

Estimating Data Processing Costs in Libraries

By HILLIS L. GRIFFIN

ELECTRONIC DATA PROCESSING (EDP) is assuming an important and effective role in the operation of many libraries. Librarians who wish to explore the possibilities of this technology in their own libraries are surprised to find that those who are using EDP systems are reluctant to disclose their operating costs. Yet those who have mechanized operations within their libraries know that their own costs are not a reliable indication of what similar operations would cost another library. Many factors influence data processing costs, and an operation which might seem costly to one library may be paying its way very nicely in another, simply because of the way it fits into the operational pattern.

In other words, costs for EDP operations cannot be extended without qualification from one library to another, because time and equipment charges vary so widely from one institution to another. If the university library tries to compare costs with a nonuniversity operation, it will find that its internal costs, as well as its equipment costs, may be considerably less than those encountered in the nonuniversity situation. If input is key-punched in the library, for example, it may cost much less than if it were sent out to a local data center for preparation. If a library adopts another library's system, it may find that it does things that are important for the originating library but unnecessary for itself. If a library does not need certain features of a system, it is senseless to pay for them.

Regardless of location or environment, one factor of the EDP operation remains constant for two libraries using the same

Mr. Griffin is Information Systems Librarian at the Argonne National Laboratories, Argonne, Illinois.

computer system and the same program. This is operating time. Each group would take approximately the same time to punch the cards to order one hundred books and, given the same computer program, to process these through the same system. Costs would then be based upon the amount paid for labor, supplies, and the use of equipment at each location. One library, for example, might be billed for equipment rental, operator labor, and data center overhead. Another library, using a similar system, might pay nothing for the use of the equipment if the institution writes off the data center as an operating expense. The cost of supplies might differ between the institutions. It should be plain that the costs of doing the job in one library are not necessarily the costs of doing it in another. The only constant factor is the amount of time required to do the job on the computer. If the amount of time required to process a job is known, and the number of items processed is given, then it is easy to determine the cost of doing the job in any library.

TYPES OF LIBRARY APPLICATIONS

Library data processing applications can be grouped in three classes. The first is the housekeeping function, consisting of applications such as journal renewal, circulation control, book ordering, and other related tasks, generally in the technical processing area. The second class is that of information dissemination. Examples of this are library announcement

bulletins of new acquisitions and printed catalogs of the library or branch library holdings. The third class is that of retrospective searching, in which the computer is used to select references or produce bibliographies of items in the collection that relate to a particular subject or subjects.

In the first two classes operations are essentially input/output. By this is meant that, in general, a card is read into the machine and a line is written out on the printer (as in writing a list of all materials in circulation). As part of the operation, cards might also be punched with expanded, updated, or rearranged data, or such records might be transferred to magnetic tape or random-access files. These are generally one-to-one operations (*i.e.*, we read, write, read, write . . .) in contrast to those encountered in retrospective searching. Here many records are read in and operated upon, but only a few lines are printed out at the end, such as a list of references (or a bibliography) which appears to satisfy the parameters of the search.

Because of the cost of organizing material for retrospective searching, and because of the cost of machine time for accomplishing such tasks, operations such as housekeeping and dissemination can, at the present time, make the greatest impact upon a library. Preparation of book orders on the computer may require little more effort than is presently being expended to do this manually. Yet the by-products of a mechanized operation, obtained by using the same punched cards over again with another program, can be useful in other ways without necessitating further input preparation. Materials that have not been received may be expedited, want lists prepared, and fiscal records provided quickly and automatically as the result of such operations. Other files may also be obtained, such as a listing, by author or title or both, of all materials on order. Such a listing might also be arranged by cost code, requestor, vendor,

etc. These listings would obviate the need to maintain certain files on a manual basis. All of these jobs may be done using the cards that were used initially to order the books, involving no further clerical effort.

The essential advantages of the computer-based system are: (1) the ability to use the same information in many ways for a variety of purposes, and (2) the ability to perform tasks quickly, easily, and accurately. This flexibility is obtained from input (generally punched cards) which may cost no more to prepare than the documents now being produced in a manual system. Punched cards lend themselves to easy manipulation for the production of other reports containing the same information in a different format, or arranged in a different sequence. Type-written documents cannot be handled in this way; furthermore, they cannot be revised automatically.

DESIGN OF THE SYSTEM

The initial step in the design of any system is to define its purpose. An important consideration in the design is the equipment available to do the job. EAM (electronic accounting machine) or unit record equipment¹ impose restrictions not met in a computer-based system. This is not to denigrate unit record equipment, but it is, nevertheless, an important consideration. Is the type face on the output printer adequate for the job, or does it lack parentheses or other special characters that may be essential? Note the word "essential"! Diacritical marks, semicolons, and question marks dress up the output, but they may not be vital to the job at hand. A decision to dispense with certain special characters or upper and lower case type can make it possible to mechanize the procedure under consideration. The output may not be aesthetically satisfying and it may not appear in

¹ Noncomputer equipment. Examples of EAM printers are the IBM 407 and IBM 402 accounting machines.

exemplary bibliographic form, but it *will* work, and it may, in fact, work rather well.

Design of the format for the printed output is extremely important if the application is to be fully effective. If the application is a new one not previously used by the library, there may be considerable freedom in design of the output format. In this case there need be no rigidly preconceived idea of a "right" output format. A fresh, new format can be used, and mistakes of the past can be eliminated. If the system is to produce an output product similar to one presently in use, it is important to give critical consideration to the content of the present form and to justify the retention of each item. Some information in the form may be obsolete or may be supplied by some other part of the system.

It is also important to consider the output format in relation to the input format (*i.e.*, the format of the punched cards) for the system. The array of printed output items, from top to bottom, should generally bear a definite relationship to the sequence in which these items will enter the system. The printing operation cannot begin until the information which must appear on the first line of the output has entered the system. If, for example, the name of the author appears on the last punched card of each item in a book ordering system, preceded by a number of title cards and cost and account information, we should not design an output form which requires the author's name as the first item of output information. This would require that we read and store the contents of several cards in the computer memory without doing any printing, and that we print several lines without reading any cards. In doing this, we take no advantage of the ability of the computer to do more than one thing at a time, and output speed (*i.e.*, over-all speed of the printing operation) may be slower than necessary.

It is important to "think big" in order

to take maximum advantage of the flexibility of modern data processing systems. It is important to consider all operations which may be served by the data entering the system, and to tailor the input data to serve these needs. Equipment limitations will also have an important part in these considerations. Unit record equipment may require special forms and be less flexible than a modern stored-program computer which may be able, in effect, to emit certain constant information (*e.g.*, page headings, date, etc.) that has been stored by the program, and to print this information as part of the output program at the appropriate time. In this way blank paper may take the place of specially preprinted forms in some operations. EAM equipment must generally print information sequentially as the cards are read by the machine and does not have the extended emitting capability of the computer.

COST ESTIMATING

With the design of the operating input and output in mind, and the available equipment known, the librarian is at last in a position to secure some idea of operating costs. Housekeeping and dissemination systems are essentially output operations. This means that we read a card and print a line (or several lines) in response to the input and the program. The limiting condition for a system is generally the speed of the output devices. By knowing the number of lines of output that we expect to generate and the speed at which our equipment will generate this output, it is possible to obtain a good estimate of what the costs of the system will be. It is only necessary to: (1) determine the number of printed lines (or punched cards) required as output for a given number of items; (2) determine the time required for the printer and/or card punch to produce this number of items; and (3) relate this to time charges for the computer or EAM equipment used.

A simple example might be the preparation on the IBM 1401 of a list of the items in a punched card circulation file. We will write one line for each item in the file, *i.e.*, one line per card. Although the rated writing speed of the 1403 printer (which is used by the IBM 1401) is six hundred lines per minute, a more realistic figure for estimating might be an output rate of five hundred lines per minute. A file of ten thousand items would take approximately 20 minutes to process. If computer time costs \$60 per hour, the cost of the job is $\frac{2}{100}$ of \$60, or \$20.00.

Consider next a job in which we wish to produce punched cards and write simultaneously. Let us assume that there are one hundred overdue items in the circulation file, and that while we are writing the complete circulation list we wish to punch a card for each of these overdue items for later processing. The 1401 will write a line and punch a card simultaneously, but it will do this only at the speed of the punch, which is 250 cards per minute. We will thus write nine thousand nine hundred lines at five hundred lines per minute (19.8 minutes) and one hundred lines will be written at the rate of two hundred fifty lines per minute while cards are punched (0.4 minutes) for a total job time of 20.2 minutes. If one thousand cards were punched the job time would increase to twenty-two minutes (eighteen minutes for writing without punching, and four minutes of simultaneous writing and punching).

The cost per item processed is the total cost of the operation divided by the number of items processed. The additional expense of preparation of the input must be added to this figure. Depending upon the format of the punched card, a key-punch operator might prepare eight hundred to one thousand cards per day. The cost of preparation of the input is the sum of operator salary and overhead plus machine rental divided by the number of cards produced.

Another type of operation which can

be used to advantage is one in which the file is searched for certain items, and action is taken upon these items as they are found. A file of journal subscriptions is an example. The file might be passed through the computer at monthly intervals, with renewal orders being written automatically only for those items expiring during a particular month. It is easy to see that if the term of any subscription is one year, and if the various subscriptions in the file expire at various times throughout the year, each will receive renewal action once during the year. The renewal action will require that several lines are written and that an updated subscription record card is punched. These cards will pass through the system with a minimum of action (probably a comparing step to test the expiration date) for the eleven monthly passes through the computer on which no action is taken. The cost of the operation is the sum of the eleven monthly passes at a nominal reading speed of six hundred cards per minute, and the twelfth pass during which the item is renewed, several lines are written, and a card is punched. It will take about two seconds over the course of a year to renew each subscription and to produce the updated subscription record.

Valuable byproducts can be obtained from this file, such as a list of the current status of all library subscriptions, bid lists, lists by branch library receiving the subscriptions, etc. These supplemental records, obtained from the subscription file, may have value equal to or greater than the value of the original operation.

Because of the wide variation in charges for computer time, it is easy to see that EDP costs in one library may be meaningless to another. This may be due to the difference in time and equipment charges between libraries, or differences in the equipment used. Another factor may well be that the library that developed the system designed a deluxe sys-

(Continued on page 481)

descriptive cataloging division at Library of Congress.

GERALDINE ZIETZ has been appointed librarian in the social sciences reference service, University of California, Berkeley.

WILLIAM E. ZIMPFER became theology librarian at Boston University on July 1.

NECROLOGY

LOUISE SAVAGE, assistant librarian of University of Virginia, Charlottesville, died on July 5.

MRS. MARY WHEELER WELLS, head of the business library at Indianapolis, Ind., public library, died on May 31.

RETIREMENTS

MAY DORNIN, head of the department of archives at University of California, Berkeley, retired on June 30.

MRS. ELIZABETH K. GUDDE, chief acquisitions librarian at Bancroft library, University of California, Berkeley, retired on August 13.

BESS LOWRY, head of the humanities reference service, University of California, Berkeley, retired this summer.

EDNA LOUISE LUCAS, fine arts librarian at Harvard University, has retired after thirty-seven years of service in that capacity.

JANNETTE NEWHALL retired on July 1, from her position as school of theology librarian, Boston University. She leaves for Manila, P.I., in October to administer a fund for the acquisition of books for a newly created Methodist theological seminary.

ALICE LEE PARKER, assistant chief of the prints and photographs division of LC retired after thirty-four years of service, on July 1.

M. EDNA VODRA, reference librarian at Jersey City State College, retired in June after thirty-five years of service.

HILDEGARDE ZIEGLER, head of the cataloging department at MIT since 1926 and member of the cataloging staff there since 1921 retired in June. ■■

ESTIMATING DATA PROCESSING COSTS . . .

(Continued from page 408)

tem, while the investigating library requires only parts of the system. Internal requirements of one library may require more input cards per item, or the input format may not be optimal for the job to be done. Many factors may influence costs, but the major factor, given good system design, is that of time charges for equipment usage.

Cost estimation for systems involving magnetic tape and random access files is also possible but will not be discussed here. The output speed (which is usually the speed at which lines are printed) is

the limiting factor for most housekeeping operations. No matter how fast the tapes, discs, or card readers, most housekeeping systems will run no faster than the printer can handle the output. It is for this reason that cost estimating is rather easy, whether it be to relate someone else's job to your operation, or to find out what a proposed application will cost in your library. What is required is a logical analysis of the input and output of the proposed system. Then, given the output speeds of the equipment used, the volume of output, and the time charges for equipment usage, the costs follow logically and easily. ■■