STANDARDIZED COSTS FOR AUTOMATED LIBRARY SYSTEMS

Mary Ellen L. JACOB: Systems Officer, Fisher Library, University of Sydney

Costs of automated library systems as currently given in published reports tend to be misleading and confusing. It is necessary to have a clear understanding of how they were derived before any comparisons can be made. Clearly defined costs in terms of time units are more meaningful than straight dollar costs and can be used as one means of comparison among various system designs and as guidelines for the design of new systems.

There is a great lack of consistency in reporting the costs of automated library systems. Cost figures given in published reports tend to be misleading and confusing; rather than indicate the true cost of a system, they tend to obscure the entire issue. Without a clear understanding of how such figures were obtained, one cannot use them for comparison against any other system (1). While it is true that no two systems are identical, use of standardized methods can make cross comparisons meaningful and give a basis for estimating the costs of new systems (2).

When all the variables affecting automated library systems are considered, it is very tempting to say that no realistic comparisons can be made. What is needed is some definite statement of just what criteria can be used to determine costs and how they are derived (3). While there have been numerous studies dealing with the cost aspects of specific functions, no real attempt has been made to define standardized cost criteria

for automated library systems. The following discussion is an attempt to identify and define some of the more common cost aspects of such systems.

WHAT IS COST?

"Cost" as defined by accountants, is not the subject of this article. Primary interest here is in cost as a yardstick for measuring the efficiency and effectiveness of a system and for its comparison with other systems. It is important to note that cost is only one criterion and not necessarily the most important one. As costs are herein described several factors need to be determined. Does cost include: 1) fixed overhead such as lighting, office space, administrative functions, etc.? 2) actual salaries or assessed salaries (some installations have a fixed figure for certain types of jobs regardless of the actual cost)? 3) equipment cost (each installation has its own methods for prorating equipment costs)? 4) material costs, paper supplies, etc.?

Cost figures in terms of dollars have little or no meaning unless their derivation is understood. More meaningful are costs in terms of time, man units for human work, and actual running times for equipment. Even for use of these units it is necessary to know something of the relative skill of the personnel involved and the equipment configuration used.

PERSONNEL COSTS

Before examination of the possible breakdown of personnel costs several pertinent points should be considered. It is necessary to state the backgrounds, skills, and levels of experience of the personnel involved. These should include extent of familiarity and experience with the system environment. This environment consists of the equipment, computer or otherwise, and the particular library application involved.

It would be advantageous if there were some objective ways of measuring system analyst and programmer performance, rather than reliance on background and experience as measures. Unfortunately there are none. This is a problem that has bothered service bureaus, software houses, and any data processing manager worth his salt. At present there is no clear-cut answer. Some try to measure efficiency by the length of time taken to code a number of program steps. This gives no measure of the efficiency of the program generated, only a guide to the translating ability of the individual involved. It is certainly no measure of the actual program performance on the computer system. In addition it is extremely difficult to estimate accurately the actual running time that a given program should take, especially for a time-sharing system. How then can one measure the effectiveness of a given programmer in achieving such a goal?

Another problem to be considered is that of the best program versus the most efficient. It is important that a program be maintainable and capable of being changed easily by another programmer. Too, if equipment changes are contemplated, it is highly desirable that the program be written

in a higher-level language, such as COBOL or FORTRAN, which can be used on another machine. Higher-level languages are not as efficient as assembler languages, but generally take less time to write and debug and are usually transferable from one machine to another with only minor modifications.

While it is not possible to measure analyst or programmer efficiency accurately, it does help to know the level of experience and the general background of personnel. While an inexperienced analyst or programmer may occasionally be more efficient than an experienced one, this is not generally true. Normally the more experienced man will know a variety of standardized methods or shortcuts that can be used effectively to either shorten running time, coding time, or both. More important, he knows where to start.

A sample personnel description might read:

Systems and Programming Staff:

1 systems analyst, B.A. in Business Administration, five years' experience with various makes of computers, two years of which were spent as a programmer working with COBOL and FORTRAN, no library background, worked on a part-time basis.

2 programmers, both with high school diplomas, one with one year's experience with COBOL, one trainee with no experience, but high aptitude, both with no library background.

1 library data processing coordinator, Masters in Library Science, manufacturer's course in systems analysis and programming, knowledge of COBOL.

Data Preparation Staff:

2 professional librarians with Masters in Library Science.

5 clerk-typists, high school diplomas, two with keypunch ability, three typists with 60 wpm.

A breakdown of personnel costs should include:

1) planning;
2) actual design (both systems and individual programs);
3) coding (writing of actual programs);
4) testing/debugging;
5) file conversion;
6) actual data preparation and correction (includes new file; preparation and maintenance of existing files);
7) program maintenance.

In the planning, design, and coding phases both total time actually spent and the elapsed project time should be given. Testing/debugging and conversion costs are normally one-time costs, but both can amount to a sizable portion of the system cost (4). Ideally, if conversion costs can also include file cleanup, it helps make that portion more valuable and easier to accept (5).

Once the system is in operation, data preparation and correction times become major cost factors. These are usually highest during the initial installation when personnel are learning the new system. Care should be taken not to let the size of initial costs bias the entire cost figure. Once initial training is over, these will reach a more realistic level and will be more indicative of system requirements and actual costs.

EQUIPMENT COSTS

Just as there are difficulties in determining analyst and programmer efficiency there are similar, though less severe, problems in comparing the efficiency of various machine configurations. Even in comparison of two identical machine configurations, different run times are possible for the same job. The operating system or monitor must also be considered, as must the experience and efficiency of the computer operator. Systems are improving to the point where operator performance is less critical than it once was, but it can still be a significant factor.

Equipment costs are largely determined by the machine configuration used. A tape system may have a totally different running time from that of a disc or drum system. The configuration also affects what types of systems may be implemented. For comparison purposes it is necessary to state the make, model, and memory size of the computer used. Memory size should be given in either words or bytes. A byte is the amount of storage required for one alphabetic character, one special character, or one or two numeric characters. If the memory size is specified in words, the word size should also be given. Details of the computer peripherals, such as the general type (i.e., tape drives, disc, printer, card reader, punch, etc.) make, model, and number should be given. For printers, card readers, punches, paper tape units, etc. the speed should also be given (i.e. lines/minute, card/minute, characters/second, etc.). For storage media, such as disc and drum, the storage capacity should be given. For tape units the tape density should be included.

A sample   description might read:
    1-IBM 360/20 submodel 5, 24K bytes
    4-IBM 2415 tape drives, model 3, 800 bpi
    1-IBM 2560 card reader/punch, reader-500 cpm, punch-160 cpm
    1-IBM 2203 printer 450 lpm
    2-IBM 2311 disc, model 12, 2.7 million bytes
Equipment costs can be subdivided in many ways. A possible method includes:
    1) Computer Costs
        a) Compile times (highly language and computer dependent)
        b) Test/debug time (should include the entire system as well as individual components)
        c) Actual run times
        d) Maintenance or debug after installation

2) Additional Equipment Costs
   a) Keypunch, paper tape, optical character, other input devices
   b) Interpreting punched card output
   c) Sorting/collating
   d) Listings
   e) Bursting, binding, etc.
3) Special Forms or Material Costs
   a) Input or work forms
   b) Punched cards (pre-printed or blank)
   c) Pre-printed forms
   d) Carbon sets or NCR forms
   e) Pre-punched badge or ID cards
   f) Masters for reproduction
   g) Special computer printer ribbons

While all of the above items may not be applicable to a particular system, all those that are should be included. Large or real-time systems might need additional categories.

Compile and test/debug times are of interest to the systems designer, but are less important than actual run times. Compile times are a function of the computer, the language used, and the complexity of the program. They are more indicative of the compiler efficiency than the system performance. Test/debug times must be allowed for in any system, but data on them will be useful to those having had little experience with automated systems. Experienced designers will be aware of the problem and make adequate allowances for it.

Actual run times are a primary cost factor in any system and representative samples should be given. Details concerning the type and volume of input, type of processing, and the type and volume of output should also be stated.

Program maintenance costs usually do not appear until after the system has been up and running for some time, and are usually not included in reports of system costs, since most reports are written before, or soon after, system installation. They are important, however, because they represent a part of the continuing cost of the system.

Conversion costs can represent a sizable portion of the installation cost of the system. This is especially true if the data must be converted to machine readable form. These costs are of great interest to others engaged in similar conversions and care should be taken to ensure these are accurate.

The most obvious type of file conversion is from a record such as a typed list or a catalog card into a machine readable record through keypunching or keytape conversion. File conversion may still exist for a file already in machine readable form if there are differences between it and the files used by the system. Normally such costs are considerably less than conversion from a non-machine-readable form. Exceptions may occur if the file lacks much of the necessary information or if extensive character

manipulation is required before the information can be used. An example of a file having insufficient data to warrant conversion might be a card file with very abbreviated authors and titles used for quick listing purposes, when what is wanted is a full shelf list containing all added entries and subjects, full titles, and imprint information. Existing information may require too many corrections to expand the authors and titles to provide any really usable information. In other words it might be more economical to repunch the file from scratch than to try to edit and punch corrections.

Non-computer equipment costs should not be neglected. While capital investments in such equipment may be small, the time spent in using the equipment can often be lengthy. This is particularly true of input devices.

Non-computer equipment includes such items as keypunches; keytape units; and any unit record equipment, such as collators, sorters, interpreters, xerox machines, typewriters, guillotines, etc. Just as for computer equipment the type, make, model, quantity used, and special features should be given. A sample description of such equipment might be:

1 IBM Selectric typewriter, ASCII OCR type element
2 IBM 029 Keypunches (no special features)
1 IBM 82 Sorter
1 IBM 85 Collator

In a system using a large number of punched cards or large volumes of paper for printing, these too can be significant cost factors. Again the volume of usage may be more helpful than actual dollar cost. Special or pre-printed forms are usually more expensive than plain forms, so it is important to state types as well as quantities. The actual dollar cost should be stated as well.

An example of materials used to produce a small printed catalog with shortened entries on a six month cycle is:

 600 DiKote masters (for multilith reproduction) at $56.00/1000 masters
1000 Pre-printed punched cards at $1.50/1000 cards
   1 IBM 1403 computer printer ribbon, No. 413197
1200 pages, standard, lined, 14¾ x 11-inch computer printer paper

PRESENTATION OF COSTS

The format for presenting cost data could be divided as follows:

Personnel: a brief paragraph describing the number, types, backgrounds, and skills of all personnel involved with the system.

Equipment: Computer equipment: a brief statement of the computer make, model, and memory size; type, model and number of peripherals.

Additional equipment: a brief statement of the types, makes, models, and numbers of any other equipment necessary for the successful operation of the system.

Materials: a brief statement of the types and quantities of forms used, and for special forms an indication of the actual dollar cost as well.

Table 1. Cost Control Form For SDI System

| Functions	Elapsed	Personnel		Equipment and Material		
		Total Hours	Type	Number	Time (in hours)	Type
Planning	1 month	80	Analyst	1	—	—
		20	Librarian	1	—	—
Actual Design	3 months	98	Analyst	1	—	—
		20	Librarian	1	—	—
Coding	4 months	90	Analyst	1	—	—
Compile	3 months	2	Analyst	1	1.2	CDC3600
Testing/Debugging	2 months	5	Analyst	1	2.9	CDC3600
File Conversion	.5 months	15	Clerk	1	7	IBM 029
					1	CDC3600
Data Preparation/Correction per run		1	Clerk	1	1	IBM 029
		3	Librarian	1	.2	CDC3600
Individual Job Run						
Citations/External		.1	Clerk	1	.5	CDC3600
Citations/Internal		.1	Clerk	1	.05	CDC3600
Profile Update		.1	Clerk	1	.1	CDC3600
Decollating					.2	
Bursting					.2	
Printing						
Computer paper (14¼ x 11)						
Citations/External					80 pages	
Citations/Internal					50 pages	
Profile Update					50 pages	
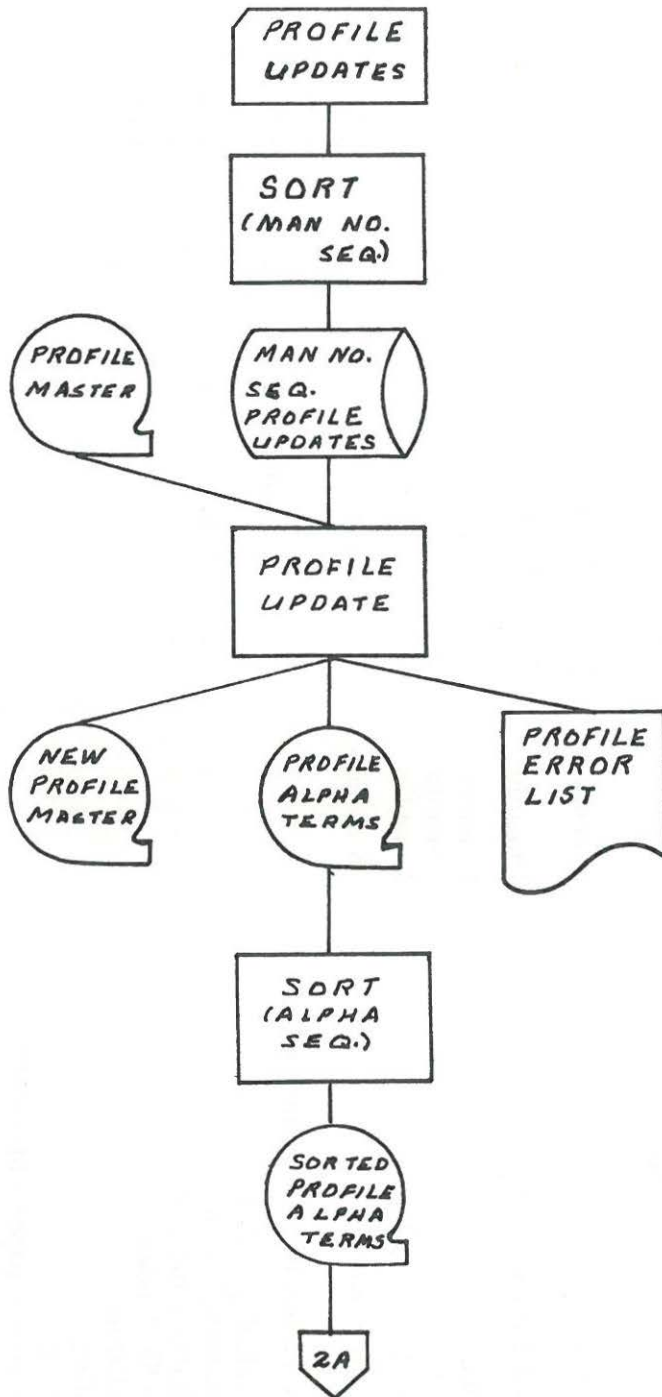
Fig. 1. Profile Update.
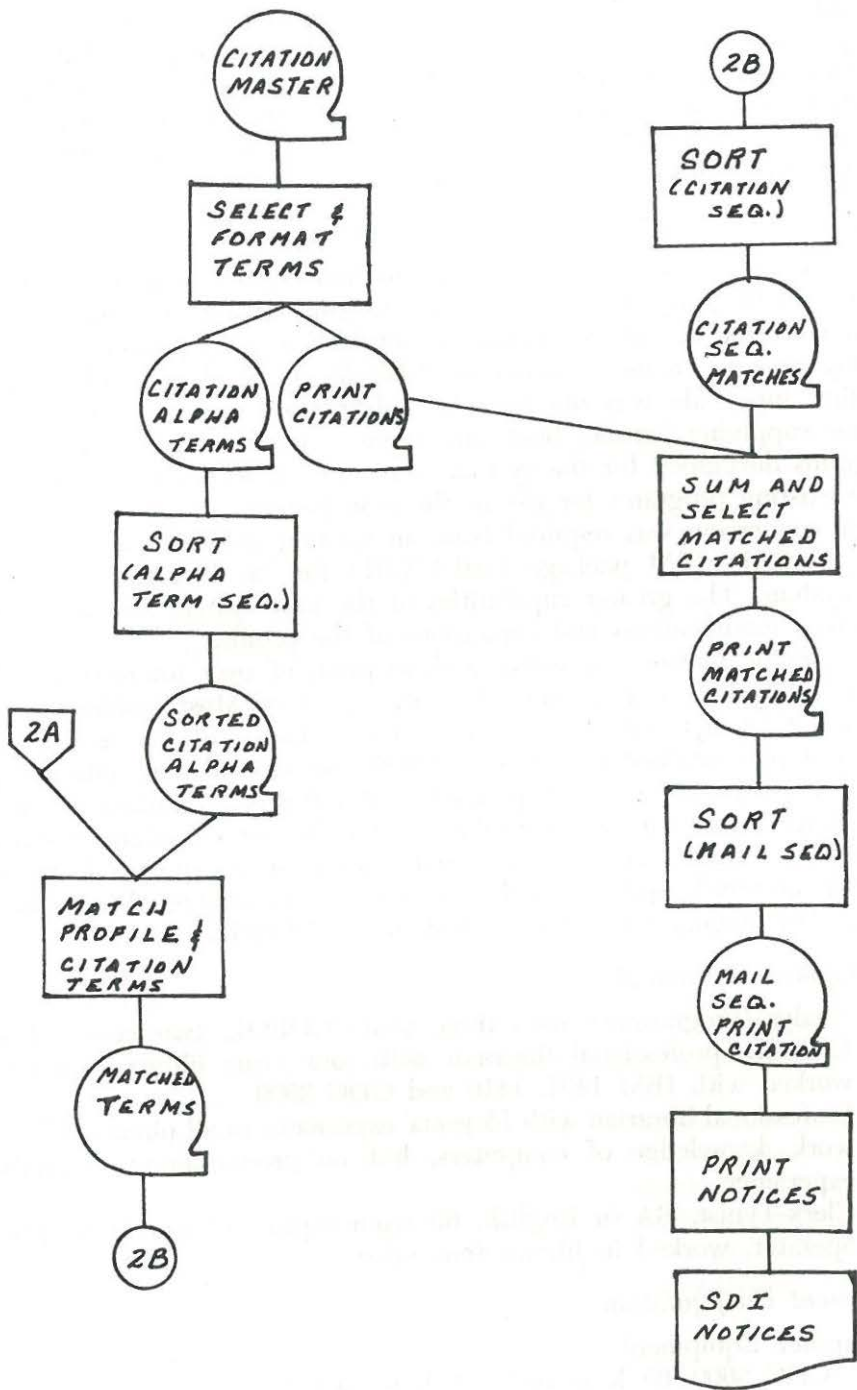
Fig. 2. Citation Run.

Table 1 shows a simple presentation of system cost. The table is a suggested form only and is not exhaustive; it can be expanded as needed for more complex systems. The information and figures given in the table illustrate the system discussed in the following section. The purpose is to provide a sample, not to describe the system in detail, and consequently the system description is very brief.

SYSTEM DESCRIPTION

A Selective Dissemination of Information System was developed to serve a small group of engineers in a scientific laboratory. One source of input consisted of current accessions obtained as a by-product of regular weekly runs to create a master shelf-list file in machine readable form. Another input file was obtained by subscription to a commercial tape service supplying journal, book and report citations. While most of the programs developed for the system were new, it was possible to modify some existing programs for use in the new system.

File conversion was required from an existing profile tape used in the previous SDI (IBM package 1401-CX-01) for the format used by the new system. The greater capabilities of the new system also resulted in numerous modifications and expansions of the profiles.

The profile master containing a description of user interests had just under 100 profiles representing 40 separate groups. Most profiles were for groups rather than individuals; these were updated only as needed. The citation tape contained slightly over 8,000 journals and book citations per week. The internal citation tape contained 180 report citations per week. An average of 400 notices per weekly run for the external citations and 200 notices per weekly run for the internal citations were generated. Systems flow for the profile update and the citation runs are contained in Figures 1 and 2. The language used for the system was COBOL.

Development Personnel

1 Analyst/Programmer with three years COBOL, two years AUTO-CODER, professional librarian with four years library experience, worked with IBM 1401, 1410 and CDC 3600

1 Professional librarian with 15 years' experience in all phases of library work, knowledge of computers, but no programming or analysis experience

1 Clerk-Typist, BA in English, 60 wpm typist, self-taught keypunch operator, worked in library four years

Equipment Configuration

Computer Equipment

1 CDC 3600, 65 K (words), 8 bytes/word

8 CDC 604 tape drives, 200/500/800 bpi, 7 track, 37.5 inches per sec.

2 CDC 861 Magnetic Drums at 4.2 million characters, 17 ms access time, 2 million cps transfer rate
1 CDC 405 Card reader, photoelectric, 1200 cpm
1 CDC 415 Card Punch, 250 cpm
2 CDC 501 printers, 1000 1 pm, 64 char. print set, 136 char. line
1 CDC 3601 Console

Non-computer Equipment
1 IBM 026 keypunch (no special features)
1 Decollator
1 Burster
1 Hand perforator

Materials
Standard (14½ x 11), lined, computer printer paper
Blank punched cards
Magnetic tape subscription @ $5000./year

GENERAL CONSIDERATIONS

How well the system attains its intended goals within the desired limits of design, development, and operating costs is the most important consideration. Design and development costs are usually initial costs only, but operating costs continue as long as the system functions. Operating costs must include the cost of data preparation, computer run times, cost of program maintenance, additional equipment costs, and cost of special forms or materials needed. Careful consideration should be given to allowing sufficient money to be spent in design and development so that overall operating costs, especially those of data preparation and computer run times, can be reduced.

REFERENCES

1. Griffin, Hillis L.: "Estimating Data Processing Costs in Libraries," *College and Research Libraries,* 25 (Sept. 1964), 400-03, 431.
2. Fasana, Paul J.: "Determining the Cost of Library Automation," *A.L.A. Bulletin,* 61 (June 1967, 656-61).
3. Landau, Herbert B.: "The Cost Analysis of Document Surrogation: A Literature Review," *American Documentation,* 20 (Oct. 1969), 320-310.
4. Gregory, Robert H.; Van Horn, Richard L.: *Automatic Data-Processing Systems: Principles and Procedures* (Belmont, Ca: Wadsworth, 1963).
5. Hammer, Donald P.: "Problems in the Conversion of Bibliographic Data: a Keypunching Experiment," *American Documentation,* 19 (Jan. 1968), 12-17.