The Costs of Data Processing in University Libraries

ON JULY 18 the membership of ACRL's University Libraries Section heard a panel discuss the costs of data processing in university libraries. Don S. Culbertson, head of data processing at the University of Illinois library, Chicago Undergraduate Division, discussed data processing costs in book acquisitions and cataloging; Melvin J. Voigt, director of libraries at the University of California at San Diego, compared the costs of manual serial records with those of computer-based serial records; James R. Cox, head of the circulation department at UCLA, examined the costs of operating circulation activities by manual, semi-automated, and fully automated systems. Since this panel developed some new cost data, not previously published, it was deemed desirable to print its papers here in full.—*D. K.*

In Book Acquisition and Cataloging

A STORE WINDOW on Michigan Avenue in Chicago recently displayed several penguins and a small pool. These birds have developed their strange, unique shape through living for generations on end in an isolated place with few natural enemies. When enemy finally appeared in the form of man, they were unable to adapt to this new danger and were easily slaughtered. In a new environment these birds were obsolete.

John G. Kemeny of Dartmouth is one of the latest in a long line of individuals who have followed the lead set by Vannevar Bush nearly twenty years ago in predicting that in their present form libraries are as obsolete as penguins, and that librarians cannot cope with information being produced at its present rate. They also predict that unless librarians mend their ways and adapt to modern conditions they and their libraries may soon be as extinct as the dodo bird.

On what do the prophets of the magic libraries base their predictions of the obsolescence of libraries? While an important recognized factor has been the obvious and much discussed increase in publication, another factor that has been virtually ignored for many years is only now coming under close scrutiny. This is the high cost of processing. We all have arrearages in acquisitions and cataloging; we have book collections parked on storage shelves or packed into boxes awaiting—frequently for years—some kind of action; we have collections of technical reports that we do not know how to handle; thousands of government documents are rarely used because access to them is made difficult by skimpy indexing.

Why do such conditions exist? The answer is readily known. They exist because we need more staff members to process more items. We need more acquisitions librarians, more catalogers, more clerks, more of everything. But where will this end? Eighty years ago three out of every four library dollars

were spent for books; today three-fourths of our expenditures are for staff. Admittedly public services have been much expanded during this period, but they cannot account for such a complete reversal of the ratio.

How much out of our total budgets should we be spending for books? Who has the magic formula to which we can compare our library to see if we are good, bad, or indifferent? I checked to see if answers had been published in the literature. My findings were that little really helpful information has been written on library costs. Much of what exists has appeared in periodicals which are not widely circulated; methodology is uneven; data are totally incompatible from one study to another; and information more than ten years old is of relatively little value. I would like, however, to mention four studies which have been made fairly recently and which are of four different types.

In July 1956 Bella Shachtman reported a study made at the USDA library on the collection of cataloging statistics. This study, while it does not contain cost information, is the raw material which must be first collected and summarized before a cost analysis can be attempted. Quantitative data, such as the percentage of titles requiring original cataloging (99 per cent) compared with those for which Library of Congress copy is available; the number of titles which can be searched in an hour (fifteen); the number of titles which can be cataloged in an hour (two); and the number of cards a professional cataloger can file in an hour (eighty), are needed to complete the picture. More information is available in this article including the checking form each cataloger filled in at the end of the day.

Mrs. Catherine MacQuarrie was chairman of a committee which made a cost survey of southern California libraries. Its report, made in the spring of 1961, covered public and college libraries in the area. Of six college libraries reporting, cataloging costs ranged from $1.08 to $2.88 per title, and processing costs ranged from $.11 to $.91 per title. Among twenty-three public libraries, cataloging costs varied from $.76 per title to $7.41! to this may be added book preparation costs from $.09 to $.54. A detailed cost computation form is included in this report which might be an excellent starting point for a library considering a similar study.

Fred Heinritz recently compared the cost of a book catalog to the cost of a card catalog as his doctoral dissertation at Rutgers University. His comparison was made between the production of the *Engineering Index* and its cumulation in the bound form as the book catalog, and the cost of producing the *Engineering Index* on cards. Whether or not these two catalogs are actually comparable to a library book catalog and a library card catalog only additional studies of this nature will tell. His conclusions are that book catalogs are much cheaper to make and more difficult to use.

Louis Schultheiss and I made a cost study in early 1961 of the University of Illinois library of the Chicago Undergraduate Division. Our cost of cataloging at this time was $8.07 per title, and our cost of acquisition was $4.55 per title. We did not include the cost of maintaining our building in these figures.

What do these studies indicate? First, they indicate that we need some kind of standard method for making reports of this nature; standard report forms, standard procedures for collecting data, and standard ways of cumulating and disseminating this information. Second, they seem to indicate that we need a central agency which can be responsible for prodding librarians who do not turn in their statistics. Third, they show that we need norms to which each librarian can compare his own performance so that he knows whether or not changes in technique or personnel will improve or harm

488

the library. The ALA now has several clearinghouses for information of this kind under consideration. We must encourage the expansion of these efforts into the establishment of one large interdivisional agency which can perform all the functions which might go into the several smaller ones.

James Skipper, now president of RTSD, stated when he accepted the gavel, that the future belongs to technical services. He is right. Unless we get our costs under control and become able to keep them there, the whole library risks belonging to technical services.— *Don S. Culbertson.*

In Serials Handling

THE USE of data processing methods for various aspects of library operation is being considered by many libraries today. In most areas—acquisitions, cataloging, circulation, for example—the activities are interrelated to such a degree that if advanced data processing methods are to be used efficiently, it appears that they must be applied in a coordinated system. The processing of serials and the maintenance of serial records are sufficiently independent areas of library activity to make a system feasible which is not necessarily tied to other aspects of library operation. Also, the repetitive nature of much of serials work lends itself to mechanization. The difficulty large libraries have in providing holdings and receipt information for their users suggests that by the use of data processing, needed services can be provided at little or no additional cost.

At the outset, I would like to make clear that this discussion will be limited to serials processing using computers. It is also possible to handle serials with IBM cards and standard sorting and printing equipment. Such procedures may be satisfactory for small libraries but mechanical card equipment does not have the capacity, speed, or efficiency necessary for serials control involving the thousands of titles found in most

university libraries. To maintain up-to-date holdings records, produce receipt and holdings lists, and to create records for the receipt of current issues, for binding, and for subscriptions, requires a computer with a large memory and which, for efficiency, operates at a high speed.

This panel has been asked to discuss the cost of data processing operations and to compare them with those of manual systems. For serials, as in the other areas, this is a difficult task for a number of reasons. First, only a few libraries have started serial computer operations and none have applied them thus far to more than a few thousand titles; second, no thorough studies of costs have been completed; and third, as in most areas of library work, there are no accepted standards or even any experimental studies of work output or unit costs for any of the parts of traditional serials processing. Thus what can be reported here must be considered as tentative. Cost data are based primarily on evidence from a single operation, which is just emerging from the pilot plant stage.

In this brief report I do not intend to describe serials operation through computers, but I would refer those who would like information on its possibilities and development to the article pub-

lished in *Library Resources and Technical Services,* in the Winter 1963 issue.[1] A further report on this project will be issued in about six months.

The costs involved may be divided into three categories. The first is the labor cost of serials processing and maintenance. The second is computer cost in terms of computer time and cost per unit of time. The third is the cost of computer program development and maintenance.

Looking first at labor costs, there are a number of processes for which it should be possible to make comparisons. The cataloging of a new serial title or the recataloging of an old title is virtually the same process in a manual system as in one utilizing computers. Entry must be established and holdings determined. If reference cards are to be maintained in the catalog for main and added entries, these must be produced. Instead of a holdings card for a central serials record, information must be prepared for the key punch operator. We call the form which results the "intermediate serials record." Producing this form from an already completed serials record card takes a serials cataloger about three and one-half minutes. At three dollars per hour this would be about $.175 per title. This includes the time required to assign a mnemonic title and a serial number. The only additional task requiring labor prior to converting the record to tape is the key-punching time. In our operation, entries average 2½ cards, punching time for an operator whether on the library staff or at the computer center is about 2/5 minute per card, or one minute per entry, which at $2.40 per hour is $.04 per entry. Key punch machines are available without added charge, as they are in many computer centers. Thus, the conversion cost is about $.215 per entry. For new titles the cost is very close to that of traditional methods, since the intermediate serials record can be prepared in about the same time as a central serials holdings record. In the Medical Library at Washington University, St. Louis, three hundred hours of key-punching time went into the development of their system for some twenty-five hundred titles.[2]

Studies of the time required for checking in periodicals using the computer-produced arrival file as against a manual system of recording on files of 6" by 8" cards have given the following data. In brief tests carried out for thirteen days involving 383 issues, the total check-in time on holdings cards was 369 minutes or .96 minutes per issue. The check-in time using the computer-produced arrival file was 313 minutes, plus 60 minutes for key-punching additional information on the cards when required. Thus, the total check-in time by computer was 373 minutes or .97 minutes per issue. Thus, it appears that there is no advantage in one system over the other in the cost of recording receipts, both requiring approximately one minute per issue.

Some tentative computer costs can be reported, although these will be modified in the next few months. These costs are based on a computerized serials record of 1500 titles using the original program developed in a computer language known as FORTRAN '60. A new, simplified, and more efficient program using FORTRAN '62 is now being tested. It is estimated that computer time will be reduced by one-half to two-thirds by the use of this new program. In addition, unit costs will decrease as the number of titles increase, since setup time will remain the same, and because, in some cases, longer lists will take only slightly longer operating time.

These costs are based on the use of

[1] George Vdovin; Melvin J. Voigt; David Newman; and Clay Perry, "Computer Processing of Serial Records," *Library Resources and Technical Services,* VII (Winter 1963), 71-80.

[2] This system is described by Irwin Pizer, *et al.,* "Mechanization of Library Procedures in the Medium-sized Medical Library: I. The Serial Record," *Bulletin of the Medical Library Association,* LI (July 1963), 313-38.

the CDC 1604 computer at two dollars per minute and the CDC 160A at fifty-five cents per minute. These are university rates which apply to nonscheduled time. Time on commercial computers would be nearly twice as much.

The cost of monthly updating, production of the new arrival file, production of three copies of the complete holdings list, and of auxiliary lists for 1500 titles now totals thirty-nine dollars per month. This includes fourteen minutes on the 1604 and twenty minutes on the 160A. As indicated earlier, this cost increases when additional titles are added, but it does not increase proportionately. For example, up-date cost for 175 titles was twelve dollars while for 1500 titles it was only four dollars more, or sixteen dollars.

The only other outputs produced at the present time are triweekly receipt lists for each of several locations. These now cost $2.20 per set or $6.60 per week.

The total computer time cost of about sixty-five dollars per month is a cost which cannot be compared directly with costs of a manual system. The availability of a complete up-to-date holdings list reduces desk time required to answer users requests for holdings information by several hours a day in an institution with a small public service load. We estimate that this savings is nearly double the entire cost of computer time. Claiming costs are reduced, but more important, claiming becomes almost automatic since unreceived issues are represented by cards left over in the arrival file. Other savings are in equipment for storing holdings records and, if serials are eliminated from card catalogs, in card reproduction, filing, and storage costs. At Washington University, St. Louis, machine time for the outputs developed there costs forty-seven dollars, while total costs including reproduction and distribution of holdings lists are two hundred sixty dollars.

It is difficult to provide any useful information on the cost of developing a computer program. At UCSD flow-charting the operation, developing system requirements, programing and re-programing, testing and debugging has required about six months of a programer's time, six months of a librarian's time, and twelve months of nonprofessional assistance of high calibre. The Washington University program, utilizing some of the results of this project, became operational with less staff time, using eleven hundred hours for systems analysis and design, programing, report design-card layout, and documentation of procedures. As programs are developed, standardization should become possible. While libraries may need to adapt computer programs to fit their needs, the cost of installing a system should be much less when a number of libraries have successful systems in operation.

As in any system, there are management costs. Those of keeping the system going and of training and supervising staff are probably somewhat higher at UCSD at this time than they will be later. Ultimately they should not be greater than in a traditional serials operation.

Additional savings will be gained when bindery preparation forms and recording of bindery receipts are computerized. The new program will include this procedure. Subscription payment can also be incorporated in the program.

In summary, it seems evident that labor costs for the data processing serials system are approximately equal to those of a manual system. Costs of computer time are not high and can be justified in savings in public service costs, in auxiliary activities such as bindery preparation, as well as in the provision of greatly improved services to library users. Programing costs are relatively high but should be reduced as computer programs become standardized and can be adapted to other institutions with little change.— *Melvin J. Voigt.*

FOSTER PALMER, associate librarian at Harvard College, wrote to me recently in answer to questions I had put to him: "It is difficult to write very intelligently about cost figures, although certain items are fairly concrete." The two parts of that sentence are most significant in any discussion of costs; it is indeed difficult to write intelligently about cost figures respecting library circulation systems. In the first place, they are difficult to determine. What does it cost to circulate a book? What does public service in a library really cost? Secondly, the variables which affect library operation cost figures are numerous, making applications to a particular library situation difficult. And thirdly, the interpretation of cost data, when obtained, can be dangerous and must be done carefully, with local factors in mind.

With further reference to the quotation, it is also clear that certain items are indeed fairly concrete. There are two generalizations often made about data processing in libraries which are worthy of attention, particularly as they may apply to circulation procedures. They are at opposite poles and represent extreme viewpoints. One is that university libraries should automate their circulation systems because they will save money. The other is that university libraries should not automate their circulation systems because it costs too much and is not worth it in the end.

Both viewpoints are probably wrong, or, at the least, they are inaccurate. As with most generalizations, it depends upon what is being discussed. If one is talking of dollars and cents, the first statement is particularly fallacious. As more and more libraries become involved to a greater or lesser degree with automatic systems or plans for them, it is becoming more widely understood that it does cost more to automate, from the pure dollars and cents standpoint. On the other hand, if we could translate faster book shelving and filing, more accurate records, and in general, therefore, better public service, into dollars and cents, automation does clearly save money. But it must be understood that this is what we mean, if we are to take this view.

It follows then that the opposite view —the view against automation because of its cost and lack of value in the end— must be a matter of local judgment after careful weighing of the facts of a particular situation, in combination with experiences of other similar institutions.

In this paper I will attempt to give some concrete examples of comparisons in operating costs between what I call manual, semi-automated, and fully-automated circulation systems, and I will attempt to draw some conclusions from this information. I will draw largely from our eight months' experience at UCLA where we installed an IBM circulation control system at the main loan desk in November 1962 and where we have had long experience with the manual system and with a semi-automated, or Keysort card, system. Considerable additional experience has been had by other college and university libraries, and many systems are now in various stages of planning throughout the country, ranging from semi-automated punched card systems to fully-automated computer-based systems. It has not been possible yet to call upon all of these institutions for cost data. Indeed, many do not have anything but the barest of estimates. However, I will present data and information from the UCLA library, the University of California library at Berkeley, the Harvard College library, the Brooklyn College library, and the Southern Illinois University library, in the hope that such information as is available will

be useful when applied, with extreme care, to situations at other libraries.

In order to do this profitably a few bench marks must be established, some terms defined, and a general outline set so that these comparisons will at least have some meaning.

For our purposes here, I define a circulation system as being manual, semi-automated, or fully-automated. A *manual system* is one in which all filing and discharging of loan records and the sorting of overdues is done by hand. A *semi-automated system* is one in which edge-notched cards, such as Keysort, are used, or where there is some punched-card equipment and machine-readable cards, such as IBM; here filing may be done either by hand or by machine, but discharging of records is done entirely manually; in such a system overdues are sorted automatically, either by punched-card machine or by needling. A *fully-automated system* may consist either of entirely punched-card equipment or a combination of card equipment and a computer; in such a system, filing, discharging, and sorting of overdues is done entirely automatically.

The only procedures dealt with here for comparative purposes will be those which are primarily affected by automation. These are:

1. Gathering and filing of loan records.
2. Counting of statistics.
3. Discharging of loan records and the handling of book returns prior to shelving.
4. Sorting of overdues.
5. Handling of renewals and long-term loans.
6. Handling of snags, or record problems.

Staff hours and costs will be expressed only in terms of these procedures. Other related procedures, such as paging and shelving, are relatively unaffected by the actual circulation system automation, and their costs may normally be considered as remaining relatively constant despite changes elsewhere. To put it another way, technological advances in the other areas of a circulation system may often be made, independent from the basic circulation system itself. Therefore, I caution my audience not to be misled into thinking that the cost figures I will give represent total costs of complete circulation operations. They do not. They are merely used to illustrate cost differentials when partial or complete circulation automation is undertaken.

I have chosen an annual circulation figure of about three hundred fifty thousand volumes as representative of a large university library's work load crossing a main loan desk. It also happened to be convenient, since it is somewhere near the UCLA and Berkeley figures.

A final word of caution—no discussion of costs would be complete without mentioning varying wage rates and equally variable supply and equipment costs, which may be higher or lower around the country. Also, the relatively poor or good physical arrangements in library circulation departments can make enormous differences in costs. I urge you to keep these things in mind as you shudder at some of the costs or as you think, "Well, that's not so bad."

At UCLA, prior to the installation of our IBM circulation system in November 1962, we used an antiquated double-charge system, employing both a call slip and a book check for each charge. This was a completely manual system. The inadequate performance of the circulation procedures outlined above for a circulation of three hundred fifty thousand volumes required around fourteen thousand six hundred hours per year at an average hourly rate of $1.73, or a total expenditure of over twenty-five thousand dollars. Additional annual costs of thirteen hundred fifty dollars for one and a half million call slips and three hundred dollars for one hundred thousand book checks brought this figure to a total of twenty-six thousand dollars.

In November we shifted to a fully-automated IBM circulation control system utilizing five punched-card machines, in which all filing and discharging of records and sorting of overdues is done by machine. Manual labor costs dropped to four thousand hours at $1.71 per hour, or $6,840 annually. Machine labor costs amount to seven thousand hours at $1.91 per hour, or $13,370 annually. Total labor cost then is eleven thousand hours at an average of $1.84 per hour, or $20,210. Comparing this with the manual system we note that total annual hours dropped by thirty-six hundred hours; hourly wage rates increased, reflecting the need for trained and experienced personnel; and total labor cost dropped by five thousand dollars.

Now we come to the crux of the matter. Card costs increased. We now pay nearly twenty-five hundred dollars yearly for one million seven hundred thousand IBM cards, an increase of nearly twelve hundred dollars. We rent five IBM punched card machines for a total of $8,200 per year. This then raises our total annual operating cost in the areas described to thirty-one thousand dollars, a net increase of a trifle more than four thousand dollars over the annual cost of the completely manual system. It is clear that the machines did the trick.

How does all this compare with a semi-automatic system? UCLA has had long experience with the Keysort single charge system, in its branch libraries and the undergraduate library, on a much smaller volume scale. I have made a rash attempt to extrapolate the smaller figures out to the rate of three hundred fifty thousand volumes circulation per year, and the results show about eight thousand hours per year at an average of $1.75 per hour, for a total labor cost of fourteen thousand dollars. Card costs of nearly nine thousand dollars (Keysort cards are expensive) bring us to a total of twenty-three thousand dollars.

These comparisons make sense, I think. As far as labor costs are concerned, it is clear that the old manual system was the most expensive. The semi-automated system improves upon this. And finally the fully-automated system provides the lowest labor cost. But the mechanical equipment rental costs force the total cost of the automated system above that of the two others.

At UCLA we are now studying a proposal to alter our present IBM system to make it less complex and provide more prompt and direct input into our loan file and more rapid discharging from that file through the use of IBM book cards and direct input units at our charge desks. By eliminating a great deal of key-punching and certain steps in our discharge procedures we estimate a possible saving of four thousand dollars annually in labor costs and five hundred dollars in card costs. These changes would require additional equipment but we could realize a net saving of about two thousand dollars per year and have an improved system.

What of other experiences? The University of California library at Berkeley has made use of a semi-automated IBM circulation system for a number of years. They hand-file and hand-discharge their loan records and sort overdues by machine. Total labor cost involved in the procedures we are discussing amounts to seven thousand hours annually at $1.85 per hour, or nearly thirteen thousand dollars. Card costs of fifty-four hundred dollars and IBM machine rental (three machines) of twenty-five hundred dollars brings the total cost to more than twenty thousand dollars. This compares quite favorably with the cost of the manual, semi-automated, and automated systems thus far described.

Harvard College library is about to install an IBM circulation system. It is estimated that machine rental costs will be fifty-five hundred dollars annually

and card costs thirty-two hundred dollars. As far as salaries are concerned they expect to add about three thousand dollars per year more for IBM skills. It is estimated that it would have been necessary to add six thousand to nine thousand dollars in salaries annually "to make the former system limp along less haltingly." To quote Mr. Palmer again, "We have not gone into this with any idea that it is going to save us a lot of money. We know, on the other hand, that it will cost us more at first, but we hope that it will result in greatly improved service. The old system was so bad that there was general willingness to invest something in a possible improvement."

I do not yet have cost information on the proposed automated circulation control system at Southern Illinois University. The system utilizes twelve pieces of IBM punched-card equipment which could be expected to rent for several thousand dollars per year, but all charges will be processed on the 1401 tape system located in a centralized University Data Processing and Computing Center.

This leads me to mention what may be the future of fully-automated circulation systems, provided its great cost can be spread over many library procedures, providing multiple-usage. That is the computer-based system involving, for example, the use of a 1440 computer and card equipment, where the library itself rents the equipment, housing it in the library, instead of using it on a time charge basis. Such systems are being considered and planned by libraries. A 1440 configuration can cost variously between twenty-five hundred dollars and four thousand dollars per month. Here, additional aspects of the circulation procedure, such as overdue notices and bills, can be automated at some saving. The cost of such systems make it essential that as many additional library operations as possible and practical be computerized along with circulation.

What conclusions, then, can be drawn from all of this? It is clear that large university libraries throughout the country are faced with circulation loads that are increasing at almost geometric rates as student enrollments increase and research demands become more intense. Manual or semi-automated systems can no longer absorb this ever-increasing growth, even if great numbers of staff could be added. Such circulation systems ultimately break down because they cannot provide fast enough service, which in a library really means that the book cannot be returned to the shelf and circulated again fast enough to meet the demand, and the loan record cannot be filed fast or accurately enough or consulted fast enough to provide good service.

M. Allyn Fox, executive assistant at the Brooklyn College library, wrote to me the following: "In terms of costs and productivity, the Gaylord system we used cost practically nothing when compared with IBM. However, with a circulation much less than the four hundred thousand we have now, we found books piling up to the ceiling and a constant backlog in the manual filing of book cards. Therefore, while we are spending a lot more money, we can keep up with the demand. In general, what has happened here has occurred wherever data processing equipment is installed. The system is more expensive in dollars, but you get a lot more speed and control for your money."

From the UCLA experience I would certainly agree with him. There is no question in my mind that the quality of our service has improved immensely as a .result of automation. In all of our thinking about these problems the question of cost must be related to the total job, a part of which certainly is improved service. The application of data processing equipment to the solution of circulation problems must be seriously and carefully studied, from all angles, before it is accepted or rejected.—*James R. Cox.* ■ ■