POSSIBLE APPLICATIONS OF DATA PROCESSING EQUIPMENT IN LIBRARIES

John A. Wertz

Conferences such as this one have made it increasingly evident that anyone speaking of the computer and the library is no longer dealing with possibilities but with probabilities. Historically, the library profession has adapted, if at times with some misgivings, any technological advance that promised to solve its problems. The computer has been no exception to this rule. That the computer is useful in the library has already been demonstrated. The librarian is now concerned in finding new applications for the computer within the library.

The development of the computer has been extremely rapid. The digital computer and its associated technology has been on the commercial scene a relatively short time. In little more than a decade its uses have outgrown the laboratory and become commonplace in the business and academic world.

In fact, as recently as 1958 General Electric's Computer Department installed industry's first solid-state computer and first computer system utilized by a bank for electronic bookkeeping. It was called ERMA, for Electronic Recording Method of Accounting. It represented the largest single order ever placed for computers—some $60 million—and set the stage for a complete new generation of solid-state computers.

Then in 1961, Western Reserve University installed a General Electric GE-225 general-purpose computer for information storage and retrieval. It is used by the Center for Communication and Documentation Research to perform literature searches within various technical and scientific areas.

In the process, a whole set of new technologies and new epistemologies have been engendered. It is these subsidiary effects of the computer that have had the greatest import for the librarian. The relationship of the librarian to the new theories of information

John A. Wertz is a Programmer at the Programming Research & Development Computer Department, General Electric Company, Phoenix, Arizona.

propagation and dissemination is a major challenge to the profession. But this is not the time to investigate the metaphysic of librarianship.

What we are concerned with today is the computer's impact upon library technology, not theory. We may take comfort from the fact that the computer is not the first device to have affected library technology. The steel pen, the typewriter, the camera, the printing press, all of these have been assimilated and have become tools for the librarian. In this respect the computer is no different from the others; it too will eventually be looked upon as casually as the typewriter is today.

Before we can explore the possible utilization of the computer in the library we must define the limitations of the device. The computer is a computational device. It was invented to perform numeric calculations. The computer, in its mathematical function, is also capable of quantitative comparison. It is also capable of being programmed to take action depending upon the result of a comparison. It can be made to rearrange data within itself. This is the sum total of the capabilities of the computer, oversimplified; but we must realize that we cannot ask the computer to think, to be intuitive.

From these simple mathematical functions have evolved the techniques of general data processing. As this concept of generalized data manipulation expanded, the librarian and the computer have been brought inexorably together, the basic function of the librarian being, after all, the rational organization of data.

If the librarian is to use the machine, he must be able to express the library problem at a logical level compatible with the capabilities of the machine. This means that operations must be reduced to the simplest possible logical steps. And it is here, in the definition of the process or the problem, that the real difficulty lies.

Library utilization of the computer is limited only by the ability of the librarian to define his process at a sufficiently simple level. In this the librarian has no greater problem than any businessman who might wish to use a computer. It is immaterial to the computer what it is doing. The computer has no awareness of whether it is doing simple accounting or very sophisticated mathematical analysis.

The librarian's problem is reduced, then, to one of defining library techniques. This is the point at which the profession now finds itself, struggling with the reduction of its technique to the machinable level. In a sense this is not a new struggle, it is as old as the profession. There has been a constant dialogue within the field trying to define its intellectual content. This had usually taken the form of trying to define and separate the "clerical" tasks from the "professional" operations. The computer has forced a new rigor into this dialogue. Now for the first time the librarian is faced with the opportunity of the perfect clerk, a clerk with no initiative or judgment but with an infinite capacity for letter-perfect adherence to instructions.

It is this dialogue which gives the first clue as to the possible uses for the computer. Search the library for operations which consist only of the clerical task of rearranging the format of information, the simple comparison of one datum to another, or the creation of ordered lists of data. These are the things a computer can do.

There is little in the operation of a library that does not fit into one of these three categories. There are very few actions that require actual interpretive decision on the part of the library personnel. In fact I can identify only three: (1) the decision to acquire a certain document, (2) the classification of a document, and (3) the reduction of a reference request to terms suitable for effective searching of the library resources. All other tasks—at the moment excluding from consideration the entire area of administration—are clerical dependencies upon these three decision points. There is no reason, therefore, why the abilities of the computer cannot be utilized in every aspect of library technology.

Given these three decision points, the intervening processes must be reduced to computer terms. It may at the moment seem to the librarian to be a Gargantuan task. In some ways it is. There exists, however, a large corpus of technical and intellectual know-how within the computer industry, developed primarily through the analysis of the operation of business and manufacturing firms. But is there very much difference between the ordering of raw materials and the purchase of books, or between the problem of inventory maintenance and the problem of circulation control? What matters here is not the name of the process but the ability to reduce a type of process to machinable form.

There is another aspect to this analysis of processes, the need for a parallel analysis of the forms of data involved. Thus the librarian who would achieve automation is faced with the simultaneous tasks of systems analysis and source data automation. The two studies complement one another, however, contributing to each other's successful completion.

Needless to say, the same problem has long faced industry, and a method of approach has been developed by the professional systems analyst. I believe that the current state of the art on the part of both the librarian and the computer industry is such that a fully integrated computer system for the library is well within our grasp. The existence of such a system for any particular library is only a matter of time. In fact several libraries have already begun the process.

What will the completely integrated computer system for a library include? Everything. It has to. The process of computerization can provide too many efficiencies to be limited in scope. This is not to say that the changeover might not be gradual and absorb one area of the library operation at a time, but the conversion must be complete to be effective or profitable.

Computerization will have to be a co-operative venture. Both the librarian and the manufacturer must be allowed to contribute to the process. The librarian should insist that no violence be done to his technology under the guise of conforming to the computer's needs, but he should also be ready to look at some of his hallowed traditions with a critical eye. Respectful co-operation should be highly beneficial to both parties.

Perhaps the first question is where should automation begin. There has been much talk about the use of the computer as an information storage and retrieval tool. All too often this is conceptualized as a simple putting of the card catalog into the computer and then asking it questions. If we accept the principle of source data automation as a critical criterion, this approach to automation breaks down. True, the card catalog is the source document for reference service, but in the larger view of the library process it is only one of the many intermediate documents. Or for that matter, it can be viewed as the end product of the cataloging function.

All this discussion re-enforces one point: automation must be planned with the total system in mind. The librarian must prepare to find some way to integrate the data used in every process, from the initial purchase request to the final discard notation. In this goal the librarian has the advantage over many prospective computer users, as he is already well versed in the concept of the unit record, a basic computer technique.

It is easy to forecast the integrated computer library system. It is close enough to reality to need no Jules Verne as its prophet. The impact of automation will be greatest upon the technical service and administrative areas of library technology. This is partly because the computer cannot change the intellectual process of questioning (and decoding the question) and partly because the technical processes are most open for improvement.

In this library of the near future, automation will be invoked from the minute a purchase decision is made. From that moment a unit record will accumulate all the pertinent information about the transaction and the document. As each new fact is developed it will be added to the unit record through the medium of punched cards or paper tape or both. At any moment the main files may be interrogated for the status of single items or for batches of data, such as orders outstanding, encumbered funds by department or by vendor, etc. The process of accretion of information to the unit record will continue throughout the usual process of order, receipt, and cataloging. The unit record, replete with cataloging information, will be used in many machine files comparable to the shelf list, author catalogs, etc.

One element which should not disappear from the library scene is the card catalog. It is still the easiest and cheapest way to interrogate the library collection. The cards will be prepared as a

by-product of the computer processing, however, and will be able to attain a new level of accuracy and completeness. By allowing the normal card catalog to carry the burden of the average, simple request, the computerized unit records can be reserved for the more challenging aspects of information (or reference) retrieval. This is where the complex, multi-faceted reference question will be answered. This file will also furnish the comprehensive demand bibliography. Here also is the source of the recurring bibliography of the library's specialty. But note the difference between this master record and the ones now used in much information storage and retrieval. Whereas the latter are created especially for the purpose, involving a duplication of effort, the master record in the integrated systems library will have been the result of a gradual and programmed accretion of knowledge. Its creation will not have involved duplication of human effort.

Needless to say, the circulation records will also be automated. Here it is harder to project an image of the possible system as every library will have highly individual needs. Here also it is difficult to predict the effect of possible developments in charging machines.

The most fascinating prospect for the automated library lies in the administrative field, however. It is a rare library that has either a sufficiency of administrative statistics or a means of utilizing them efficiently. The systems automated library would be in an excellent position in terms of statistical records. Statistics would be extracted via computer from the various daily working records. They could then be correlated and printed in usable form by the computer.

The possible types of analysis are manifold. For instance, helpful studies might be made of the rate of document use against subject area as a means of maintaining an up-to-date collection, or perhaps a statistical analysis of borrower patterns in order to determine more intelligently the location of a new branch or bookmobile stop. Studies could be made of the internal operations also. Just a few possibilities might be: (1) analysis of vendor or binder performance on the basis of cost or service, (2) programming of personnel scheduling to even out work loads, and (3) the programming of serial binding to take advantage of both cost and time factors. These are just some of the many analytical possibilities in the systems oriented library.

Now one last note on the advantages of the systems approach to library automation. If the library has been carefully analyzed, its size makes little difference to the computer system. There will be certain differences in technique between libraries of widely divergent sizes, but once a particular library system is developed it will be much more elastic for library growth than would any manual system. Another consideration is the slowly changing emphasis in libraries

from the treatment of books to the treatment of report and serial literature. A system which has been carefully automated would be much more amenable to shifts to deeper levels of subject cataloging or indexing than is a manual system. The computer can become a tireless inter-filer of subject lists.

These are not visionary schemes; such totally integrated library automation systems are just around the corner. There is nothing I have spoken of that is beyond the present state of either the librarian's or the computer manufacturer's art. There is no technical reason why such a system could not be operable next year. It is my firm conviction that the librarian, pressed on one side by the information explosion and enticed upon the other by the increasing availability of computers will soon turn to library systems automation.