDS, ARCTIC CANADA; p = JOURNAL OF PALEONTOLOGY, V. 51, P. 459–479; per = SIL; s = 1; fg = 14; 0 = 1; fs = 1; foot = 260 FT; pos = AFB; 1 = R; lp = RACK 20A; sum = L; i = COLLECTION ON LOAN TO A. N. OTHER, DEPT. OF GEOLOGY, UNIVERSITY OF SOMEWHERE; dl = 10 MARCH 1977; dd = 1 JAN 1978;

In addition to the coding achieved through the use of subfiles, additional coding is programmed in the main file through the use of system functions (procedures that are used repeatedly are part of the SPIRES system and can be implemented in the main program by coded actions). Company Name, Period, Fauna Group, and Summary Status are coded in this manner. The advantages of coding in the main file are the same as those discussed with reference to the subfile. The difference is that input can be achieved either by using the code or by using the full name (input using subfiles permits use of code only). Although this dual input facility is potentially very useful the system function that permits coding in the main file is limited to 128 codes and does not permit the input of accessory elements.



TEXT-FIG. 3. Structure of the main file showing elements and their aliases (in italics).

INDEXING AND SEARCH PROCEDURES

Of the 25 possible categories relating to each record the following 10 categories composed of 17 elements are indexed and thus available for searching (aliases in brackets):

1. Palaeontological Collection Catalogue Number (pcc)
2. Company Catalogue Number
   (*a*) Company (comp)
   (*b*) Number (CN)
3. Geologist (G)
4. Field Season (FSea)
5. Locality Number (LN)
6. Formation (Form)
7. Age
   (*a*) Period (Per)
   (*b*) Stage (S)
8. Fauna
   (*a*) Group (FG)
   (*b*) Order/Suborder (O)
   (*c*) Species (FS)
9. Summary Status (Sum)
10. Locality (NTS-location)
   (*a*) Primary Quadrangle (Prim)
   (*b*) Lettered Subdivision of Primary Quadrangle (Let)
   (*c*) Number Subdivision (Num)
   (*d*) Air-photograph Number (APN)

Search procedures can be conducted on any single index or any number of indexes in combination. For example:

REQUEST: Find Palaeontological Collection Catalogue Number 1000.
?find pcc 1000 (Request given to computer)

RESULT 1 (Result is response from computer giving number of records satisfying the above request)
?Type (type command causes output of result, in this case all the information pertaining to Paleontological Collection Catalogue Number 1000)

REQUEST: Find all Pridolian brachiopods from the Read Bay Formation of Cornwallis Island
?find S 1 and FG 14 and Form 1 and Prim 58 and Let G and Let H

RESULT 10
?Type (lists all information from the 10 records that satisfy the above request)
or

?Type pcc (lists Paleontological Collection Catalogue Numbers of the 10 records satisfying the above request)

REQUEST: Produce listing of fossils donated by Shell Canada Limited
?find comp 1

RESULT 10 000
?Type (all 10 000 records output in full)

REQUEST: Find record number Shell 12567
?find comp 1 and CN 12567

RESULT 1
?Type (lists single record in full)

FORMATS

On output the input data will be set out according to the default format (a default format is part of the SPIRES system and is automatically used if a specially written format is not specified). However, the data can be manipulated using formats to suit whatever requirements are needed.

This can be illustrated by considering the subfile 'Formation Names'. By default format the output shown in text-fig. 4 is obtained. However, an output of this form is of little use. By formatting, the data can be manipulated to fit on index cards as

```
FORMATION = 1;
NAME = READ  BAY  FORMATION;
POINTER = " ";
GEOG-LOC = READ BAY,  CORNWALLIS ISLAND
  LATLONG = 96W  75N
  NTS = 58F15
AUTHOR = THORSTEINSSON, R;
  YEAR = 1958;
  TITLE = CORNWALLIS  AND  LITTLE  CORNWALLIS  ISLANDS,
            DISTRICT  OF  FRANKLIN,  NORTHWEST  TERRITORIES;
  PUBLICATION = GEOLOGICAL  SURVEY  OF  CANADA  MEMOIR 294,  134 P.
```

TEXT-FIG. 4. Default format for the subfile 'Formation Names'. Note that the element names are in abbreviated form and ordered according to the order of input. The resultant output is not aesthetically pleasing.

shown in text-fig. 2. Such cards are sorted prior to output and can thus be placed directly into a card index. In the example shown in text-fig. 2 the index card is designed for a system keyed on Formation Name. However, if additional indexes keyed on other elements are needed other formats can be designed. For example, text-fig. 5 shows examples of two formats used to produce dictionaries of species names keyed on their code (text-fig. 5*a*) or name (text-fig. 5*b*).

The other subfiles containing geologists' names, order/suborder names, species names, and stage names each have associated formats which can output the data on index cards or in dictionaries. Although the primary use of these cards or dictionaries is for determining codes during input and search procedures, a very useful 'spin-off' value is becoming apparent. For example, using the cards relating to the subfile 'Formation Names' it will be relatively easy to find the location of the type section of that formation and the reference in which it is described (text-fig. 2). Similarly, the index cards produced from the subfile 'Species Names' contain information about the authorship of the species, the type species of genus, and the references in which the type species and the species are described.

Similarly, the main subfile containing all the record data can also be formatted. To date, two formats have been designed. One format produces a comprehensive listing of all records and contains all the available pertinent information (text-fig. 6).

A.

CODE	SPECIES NAMES	ORDER CODE	FAUNA GROUP
20	ORBICULOIDEA		INARTICULATE BRACHIOPOD
21	DICTYOCLOSTUS		ARTICULATE BRACHIOPOD
22	PRODUCTUS		ARTICULATE BRACHIOPOD
23	LINOPRODUCTUS		ARTICULATE BRACHIOPOD
24	CHONETINA		ARTICULATE BRACHIOPOD
25	CHONETES		ARTICULATE BRACHIOPOD
26	SPIRIFER		ARTICULATE BRACHIOPOD
27	MEAGENOCONCHA		ARTICULATE BRACHIOPOD
28	BELLEROPHON		GASTROPOD
29	STENOCISMA		ARTICULATE BRACHIOPOD
30	DIELASMA		ARTICULATE BRACHIOPOD

B.

SPECIES NAMES	SPECIES CODE	ORDER CODE	FAUNA GROUP
BELLEROPHON	28		GASTROPOD
CHONETES	25		ARTICULATE BRACHIOPOD
CHONETINA	24		ARTICULATE BRACHIOPOD
DICTYOCLOSTUS	21		ARTICULATE BRACHIOPOD
DIELASMA	30		ARTICULATE BRACHIOPOD
LINOPRODUCTUS	23		ARTICULATE BRACHIOPOD
MAEGENOCONCHA	27		ARTICULATE BRACHIOPOD
ORBICULOIDEA	20		INARTICULATE BRACHIOPOD
SPIRIFER	26		ARTICULATE BRACHIOPOD
STENOCISMA	29		ARTICULATE BRACHIOPOD

TEXT-FIG. 5. Examples of parts of the dictionaries produced from the subfile 'Species Names' ordered by (*a*) code number and (*b*) alphabetical ordering of generic names. Such dictionaries are used for determining codes to be input into main program and are produced on $8\frac{1}{2}$ in. $\times$ 11 in. paper.

```
                         PALEONTOLOGICAL COLLECTION CATALOGUE NUMBER:- 10000

COMPANY CATALOGUE NUMBER:- UNIVERSITY OF ALBERTA 6789

GEOLOGIST:- BRIAN JONES              FIELD SEASON:- 1975          LOCALITY NUMBER:- T32          LOT NUMBER:- N/A

GEOGRAPHIC LOCATION:- GREEN VALLEY, 14.5KMS SE PRESSURE POINT, NW SOMERSET ISLAND
    LATITUDE LONGITUDE:- 73 52N 94 57W                NTS LOCATION:- 58C13                 UTM LOCATION:- VM393985
         AIR-PHOTO NUMBER:- A13456                         CO-ORDINATES:- X-425:Y-072

FORMATION:- READ BAY FORMATION
    POSITION IN FORMATION:- 136M Below formation top

AGE OF FAUNA:
         PERIOD:- SILURIAN                    STAGE:- LUDLOVIAN                BIOZONE:- NOT SPECIFIED

REFERENCE IN WHICH PART(S) OF COLLECTION DESCRIBED:-  BRIAN JONES AND O. A. DIXON  1977  STRATIGRAPHY AND SEDIMENTOLOGY OF UPPER
SILURIAN ROCKS, NORTHERN SOMERSET ISLAND, ARCTIC CANADA  CANADIAN JOURNAL OF EARTH SCIENCES, V. 14, P. 1427-1452

                                                      FAUNA

                   FAUNA GROUP                      ORDER/SUBORDER                      SPECIES

              Articulate Brachiopods                 PENTAMERIDA                        GYPIDULA GALEATA

                                                    RHYNCHONELLIDA                      STEGERHYNCHUS BOREALIS

                                                     SPIRIFERIDA                        ATRYPOIDEA FOXI
                                                                                        ATRYPA AP.

                                                    STROPHOMENIDA                       FARDENIA SP.


              Nautiloids

              Pelecypods

    STORAGE POSITION OF COLLECTION:- RACK 20A, BOX 5077

    GENERAL STATUS OF COLLECTION:- Present     QUANTITY PRESENT:- ALL PRESENT
         COLLECTION ON LOAN TO:-
                   DATE LENT:-          DATE DUE BACK:-
```

TEXT-FIG. 6. Example of formatted output from main file that includes all available information. This layout (format Test—see text-fig. 3) is used for producing the main catalogue for the collection on 15 in. × 11 in. paper.

The second format produces small cards which are used in the trays housing the fossils (text-fig. 7).

During formatting procedures, the coded input is transformed to its equivalent full name.

PALEO-COLLECTION NUMBER:- 2000	PALEO-COLLECTION NUMBER:- 2000	PALEO-COLLECTION NUMBER:- 2000
COMPANY NUMBER:- U. OF A. 23456	COMPANY NUMBER:- U. OF A. 23456	COMPANY NUMBER:- U. OF A. 23456
GEOLOGIST:- BRIAN JONES YEAR:- 1977	GEOLOGIST:- BRIAN JONES YEAR:- 1977	GEOLOGIST:- BRIAN JONES YEAR:-1977
LOCALITY:- SOMERSET ISLAND	LOCALITY:- SOMERSET ISLAND	LOCALITY:- SOMERSET ISLAND
FORMATION:- READ BAY FORMATION	FORMATION:- READ BAY FORMATION	FORMATION:- READ BAY FORMATION
AGE:- SILURIAN PRIDOLIAN	AGE:- SILURIAN PRIDOLIAN	AGE:- SILURIAN PRIDOLIAN
FAUNA:- ATRYPOIDEA FOXI	FAUNA:- STEGERHYNCHUS BOREALIS	FAUNA:- ATRYPOIDEA PHOCA

TEXT-FIG. 7. By using another format, cards suitable for use in specimen trays are produced from the main file. It selects pertinent information from the full record shown in text-fig. 6 and prints it on suitable-sized cards. The solid lines indicate the position of the perforated lines on the computer output card.

CONCLUSIONS

Although SPIRES was originally developed for use in libraries it can easily be adapted for use in the museum, and storage of data pertaining to fossils. The system described in this paper permits the storage and easy, accurate retrieval of vast quantities of data. Moreover, since information is only input once, considerable processing time is saved. Data processing using a computer based retrieval system certainly leads to a more effective utilization of the available fossils.

Once designed, the system is very easy to use. Staff members can be taught to use the system in less than a day since commands are logical and not entwined in complex computer terminology. A handbook, being compiled by the author specifically for the program described in this paper, will list all procedures needed to operate the system.

REFERENCES

HOGG, J. F. and TENISCI, T. 1976. The File Editor. *The University of Alberta, Edmonton, Computing Services Reference Manual*, **15**, 1–77.

JACKSON, G. R. 1977. SPIRES/370 File Definition. Ibid. **62**, 1–293.

—— and DAVIES, A. 1976. SPIRES User's Guide. Ibid. **29**, 1–69.

BRIAN JONES

Department of Geology
University of Alberta
Edmonton
Alberta
T6G 2E3
Canada

DISCUSSION

S. Turner. The Hancock Museum (University of Newcastle upon Tyne) decided recently to use the SPIRES system for data storage and retrieval to replace their incomplete card files, to better utilize the information put directly into the standard running catalogues, and to store site records collected over the past two years. SPIRES had just become available on the University computer in August 1977, and with the aid of Dr. N. Rossiter, Peter Robson of the Northumberland Wildlife Trust and I planned the original file definitions.

Since September 1977 a Job Creation team of one graduate and two clerical assistants have been putting data on to SPIRES. To date there are 218 site records on file, and some 5000 geological specimen records of fossils, rocks, and minerals. The system is most effective for almost immediate retrieval of records, either whole or as lists indexed according to various criteria. This replaces the tedious searching of all the running catalogues. Searching and retrieval on relatively obscure criteria is just as fast, and of course so much easier than doing the same by hand on a card index. This advantage will become of increasing importance as the amount of data being stored increases.

One major project will be the generation of a catalogue of type and figured specimens. A major feature is the ability to update any record as new information is available, and to format the output in any way one wishes.

M. G. Bassett. I wonder if curators have thought seriously of the ways in which they will use the computer to answer enquiries in future. At present virtually every type of enquiry that we get in the National Museum of Wales can be answered by simple, direct reference to the collections—they certainly do not require a computer. Is this because we do not have a computer, and is it possible that our enquiries will become more complicated if we develop such a system. I am really asking if we are setting up complex systems that are capable of answering the kinds of questions that will never be asked.

S. Turner. As I mentioned above, we have both site records and catalogue information on computer. We now receive regular enquiries from people such as planning officers about the use of sites. In that respect our system has certainly been useful.

W. D. I. Rolfe. It seems to me that once information of this kind is produced it leads to an increase in the number of enquiries.

B. J. Pyrah. Many of the questions associated with the use of computers for curating seem to me to be insoluble at present. It is a pity, however, that there are numerous very capable graduate students who are out of work who could well be employed to do many of the basic jobs.

W. D. I. Rolfe. Surely the main aim of the computer is to speed up or remove the repetitive and time-consuming tasks. Should not Dr. Jones and others be employing graduates to do research, with the computers producing the cards and generating the indexes.

B. Jones. I would agree entirely with this approach suggested by Dr. Rolfe. In justification of computerization, I should simply like to repeat one example that I have quoted; instead of the secretary taking fifty years to type the cards, the computer-based system could do it in ten years.