COMPUTER-BASED CATALOGUING IN BRITISH MUSEUMS

by R. B. LIGHT

ABSTRACT. The data standards evolved by IRGMA (Information Retrieval Group of the Museums Association) and later by MDA (the Museum Documentation Association) enabled the production of a recording form, suitable both for a manual record card and for computer input. By using this MD System a manual catalogue can be translated into a computerized one, without re-cataloguing. MDA is finalizing a package of programs (known as GOS) to generate a variety of indexes and lists from the large hierarchical structure of the MDS, and pilot tests are being run on data from several British museums. The economics of publishing bulky print-out indexes and catalogues suggest that microfiches may have to be employed by museums for this purpose. Trials are under way for type-setting direct from the computer magnetic tape, which is another way to reduce unit costs. The interactive use of computer terminals for information retrieval is not currently justifiable in terms of economics, or demand.

COMPUTERS do not provide a magic solution to a museum's documentation problems. They require inputs of staff time and expertise, just as in a manual system. Also the type of expertise needed is rather different, and this can be a disincentive to adopting such a system. The real advantage of computers is that they allow the generation of a wide variety of catalogues and indexes from a single computer-based file, without additional involvement by curatorial staff. This paper outlines the special types of expertise needed to run a computer-based documentation system, and surveys the benefits which can accrue from such a system. Finally, the present situation in British museums is outlined.

THE MUSEUM DOCUMENTATION SYSTEM

Before one can produce any output from a computerized catalogue, the data must be prepared in a form suitable for input to such a system. Even more basically, the system adopted must be capable of accepting those data! In the case of the Museum Documentation Association (MDA), the basis of such a system has been provided by the work of the Information Retrieval Group of the Museums Association (IRGMA) on data standards, which led to the publication of the first IRGMA record cards in 1976. By starting from the IRGMA (later MDA) standards, a recording form can be derived which is suitable both as a manual record card, and as a form for computer input. The data standards are also used to define the form of the computerized files when they are produced. All these elements merge together to give a single Museum Documentation System, in which one can move from a manual cataloguing system to a computerized one without any need for recataloguing the material.

The Museum Documentation System has been designed from the outset to be flexible. The present set of published record cards is not meant to be exhaustive, and users of the system are free to design their own version of any card, so long as the headings on that card are derived from the MDA Data Standards. Similarly, the

[Special Papers in Palaeontology No. 22, pp. 149–157.]

recording conventions issued with each card place few restrictions on the way that data can be recorded on the card. The one restriction, implied throughout the instructions for using the cards, is that data must not be recorded in any way that is genuinely ambiguous. Thus, for example, a British user cannot record dates in the order 'month. day. year' (i.e. the American way of recording dates) since dates such as '10. 2. 1954' would be genuinely ambiguous.

One of the main aims of the documentation system is to make indexing a straightforward task, which could be carried out by a clerical assistant, or even a computer. In order to achieve this the system requires that the recorder denotes potential index headings (keywords) by special symbols. Text-fig. 1 shows a completed Geology Specimen card, with examples of such notation. The principal symbols employed are the ampersand '&', which is used to separate keywords, and round brackets '()', which are used to surround additional information (detail) not needed for indexing purposes. Thus in text-fig. 1, 'type & fig'd' indicates that 'type' and 'fig'd' are separate status keywords, each of which is worthy of retrieval.

The GOS program package

Having produced some records, access is then required to a set of computer programs which will take the data in question, and from it produce various listings. MDA is currently finalizing such a set of programs, collectively known as GOS (a fictitious name—not an acronym). Like the record cards, this program package has been designed to be as flexible as possible. It will deal with almost any files which have a well-defined structure. For example, GOS has been used to produce word lists for use in thesaurus construction, for lists of documentary references, lists of museums, and even for inventories of museum equipment! In each case, the package is told how the data should be structured, and it proceeds to file it according to this structure.

Once the details of the MDA data standards have been finalized, their structure will be described to GOS, which can then be used to generate computer records which conform to the same standards as the MDA/IRGMA cards. No other program package of which MDA is aware has the ability to cope with the large, complex hierarchical structure which the data standards will present.

GOS gives the user the ability to manipulate data at any stage of the indexing process. For example, a date could be fed into the system as a string of text, e.g. '19/3/1977', and then analysed by GOS (using the '/' symbol as a delimiter) as 'day = 19; month = 3; year = 1977'. This analysed version of the date is much more useful for index generation and information retrieval than the original string of text. Similarly, data can be altered prior to indexing in order to give a useful index. In this way, taxonomically based indexes can be sorted into taxonomic, rather than alphabetical order.

It should be emphasized that, while GOS is easy to use once it has been 'set up' at a particular computing centre, the process of getting the programs to work on a new computer is not so straightforward. This operation requires the skills of a computer programmer, and it is possible that someone with such skills will be needed at each centre where GOS is used. Thus the comments on 'staff' below assume the existence of a GOS implementation, adequately staffed.

Card of

File

Institution : identity number: CAMSM : A.18726 — Part: a-b

IDENTIFICATION

Simple name — D — Form: part and counterpart — Number

Classified identification or full name: GRA Climacograptus antiquus Lapworth var. lineatus Elles and Wood

C

Current / ~~Label~~ / ~~Other~~ — System: Linnaeus — Status: type & fig'd — D — Identifier : date: 1 — D

COLLECTION

Place names/detail: Carnarvonshire — Locality number

Lat Long / NGR — Other co-ordinates — value & units/accuracy — Altitude / Depth — Other position — value & units/accuracy

~~Complex~~ / ~~Rock~~ — ~~Zone~~ / Age — Stratigraphy keyword/detail: Bala

Complex / Rock — Zone / Age

Complex / Rock — Zone / Age

Stratigraphy detail

Locality detail — D

C

Collection method — Collector : date — Collection number — D

STORE

Store : date — Recorder : date: Light, R.B. : 1/4/1978

GEOLOGY

© IRGMA 1975

1/12/75

ACQUISITION

Acquisition method — Acquired from : date

C — D — Price — Conditions D Yes/No — Valuation : date — D

DESCRIPTION

Condition keyword/detail — Completeness keyword/detail

Dimension measured — value & units/accuracy — Dimension measured — value & units/accuracy

Part : aspect : description keyword/detail — D

C

PROCESS

Conservation / Reproduction — Other process — Method/detail : operator : date : detail — Cross-reference — D

C

Conservation / Reproduction

DOCUMENTATION

L — Class: 1 syntype & fig'd — Author : date : title : journal or publisher : volume : detail: Elles & Wood : 1906 : Brit. Graptolites : Mon. Pal. Soc. : : p.201 pl. xxvii fig. 5c (A.18726a) fig. 5e (A.18726b) — Cross-reference

L — Class — Author : date : title : journal or publisher : volume : detail — Drawing or photo

C

NOTES

Notes

C

Published by the Museums Association, 87 Charlotte Street, London W1P 2BX

TEXT-FIG. 1. Example of a completed MDA/IRGMA Geology Card.

Staff

The highest level of expertise needed during the process of computerizing a museum record occurs when the original MDA/IRGMA card is filled in. Curatorial skill is perhaps taken for granted by museum staff, but nonetheless the quality of the original manuscript record will be directly related to the quality of the computer products generated from it. The next stage in the process—retyping the data in computer-readable form—requires a typist with patience and accuracy, but the skills involved in using a paper-tape typewriter or its modern equivalent can be mastered in an afternoon. The processing of the data to produce catalogues and indexes could be controlled by someone with some knowledge of computers—not necessarily a programmer—who had been given time to study the programs being used, and was employed wholly or partly to do such work. In university terms it could be a research assistant post. Such a person would have to read the data into the computer's memory, and produce a listing of it for proof-reading. The actual proof-reading could be done by the recorder him/herself, or could be delegated to the research assistant. In the former case, intellectual errors as well as transcription errors will be spotted, and one has to decide whether to allow 'cheap' proof-reading, bearing in mind that the result may contain a higher proportion of errors. Incidentally, a feature of computer systems is that they constantly remind one that there is no such thing as a perfect script—the slavish accuracy of the computer shows up the slightest inconsistency.

Once the errors in the original transcript have been marked up on the 'master' copy, they must be corrected, and this would be carried out by the research assistant. There is no need to re-type offending passages—individual letters can be corrected using the computer's 'editor'. Once the file is correct, the research assistant can run various standard programs on it, which will generate the listings needed for the museum's work. In most cases the museum will not need to write these programs, since MDA will provide a standard implementation of the GOS package, which will generate most of the listings commonly needed by museums. Any museum staff needing a special listing will either have to write it themselves, or more probably ask the MDA to write a special program.

So much for the theory—but what is happening in practice? Since January 1976 MDA have sold a great many cards (approaching 750 000) to a large number of British museums (around 150). What effect has this influx of cards had on the organization of museum cataloguing? In terms of computing expertise, the answer is—very little. Most museums have adopted some internal conventions to improve the consistency of their records, and a fairly large number have appointed an existing member of their staff as 'documentation officer', but very few, if any, have actually employed someone to co-ordinate operations with a view to eventual computerization. But this does not mean that the skills are not there. Several museums are currently undertaking computerization tests, and the person in charge of the museum's end of such tests will inevitably gain some familiarity with the day-to-day use of the GOS package. Assuming that the tests are successful, the museum will presumably undertake larger tests, or a complete computerization program, and the original co-ordinator will gradually acquire the skills I have outlined. Of course, one or two museums already have their own programmers working full time on computerized

documentation systems (the National Maritime Museum and British Museum (Natural History) spring to mind) and such institutions can provide adequate computing facilities without recourse to the MDA.

POSSIBLE RESULTS FROM A COMPUTERIZED SYSTEM

Catalogues

What products will the computer yield? The first and most obvious demand is for a 'catalogue' listing, giving each record in numerical order with the full information about each object. Having this listing at all is a useful security measure, since it can be kept separately from the manual (inflammable) card-based catalogue. Such a listing will usually contain all the information that was present on the original cards, and if so it could act as a replacement for them. However, if some information, such as the value and location of each specimen, is left off the computer-produced catalogue, then the original cards should always be kept to provide the additional information.

The computer can be asked to print out the catalogue as many times as requested, and at first sight it appears to offer a straightforward way of producing multiple copies. However, two words of warning are in order. Firstly, computers print out their results on what is called line-printer paper, which is a high-quality thin paper, folded in concertina fashion, with holes down each side. It is designed to go through a line-printer at high speed. This it does very successfully, but the result is hard to treat as a conventional document. The pages open from the bottom, rather than the side, and the page size is unusual. These difficulties can be overcome, either by putting the catalogue into a special binder, or by tearing off each page, turning every other page over, and binding the result in the conventional way.

Secondly, this could prove to be a very expensive operation. The largest and most complete computerization test in this country has been carried out at the Sedgwick Museum, Cambridge. Around 400 000 records of fossils have been transcribed into computer-readable form (Porter, Light, and Roberts 1977), resulting in about 200 000 computer records. The complete Sedgwick catalogue, if printed out on line-printer paper, would take up some 30 000 pages. This amounts to a pile of paper about 3 ft 6 in. high, which would take over 40 hours to print and, if unravelled, would be nearly 4 miles long! Even at the moderate rate charged to MDA by Cambridge University's Computer Laboratory, every such print would cost around £250. While it could be argued that this is a fairly compact format, and at a fairly low cost compared with the size of the original catalogue, none the less it is a very clumsy and expensive way of producing multiple copies.

A much neater and cheaper solution is to print the catalogue on to microfiche. This is a facility which is provided by an increasing number of computer installations and bureaux. Any computer output can be printed in this form—it is not in any way a special property of the GOS package. To continue the example of the Sedgwick catalogue, each fiche costs around £3–£4 to produce, but it can be copied cheaply, at around 15p per copy. Each fiche contains 13×16 images, which gives just over 200 pages on each fiche. Thus the Sedgwick catalogue could be printed on to about

150 fiches. The original copy would cost £450–£600 to produce, but each subsequent copy would only cost £25. Thus a run of ten copies could be produced at £75 per copy, 100 copies at under £30 per copy. In addition to being cheap, fiche is also compact. The 150 fiches containing the Sedgwick catalogue could easily be kept in a small box on a desk top.

Obviously special readers are needed for fiches, costing in general £100–£150. But it might be worth this amount just to save the £250 needed for one line-printer copy of the Sedgwick data.

The image on a fiche will be identical to the image which would have come out on a line-printer, which imposes some problems of type-setting. It is, for example, impossible to achieve a beautifully justified document on a line-printer, since one is dealing with a fixed grid of characters, typically ten to an inch. Such objections can only be overcome by producing a properly type-set document. Traditionally this has involved starting from scratch again, with all the associated costs of typing, proof-reading, etc. Techniques now exist which allow the data to be taken directly from a magnetic tape, and then set up as a block for printing completely automatically. The latest *Museums Yearbook* was printed in this way. MDA are currently working with an Oxford firm to set up a procedure whereby GOS will produce magnetic tapes which are suitable for such a process. It is hoped that this technique can then be tested on the catalogue and indexes to Ulster Museum's collection of minerals. It has been estimated that a 350-page book, generated in this way, would cost £3500 for a 1000-copy run, which works out at £3·50 per copy. Once this process has been established, it will be applicable to any data recorded on MDA/IRGMA cards.

Indexes

I have dealt at some length with the alternatives to line-printer paper as a means of displaying a computer-based catalogue, without mentioning any of the other benefits to be gained from computerization. The main benefit lies in the fact that as many different indexes as are deemed necessary can be produced without having to do any additional data input or checking. Text-fig. 2 shows a page of the main index produced for the Sedgwick Museum. This is ordered primarily by taxonomic names, but additionally is sorted by stratigraphy and then by place. The latest taxonomic name in the record, plus any publication details, are given as additional information. If the specimen has type or figured status this is also noted.

Another four indexes have been produced for the Sedgwick, ordered by stratigraphy, locality, museum store, and bibliography. Indexes of donors, and possibly of descriptive terms, could also be generated. The taxonomic names given could be treated slightly differently. Instead of ordering from broad to specific terms, a list of species could be produced, together with the genera into which they have been placed. Similarly the bibliography could be ordered by journal rather than by author, which might help with systematic literature searches.

Information retrieval

However many carefully thought-out indexes might be produced, there will always be questions which they cannot answer. For example, the question 'do you have any *Monograptus priodon* from the Denbigh Flags?' can be answered by reference to the

```
        Ordovician
            Eire
                (Climacograptus antiquus Lapworth lineatus Elles and Wood)
                    identified, Rickards, 1969.                                  A.61671-61673
                                                                                 A.61906-61926
                    identified, Rickards, 1969.                                  A.62065-62066
                                                                                 A.62100-62102
                                                                                 A.62128-62129
                                                                                 A.62150-62159
GRA     Climacograptus antiquus Lapworth var. lineatus Elles and Wood
            Carnarvonshire
                (Climacograptus antiquus Lapworth var. lineatus Elles and Wood)
                    syntype, fig'd, Elles and Wood, 1906, Brit. Graptolites, Mon. Pal. Soc.  A.18729 type; fig'd.
        Bala
            Carnarvonshire
                (Climacograptus antiquus Lapworth var. lineatus Elles and Wood)
                    syntype, fig'd, Elles and Wood, 1906, Brit. Graptolites, Mon. Pal. Soc.,
                      p.201 pl. xxvii fig 5 c (A.18, 726a) fig.5 e (A.18, 726b).  A.18726 type; fig'd.
                    syntype, fig'd, Elles and Wood, 1906, Brit. Graptolites, Mon. Pal. Soc.,
                      p.201 pl. xxvii fig 5 d.                                   A.18727 type; fig'd.
                    syntype, fig'd, Elles and Wood, 1906, Brit. Graptolites, Mon. Pal. Soc.,
                      p.201 text-fig.132 c.                                      A.18728 type; fig'd.
        Caradoc
            Malaya
                (Climacograptus sp.)
                    identified, Bulman, O.M.B., 1956.                            A.46159-46160
        Llandeilo
            Carnarvonshire
                (Climacograptus antiquus Lapworth var. lineatus Elles and Wood)
                    identified, Elles, G.L.                                      A.18730
                                                                                 A.18740
GRA     Climacograptus aff. bicornis (Hall)
        Ordovician
            Pembrokeshire
                (Climacograptus aff. bicornis (Hall))
                    identified, Rickards, 1971.                                  A.78377
GRA     Climacograptus cf. bicornis Hall
        Ordovician
            Pembrokeshire
                (Climacograptus cf. bicornis Hall)                               A.79631-79635
GRA     Climacograptus bicornis (Hall)
            Carnarvonshire
                (Climacograptus bicornis (Hall))
                    identified, Elles, G.L.                                      A.19593-19596
            Dumfriesshire
                (Climacograptus bicornis (Hall))
                    recorded, Elles and Wood, 1906, Brit. Graptolites, Mon. Pal. Soc.,
                      pp.193, 195.                                               A.18703-18705
                    identified, Elles, G.L.                                      A.19592
            Hartfell
                (Climacograptus bicornis (Hall))
                    identified, Elles, G.L.                                      A.19861
            Ireland
                (Climacograptus bicornis (Hall))
                    recorded, Elles and Wood, 1906, Brit. Graptolites, Mon. Pal. Soc.,
                      pp.193, 195.                                               A.19598-19600
            Moffat
                (Climacograptus bicornis (Hall))
                    fig'd, Elles and Wood, 1906, Brit. Graptolites, Mon. Pal. Soc., p.193
                      pl. xxvi fig.9 a.                                          A.19585 fig'd.
```

TEXT-FIG. 2. A page from the Sedgwick Museum's computer-produced taxonomic index, showing the graptolite section.

taxonomic index, while the question 'do you have any graptolites from Pembrokeshire published before 1935?' cannot. In the latter case the retrieval facilities in GOS can be used to pick out any records which satisfy the query. The result of such a request will be a small file containing only those records which satisfy the original request. This can either be displayed as it stands or indexed in some appropriate way. For example, the pre-1935 Pembrokeshire graptolites could be sorted into taxonomic order.

This paper has reviewed the major techniques of computer-based cataloguing which are being developed for use by British museums. One important omission is the interactive use of computer terminals for information retrieval. Since many people think of computer-based information retrieval in such terms, it is worth explaining why it has not been mentioned. It is not because it is not possible. GOS, as mentioned above, has an information retrieval facility, and this could be used interactively. The problem is one of size, i.e. of cost. The Sedgwick files take up two or three complete magnetic tapes, and very few computer services are going to allow that quantity of information to sit for ever on 'instant access' storage space, awaiting the arrival of a palaeontologist with an unusual question. In general, museum data are too bulky, and are not accessed sufficiently often to justify on-line retrieval.

Other results

The techniques outlined above provide considerable scope for collections management as well as information retrieval. A listing of all specimens valued at more than, say, £500 could be useful for insurance purposes, while a systematic index to an entire collection might point out where changes could profitably be made in collection/acquisition policy. An index giving the contents of each drawer could be used as an inventory.

More generally, data can be easily copied on to another magnetic tape and sent to another museum using the GOS package. Thus information exchange between museums, and possibly the collection of data into a central museum data bank become merely administrative problems.

THE CURRENT STATE OF COMPUTER-BASED CATALOGUING IN BRITISH MUSEUMS

As the GOS package is still not generally available, MDA are conducting most work with GOS themselves at Cambridge, using a pre-release version of GOS. The Sedgwick and Ulster projects have already been mentioned. Tests are also being conducted for the Science Museum (Pictorial Collection and Wellcome Collection), Lancashire County Museum, the Imperial War Museum, Tyne and Wear County Museum Service, St. Albans Museums, the Hunterian Museum, and Bristol City Museum. The latter four museums are all including geological records in their tests. To do this work the MDA have taken on the services of a part-time typist to do data preparation. Once GOS is on general release it is hoped that individual users will take some of the load off our hands by setting up their own data-preparation facilities. Several national museums are actively seeking to establish a common computing centre, and a large number of local authority computer managers have been prompted by their local

museum to contact MDA about the possibility of using GOS. It is too early yet to say how regional computing facilities will be established for museum use, but the current level of interest leaves little doubt that they will be.

REFERENCE

PORTER, M. F., LIGHT, R. B. and ROBERTS, D. A. 1977. A unified approach to the computerization of museum catalogues. *British Library Research and Development Report*, **5338** HC (ISBN 0 905984 01 3).

R. B. LIGHT

Museum Documentation Advisory Unit
Imperial War Museum
Duxford Airfield
Duxford
Cambridgeshire CB2 4QR
U.K.

DISCUSSION

E. L. Yochelson. I am concerned about the use of computers for cataloguing since I believe that in the United States the computer has been 'oversold'. It is a useful tool, but it is only a tool. I grant that computers can print labels faster than writing with a quill pen, but this does not solve all curatorial problems. A collection should be arranged in a systematic way so that one can find specimens without recourse to lists. The logic of a collection should be human logic and not machine logic. To give up human attributes in order to make systems easier for a machine is a retrograde step. Although lists might be made by the computer, we must surely consider whether the lists are justifiable in terms of the time required to enter the data in the first place.

R. B. Rickards. Museums must be stable for a very long time, at least a quarter of a century, to be really working, and considering the way that computer techniques are evolving so rapidly we must surely think seriously about the money that we have available to spend and on what to spend it; and whether the computer techniques that could be adopted now might be hopelessly out of date in a few years.

J. Cooper. I agree with Dr. Rickards's comments on the stability of museums and would like to stress that what really matters is the quality of curation and information recorded, irrespective of whether the recording format is manual or mechanical. The computer is fine when it can be justified financially, but it must be done thoroughly and ideally from scratch.

W. D. I. Rolfe. I am concerned that different forms of computer systems are being developed in different museums, all of which are working on their individual problems. At the Hunterian Museum we have had to devise our own vocabulary, whereas a lot of this kind of basic work could have been provided by custom-made systems. Far too many places are experimenting and we never seem to get beyond the experimental stage. The original IRGMA Project at Cambridge spent years doing all this and yet we are still acting individually. All museums want is a simple system that will yield an answer to all our questions; we want it to be beautifully packaged and we do not want to be guinea pigs for testing systems since this will lead to disillusionment.

R. B. Light. When the original instructions for using the IRGMA cards were produced in 1976, the team at the Sedgwick Museum did not have the resources to produce a simple 'custom built' system. With the facilities available to them, the wisest course was to produce something that could be used in different ways by different users. We did not want to force everyone down the same path without the experience, which we certainly lacked, to back our instructions. Things have changed: MDA now have the staff and time to collate what people have done in different museums, and possibly to produce a single recommended formula. Until museums, each knowing what they want to do, have tried several alternatives, then MDA will not be in a position to lay down the line to be followed by museums curators. We certainly cannot begin to collate different systems until individual museums send us copies of their own conventions.