Determining Costs of Information Systems*

MONROE E. FREEMAN
Science Information Exchange, Smithsonian Institution, Washington, D. C.

Received August 16, 1966

Little has been reported on the cost of information systems, partly because of the newness of such systems and partly because of their complexity. This paper separates the various activities of the system used at Science Information Exchange and presents a critical cost analysis of these activities. The Exchange has some unique features, including constant input and updating, an annual corpus of nearly 100,000 research summaries which yield about 2,000,000 items of information, and input from both Federal and non-Federal sources. Illustrations and tables show the various steps followed in determining the costs and productivity at the Science Information Exchange.

In seeking or justifying financial support for a scientific information service or center sooner or later two questions arise: How much does it cost? How much is it worth? The algebraic difference is the expected profit or loss. If not in dollars, it will be an advantage or disadvantage in other terms; but in any term it will influence the decision to support or not to support the center.

In scientific research the prediction of future accomplishments is even more difficult than in some other enterprises. In the same way, the actual value of the information provided to or acquired by a research scientist to support his research work cannot be determined, at least before the fact. While the value cannot be determined, the cost of the information can and should be. There are two cogent reasons why it should be.

The first reason is that good cost and production data will do as much as anything else to inspire the confidence of the prospective donor of funds or support. He may be inclined to think that the supplicant really knows what he is talking about, and if the prospective donor should happen to know anything about the information business, he will be even more impressed.

The second reason is that good cost and production data for each unit process of the system and for each unit product or service are essential for positive control

* Presented before the Division of Chemical Literature, Symposium on Resources for Information Systems, 151st National Meeting of the American Chemical Society, Pittsburgh, Pa., March 23, 1966.

Table I. Input Flow and Backlog Report for One Sample Month

ACTIVITY Station or Section		Number of Research Task Records IN	OUT	BACKLOG	EFFECTIVE STRENGTH, %[a]
Registration		5397	5290	1144	85
Administrative Coding		5290	6924	1481	96
Key Punch (Adm.)		6924	5255	2315	90
Reproduction		6219	6482	568	95
Master Routing		6483	6043	1268	77.6
Preliminary Indexing		3612 (O)[b] 5426	2633 (O) 4384	1207 (O) 1622	97
		1814 (C)[c]	1751 (C)	415 (C)	
Science Analysis	L.S.D.	2359 (O) 3880	1462 (O) 3331	1500 (O) 3831	77
		1521 (C)	1869 (C)	2331 (C)	
	P.S.D.	459 (O) 962	352 (O) 886	142 (O) 169	91.3
		503 (C)	534 (C)	27 (C)	
Science Coding	L.S.D.	1428 (O) 3261	1553 (O) 3162	0 (O) 502	92.5
		1833 (C)	1609 (C)	502 (C)	
	P.S.D.	231 (O) 574	254 (O) 1575	0 (O) 19	92.5
		343 (C)	1321 (C)	19 (C)	
Key Punch (Sci.)	L.S.D.	1553 (O) 3162			
		1609 (C)	3255	2378	90
	P.S.D.	250 (O) 1545			
		1295 (C)			

[a] Percentage of assigned personnel actually present for work. [b] (O) = Original Grant. [c] (C) = Renewed Grant. (These have second priority since there is a record on file for the previous year.)

of the work flow, for the efficiency, and for the economy of the total system. Without such data, management is little more than guesswork, uninformed and somewhat irresponsible. We have had impressive examples of guesswork by outside consultants who recommended improvements in our system. What was not noted was the fact that the recommended improvements would have cost \$60,000 or more, and would have saved about \$4000 a year. Fifteen years to amortize a \$60,000 investment in computer software does not seem to be very sound economic practice, especially in the unsteady state of the information business today. Expert advisers sometimes are curiously indifferent to such mundane details. Responsible managers, on the other hand, cannot afford to be.

At Science Information Exchange (SIE), it was desirable and necessary to have positive control of the continuous flow of work through an assembly line system. This system now handles an input of 100,000 records (2,000,000 information items) a year and an output of 45,000 reports including more than three-quarters of a million documents. It grows steadily every year. Positive control of a process stream of this dimension demands something more than guesswork.

Two years ago we were unable to find a satisfactory cost and production accounting method that could be applied to scientific information systems in general, or to ours in particular, so we designed one. While it was designed to meet some of the unique features of SIE, there may be elements of interest and general applications useful to others.

Standard business accounting practices are not applicable to documentation and scientific information centers and services without considerable modification. Therefore, I will jump immediately to the assumption (tentative, of course) that one practical approach is to separate the input or inventory cost from the output or service cost, treating them, initially at least, as independent operations. This will yield all the data needed for internal management purposes, decision, control, efficiency, and economy. If

Table II. Number of Output Units Completed

Period SAMPLE MONTH

Science Divisions	Questions: D	E-1	E-2	Jt. D	Jt. E-1	Jt. E-2	Total Questions	Total Requests
Life	26	215	118	3	35	16	413	243
Physical	56	31	72	9	15	15	198	153
TOTALS	82	246	190	12	50	31	611	396

	No. of Requests: C	G-1	G-2	Invest.: F
Operations Division	7	4	9	3615

Notice of Research Projects Distributed

D & E	24,998
C, G-1, G-2	884
F	21,099
H	19,036
TOTAL	66,017

Projects in Process

A	13
B	2

OUTPUT CODE

A	Subj. Field Special Project
B	Agency Program Special Project
C	Compilation, Administrative[a]
D	Compilation, (including Subject)
E-1	Retrieval, Subject by Computer
E-2	Retrieval, Subject by Manual
F	Retrieval, Investigator
G-1	Retrieval, Agency Accession
G-2	Retrieval, SIE Accession
H	Automatic Distribution

[a] Source, Support, or Location

it is necessary to price and sell the products at a profit or on a cost recovery basis, then an arbitrary formula can be chosen later to allocate the inventory costs against the output products. The method of allocation, however, depends so much on the characteristics, circumstances, and objectives of each individual system that discussion in general terms here would not really be useful.

SIE does have some peculiar, if not unique, characteristics and circumstances, so a brief description of its information handling system and its products may help to understand why and how this cost and product accounting scheme was planned as it was and put into operation.

First, SIE deals only with records of currently active research. These are updated annually with a probable half-life that is less than 12 months. Therefore, speed in processing the input and the output is especially critical for SIE. The input fluctuates widely and unpredictably, as these records come in throughout the year from over 500 different research organizations. They pass through 10 unit processing sections in a continuous flow. These various sections are under the purview of the Life Sciences Division (LSD), the Physical Sciences Division (PSD), or the Operations Division (OPD). The sections perform about 60 to 70 operations altogether, including the coding, analysis, filing, and storage of about two million items of information a year. The result is an input inventory stored in essentially two basic files. The computer file is in code for easier manipulation of the data. A parallel file of full text is in hard copy and microfilm. This is the merchandise on the shelf and ready for use. All cost elements up to this point are computed as inventory or input cost.

All the output products and services of SIE are on demand and are more or less custommade to the specification of the user. This service arises from the fact that SIE's users are mostly active research scientists, the research administrators, and program managers responsible for multimillion dollar programs. Their information needs are sophisticated and more often than not are individually and specifically related to their own special problems. Their demands, therefore, cover a rather wide variety

Table III. Individual Job Cost Sheet

Job No. Date Started

Requestor Date Completed

	MONTH				
	JULY	AUGUST	SEPTEMBER	OCTOBER	TOTAL
Operations Division					
Direct Labor Hours[a]					
Direct Labor Dollars[a]					
Direct Support Dollars[b]					
Life Sciences Division					
Direct Labor Hours[a]					
Direct Labor Dollars[a]					
Direct Support Dollars[b]					
Physical Sciences Division					
Direct Labor Hours[a]					
Direct Labor Dollars[a]					
Direct Support Dollars[b]					
Special Projects Division					
Direct Labor Hours[a]					
Direct Labor Dollars[a]					
Direct Support Dollars[b]					
Computer Unit					
Direct Labor Hours[a]					
Direct Computer Hours[a]					
Direct Labor Dollars[a]					
Direct Computer Dollars[a]					
Direct Support Dollars[c]					
Total Direct Cost[d]					
Director's Office					
Direct Support Dollars[e]					
Total Cost					

[a] Obtained from divisional time reports and average labor pay rates. [b] Developed from the Unit Cost Accounting System expressed as direct support dollars per direct labor hour. [c] Developed from Unit Cost Accounting System expressed as direct labor support dollars per direct labor hour, and direct computer support dollars against direct computer hours. [d] Overhead and general support costs, such as rent, insurance, supplies, etc., are allocated in proportion to the direct labor cost element. [e] Computed based on direct support dollars per total direct cost dollar.

of products and services. In practice, these range up to catalogs and data compilations that may cost up to 10 or 20 thousand dollars each, and down to simple requests for a few records that cost only one dollar or less. Fortunately, a goodly number have common cost elements and fall into fairly narrow cost ranges, so that several groups or categories can be calculated to average unit cost with sufficient accuracy to be meaningful.

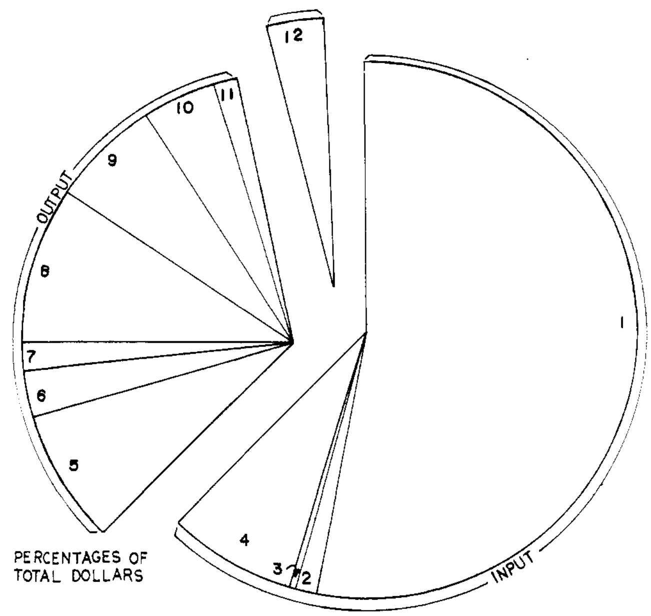
On the other hand, other products do not have common cost elements and their total cost varies widely and upwards. If large enough, these must be individually job-costed in order to get meaningful data. The problem now was simply to find the significant cost elements; but this was easier said than done, since each of the 10 operating units contributed only a part of its effort to one, or to some, or to all of each of the different products or services. Detailed analysis was needed to identify these cost elements in each of the 10 unit processing sections. Once identified, these elements could be summed for evaluation and management control of each operating unit itself. They could also be sorted in another way so that the cost elements in each operating unit were identified with each input or inventory item and with each of the output products or services.

Labor is a critical cost element and requires a good deal of timekeeping and bookkeeping. In this regard there are some complications at SIE that may or may not be important in other systems and centers. Rather than detail these matters, I will simply mention a few. First, the number and variety of output products and services are probably more complicated at SIE than in systems that are oriented solely towards document recovery. Second, the unpredictable and widely fluctuating workloads of both input and output cannot be controlled by SIE and may be more of a problem. Third, the processing speed noted above is a critical factor. And fourth, there is mandatory economy of expensive and scarce category manpower. One-third of SIE personnel are scientists, engineers, and ADP specialists. Under the circumstances listed above, they cannot be held to fixed work schedules or to one process or product. They must divide their time where and when needed to meet contingencies of the moment. Their time and its cost must be divided between input and many different output services. This allocation of their time may vary from day to day, depending on the fluctuating workloads. If the timekeeping is too detailed and meticulous, it is not only resisted by human nature, but takes too much from productive time. On the other hand, guesswork and sloppy time records yield unreliable and unusable data. The happy medium is simply a matter of experience with each individual system, plus good management judgment.

This is the bare outline of a cost and productivity system that has been used for almost two years at SIE with a good deal of modification and debugging during this test period. We believe it is well worth the effort, and we offer a few illustrations of the data and its uses.

Table I is a periodic input flow report that shows the number of records that go in and out of each station in the assembly line during a selected report cycle. There is also (column 4) a backlog count at the close of the report cycle. When the system was first organized, closer controls—*i.e.*, more checkpoints and weekly reports—were needed, but as problems were resolved and the system proved itself, the frequency and number of checkpoints were steadily reduced.

Table IV. SIE Activity by Category, January–June 1965



	% of Total Dollars
Input (62.85)	
Projects	53.31
Proposals	1.50
Administrative changes	0.36
Index development	7.68
Output (33.89)	
Special projects	8.02
Administrative compilations	2.85
Subject: Adm. retrievals	1.53
Computer retrievals	9.44
Visible retrievals	6.36
Investigator retrievals	4.46
Accession number retrievals	1.23
Future Systems Support	3.26

Table II is a periodic output production report of completed units. Output products here are shown only by category or class. This is a monthly report but it could be quarterly or annual, depending on circumstances.

Although not included here, a detailed labor and time distribution report shows the man-hours distribution for each operating unit. It shows the man-hours spent by each division on each different category of input and output product that must be costed. In this table are also the elements of training, new systems development, present systems support (maintenance), general and clerical support, and other activities that may be assigned to such accounting categories as capital investment, maintenance, or overhead. These can be expanded or further detailed as may suit individual management needs.

As noted previously, many products must be individually job-costed. Table III shows the data collection for individual job-costing that may be peculiar at SIE.

All these data are furnished by clerks in the operating units, then consolidated and tabulated in a monthly summary sheet (not shown here). While all data and the information derived therefrom are now compiled by clerks, they could easily be handled by computer. However, we hesitate to invest in a computer program for this operation until cost estimates clearly predict this to be more economical than clerical handling.

The next series of tables illustrates some of the useful information that can be derived from these data.

One obvious but useful product is the distribution of total dollars spent by SIE, as shown in Table IV.

Table V illustrates a three-month unit cost summary. From a comparison of these successive quarterly reports, it becomes quite evident whether one is gaining or losing in efficiency. These comparisons can also reveal, embarrassingly sometimes, whether those changes that were made really improved things or were based on an optimistic assumption.

Table VI is a comparison of four selected unit costs over three successive quarters. It shows how these data may help to decide where to put the extra effort toward improving the system. For example, the last item, investigator reports, went down from $2.43 to $1.70, and then to $1.72. These data, taken with a critical analysis of the production methods, offer good reason to believe that more work on the production of these investigator reports would probably yield only marginal gains. Another type of product, the subject (computer) reports, went from $160 to $60, and then $54. The cost data here, together with a critical analysis of the production method,

Table V. Three Month Unit Cost Study

	Total Cost, $	Number of Units	Average Labor Per Unit, Man-Hours	Average Labor Rate Per Hour, $	Average Computer Time Per Unit, Hours	Average Unit Cost, $[a]
Input Classes						
Projects	237,213	20,941	2.16	3.50	0.004	11.33
Project Changes	2,561	...	...	3.16	...	...
Index Maintenance	30,630	...	...	5.01	...	...
Projects, Total	270,404	20,941	2.51	3.49	...	12.91
Proposals	5,557	6,147	0.17	3.41	0.001	0.90
Output Classes						
Data Compilations[b]	7,117	30	27.59	3.79	1.57	237.23
Data Compilations[c]	20,581	212	11.64	4.11	0.64	97.08
Retrieval, Subject Questions[d]	29,827	549	6.80	4.02	0.35	54.33
Retrieval, Subject Questions	16,755	479	5.91	4.46	...	34.98
Total Subject Reports[c,d,e]	67,163	1,240	7.30	4.13	...	54.16
Retrieval by Author's Name	19,792	11,521	0.70	3.32	0.0005	1.71
Retrieval by Accession Number	597	24	5.01	3.60	...	24.88
Automatic Distribution	6,004	112,054	0.03	3.28	...	0.05
Output, Individual Special Projects					[g]	
Viral Tumorigenesis	1,241	...	...	4.65	1.28	...
Water Resources Catalog	4,176	...	...	4.49	11.77	...
Solid State Physics	1,625	...	...	3.14	12.65	...
Welfare	20	...	...	4.85	...	...
All Physics	2,196	...	...	4.77	5.79	...
Entomology	23	...	...	4.30	...	...
Waterfowl	161	...	...	4.99	0.48	...
Oceanography	476	...	...	4.00	7.23	...
Reproduction	112	...	...	5.29	...	...
Rehabilitation	59	...	...	4.59	...	...
Water Resources Vocabulary	2,798	...	...	4.75	1.64	...
Fluorides	45	...	...	5.06	...	...
Pesticides	52	...	...	4.87	...	...
Environmental Biology	351	...	...	4.37	1.61	...
Recreation	38	...	...	4.56	0.15	...
Urban Research	1,600	...	...	4.99	1.44	...
Transportation	35	...	...	4.72	...	...
Building Research	210	...	...	4.91	...	...
Veterans Administration	2,624	...	...	4.84	10.22	...
Office of Aerospace Research	60	...	...	4.95	...	...
National Bureau of Standards	336	...	...	3.27	6.32	...
Department of the Interior	17	...	...	4.08	0.32	...
Science Information Exchange Quarterly	2,455	...	...	4.17	24.55	...

[a] Includes all but approximately 6.5% of the overhead. [b] All search parameters except subject. [c] All search parameters including subject. [d] Retrieval by computer files. [e] Retrieval by manual files. [f] Direct labor and computer cost for each individual special project listed. [g] Not an average per unit but the total computer time for each project listed.

Table VI. Comparison of Selected Unit Costs in Dollars[a]

	Fiscal Year 1965 First Half	Fiscal Year 1965 Second Half	Fiscal Year 1966, First Quarter
Input			
Projects	12.18	12.08	10.45
Output			
Retrieval, Subject Questions Manual	53.17	31.19	34.98
Retrieval, Subject Questions Computer	160.55	60.85	54.21
Investigator Reports	2.43	1.70	1.72

[a] Based on total cost, including all general and administrative support, except for one overhead item, 6.5%.

suggest that more savings might be anticipated by putting the extra effort into improving this one.

Table VII illustrates comparison of comupter *vs.* manual retrieval costs for subject matter information. SIE was in business 10 years before computers were well-adapted to retrieval by subject search parameters. We had then (and still have) a manual file system for retrieving subject information. With both manual file and computer file systems available, the staff scientists usually go to the file they feel is cheaper and quicker for such questions as, "all current research on viral agents in children with paralytic poliomyelitis that had been previously vaccinated with live virus vaccine."

In this comparison, about 1200 subject searches (comparable to bibliographic searches) were done by the computer over a three-month period at an average cost of $46 each. During the same period, 1600 other searches were made *via* the manual files at an average cost of $23 each. This does not mean that subject search by computer is necessarily more costly or less efficient. These data were not intended to test this point, but are simply a by-product of the data assembled for another purpose. For instance, the samples were not identical, although they were the run-of-the-mill over the same three-month period. Undoubtedly, the easier searches were done by manual files. The more difficult and time-consuming ones went to the computer. Coordinate searches were asked of the computer.*

Table VII. Cost Comparison of Information Retrieval by Manual vs. Computer File[a,b]

	Computer Retrieval (1,279 Searches) Total Cost, $	Computer Retrieval (1,279 Searches) Average Cost Per Search, $	Manual Retrieval (1,678 Searches) Total Cost, $	Manual Retrieval (1,678 Searches) Average Cost Per Search, $
Direct Labor	34,269	26.79	37,904	22.59
Computer	25,098	19.62	...	...
	Cost per Search[c]	46.41		22.59

[a] Retrieval of documents by subject (bibliographic) search. [b] Linear computer file, batched questions, and coordinate searches included. [c] Overhead and general administrative support costs not included since they are approximately proportional to the labor costs.

Nevertheless, the results of this crude, but very practical, test are so striking that we were alerted at once to a systems problem that deserved immediate attention and more accurate study.

The comparison does indicate that marginal gains are not likely to put this computer search system in a competitive position. The cost of new programs and new computer equipment can hardly be justified unless there is some hope of doubling the over-all efficiency. They also suggest that it is really naive to assume that computers are always better than people, simply because computers can also get the wrong answers, but at such incredible speeds. These data presented here are not intended as rigorous proof bearing on difficult information retrieval. They are offered only to illustrate that it is not necessary to live in an "Alice's Wonderland." Good cost production data and analysis can give solid answers, if anyone really wants to take the trouble to get the facts.

* An inverted subject file on disc storage with random access has subsequently reduced the SIE computer retrieval to a competitive basis and the manual files are being phased out after some 15 years of service.