THE PRODUCTION OF FAUNAL LISTS BY AUTOMATIC METHODS

by IAN E. PENN

ABSTRACT. Computer programs have been developed in the Institute of Geological Sciences which, after eliminating any unwanted data from original determinations, generate correctly type-set and punctuated faunal lists. These are suitable for direct publication or for easy incorporation into accounts dealing with wider geological topics.

LISTS of species found are a simple and most fundamental means of recording faunal (or floral) distribution and, since the days of William Smith and his recognition of strata distinguished by means of fossil content, they have occupied a place of special importance in geology. After the palaeontologist has recorded his observations, he may still spend a considerable amount of time in non-palaeontological work when he communicates his discoveries, even by a means apparently as simple as a faunal list. Thus the fossil names must be written, typed, and checked (often in several copies). This process may be repeated more than once when, in a large organization, the palaeontologist's report is incorporated in a larger work such as a geological account or memoir written by a colleague, which itself must be typed and checked (often in several copies). Finally, the lists must again be checked at least once on return from the printers, prior to eventual publication. When for any reason some part of the data has to be published additionally or even separately, the whole process may need to be repeated.

Within the Institute of Geological Sciences, analogous problems had been encountered in the production of the relatively more sophisticated stratigraphical range-diagram, and a package of computer programs was written to eliminate non-palaeontological activity as far as possible (Penn 1974; Farmer and Johnson 1975, in press). It was then decided to generate fossil lists suitable for direct publication in the same manner as the range-diagrams and to incorporate a program which would simultaneously eliminate unwanted data. The main features of these programs are here outlined (text-fig. 1), while full program listings may be obtained on request.

The programs are housed on the Institute's IBM 1130 computer configuration and also on the Edinburgh Regional Computing Centre's PDP-11/45 configuration. They may be installed and used to produce 'crude' line-printer output by anyone having access to such machines. Production of type-set lists requires access, however, to a more specialized instrument such as the Institute's Addressograph-Multigraph AM-747. The species dictionaries contain the names used by individual palaeontologists within the Institute, and it is intended that these will be subsequently integrated, as will the data generated by them. It is envisaged that users outside I.G.S. could construct similar dictionaries and data files which may subsequently be brought together. The Institute's facilities are, however, available at the discretion of the Director.

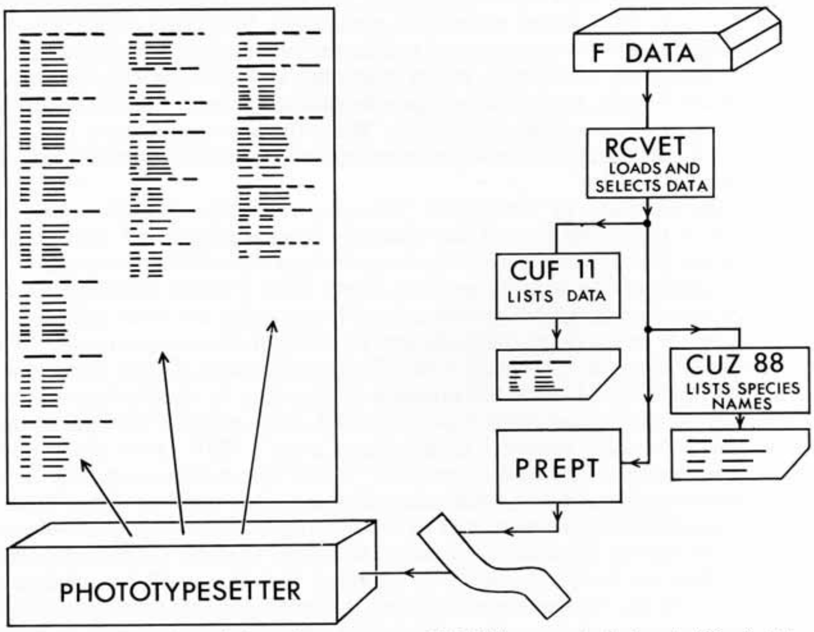
[Palaeontology, Vol. 18, Part 4, 1975, pp. 865-869.]

PROGRAM INPUT, CONTENTS AND OUTPUT

The fossil data (F DATA in text-fig. 1) are presented on punch cards in stratigraphical order such that a preliminary card lists the number of batches in the stratigraphic section being described. Then the first card in each batch states the number of species found as well as the measured stratal range (e.g. depth range in a borehole of the sampled horizon), while each subsequent card records the species code number (corresponding to the full fossil name stored in a companion dictionary) and a code denoting species abundance at that horizon (see Penn 1974). At the present embryonic stage of this data bank, each such stratigraphical section is identified manually.

Program CUF 10 reads these data and stores them on magnetic discs ready for accession by remaining programs. But program RCVET is firstly presented with a list of the code numbers of those species which it is desired to eliminate (or retain, as the case may be) from the main body of the data. Only these selected species, with their abundance codes, are read and stored by program RCVET. Indication is given if such selection results in the elimination of all species from the data. Thus in the example shown (text-fig. 2), the Bivalves have first been selected from the total data and listed, followed by listing of the remaining species.

Preliminary inspection of the data may be made by printing a list of the encoded



TEXT-FIG. 1. Flow chart of the various programs. F DATA represents the input of fossil data. The various identifiers within rectangles represent the various program decks. PREPT produces paper-tape and line-printer output.

data, and a list of all the species found in the total data, by using programs CUF 11 and CUZ 88 respectively (Penn 1974). Program PREPT, however, translates the encoded data and, after consulting the species dictionary, punches out a paper-tape of the full fossil name preceded by an indication of abundance for every determination made. This tape, which contains type-setting instructions obtained from the species dictionary, is fed into a phototypesetter to produce a correctly type-set and punctuated list (text-fig. 2) for each sampling horizon. Provision has been made for an addition and multiplication factor to be incorporated into the depth-range values, so as to

SPECIMEN DATA. ALL MACROFOSSILS

SAMPLING HORIZON 1 35.55 to 36.23

p wood [frag.]
fc *Rhynchonelloidella smithi* (Davidson)
p *Anisocardia bathensis* Cox
fc bivalve [indet.]
p *Catinula* cf. *ampulla* (d'Archiac)
p *Entolium sp.*
p *Grammatodon bathonicus* Cox & Arkell
p burrow [horizontal and straight]

SAMPLING HORIZON 2 36.23 to 37.19

p bivalve [indet.]
p *Modiolus sp.*
p belemnite [indet.]

SAMPLING HORIZON 3 37.37 to 38.12

p serpulid [indet.]
fc rhynchonellacean [indet.]
fc *Rhynchonelloidella smithi* (Davidson)
p *Rhynchonelloidella wattonensis* Muir-Wood
fc *Rhynchonelloidella sp.*
fc arcacean [indet.]
c bivalve [indet.]
fc *Catinula* cf. *ampulla* (d'Archiac)
p *Chlamys (Radulopecten) sp.*
p *Gervillella sp.*
p pectinoid [indet.]

SAMPLING HORIZON 4 38.12 to 38.39

p *Rhynchonelloidella sp.*
p bivalve [indet.]
fc *Entolium sp.*
p *Gervillella sp.*

SAMPLING HORIZON 5 38.39 to 38.60

p *Sarcinella socialis* (Goldfuss)
fc serpulid [indet.]
fc *Rhynchonelloidella sp.*
p *Ornithella bathonica* (Rollier)
p *Procerithium sp.*
fc bivalve [indet.]
? *Camptonectes sp.*
fc *Entolium sp.*
p *Inoperna plicata* (J. Sowerby)
p *Liostrea sp.*
fc *Modiolus anatinus* Wm. Smith
p *Modiolus sp.*
p *Vaugonia sp.*

SPECIMEN DATA. BIVALVES

SAMPLING HORIZON 1 35.55 to 36.23

p *Anisocardia bathensis* Cox
fc bivalve [indet.]
p *Catinula* cf. *ampulla* (d'Archiac)
p *Entolium sp.*
p *Grammatodon bathonicus* Cox & Arkell

SAMPLING HORIZON 2 36.23 to 37.19

p bivalve [indet.]
p *Modiolus sp.*

SAMPLING HORIZON 3 37.37 to 38.12

fc arcacean [indet.]
c bivalve [indet.]
fc *Catinula* cf. *ampulla* (d'Archiac)
p *Chlamys (Radulopecten) sp.*
p *Gervillella sp.*
p pectinoid [indet.]

SAMPLING HORIZON 4 38.12 to 38.39

p bivalve [indet.]
fc *Entolium sp.*
p *Gervillella sp.*

SAMPLING HORIZON 5 38.39 to 38.60

fc bivalve [indet.]
? *Camptonectes sp.*
fc *Entolium sp.*
p *Inoperna plicata* (J. Sowerby)
p *Liostrea sp.*
fc *Modiolus anatinus* Wm. Smith
p *Modiolus sp.*
p *Vaugonia sp.*

SPECIMEN DATA. NON-BIVALVES

SAMPLING HORIZON 1 35.55 to 36.23

p wood [frag.]
fc *Rhynchonelloidella smithi* (Davidson)
p burrow [horizontal and straight]

SAMPLING HORIZON 2 36.23 to 37.19

p belemnite [indet.]

SAMPLING HORIZON 3 37.37 to 38.12

p serpulid [indet.]
fc rhynchonellacean [indet.]
fc *Rhynchonelloidella smithi* (Davidson)
p *Rhynchonelloidella wattonensis* Muir-Wood
fc *Rhynchonelloidella sp.*

SAMPLING HORIZON 4 38.12 to 38.39

p *Rhynchonelloidella sp.*

SAMPLING HORIZON 5 38.39 to 38.60

p *Sarcinella socialis* (Goldfuss)
fc serpulid [indet.]
fc *Rhynchonelloidella sp.*
p *Ornithella bathonica* (Rollier)
p *Procerithium sp.*

TEXT-FIG. 2. Specimen output from a Middle Jurassic borehole near Bath. ? = possibly occurring; p = present; fc = fairly common; c = common. The total data on the left of the diagram has been split into the Bivalve and non-Bivalve sections shown on the right-hand side.

eliminate the punching of unnecessary digits where, for example, closely spaced samples have been taken from considerable depth in a borehole. Such factors may also be used as a security 'link' on the depth range of the sample in the case of confidential material.

PROGRAM PERFORMANCE

The size limits for the data are at present 100 (species) $\times$ 100 (horizons) and the number of species at any one horizon must be limited to 50 as in the range-chart program (Penn 1974). Data preparation (including the noting of the preliminary raw observations) and checking for the computer takes about the same time as preparing a first draft of a clean-copy manuscript for typing. Typing the first copy, however, takes about as long as drafting the manuscript, and checking the first and each subsequent typescript is conservatively estimated at 50% of the time taken to prepare the first manuscript. Typing of subsequent copies is slightly quicker (perhaps by about 25%). Thus, after the first draft of the manuscript, the time spent by the palaeontologist in non-palaeontological work increases by 50% of his original time, and similarly that of the typist by 75% for each successive 'round' of typing. The final 'round' is that done by the printer and is estimated to be 120% of the time taken by an ordinary typist. Thus a manuscript taking 1 hr for the preparation of the first draft of a clean-copy manuscript would take (assuming two further drafts) 2 hr 45 min to submission and 1 hr 20 min work by the printer as against 1 hr for punching and checking and 4 min computer time. Program RCVET, which selects data during the loading process, performs basically the same functions as the normal loading program and, indeed by setting it so that no selection is made, may be used as a substitute for program CUF 10. The operational time taken by program PREPT is dominated by the number of species determinations in the data. Since most of the activity is in punching paper-tape, the time taken to produce the output is almost entirely dependent on the size of machine, the speed, and the arrangement of output devices. Thus the IBM 1130 configuration, on which the program was established, does not conveniently allow separate operation of the central processor and paper-tape punch. The speed of the operation was therefore determined by the speed of the punch which, at 14·8 characters per sec, is very slow. Operation on a larger computer (a PDP-11/45) with a faster paper-tape improved the run times, in a conservative estimate, by a factor of five times. In practice this means a job of around 250 determinations is run in around 5 min. In fact, in a multi-user environment, output devices would be operated simultaneously with entirely different operations, meaning that the time spent producing the fossil list would be almost negligible. The advantage is even more marked if the palaeontologist needs an accompanying range-chart, since the same data input is used, and producing the fossil list simply involves a small amount of extra computer time.

It is thus possible to free the palaeontologist from a very considerable proportion of non-palaeontological work once he has initially recorded and checked his data. In addition, his data can be stored in computer-processable form ready for the performance of other analytical techniques and, in the long term, ready for incorporation into a computerized data bank.

Acknowledgements. The programs were written by Dr. D. G. Farmer (Computer Unit, I.G.S.) and Dr. T. J. Dhonau (Editorial and Publication Section, I.G.S.) advised on the use of the phototypesetter. The paper is published by permission of the Director of the Institute of Geological Sciences.

REFERENCES

FARMER, D. G. and JOHNSON, L. 1975. *Rep. Inst. geol. Sci.* (In press.)

PENN, I. E. 1974. The production of stratigraphical range-diagrams by automatic methods. *Palaeontology*, **17**, 553–563.

IAN E. PENN

Department of Palaeontology
Institute of Geological Sciences
Exhibition Road
London, SW7 2DE

Typescript received 23 December 1974
Revised typescript received 24 February 1975