computers & systems

an introduction for librarians

computers & systems

an introduction for librarians

JOHN EYRE ALA & PETER TONKS MBCS

CLIVE BINGLEY LONDON

FIRST PUBLISHED 1971 BY CLIVE BINGLEY LTD
16 PEMBRIDGE ROAD LONDON W11
SET IN 10 ON 13 LINOTYPE PLANTIN AND
PRINTED IN GREAT BRITAIN BY THE CENTRAL PRESS (ABERDEEN) LTD

0 85157 120 4

contents

list of illustrations

introduction

The continual cry from those attending the data processing courses at the School of Librarianship of the Polytechnic of North London has been for a simple and uncomplicated book on computers. Such a work was to outline what a computer does, especially in the library context, and explain in general terms what systems work is about. After four years of running these courses for both non-graduates and graduates, we decided that the time had finally come to attempt to meet this need. At the risk of being charged with adding to an already prolific area of literature we have produced this book as an introduction specifically written for librarians or library students. Most works on computers seem to delve into the technicalities too deeply for the requirements of those wishing simply to know what are the potentials and limitations of a computer in the library. The other objection to many of the presently available books was that the accent was placed on commercial examples used to illustrate computer applications. These do not help much when the reader is thinking in terms of bibliographic records and information retrieval; areas quite outside commercial practice.

We are conscious of the excellent works which have been published on computers in libraries, but all these seem either to assume a knowledge of computer workings or are intended to instruct the reader on methodology. A selection of such works has been given in a bibliography at the end of this book so that those wishing to pursue the matter further can do so profitably. There are of course many periodical articles on the computer in the library, but these tend to give very individualised views of systems in the making and, though valuable, often defeat the uninitiated.

With all these points in mind, we have tried to cover the subject of the computer in a way which will be at once simple and informa-

tive. In order to keep the length of the book realistic in the light of its intended purpose we have had to simplify many aspects which might otherwise have required considerable technical detail to cover fully, and we have therefore contented ourselves with a presentation of what may be termed the overall view.

As may be expected, no work of this kind can be produced without help or encouragement from colleagues. It is appropriate to record our sincere thanks to them, and especially to Mr A Croghan of our teaching staff and Miss M Ritchie of our research staff for their invaluable criticisms and suggestions. Our sincere thanks must also go to the teaching staff of the School of Librarianship, for the farsighted support of our data processing courses which have now become a regular part of this school's curriculum.

April, 1971

JOHN EYRE
PETER TONKS

1
introducing the computer

Ever since the second world war the computer has increasingly become used as an everyday tool, to the extent now that the majority of people have some connection with it in either their professional or private lives. While it is obvious that not everyone wishes to be associated with computer operation, there are numerous reasons for understanding its workings and the contribution it can make in various situations. A knowledge of the computer offers a certain degree of protection against its abuse and misapplication, while at the same time enabling one to recognise those situations where its use would be beneficial. It is equally important to know enough to justify its use as it is to explain its inappropriateness. We shall therefore begin by considering the computer and how it works, and then go on to discuss systems analysis and design. As stated previously, we intend our book as a simple introduction and not as a comprehensive textbook on the subject.

TYPES OF COMPUTER

There are two types of computer and although library applications demand the use of only one of those it will be as well to be acquainted with both types. The two types are a) analogue and b) digital.

THE ANALOGUE COMPUTER accepts information for processing in physical form, *ie* in terms of weight, volume, electric current, etc. It performs calculations on the incoming information according to a predetermined pattern which is usually represented by mechanical, electro-mechanical or electronic units. The results of the calculations usually appear in a similar form to the input. This computer is built to satisfy a particular requirement and can usually only be used for this one purpose. This makes it unsuitable for library requirements which necessitate that data be processed and recorded

2
data carriers and the peripherals

Since we have not yet arrived at the happy situation where the computer could accept data by our speaking directly to it or scribbling on a piece of paper, it is necessary to understand the way in which information can be fed into, and received back from the computer. We shall therefore look at the media which carry the data and the machines which handle them. Not every computer requires the data to be coded onto a special input medium before being fed in. Some permit data to be keyed in direct to the computer. However, the former method is the more usual, particularly in cases where a large amount of data is concerned, since preparation can be done away from the computer installation and can be batched. Batches are fed into the computer when they have reached a size sufficiently large to warrant a processing run. The three most common media are: Punched cards, Punched paper tape, Magnetic tape.

PUNCHED CARDS

Since punched cards have been a popular data carrier for some 80 years, there is a considerable range of equipment available besides the computer to process them. The card is divided into 80 columns, each of which may be used to record one character whether a letter, number or symbol. Most characters have a standard punching code and an example is shown in FIGURE 3. It will be seen that one or more holes in a column will be punched out to represent the desired character, and that there are in all twelve such punching positions in each column.

A number of columns can be grouped together for recording a multi-digit number or a word. Where such numbers or words relate collectively to one part of the data, as might be the case with an

3
the peripherals continued

BACKING STORE UNITS

The backing store units of most computers are either magnetic tape decks or disc drives. Unless the files of data are kept permanently on a fixed disc or a drum, it will be necessary for the operator to select the required files from the file library, and load these onto the appropriate unit. The program can then extract data from the files when necessary and read this into the working store. When we say data is ' read ' from disc or tape we mean that it is copied into the store so that the original remains unaltered. Because systems using computers have to recognise the dictates of magnetic tapes and discs we shall now go on to look at these two media in some detail.

MAGNETIC TAPE DECKS

The data on magnetic tape is grouped into fields, records and blocks, as shown previously in FIGURE 6. Characters are grouped together to form a *field* in the same way as with cards or paper tape. This group may represent a number or words. In most cases it is necessary to use a number of these fields to form what is called a *record.* An example of a record might be an accession number followed by author, title and imprint. In turn these records are grouped together to form *blocks.* There are two good reasons why records, if they are short and do not contain many characters, should be treated in blocks. Firstly, the working store may well be capable of accepting more information than is contained in one record, thereby wasting time if only one is read at a time. Secondly, the way in which the tape deck operates could make the reading of single records a slow and uneconomical process. Like the fields and

4
software

The function of the program has been mentioned earlier, and it is important to appreciate that the efficient use of the hardware depends on programs of different types. These programs, together with all documentation associated with computer systems, are known as *software,* and can be classed under the following headings: user programs, programming aids, supervisory or executive programs and operating systems, and applications packages. User programs would be written by programmers on the library staff or in specialist agencies, in order to produce the particular output needed by that library. The other three types are termed standard programs as they can be used by anyone who has access to the type of computer for which they have been specifically written. Since they are related more to a make of computer than a particular job, the manufacturers generally prepare these programs and make them available to their customers, sometimes without charge. Such programs may also be developed by specialist organisations such as universities, government establishments or industry, and at a later stage made available for purchase. The value of such programs is that they satisfy many of the basic requirements of computer users, thus avoiding duplication of programming effort. Whenever considering the use of a computer, it is therefore wise to find out what software is already available from the manufacturer or elsewhere, to determine what may be used or adapted to meet expected requirements. Let us now look at the four types of program.

USER PROGRAMS

These programs are prepared by the computer user to suit the special requirements of his system. Many such programs may be

5
concerning systems analysis

Since we have explained the way in which a computer performs the tasks we give to it, there may be a temptation to look upon this machine as the obvious answer to virtually every work situation. Certainly when time on the computer is offered, and general opportunities for computer use are presented, there would seem to be every reason to seek out jobs which could be turned over to the machine. At the time such adventurousness may seem to have merit, if only to show that the library is keeping up with the twentieth century or to relieve immediate work loads. As the use of the computer is inevitably extended, the limitation of such a piecemeal policy may become apparent and lead to the realisation that the way in which the library operates has to be viewed as an overall concept if work is to flow smoothly and services are to run at their most effective. It is for this reason that the design or re-design of the whole library system is a matter for the systems analyst. It should be realised immediately that an investigation does not at all presage the use of a machine of any kind. It is therefore not the case that systems analysis is concerned solely with the use or introduction of computers. The purpose of analysis is to optimise design in whatever way this may be achieved. But having said this, it must also be admitted that the systems analyst is associated more with those situations likely to require a computer than with any others. This arises because the efficient use of the computer rests on the satisfactory exploitation of its potential.

What is meant by the words system and analysis? A system may be defined as a network of operations involving people, equipment and documents, which processes given inputs to produce required outputs. The difficulty which arises here is the identification of the limits of any network in order to define our system. The levels at

6
preparatory steps to analysis

The whole approach to analysis has to be carefully planned, and executed with some regard to time, cost, manpower and the possible disruption to current work flow. To begin with there is good reason for taking a rather cautious approach. Not everyone in the organisation, even at the higher management levels, may agree with the need or see the desirability of a new system. This feeling may be aggravated if it is intended that the new system will make considerable use of the computer, particularly so if this move results in redundancies. Computer use has not been without its difficulties, and certainly in libraries there has been rather less than enthusiasm for its installation. It is thus a wise policy for the analyst to detect at an early stage the quarters from which opposition or unhelpfulness might be expected during the course of investigations. One might therefore give four headings as the first steps in analysis:

1 Determine the objectives of the new system.
2. Draw up a preliminary report.
3 Obtain management agreement.
4 Publicise the analysis programme.

1 DETERMINING THE OBJECTIVES

A tale is told of an organisation which convened a meeting to discuss in detail the objectives of the library system, so that the analyst could determine the overall purpose of the system he was expected to design. Some weeks later the meeting was adjourned, it having been decided that to outline the objectives of a library system without knowing the design was impossible. It will be appreciated that the commercial enterprise or the government department has an inherent purpose around which the system will be designed. In the case of a library service there is no ideal except

7
collecting data about the system

Once the preliminary steps have been taken the analyst is ready to plan the way in which the data will be collected. There are three methods of collection, and the choice will be governed as much by the availability of staff to undertake the work involved, as by the accessibility of the various parts of the organisation to be investigated. In most situations the survey is conducted by a mixture of all three methods, since this way lends itself to a more balanced view of the existing system. The three methods are: interview, questionnaire, observation.

INTERVIEW

Those who have already attempted to conduct interviews eliciting information about people's work will know that the whole exercise is fraught with difficulty. Each interview is different and there is nothing to be gained from trying to confine these to a single form. The questioners certainly will have an interview sheet on which are reminders of the points which need to be covered, the remainder of the sheet being used to make notes as the talk progresses. Textbooks on the subject rightly emphasise the need for putting interviewees at their ease in comfortable and at least semi-private conditions. One never knows just what will come out of these meetings and it is essential to avoid stifling anything important through lack of privacy. Few people will discuss their job in the sort of detail which the analyst needs and it will be for him to guide the conversation as unobtrusively and unobstructively as possible. This may mean that the envisaged plan for an interview may not be adhered to, and that the result will have to be written up afterwards. In this connection it is customary to make brief notes only, since an interviewer who appears to be writing down every word will cause considerable suspicion about his motives. A good interview is conducted

8
stages of systems design

With the analysis complete, the analyst is now ready to begin tentative designs for a new system. The order in which he approaches the whole operation has implications for the success of the scheme, and if the order is wrong he may find that some parts of the system prove inadequate. We shall therefore consider the stages of design in the order which the analyst should be expected to take them.

OUTPUT

Although earlier meetings with the management will have elicited some expectations concerning desired outputs, the analysis will have thrown up a number of possibilities which may not have been foreseen. These should be clarified and discussed further with the management to ascertain their economic acceptability. Thus, for example, the claim that analysis had revealed that an on-line service should be provided at two or three key points may be legitimate, but nevertheless unacceptable. Similarly, the analyst may conclude that the equivalent of a catalogue should be made available at many more points and that this would be best achieved by the use of microforms. However, early consultation with the management may establish that they would be unwilling to vote the money required for special readers. Obviously that which is to be derived from a system will govern the design. It is therefore a prerequisite to finalise, with general approval, the output from the new system. Having started on the output, the designer works backwards through the system so that sectors which are dependent are designed before those on which they depend.

DATA BASE

Once we know what is wanted from the system, the data base content can be specified. In the library situation it is possible to have a

9
input procedures

Just as a library will standardise its catalogue entries in order to promote accuracy, uniformity of practice and ease of use, so do the input procedures of a mechanised system need to be standardised as much as possible. There are other benefits to be derived from such action in that errors are more easily detected and the speed of processing is much assisted. However, the method evolved will depend initially on whether the bibliographic data is to be encoded by library staff or by professional punch operators. It should be stressed here that while both situations permit some flexibility, the professional punch operator is trained to depress keys at rates exceeding 8,000 per hour, so it is not normal practice for her to check data before punching it out. This means in effect that the operator will encode exactly what she sees before her—mistakes as well. Often those using punching services are surprised, if not pained, to find all their own errors faithfully reproduced. The moral is that if encoding is to be done correctly then the system must be designed to provide the information in a way which is quick to read and easily checked.

Once the matter of who is going to do the encoding has been resolved, we can then decide on the appropriate way in which original data will be presented for encoding. Obviously the less writing to be done the better; cataloguing from source can offer appreciable savings in this connection.

There are thus two methods of data presentation: *a*) use of originals with annotations; *b*) use of punching sheet. For the present we are ignoring the possible use of optical character or other special readers since it would still be necessary to provide an entry which could be fed into the reader, which brings us back to the equivalent of the two options given. In the case of the first it is obvious that

10
the record

FORMAT

Before discussing record format in general, it may be helpful to explain what is meant by ' fixed ' and ' variable ' fields. If you look at FIGURE 21 it can be seen that a certain number of columns or characters have been allotted to each data element or field. Not only is the number of characters fixed but the position of each field in relation to others may not be changed. There is one variation possible, however, and that relates to the number of characters which actually appear in each such field. The accession number might always be the same length, but author's names would obviously vary in length. In the case of variable fields, the number of characters which may be required in each field is unpredictable and therefore without restriction. So in the case of a fixed field, unless the data element is always of the same length, such an element will often be too long or too short to fit the field provided. With the former the field content has to be truncated, while with the latter unused character places are wasted. Neither of these situations would be acceptable for bibliographic files. It is for this reason that the variable field format is to be preferred. It permits a field to be as long as the data in it when the amount of data can vary unpredictably.

It is necessary to indicate the end of each variable field, as well as the end of the entire record by unique characters, that is, used for this purpose only when occurring by themselves or with other predictable characters. In the case of card input, data elements can be punched on separate cards so that the author's name appears on the first, the title on the next (abbreviated if necessary) and so on (FIGURE 22), or the data can be punched on in one sequence on as many cards as required (FIGURE 23). With paper tape there need be no wastage, since there is little to be gained by leaving spaces between elements

11
processing

For many libraries the question does not arise of whether to operate the system on a continuous basis or by batching. Computer facilities are made available to the library without particular regard to its needs, and it has to make use of what is left after those with higher priorities have made their claims. Batch processing can be done with both off-line and on-line facilities. Much will depend on the volume of work to be done and the number of staff. But whatever considerations determine the way in which the system will operate, the ability to keep up to date with on-line files offers many advantages. However, quite apart from the cost of the terminals needed, and the processor and storage devices necessary to hold large files on direct access, there are practical considerations which discourage wide use. Even with the large third generation computers, store size still appreciably limits the number of users who can be served at the same time. A library system which was wholly mechanised could monopolise the computer at peak periods, a situation unlikely to be tolerated for long by other users. A further point is that teleprinter terminals are noisy and not particularly fast (speeds average about 30 characters a second) while visual display units although fast and noiseless do not produce printed copy. Lastly, where a number of departments have terminals it is normal practice to determine a priority rating. Under this scheme terminals with a high priority can obtain access to the store almost immediately, which in turn means that lower priority users experience some delay. Although such delays are not likely to be more than three minutes, these could be rather disruptive during busy periods.

However, looking at the two ways of processing, it can be seen that batching allows the amount of data to be accumulated, so that the computer's time is given entirely to processing. Presented with

12
the data base as an index

CRITERIA

Once the data are input and processed, the system users are naturally interested in what type of question it is possible to put to the file. Theoretically it should be possible to put any type of question, but limitations do arise from both the way in which the programs have been constructed and the record format. The former governs the way in which file interrogation can be conducted, *ie* the number of parameters it is possible to cite, and the complexity of relationship which can be demanded. The question 'Have you a book by B C Vickery?' is a single parameter. On the other hand if one asks for the book by B C Vickery entitled *Techniques of information retrieval* published by Butterworth in 1970, we then have four parameters joined by the logical *and* device, for all four must be presented to effect a match. More complex relationships would obviously arise where subject terms were being used to formulate the question.

In the case of record format, we have already looked at the MARC structure and discovered that only by the provision of tags and subfield indicators in a variable field record is it possible to extract parts of the record without having to compare the question and record content character by character. The computer is quite capable of this, but in the case of very large files the time required would probably be unacceptable. The record may also carry a key; this, however, is not essential since the need for it depends on the way in which the file is interrogated at the first pass. If records are called up by author on the first pass, the placing of the author at the front (as the key) of each record allows this to be effected with maximum efficiency. It is also possible to have in place of, or in addition to, the key a string of bytes at the front of the record to give a binary coded

13
output by mechanical line printers

It is perhaps in the area of computer output that there is the most controversy amongst librarians. As was explained in the chapters covering peripherals, the line printer usually provided for outputting data from the computer, can work at anything from 1,000 to 1,500 lines a minute allowing up to 160 characters a line. But these speeds require a limited character set to keep the barrel carrying the type to the minimum possible circumference. Thus one finds that all high speed barrel printers have upper case only. This together with the need to print everything simultaneously on one line, undoubtedly deters many users who expect the quality of output which they are accustomed to receiving from conventional printers.

First we should distinguish between what may be called 'work prints' and those intended for library use, which we shall call 'user prints'. Under the term work prints would be classed such lists as those of items on loan, book selection lists and books on order. User prints include the catalogue, bibliographies, selective lists provided to individuals or groups on a regular basis (known as SDI—selective dissemination of information), and the like. Quite obviously the work prints have a relatively short life and are intended for the use of staff who can become acclimatised to their rather utilitarian appearance. Since these lists will require frequent updating and therefore reprinting, speed of output has a definite merit. In the case of user prints certain qualities of legibility and attractiveness are desirable.

Although the speed at which a line printer is run sounds considerable, it is nevertheless unable to keep pace with the operating speeds of the computer. It is therefore usual to buffer the data to be printed so that the processing run does not have to be restricted to the speed of the printer. All the data for printing must be formatted in the correct sequence for output. This of course is part of the

as low as 100 lines a minute according to the number of characters in the set. Such a reduction is no small matter when high volume output is required. To regain the speed and still permit the use of an extensive range of non-standard characters takes us into the realms of photocomposition and the like, which will be discussed in the next chapter.

One vexing problem has also to be faced in connection with the use of printers. It was pointed out earlier that only four usable copies can be produced at one pass. In cases where multiple copies are needed use can be made of the various duplicating methods available. One point must be made here, and that is the rather indifferent quality of computer continuous paper. Many computer installations are at pains to buy paper at a competitive price. This leads to bulk ordering of one type only, which in turn may mean all output without special stationery is issued on lined paper. Such lines aid the eye across a sixteen inch page, but can also interfere with the readability of the text. The paper is often rather thin which makes handling difficult when collated into book form.

As inspection of any computer paper will reveal, it is necessary to have sprocket holes on each side at half inch intervals to ensure the paper passes through the printer without slipping. Once printed, the paper can be guillotined in various ways. It is quite possible to cut off the perforated edges, and make cuts down the page, as well as across. The same machine will also stack as it cuts. Note that this is not the same as collating, since if the pages are split down the middle there would be two stacks produced which would have to be merged manually. But at least as each page is cut the resultant two would be placed on top of the two before.

If computer paper is to be kept between covers having twin or triple post anchors, it can be useful to leave the perforations on the lefthand side. Sometimes, though, the paper is scored just inside the perforations to assist their removal, in which case one may find that pages begin to come adrift if they are in service for too long a period.

bibliography

Suggestions for further reading:

Batty, C D *ed*: *Libraries and machines today*. Scunthorpe: North Midland Branch of the Library Association. First edition 1966 (56 pages), second edition 1967 (50 pages).

Bourne, C P: *Methods of information handling*. New York: Wiley, 1963. 241 pages.

Chandor, A *et al*: *Practical systems analysis*. London: Hart-Davies, 1969. 388 pages.

Coward, R E *ed*: *MARC record service proposals*. London: British National Bibliography, 1968. (BNB/MARC documentation service publication no 1.)

Cox, N S M *et al*: *The computer and the library*. University of Newcastle upon Tyne; Hamden (Connecticut): The Shoe String Press, 1966, 96 pages.

Cox, N S M and Grose, M W *eds*: *Organisation and handling of bibliographic records by computer*. London: Oriel Press; Hamden, (Connecticut): The Shoe String Press, 1967. 191 pages.

Daniels, A and Yeats, D *eds*: *Basic training in systems analysis*. London: Pitman/NCC, 1969. 275 pages.

Dolby, James L: 'An algorithm for variable length proper-name compression' in *Journal of library automation* 3(4), December 1970. Pages 257-275.

Dougherty, R M and Heinritz, F J: *Scientific management of library operations*. New York: Scarecrow Press, 1966.

Gilchrist, A: 'Further comment on the terminology of the analysis of library systems' in *Aslib proceedings* 20(10), October 1968. Pages 408-412.

Harrison, J and Laslett, P *eds*: *The Brasenose conference on the automation of libraries*. London: Mansel, 1967. 175 pages.

Harvey, J *ed*: *Data processing in public and university libraries*. London: Macmillan, 1966. 150 pages.

Hayes, R M and Becker, J: *Handbook of data processing for libraries*. New York: Becker and Hayes Inc, 1970. 885 pages.

Kimber, R: *Automation in libraries*. Oxford: Pergamon Press, 1968. 140 pages.

Laver, F J M: *Introducing computers*. London: HMSO, 1965. 69 pages.

Licklider, J C R: *Libraries of the future*. Massachusetts: MIT Press, 1965. 219 pages.

Mitchell, R K: *Information science and computer basics: an introduction*. London: Clive Bingley; Hamden (Connecticut): The Shoe String Press, 1971. 101 pages.

Phillips, A H: *Computer peripherals and typesetting*. London: HMSO, 1968. 665 pages.

Robinson, F *et al*: *Systems analysis in libraries*. London: Oriel Press.

Ruecking, F H: 'Bibliographic retrieval from bibliographic input' in *Journal of library automation* 1(4), December 1968. Pages 227-238.

Scientific American: *Information*. San Francisco: W H Freeman & Co, 1966. 218 pages.

Periodicals frequently carrying articles on computer applications in libraries:

Aslib proceedings. 1949-. (M) London: Aslib.

College and research libraries. 1939-. (BiM) American Library Association.

Journal of Documentation. 1927-. (Q). London: Aslib.

Journal of library automation. 1968-. (Q) American Library Association.

Library resources and technical services, 1957-. (Q) American Library Association.

Program. 1966-. (Q) London: Aslib.

index

This index has been compiled after the style of PRECIS developed by the British National Bibliography. The facet order for the main entries is ACTIVITY: THING: TYPE: ASPECT. All added entries are derived from the main entries in this manner:

Main entries — A:
B:C:D.

Added entries — 1. B:A
C:D.
2. C:B:A
D.
3. D.C:
B:A

Where a facet would present an unsought term if brought to the indexing position, it is missed out. Thus entries 1 and 3 might be given, but entry 2 omitted as an unnecessary entry term. The suffixes a, b & c indicate the upper, middle or lower part of the page.