PARAMETRIC COST MODELS OF NATIONAL DATA BANK NETWORKS - VOL. I

McGILL STUDY

RESEARCH AND DEVELOPMENT OF PARAMETRIC COST MODELS TO EVALUATE STRATEGIES IN THE DESIGN OF A NATIONAL DATA BANK NETWORK

Report on a study carried out
under the direction of Professors
H.R. Howson (director), L.R. Amey
and W.D. Thorpe at the Faculty of
Management, McGill University.

May 1974

CONTENTS

PART II (cont'd.)

PART II (cont'd.)

PART III CENTRALIZATION vs. DISPERSION

FOREWORD

This report represents the work of a number of people, staff and students, at McGill University during academic year 1973-74. Apart from the three principals, who supervised the whole undertaking and made their own personal contributions, the two main participants were:

Georges Sudarskis, Diplômé de L'Ecole Supérieure de Commerce de Paris, who contributed Part II, sections 2 and 3 of Part III, in fulfillment of the thesis requirement for the M.B.A. degree.

Professor T.H. Merrett, B.Sc. (Queen's), Ph.D. (Oxon.), assistant professor, School of Computing Science, who contributed the whole of Part IV with the exception of section 3.

The other contributors were :

John M. Meaker, B.A. (Hons. Economics), Sheffield, part of whose thesis towards the M.B.A. degree at McGill forms the Appendix to section 2 of Part I.

P.S. Migicovski: an undergraduate (B.Com.) student at McGill, who under the supervision of Professor Thorpe was responsible for section 4.2.3. (Intensity of usage) of Part II.

PART I: INTRODUCTION.

1. Objectives

As set out in Contract No. OSP3-0125 dated June 11, 1973, the title of the present study is "Research and development of parametric cost models to evaluate strategies in the design of a national data bank network". The objectives set were as follows :

1. The purpose of the project is to develop a model for evaluating strategies in the design of national data banks. Two basic strategies in designing such a system are identified :
   - (i) One central data bank accessed directly by all users,
   - (ii) One central data bank which periodically updates satellite banks. The satellites may contain complete contents of the central bank or a subset of most frequently required information, with the user accessing the closest data bank containing required information.
2. The major factors for consideration in selecting the appropriate design for a particular application will include :

   hardware vs. communication costs,

   user requirements for current information (hourly, daily, weekly,etc.; batch vs. immediate enquiry),

   patterns of user activity (uniformly distributed through time, or peak activity at particular periods),

   user requirements for information access (the size of upset that would satisfy needs of a regional group of users).

   The project will develop parametric cost models reflecting these factors. The comparative analysis will identify the parameters of critical importance in system design and the effects of varying such parameters, such as by government policy or subsidy. Use of the model in planning the design of a specific data bank will identify minimum cost configurations. The model envisaged resulting from this initial project will necessarily be restricted in scope and flexibility, but can be the basis of further development and refinement as additional data bank requirements are identified.

3. The project work will be performed by a team of three McGill faculty from the Faculty of Management and the Computing Centre and two graduate students. The primary source of data for constructing the model will be data bank systems operating at the McGill Computing Centre, in particular the Financial Research Institute data bank. A survey of other major data banks in operation or planned will be conducted, based on published information.

Statement of work

The work shall consist of the following steps :

1. Define hardware models of each system [1(i) and 1(ii) beforementioned]. Develop a cost model of each system, including development, operating and maintenance costs, and communication costs.
2. Define relevant user profiles reflecting demands for currency of information, information access and scope of information.
3. Validate the models with known parameters and user profiles.
4. Perform a comparative analysis of the two basic models, varying parameters reflecting hardware costs, communication costs and user profiles, as a basis for selecting the appropriate design strategy for a specific type of data bank.

2. Scope

Almost inevitably, the fit between what has been accomplished and the set objectives is not perfect at all points. In contrast to the treatment suggested in paragraph 1 of the objectives, our approach in 1(i) has been to describe and model the F.R.I. data banks (Parts II and IV), while in 1(ii), after exploring the conceptual issues involved (the latter was a major part of the project), we have formulated the centralization vs. dispersion problem by means of a general parametric cost model which could apply to F.R.I. or to any other data bank system, including one yet to be established or computerized, and have carried out a number of sensitivity analyses on this model.

Time and circumstances - mainly circumstances - did not permit us to go as fully as we had hoped into questions of user needs and characteristics : the user profiles described in the appendix to section 2 of Part II are based on a limited sample. We have not, at a practical level, gone into the problem of partial dispersion of a central data bank (dispersing only the most frequently required information), though we have discussed this problem conceptually. See also our further remarks under section 5, "Other findings". Finally, time has not permitted us to conduct a survey of "other major data banks in operation or planned". We did make contact with the Canadian Construction Information Corporation, Ottawa, and had some conversation with them in August, 1973; but as their system was not coming into operation until early in 1974 it was too late for us to follow up with a visit. Mr. B.W. Holmes, Operations Manager of the Corporation, expressed interest in our project, and felt that its benefits could be extensive.

Conceptually, we feel we have covered most of the ground set out in the objectives. It would be useful to gather more information than we have been able to about data bank users, preferably in relation to a data bank other than F.R.I. (we encountered some difficulty in gaining entrée to F.R.I. members). This would enable further sensitivity analyses to be carried out on the parametric cost model of Part III, or on minor modifications of it.

Finally, we have (in Part I, section 6) offered gratuitous comments on a matter which, as a result of undertaking the study, seemed to us important in relation to government information and communications policy.

3. Methodology

From the point of view of methods employed, the study divides into three fairly distinct parts. The first (Part II and certain sections of Part III) is concerned with familiarization and information gathering. Under this heading we obtained information by personal visits from the F.R.I., the McGill Computing Centre, and from Bell Canada (on communication rates). Further information was obtained by means of a questionnaire sent to F.R.I. members. We thank all those who co-operated with us, in particular Mr. Richard Hamilton (director) and Mr. W. Calder of the Financial Research Institute, Mr. Gerald Samuel of Bell Canada, and the respondents to the questionnaire.

The second part, concerned with data storage considerations (Part IV), is a computer science study, and is exclusively the work of a specialist in that field, Professor T.H. Merrett.

The third part, which is the core of the project, is concerned with the development of cost models - notably the two "naïve" cost models developed in sections 2 and 3 of Part II, and especially the parametric cost model of Part III. In general, cost functions may be developed in two ways :

(i) By means of activity (or process) analysis : This analysis starts with production functions, expressed in physical terms and built up from engineering data or data collected by experiment. Cost functions are then derived from these, usually by assuming optimizing behaviour (i.e. minimize the cost of various output levels, given input prices and the technical relationships between inputs and outputs represented in the production function). In practice it is sometimes difficult to obtain the data required for this approach, though if it is available there are considerable advantages in approaching the production function from its technical base, becaust it is then relatively easy to incorporate the effects of technological progress. Moreover, one is not restricted to a narrow range of observations as under the second method to be described. In following this approach it is sometimes necessary to transform units to those more suited to economic analysis; this may involve specifying and introducing further activities. It is necessary to ensure that all activities are independent and additive.

(ii) The second approach is to develop production functions or cost functions statistically, either from cross-section or time-series data. The form of the models would here range from single equation least squares models at one extreme to simultaneous equation models adapted to cross-section data at the other. The statistical route is fraught with difficulties, some of which are listed below :

(a) Statistical analysis cannot prove that a certain cost relationship is the true one. It provides a procedure for the rejection of an hypothesis if the probability of a particular relationship having generated the sample observations is less than some fairly small preselected value. Several different hypotheses may well be not inconsistent with the observations.

(b) Unless carried out very carefully, statistical analysis may impart a bias towards linearity (or curvilinearity) in cost functions; e.g. in statistical estimation of long-run cost functions using accounting data, inclusion of straight-line depreciation introduces a linear bias if depreciation due to use is in fact non-linearly related to output.

(c) In statistical studies of multiproduct firms, where an output index is used to measure changes in the level of a diversified output range and the index is (as customarily) constructed by weighting quantity relatives with estimates of average variable cost for each product, this amounts to determining output by cost, and of introducing a spurious dependence where measurement of an independent relationship is really required.

(d) In the case of cost functions based on time series analysis, observations of cost coming from successive time periods during which factor prices may have changed substantially in response to influences other than the firm's purchases, there are two commonly used methods of correcting for price changes : deflation of actual cost figures by a factor price index, and recalculation of costs by applying some common set of factor prices to actual factor inputs of each period. As Johnston has pointed out[1], the adequacy of the deflation procedure depends on the form of the production function; and in fact an adequate deflation procedure may not exist in particular cases. The second method has been shown to result in an overstatement of costs in every period except that to which the selected factor prices relate. The seriousness of any bias introduced by corrections for factor price changes has to be assessed in individual cases according to the amount of variation in relative factor prices in the period concerned and the possibilities of factor substitution in the production process.

(e) With cost functions based on cross-section data a difficulty arises because of the variability of conditions between firms of different size at any given time. This makes for variations about the cost scale line.

(f) Also in the case of cross-section studies conditions favourable to the occurrence of the regression fallacy are likely to be common.

---

[1] J. Johnston, Statistical Cost Analysis, McGraw-Hill, 1960.

(g) As already noted, when observations are confined to samples of actual data, this may provide too narrow a range of values.

(h) Finally, a possible criticism : the results of most statistical cost studies have been incompatible with received economic theory (specifically, they suggest an L-shaped rather than the traditional U-shaped short-run average cost curve). Many of these studies have been statistically suspect, e.g. observations did not span the whole range of possible outputs.

Our project, as it unfolded, seemed to lend itself to the activity analysis approach[1], and this we have followed. Part III is in fact a linear and non-linear activity analysis of the problem of centralization vs. dispersion of data banks. The parametric cost model which is proposed in section 3 of Part III is an integer linear programming model which has the very useful property of unimodularity, enabling it to be solved as a continuous linear programming problem.

---

1 Though it was not clear to us that we would be able to do so for some time at the outset - which is an additional reason for discussing statistical cost functions here.

4. Basic study: Parametric cost model

The concept of data-bank implies centralization, consolidation, efficiency (lack of redundancy). However, with the multiplication of users, it appears that the communications cost due to the connection to the data-bank hinders further audience and patronage. The attempt made with Dataroute and Infodat by the two Canadian communications carriers to abolish distance is a step towards the information utility which will allow everyone to get access to the stores of knowledge.

However, even this breakthrough is not enough to eliminate entirely the communications cost. Just as time-sharing system managements tried to "distribute intelligence" (and reduce the volume of data transmitted) to save on their long-distance communications bills, it is fruitful to see how data-banks can distribute data to reduce their overall operating cost.

In the following, reference will be made to the previous work as an introduction to the proposed model. A more detailed analysis appears in part III of this report.

4.1 Previous work

Four models relevant for the analysis of centralization vs. dispersion have been discussed, one very general in scope[1], the other three more specifically related to the information science-computer field[2].

---

[1] Kochen and Deutsch: "Toward a rational theory of decentralization: some implications of a mathematical approach", Political Science Review (63), 1969.

[2] Information Dynamics Corporation, "A methodology for the analysis of Information Systems", Final Report to the National Science Foundation, 1965. Streeter, "The optimal number of computer installations", IBM Systems Journal, Vol. 12, No. 3, 1973. Casey, "The optimal allocation of files in a network", Spring Joint Computer Conference, 1972.

Kochen and Deutsch, in a brilliant study, have attempted to identify the basic economic parameters of the dispersion decision, a decision which bears on the optimal number of service centres. Assuming a uniform distribution of demand along an elongated strip, the greater the number of service centres, the less the distance travelled and thus the communications or transportation costs. Since each service centre has an operating cost, the optimum occurs when the additional cost of one more service centre is just equal to the savings in communications cost. Kochen and Deutsch develop a formula which is the mathematical transcription of this observation. When the assumption of a uniform distribution of demand is relaxed, the formula becomes more complex, and the optimal number of service centres is less than that in the uniform distribution case by a factor dependent on the geographical concentration of demand.

It is thus the relative weights of the following parameters which determine the optimal system configuration: fixed operating cost of a service centre, total demand, channel capacity, unit communication cost. If the direction of changes due to the variations of parameters are relatively easy to predict (an increase in operating costs of a service centre will make dispersion less desirable), the magnitude of changes in the solution, by contrast, is very dependent upon the interactions between the parameters.

Some qualitative factors are not incorporated into the model, but simply mentioned. The improvement of the response time and the reliability of the system, the interaction between the user and the service centres (which multiplies the distances) are such factors, and in particular, ceteris paribus, call for more service centres than initially envisaged.

Kochen and Deutsch's model is of wide theoretical applicability (warehouse location, logistics, transportation, communications) although its generality limits its use for practical purposes.

Streeter carries out a rather similar analysis, but with a particular emphasis towards the design decision for computer-communications systems. Notably, two features which are commonly applied to the computer field - the concepts of economies of scale and of service quality (turnaround time) - are incorporated in his analysis.

Economies of scale appear in three aspects of the computer operation:

(i) the hardware cost is subject to Grosch's law : computing power increases as the square of computer cost;

(ii) larger supervisory software is relatively more efficient; larger, consolidated data storage avoids duplication; large computer systems can more often be better utilized.

(iii) personnel costs demonstrate very steep economies of scale.

All other things being equal, the economies of scale thus tend to favour large installations, and consequently centralization.

The service quality is enhanced by the use of one large capacity service centre as opposed to many smaller capacity service centres. This queuing-theoretic result also supports a tendency towards a unique computer installation. However, as centralization increases, the service interruption duration also increases, since a computer failure is more and more synonymous with the total system failure. Turnaround time and service interruption time are thus very differently affected by the degree of centralization.

Both Kochen and Deutsch's and Streeter's approach aim at finding the optimal number of service centres. The tools they developed are not powerful enough, however, to deal with the class of problems known as plant location, which require the consideration of local constraints (warehouse availability, for instance) as well as of the overall optimum.

Two different models have been built to deal with this situation. Sparks, Chodrow et al., developed a simulation framework, which uses as main inputs the user data (needs, number, distribution), the information-related data (volume and forms of information) and the activity-related data (processes in the transformation of the information). The simulation combines these data and translates them into information flows and resource requirements, according to alternative designs called "structure" (involving three different degrees of decentralization and three different degrees of specialization). The information flows and resource requirements in turn determine the communications cost and the manpower cost. The inconvenience of the simulation rests in the large number of input data required (more than 47,000 pieces of data have to be fed in!). Practically, this model has been run once, for its validation. Another shortcoming is that the model does not optimize: it simply regurgitates the input data and it is the responsibility of the designer to propose the alternatives, translate each of them into the model "structure", and see which one is best. Accordingly, although the model uses a number of interesting concepts, it is hardly recommended for practical purposes.

As explained in detail in Part III of this report, the problem of the optimal degree of dispersion for a data-bank operation can be expressed as a fixed-charge transportation problem (also known in the operations research literature as the warehouse location problem). The improvement of the analysis due to the introduction of mathematical programming techniques consists in the possible addition of constraints to the initial simple maximization or minimization problem. An excellent illustration of this is a recent paper by Casey, in which he presents a fixed-charge transportation problem for the optimal allocation of files in a network. This model incorporates the transmission of both the update traffic and the query traffic. These tend to have opposite effects on the optimal file assignment: much query activity favors the closeness of the file to the user (hence dispersion), whereas the existence of update activity favors the centralization of the files. An essential parameter is thus the update/query traffic ratio. This is quite true in a network where each remote user can be authorized to update the files, or in certain applications such as airline reservation systems or national library catalogues. It is an unnecessary complication when updates originate from just one centre. In this last case, the update transmission cost can be lumped into the fixed charge associated with each satellite. We then revert to the above fixed-charge transportation problem.

With Casey's model, we come very close to a solution for the optimal degree of dispersion. We find, however, that the fixed-charge transportation problem is simpler and more appropriate to the operations of a data-bank similar to the FRI. One seeks to find the number of copies of a data-base (and associated programs), which will minimize the overall system cost, subject to local servicing and capacity requirements.

Mathematically :

$$\min \quad \sum_j \sum_k c_{jk}(\lambda_{jk}) + \sum \delta_k \sigma_k \qquad (1)$$

$$\text{subject to} \quad \sum_k \lambda_{jk} = a_j, \quad \text{all } j$$

$$\sum_j \lambda_{jk} - b_k \delta_k \leq 0, \quad \text{all } i \text{ .}$$

where λ_{jk} = the communication volume between region j and service centre k in bits.

$c_{jk}(\lambda_{jk})$ = the communications cost associated with the transmission of λ_{jk} .

σ_k = the fixed cost associated with the maintenance of the data-base copy on the host computer k.

δ_k = a binary 0-1 variable taking the value 1 if there is a copy of the data-base in host computer k.

The direct application of this formulation to a situation involving the use of the digital data communication network is not possible, however, because the pricing of Dataroute and Infodat is based on the concept of private line, e.g. the user pays a monthly charge for his line regardless of the usage which is made of it. c_{jk} is thus unrelated to λ_{jk}. Since the communication cost is a function of the line capacity, which in turn is some function of the number of end users, an alternative formulation expresses the communication cost as a function of the number of required channels. In expression (1) above, λ_{jk} will no longer represent the communication volume, but the number of end users in region j who are serviced by the installation k, and $c_{jk}(\lambda_{jk})$ is the associated communication cost.

4.2 The Proposed Parametric Cost Model

In view of the computational burden for a large-size mathematical mixed-integer program, a slightly different mathematical tool was developed, which fully exploits the nature of the Dataroute pricing scheme. Establishing a communication link between region j and computing centre k (as opposed to installing a copy of the data-base in a computing centre in region j itself) is a yes-no decision in terms of cost. Accordingly, the file allocation can be expressed as a simple integer programming problem, which is to find the least-cost combination of communication links and service centres, subject to service constraints. The basic model thus is:

$$\min \sum_{jk}\sum a_{jk} x_{jk}$$

subject to: (i) $\sum_{k} x_{jk} = 1$ for all j.

(ii) $x_{jk} \leq x_{jj}$ for all $k \neq j$ and all j

(iii) $x_{jk} = 0$ or 1

a_{jk}: when j is not equal to k, it is the cost of a communication link between service centre k and region j; when j equals k, it is the fixed cost of maintaining a copy of the data-base at the computing centre k. The a_{jk}'s can be visualized as the elements of a matrix; the diagonal elements, the a_{jj}'s, are the data-base maintenance cost, whereas the rest of the matrix elements are the service centre-to-user region communications cost.

x_{jk}: an integer variable which can take values 0 or 1, depending on the existence of a facility of cost a_{jk} (communication link, or data-base copy).

The first type of constraints shows that each user region j must be serviced by one (and only one) service centre k. The second type of constraint is logical: a communication link from service centre k to region j necessarily implies the existence of a satellite data-bank in location k.

a_{jk}, the cost of a given facility, is determined in advance, since we are considering service requirements of the type: twenty users in the Toronto area are to be connected to the data-bank. The alternatives are connection to one of the n-1 candidate locations for a service centre or installation of a copy of the data files in a Toronto-based computing centre. Thus, given the user population in each area, and the envisaged Dataroute costs, the cost a_{jk} of each alternative is first computed, then fed in as data input for the solution of the mathematical programming problem above. The communications cost includes the Dataroute line and end-of-line equipment cost; local lines, modems and terminals costs are not relevant to the analysis. The data-base copy maintenance cost (or operating cost of a staellite) includes the storage cost of the copy, the computer cost for updating the copy, the update transmission cost, and the annualized set-up cost for the initial duplication onto the host computer storage.

Since the costs of the alternatives are worked out in advance, a few assumptions are needed. Notably, it is assumed that the required line capacity (resulting from the number of end users) is given for each user region and all users must be serviced. It is also assumed that the tradeoff is between storage cost and communications cost; in other words, the usage cost is the same across all candidate computing centres, and excess capacity exists everywhere to accommodate the extra load (this eliminates the interference of differentials in the optimal location decision). Another critical assumption bears on the updates: these are all received by the satellite computing centres directly from the acquisition office and no relaying is allowed (otherwise the update transmission path and cost would depend on the network design, which it has to precede).

This mathematical programming model possesses some interesting properties, some of them are being discussed in section 3.4 of Part III. An illustrative example is given, together with the computer output, in the appendices to Part III.

4.3 Sensitivity Analyses

The proposed model is a constrained optimization problem (integer programming). The solution of the problem is through algorithmic procedures, some of which exist as computer codes. The computer routine used was "MPSX" (mathematical programming package developed by IBM). In particular, MPSX provides a parametric programming option which permitted the following sensitivity analyses, by modifying the relevant cost coefficients a_{jj} in the objective function:

$$\min \sum_{jk}\sum a_{jk} x_{jk}$$

Table 1 is a summary of the results. Because of the very schematic problem which is addressed and because of the very strict assumptions, these results must not be taken at face value. What was intended with this model was to indicate tendencies and the economic desirability of some configurations. It must also be kept in mind that this model singles out a portion only of the hardware costs of a data-bank (e.g. computer storage and update cost, and communications cost) without taking into account the other costs.[1] Notably, when we refer to total cost, it is the sum of the cost of maintaining the data-bases[2] and the cost of communications.

---

[1] Computer retrieval cost, acquisition and validation cost, local lines and terminals, development cost.

[2] The (incremental) cost of a satellite.

TABLE 1: The results of the sensitivity analyses

Operating costs of a satellite		Case 1 15 channels per area Total channels: 180	Case 2 30 channels per area Total channels: 360	Case 3 Non-unif. dist. Total channels: 90	Case 4 Non-unif.dist. (x8) Total channels: 720
$10,000	Total costs:	$35,040	$46,228	$19,437	$54,691
	# of copies:	1	1	1	1
. . .					
$ 7,000	Total costs:	N/A	43,230	N/A	59,093
	# of copies:	N/A	1	N/A	2
$ 6,000	Total costs:	N/A	42,095	N/A	46,286
	# of copies:	N/A	2	N/A	4
$ 5,000	Total costs:	N/A	40,095	14,437	42,286
	# of copies:	N/A	2	1	4
$ 4,000	Total costs:	N/A	38,095	13,437	37,878
	# of copies:	N/A	2	1	5
$ 3,000	Total costs:	N/A	35,003	11,826	31,655
	# of copies:	N/A	5	2	8
$ 2,000	Total costs:	N/A	24,000	9,826	22,900
	# of copies:	N/A	12	2	10
$ 1,800	Total costs:	18,168	N/A	N/A	N/A
	# of copies:	4	N/A	N/A	N/A
$ 1,600	Total costs:	17,216	N/A	N/A	N/A
	# of copies:	6	N/A	N/A	N/A
$ 1,400	Total costs:	15,840	N/A	N/A	N/A
	# of copies:	7	N/A	N/A	N/A
$ 1,200	Total costs:	14,288	N/A	N/A	N/A
	# of copies:	9	N/A	N/A	N/A
$ 500	Total costs:	6,000	N/A	N/A	N/A
	# of copies:	12	N/A	N/A	N/A

Twelve user areas have been considered (slightly less than the number of Dataroute Service Areas) with four assumptions about user population: two uniform user distributions with respectively 15 and 30 required channels per area; two non-uniform user distributions following the city population density.[1]

The study can be visualized as finding how much can be saved in communications costs by installing copies of the original data-base (and incurring the additional operating costs). The monthly operating cost of a satellite is decreased from a high $10,000, as a response to the net changes in various parameters (storage volume, storage cost, update volume . . .), individually or simultaneously. It must be noted that this cost would be between $4,000 and $5,000 for a data-bank similar in size and operation to the FRI.

The first observation is that a greater total user population leads to data-base duplication much earlier (at higher levels of operating costs), as expected. A second observation is that the "total" cost seems to level off as the total user population increases. A third observation proceeds from the comparison between the uniform and the non-uniform distributions. The "total" cost for a non-uniform distribution can be lower than the "total" cost for a uniform distribution of a much smaller number of users, and dispersion occurs much earlier for an uneven distribution, although there are less satellites for an uneven distribution at the same level of total user population. This can be explained

---

[1] The first non-uniform distribution assumes two data-bank users per 100,000 inhabitants, the second distribution, eight times as many.

as follows: as the operating cost of a satellite decreases, it becomes attractive to save on the communications cost of serving the most populated areas by installing a copy on one of the areas' computers, and thus one would expect a progressive cream-skimming of these populated areas. In the case of a uniform distribution, on the contrary, the dispersion would not be as gradual. A last conclusion can be made on this dispersion process: once a threshold is attained by the decreasing operating costs of a satellite, a large number of copies becomes warranted. In other words, if the storage cost (which ordinarily is the largest cost component in the maintenance of a copy) decreases sufficiently, there will be advantages in duplicating the data-base in every significant user area, without carrying out lengthy tradeoff analyses.

4.4 Scope and Limitations

The model presented aims at answering a very precise problem: what is the impact of variations in such factors as storage volume and cost, update volume and cost, user population and distribution, communications cost, upon the consolidation or on the duplication and dispersion of a data-bank's files? In particular, it sheds some light on the fact that beyond a certain point, it becomes economical to duplicate a widely accessed data-base, or at least a part of it. The merit of this model, besides its simplicity, is that it allows for a thorough sensitivity analysis, as shown in the previous section, and it can constitute one building block in a simulation model in which some relationships that have not been included in this study (user population vs. required channels, total vs. portion of the data-base, storage and update methods; batch vs. on-line enquiry) could be more deeply investigated.

The limited scope of this model is responsible for its shortcomings:

(i) the whole formulation is based upon the current pricing structure of the two communications carriers in Canada, and particularly on the concept of private line. Should the pricing scheme turn to a switched pay-as-you-use offering, due to the development of digital data communications, the approach would have to revert to the fixed-charge transportation problem.

(ii) the model restricts itself to the consideration of one acquisition centre updating all the other satellites, which is typical of certain applications only or of a process of dispersion, rather than of centralization (in which satellites also have acquisition autonomy).

(iii) the model does not allow for the servicing of one area from more than one satellite (this would destroy the all-or-none type of decision assumed in the model).

(iv) although we mention above that this may constitute a possible improvement, the model as it now stands does not consider the queuing problem which arises when the notion of end user is more closely examined. The user distribution is simply specified in terms of required channels.

(v) the model assumes there is no midway between the complete duplication of the data-base on a local computer and the remote access from a terminal. In other words, either the terminal is connected to a local computer, or it is connected to a remote computer. In some cases, notably with the advent of computer networks, this is no longer an accurate representation of reality, since terminal-to-computer communications will be superseded by computer-to-computer communications, which will make feasible the data transfer with various degrees of preprocessing, from a complete data-base transfer to a data-base subset transfer or just a print-out file transfer. However, the concepts and the method underlying the model are still applicable in the context of computer-to-computer communications which also revolve around the very common storage-computation-communication tradeoff (the concept of distributed intelligence is such an example).

(iv) administrative costs arising from the control of operations have been neglected. It is possible that the incremental management time spent on smooth decentralized operations is so considerable that this fact more than offsets any other consideration.

5. Other findings

5.1 Communications Cost

Our attention being more particularly directed towards the recent introduction of the digital data communications networks, a few observations concerning our findings are made here.

Figure 1 depicts on the same graph the average cost per channel (when multiplexing is used[1]) for communication links between Montreal and Toronto, and Montreal and Vancouver for a range of required channels. It shows that :

(i) the cost per channel is not very sensitive to variations in channel capacity (going from a 110 bps channel to a 300 bps channel increases the channel cost by 25%).

(ii) the cost per channel is not very sensitive to distance (there is a 30% difference in cost between a channel from Montreal to Toronto and a channel from Montreal to Vancouver).

(iii) there exist very steep economies of scale (within some ranges), due to the fixity of the end-of-line equipment cost.

Observations (i), (ii) and (iii) stem from the fact that the end-of-line equipment cost (notably the multiplexor cost) constitutes a non-negligible portion of the communications cost, with the result that other cost factors (channel capacity, distance) almost vanish from the cost equation.

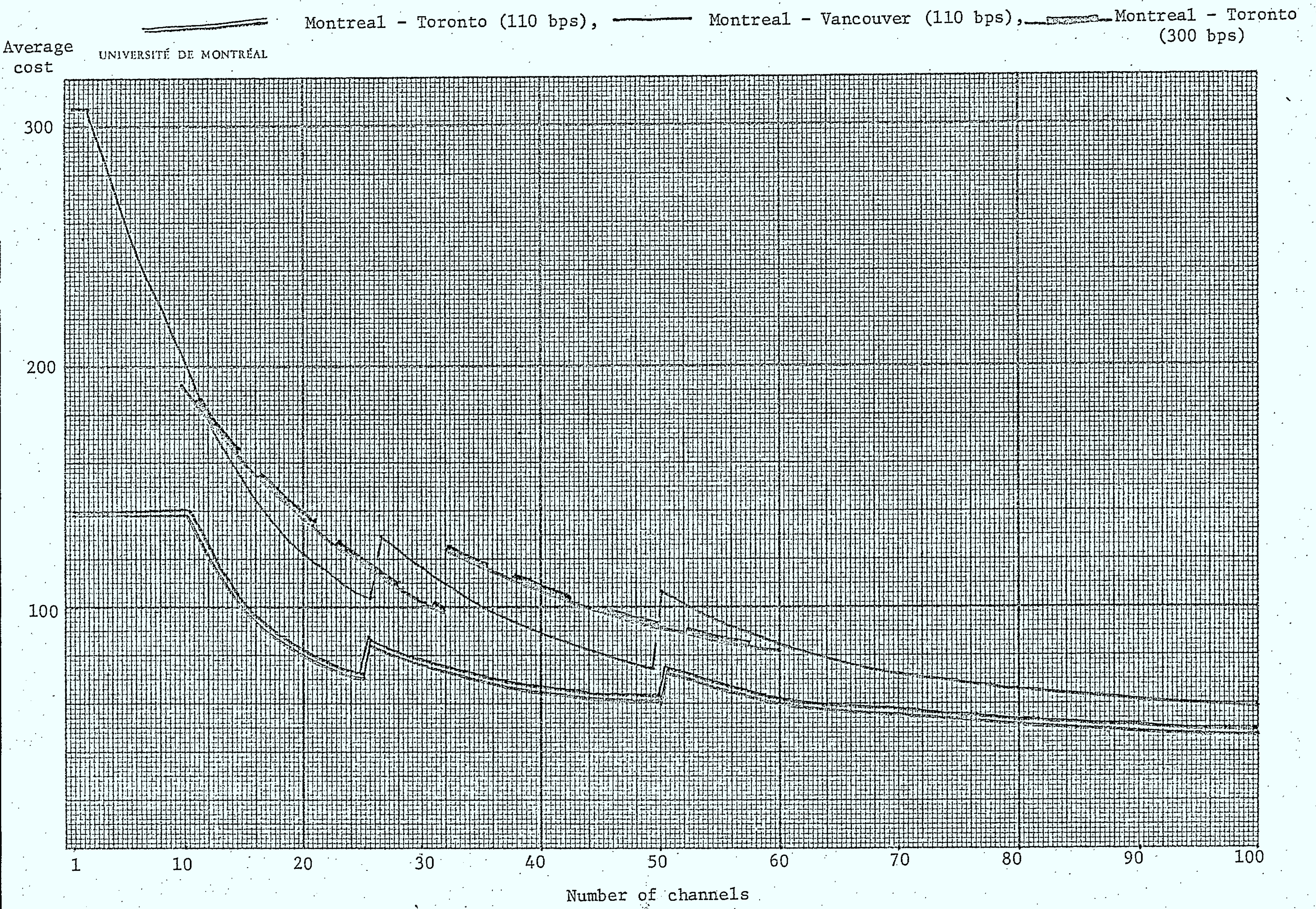
It is interesting to relate the above observations to the responses to questions in the 1969 "Study of the Relationships between Common Carriers, Computer Services Companies and their Information and Data System"[2].

---

[1] Cost per channel is defined as average cost (total cost + number of channels) including the line, access arrangement and multiplexor costs.

[2] Telecommission Study 5, Appendix to the volume. The Department of Communications.

Fig. 1: Average Cost of a Channel



The two questions were :

> Q.6: "What new telecommunications and processing services are or will be required to meet the present and anticipated needs of the computer industry and its customers?"
>
> Q.7: "In what respects and to what extent are present day transmission facilities of common carriers inadequate to meet the requirements of computer technology, including that of accuracy, speed, and bandwidth?"

As far as costs were concerned, the general feeling among users of data-processing services was that teleprocessing services were too costly, and the comments suggested that charges should be made more in relation to data volume and mileage than to channel time usage (in case of switched lines). In the case of leased, private lines, these often exceeded the usage time requirements. The user comments regarding the improvement of the telecommunication services and equipment (channel speed, error rate, etc.) are now almost matched by the technical achievements realized by Dataroute and Infodat.

Another category of respondents was computer services organization which, besides levelling the same criticisms as the users, pointed to "an artificial pricing structure". The example is given of the multiplexor charge which provides no incentive for the multiplexing of channels.

The negative comments thus focus on two points : the pricing concept (leased private line; switched line; packet switching) on the one hand, and the pricing structure on the other. If we can assume that these criticisms indicate what the users' needs are, it is possible to see how Dataroute, for instance, performs against them.

Dataroute is a private line offering. Time-sharing applications and data-bank enquiry are characterized by relatively long holding times (terminal sessions) combined with a low line utilization rate[1] due to short bursts of transmission of information. These characteristics call for efficiency gains by channel sharing, e.g. circuit switching and even message-switching as opposed to private lines as a pricing concept. The case where leased lines would make sense is the possibility of sharing the communication facility, by multiplexing, in which a channel is divided into a fixed number of subchannels (although this would not take advantage of the statistical nature of data communication). As will be shown, multiplexing is to some extent encouraged.

Our observations (Fig. 2) confirm that the pricing structure of a component as critical as the multiplexor is an incentive to its use. It is time that the multiplexed channel average cost is half of the cost of the equivalent low-speed channel.

5.2 Computer operations

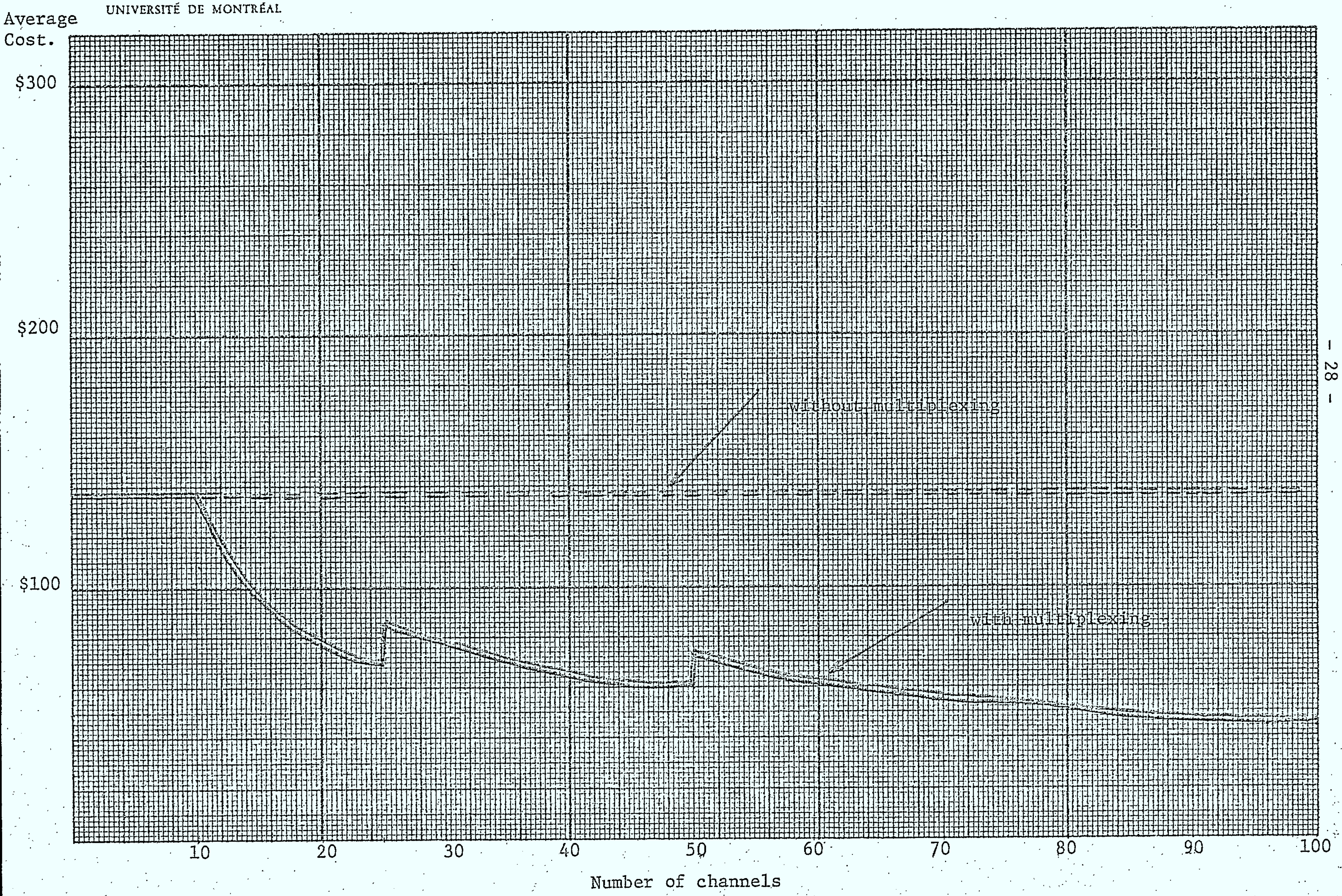
In Part IV, the computer operations of acquiring, maintaining and using a databank are analyzed. The work relates the databank as the computer sees it to the databank as the user sees it and develops analytical tools to calculate the cost to the user from the basic operating costs of the computer. The final results are total costs, expressed as a polynomial in t, the age of the databank, of acquiring, storing, maintaining and using the databank throughout its lifetime. This formulation of the cost as a function of t permits several useful transformations that can aid databank planners and analysts, and is original. It also allows growth of the files to be taken into account.

---

[1]Telecommission study 5(g) p. 42, also our study in Part II, section 2.4.2.3

Fig. 2 : Average cost of a channel

(Montreal - Toronto, 110 bps)



The concept of the usage distribution is developed in the work, and general analyses are given for sequential and direct access file organizations. While the treatment of different file organizations is necessarily incomplete, many practical databanks use combinations of sequential and direct organizations and the analyses given can be applied.

Two extensive applications of the analytical methods developed are given. An application to a simple sequential file illustrates and helps develop the concepts. The methods are shown to be viable by an application to a large existing databank with a complicated organization and growth pattern. Specific suggestions to improve the efficiency of this databank, which can be made as a result of analysis, are not presented here but reserved for future work.

5.3 User Needs and Characteristics

5.3.1 User Needs

A user needs survey was carried out with the purpose of ascertaining the relationships between the data-bank and the other sources of information for the financial community, and between the data-bank and the user's expectations.

According to the respondents, although the data-bank is "invaluable", it is the least used information source. Respondents add that they foresee an increased use of data-banks in the future. Among other things, two points are mentioned.

(i) the timeliness of data is essential, and there is much room for improvement in this field, for financial data-banks.

(ii) international data and a greater breadth of company fundamental data to be needed. Of course, these suggestions are examples of what are some of the user needs.

Numeric data banks will only compete against other sources of information insofar as they capitalize on their inherent advantages, notably size, speed and data manipulation capabilities. The development of a number of programs or a significantly sophisticated language supporting the manipulation of data is thus a definite requirement besides the data acquisition and storage function.

5.3.2 Usage Over Programs and Files

The characteristics of usage were studied, and in particular, distribution-of-usage curves were drawn for both the FRI programs and files. Fifty percent of the programs are almost never used: this has clear implications for program development in spite of what was remarked earlier (the need for more and more programs). Tight control of the program development activity has thus to be exercised.

As for the usage distribution over the files, it strongly suggests that the data-bank files could be split according to the access intensity, and that the most popular data-files could be duplicated to serve the demands more efficiently.

The sensitivity analyses carried out for a wide range of operating costs demonstrate that beyond a threshold, it becomes economically desirable to make copies of a data-base. One obvious way to achieving lower operating costs (in terms of storage cost and update transmission cost) is to make a copy of only a segment of the total data-base. The least accessed data could then be stored centrally and searched through remote batch processing.

With the development of computer networks, there has recently been much discussion about the wholesale-retail specialization of the computers in the network. If a computing centre is specialized in the financial field, and if other computing centres are interested in its files, then part of them could economically be duplicated onto the retail computing centres. We call them retail computing centres because they would presumably also store segments of other data-bases (medical, legal, bibliographic . . .) to provide users with extended services.

As pointed out among its limitations, the parametric cost model developed earlier does not permit us to make judgments concerning the allocation of files in a computer network, since it assumes that the communication links are not shared[1] by an activity other than that related to the considered data-base. However, two pieces of information - our sensitivity analyses and the distribution-of-usage curves - allow us to think in terms of a computer network, in which some computers will perform very specialized data acquisition and validation functions whereas the other computers in the network would be repositories of retail files suiting the needs of the local users.

5.3.3 Usage Patterns of Terminals and CPU

The outstanding result of a study of the data-bank user behaviour is that a very small number of users account for the use of most of the resources in communications and CPU: 25% of the data-bank users use up more than 75% of the total terminal session time, and 25% of the users (presumably the same) account for 75% of the total CPU time.

This has interesting implications for at least two aspects of the operations of satellite data-banks. For the parametric cost model, it was assumed that it was immaterial[2] which computing centre the data-bank user was connected to. In fact, the attitude of the computing

[1]Resource sharing would present considerable difficulties for cost allocation.

[2]All the computing centres providing the same advantages.

centres' management[1] very much depends on the additional revenues the data-bank users will provide, that is, on the usage. By contrast, the data-bank management sees the situation in terms of number of users, and it was emphasized that the relationship between user number and load was far from proportional.

Another conclusion is that most users (the least active) do not need a private line since they have very low communication requirements (75% of the terminal sessions are very short, less that 25 min.). However, if one turns to a dial-up access to the Dataroute line (where a number of users have access to the same line, one after the other), the waiting time for the free line may be a serious deterrent to a system of fast queries. This situation calls for techniques which enable these users to share the same channel at the same time (concentrators or message-switching).

---

[1]For such matters as storage cost, and maybe, maintenance and supervision services.

6.1 Ownership of major data banks

A second comment which has important policy implications concerns the ownership of major data banks. Should they be owned by private computer bureaux or by the Government? And if by the former, should the government regulate the activities of such institutions? Some fairly compelling arguments can be adduced as to why governments should have a strong interest in information policy and exercise some control over major informational activities. Whether these imply public ownership or merely public regulation is a question which would require further analysis.

First, let us affirm that the production of information and knowledge, and its dissemination, are important economic activities. In 1962 the American economist Machlup estimated that the "knowledge industry", which he defined as comprising such activities as education, research, publishing, broadcasting and other information services, accounted in the U.S.A. for as much as 29 per cent. of GNP.[1] Mainly based on the years 1947 to 1958, he further estimated that these activities were collectively growing at an annual rate of 10.6 per cent. Machlup's estimates were generally regarded as being somewhat inflated, especially the growth rate. There would, however, be general agreement with the more moderate view that:

---

1 F. Machlup [1]. In more detail, the activities which Machlup distinguished as making up the "knowledge industry" were education at all stages (the largest), research and development, the media of communication (books, periodicals, newspapers, broadcasting, etc.), information machines (electronic computers, automatic control systems) and information services (legal services, engineering services, accounting services, medical services; financial, insurance, banking and real estate services, stock and commodity market services, credit agencies and loan associations, employment agencies, auctioneers; postal, telegraph and telecommunication services, the judiciary, parts of the services of national security and international affairs and finance.

> "Trends in research and development expenditure, the employment structure, educational activity, and consumer expenditure patterns suggest that the share of the information and knowledge industry [in the GNP's of advanced industrial countries] may well be increasing". [1,2]

The first reason why this important and growing activity should be of especial interest to governments is that investment in the production of knowledge and information is one of the main sources of economic growth in advanced industrial societies. One American researcher believes it to be the fundamental source of economic growth (see Denison [3]). In other words, it is now thought by a number of economists that technological innovation rather than investment in real capital is what chiefly distinguishes fast-growing and slower-growing and stagnant economies, economic growth being measured as the change in the value of the aggregate output of goods and services of the economy, measured in constant prices (i.e. changes in real national income). The "information explosion" has meant that "machines and equipment are more the external evidences than the real core of a technology". [3]

It is generally recognized that as the degree of industrialization in an economy increases, the greater becomes the skill content of its output of goods and services: the production of information and knowledge leads to changes in occupational structure, as well as to changes in the form and the size of the gross national product. Much of the early work on economic growth by contrast started with concepts and estimates of labour and capital which abstracted from their productivity: labour input was measured in man-hours, undifferentiated for differences in skills and knowledge, and the capital inputs similarly. Progress has been made in identifying

---

1 Lamberton [9], p.7 .

2 Further evidence in support of this contention is provided by J. Marschak [12]

3 Simon [14], 1964, p. 94.

and measuring the information that becomes an integral part of labour input. Measurement of the productivity of non-human capital forms that comes from increased knowledge has proved less tractable.

To be useful, knowledge and information must be produced and disseminated: like other activities, informational activities have both production and distribution aspects. As Machlup pointed out,[1] governments already finance or otherwise support a large part of these processes in various ways: by conducting scientific research in their own establishments and supporting industrial research and development and academic research, indirectly through the patent system, by financing education and training, providing tax incentives to undertake research in certain defined areas, in some countries even paying individuals and firms to file certain kinds of information, such as the results of scientific research. Many of the items listed are concerned with the production of information. In view of their large financial stake in these activities, and awareness of their economic effects, it seems natural that governments should also be concerned to see that maximum benefit is obtained by widespread dissemination of this knowledge and information. As the Organisation for Economic Co-operation and Development has observed [13] :

> "As research develops and specialities multiply, and as the number of researchers grows, the reciprocal information of scientists and the transfer of discoveries raises increasingly important problems. Information policy is rapidly assuming a paramount place in modern national and international science policy. And all the more so since, as information grows and its transfer becomes more essential to the development of science, the task of diffusion itself grows more difficult. The individual on his own cannot succeed in satisfying his need for information, and government is bound to intervene in this field. Universities, libraries,

1 [11], pp. 9-10

business and administration must be able to call upon a vast network of specialized institutions with costly equipment, qualified staff and common rules. The government is the natural coordinator of this network. It alone can bear certain costs. It must compel its agencies to search for, acquire and supply the maximum of information. It must initiate rules and methods that will make information cheaper and easier to use. It is the normal correspondent of national and international organizations in the field of scientific and technical information".

The OECD report goes on to discuss the two-fold function of information policy (in addition to the encouragement of the production of information and knowledge) : documentation and diffusion, and encouraging users to utilize the transferred knowledge. The methods used for transferring knowledge and discoveries in fundamental research to the scientific community are often very close to those used for the transfer of technology results to industry. Presumably there is also a point beyond which computerized data banks become more efficient than conventional libraries.

The <u>efficiency</u> of various sectors or activities in the economy interacts with the growth rate of the economy, and this leads to a second argument. Even in an economy not especially wedded to an objective of rapid economic growth it would still be desirable that all activities should be carried out as efficiently as possible : the economic welfare of the community depends on this (and on the existing distribution of wealth)[1], and the government of most free societies have the raising of economic welfare as a foremost (or at least high) objective. A particular distribution of resources in the economy is said to be economically efficient if and only if it is not possible to make someone better off without making anybody else worse off - assuming there are no externalities (i.e. that one individual's welfare is unaffected by the welfare of others, and that the costs and benefits of an action all fall to the decision maker).

---

[1] I.e. economic efficiency is a necessary, but not a sufficient, condition for maximizing the economic welfare of the community.

Any configuration of the economy other than this economically efficient (or Pareto-optimal) one is undesirable on welfare grounds. If there are no externalities present it is no longer necessarily true that we can consider a configuration better because one individual has more goods and no others have less.

The desirability on efficiency grounds of optimizing the production and distribution of knowledge and information in an economy is fairly obvious. Thus educational background is a crucial determinant of the quality of labour, conditioning both the types of work an individual is able to do and his efficiency in doing them. The output that can be obtained from a given amount of labour, capital and land is enlarged .ith each advance in knowledge of how to produce and distribute goods and services at lower cost. Such advances may occur either in the field of technology or in that of management and business organization.[1]

Here we refer not only to the application of new knowledge but to achieving the maximum use of existing knowledge. The best practice possible with the knowledge available at any given time differs, often quite substantially, from the average practice actually in use. Consequently a contribution towards greater efficiency and economic growth may be made by advances in knowledge or by reducing the lag between average and best current practice.

It remains to be shown why this desirable state of affairs - increased efficiency - is likely to be achieved more fully if the government is strongly represented at the distribution, as at the production, end of the "knowledge industry".

---

[1] In a pioneering article in 1945, Hayek [6] emphasized the importance of "knowledge of particular circumstances of time and place" (information of particular applicability) as opposed to "scientific knowledge" (information of general applicability).

(i) First, the government is already the repository of a large amount of information in areas such as science, technology, economic surveys and statistics, to name a few. Under the National Library Act, 1968-69, copies of all "books" published in Canada must be deposited, "book" here including any "document, paper, record, tape or other thing published by a publisher, on or in which information is written, recorded, stored or reproduced". The provinces have similar requirements.

(ii) Reverting to the economic efficiency argument,[1] and still assuming there are no externalities, for optimal resource allocation within the economy it must be impossible to make someone better off and none worse off by :

(a) taking the goods actually produced and allocating them differently among consumers;

(b) taking the existing resources and allocating them differently among producers, so as to produce more of all goods;

(c) changing the composition of total output, i.e. the relative amounts of different goods, or of goods and leisure.

Paragraph (b) requires that production be technically efficient (i.e. that no reallocation of resources between producers will increase the total output of one product without decreasing the total output of others). Because increasing returns to scale are likely in the activities at the

---

1 See [1] and [2].

distribution end of the information industry (computing, communications), it is possible that any single private firm would operate at too small a size to take advantage of large-scale processes : a larger output of the service could be obtained, with the same resources, by operating with large-scale methods. Many kinds of information are produced inefficiently at the present time.

(iii) For distribution of information services to be efficient, all consumers must be able to exercise free choice, subject to any budget constraints, at prices which are uniform for all. A private firm might practice price discrimination - between different classes of users, large and small users, temporally or spatially.

(iv) In determining the price at which information is supplied to users the government is more likely to act in a manner consistent with efficiency considerations than private firms would be.[1]

---

[1] The argument here is by no means straightforward, and has to be qualified in several ways. Provided the allocation of resources among producers and the allocation of goods/services among consumers are both efficient, the optimal outputs of information will be those that equate marginal cost to price. In the absence of divisibilities, increasing returns to scale, of resource limitations to individual producers, and of corner solutions, this condition would be realized in a perfectly competitive economy. The existence of many imperfections in the real world means that we are faced with a 'second best' situation and marginal cost pricing is no longer the appropriate rule. All that can be said is that to the extent that imperfections (e.g. monopolistic elements, taxation) lead to prices greater than marginal cost in the rest of the economy, the 'second best' price may be $\gtrless$ MC : on the assumption that market imperfections elsewhere in the economy raise rather than lower this 'second best' price, the theoretical rule becomes price $\geq$ MC. For further details see Lipsey and Lancaster [10] and Farrell [4].

(v) There are strong grounds for believing that the production of knowledge yields social benefits in excess of the private benefits accruing to the recipients of the knowledge.[1] Some of these "third-party" benefits (externalities) would probably be lost if (in some cases) the production of information and (in others) the dissemination of information were left in private hands.[2]

(vi) Search and advertising are complementary informational processes. In terms of exchange transactions, the searcher locates specific offers; the advertiser "pushes" the fact of his existence and possibly some details about his terms for dealing. Regarded as an information-transfer process, resources are wasted in advertising which conveys unauthentic information. Greater government control over the dissemination of information might prevent some of this social waste.[3]

(vii) Expenditure by users in acquiring at least some kinds of information is one of the first items of expenditure likely to be cut in time of slump, making the private operation of such services hazardous in the long run. This makes the public operation of such services a natural candidate for inclusion in government counter-cyclical economic policy.

---

1 See Machlup [11], pp. 115-7, for a discussion of the externalities arising from educational expenditure.

2 Some would argue that externalities are, in fact, of predominant importance. For example, Galbraith [5], who speaks of "private wealth and public squalor", believes that much that is wrong with the United States is due to having neglected them.

3 See Hirshleifer [7].

If these very briefly stated arguments in favour of a high degree of government control over major information services were to be elaborated, besides putting the welfare arguments more rigorously, a distinction would have to be drawn between knowledge (data that can be put to general use) and information (data which is useful for some particular purpose), and between private and public information. Information is of value only if it can affect action. It would be socially valueless to inform everyone in the community that a particular state of affairs would obtain with certainty in 1984.[1] All our remarks in this section are to be interpreted in this light. Also certain 'sensitive' information of which the government is custodian must be kept private in the public interest.

---

[1] See Hirshleifer [8].

REFERENCES

[1] Arrow, K.J., "Economic welfare and the allocation of resources to invention", in The Rate and Direction of Inventive Activity : Economic and Social Factors, Universities-National Bureau of Economic Research series, Princeton, 1962.

[2] Demsetz, H., "Information and efficiency : another viewpoint", Journal of Law and Economics, 12, April 1969, pp. 1-22.

[3] Denison, E.F., Why Growth Rates Differ : Postwar Experience in Nine Western Countries, Brookings Institution, 1967.

[4] Farrell, M.J., "In defence of public-utility price theory", Oxford Economic Papers, New Series, 10 (1958), pp. 109-123.

[5] Galbraith, J.K., The Affluent Society, Houghton Mifflin, 1958.

[6] Hayek, F.A., "The use of knowledge in society", American Economic Review, 35, Sept. 1945, pp. 519-530.

[7] Hirshleifer, J., "Where are we in the theory of information?", American Economic Review, 61, 1971, pp. 31-39.

[8] Hirshleifer, J., "The private and social value of information and the reward to inventive activity", American Economic Review, 61, 1971, pp. 561-574.

[9] Lamberton, D.M., ed., Economics of Information and Knowledge, Penguin Modern Economics Readings, 1971.

[10] Lipsey, R.G. and Lancaster, K., "The general theory of second best", Review of Economic Studies, 24, Dec. 1956, pp. 11-32.

[11] Machlup, F., The Production and Distribution of Knowledge, Princeton University Press, 1962.

[12] Marschak, J., "Economics of inquiring, communicating, deciding", American Economic Review, Proceedings, 58, May 1968, pp. 1-18.

[13] Organisation for Economic Co-operation and Development (OECD), Reviews of National Science Policy : United States, 1968, ch. 11, pp. 225-241.

[14] Simon, H.A., "Decision making as an economic resource", in Seltzer, L.H.,ed., New Horizons for Economic Progress, Wayne State University Press, 1964.

PART II: COSTS OF A DATA BANK.

CHAPTER 1

THE FRI: GENERAL DESCRIPTION

INTRODUCTION

Data-banks already are a pervasive institution in our computer age, and will become more and more so in the future. The computerized information services made possible the dream of the world's entire store of knowledge accessible to everyone, with a corresponding upsetting of the traditional educational methods; electronic libraries can reduce the considerable wastage of time due to bibliographic searches and duplication of research; the banking networks may realize the cashless society. There are no limits to the developments of the data-bank applications, only those that public policy imposes.

We feel very strongly that although the concept of data-bank already has emerged, in practice at least, it is too interdisciplinary by essence to be simply categorized as an offspring of the computer, as it may be older than the computer itself. Notably, the documentation science has yet to contribute a lot to the development of data-banks. The communication technology also has a synergistic effect when associated with the computer technology, and here, new problems arise for the computer-communication utility: in particular, the degree of centralization and consolidation of data-banks which is economically and socially desirable is greatly affected by the telecommunications industry offerings (in terms of technology and rates).

These remarks suggest some avenues of research in the economics of data-banks (as opposed to an engineering viewpoint): what is the cost of a computer system and how to price its services? What is the cost of processed information and how should it be spread among the users? Given the cost of communications, what is the optimal degree of dispersion and how does it vary with changes in rates and usage load?

These are typically the questions we address to in this paper, with particular reference to the experience of the Financial Research Institute in Montreal, a financial data-bank, and to the present state of the Canadian digital data network. The first two chapters are essentially descriptive: Chapter One is a general presentation of the Financial Research Institute; Chapter Two goes a little more into the detail of its operations; Chapter Three contains a discussion of the conceptual elements of the costs of a data-bank: computer and communications system costs, and information storage and retrieval system costs.

1.1 Historical Development[1]

The idea of creating a financial data-bank in Montreal appeared in February 1968. Professor Donald Armstrong, from the Faculty of Management at McGill was the originator of the project. He launched the idea in a speech delivered to the business community in Montreal: it stemmed from a growing frustration among the financial community about the lack of a financial data-bank. After the speech, high-ranking officers from the Montreal Trust, Royal Trust and Nesbitt Thomson showed interest in the proposal and they were joined later by Canadian National, Canadian Pacific, The Bank of Montreal, The Royal Bank, Burns Brothers and Denton, Bell Canada and Sun Life. Together with McGill University, these organizations set up the Financial Research Institute, with a goal of developing a data-bank for the usage of its members and the application of financial techniques which needed to be computerized. The Financial Research Institute was set up as a non-profit organization (section 2, of Companies Act) in August 1968, with ten members.

The start-up of operations was made with the Compustat tapes,[2] donated to McGill University by Standard and Poor's in 1967.

---

[1]This section and section 1.4 have their source in private communications with Professor Donald Armstrong and Mr. Richard Hamilton, presently Chief Executive Officer at the FRI.

[2]Financial data on U.S. and Canadian companies, compiled by a service of Standard and Poor's and made available on magnetic tapes.

Until the beginning of FRI operations, the tapes were not utilized because of the lack of an access program. One programmer was hired, then a second, to write the programs for accessing the data (the very first program, PANCA, ranked and screened the companies).

For some time, the FRI had not even an office. Initially, the data-bank was not available on-line through time-sharing and terminals. Users used to fill in forms with their request; the requests were then batched and processed, and the users obtained a report after a few hours. Although this system was unsatisfactory, it lasted almost one year, because it was cheaper[1] for the members than to go to the U.S.A., where the source of data was.

The next move was to develop a sizeable Canadian data-base. An annual fee paid to Standard and Poor's enabled the FRI to get the special service of Canadian company data. The stock price data were picked up by bits and pieces from the Montreal, the Toronto, the New York and the American Stock Exchanges, which all had the daily prices recorded on magnetic tapes. The last part of the data-base, the economic data, was to be obtained from Statistics Canada, after long negotiations: Statistics Canada had never entered this kind of contract, and the FRI was their first customer.

---

[1]At that time, the annual subscription fee for the Compustat tapes was around $30,000, while the membership fee for FRI was $45,000 per three year period, or $15,000 on an annual basis.

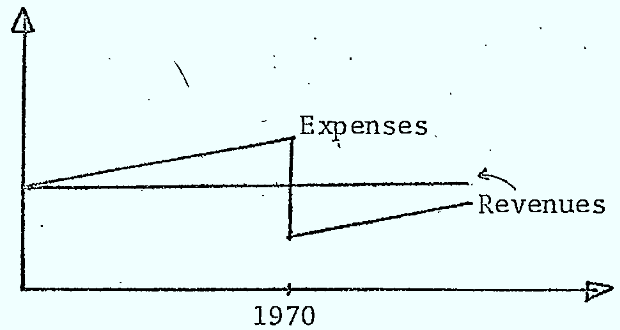
When in 1969, the FRI changed operations from batch processing to on-line time-sharing processing, it benefited from the operating system (RAX) of the McGill Computer, which was one of the best that could be found in Canada at that time, according to the words of the then Chief Executive Officer, Professor Armstrong.

However, the FRI was a more successful venture in sales than in production. The main reason was that the FRI services were a highly saleable product, but a product so vast that it did not always satisfy the users' wants. A measure of FRI's success can be found in the growth in membership: from ten, it grew to forty within two years.

The FRI also became attractive for its main competitors. The FRI was unique in its kind, and offered an ever-growing range of more and more sophisticated services. The FRI sold more of Standard and Poor's material in Canada than Standard and Poor's itself did in the U.S.A. An offer from McGraw-Hill-Standard and Poor's to buy the FRI and set up a worldwide network of similar data-banks with FRI's present know-how and personnel, was turned down after long negotiations. Instead, the FRI members agreed to put down the funds the FRI needed for its expansion.

What was a booming environment suddenly became gloomy in 1970 when the stock market collapsed, and some stock brokers disappeared. The market for the FRI services fell, and within a few months, "the business evaporated," according to Professor Armstrong. Indeed, this is an illustration of the volatility of the information service industry prospects. Not only are the particular user needs a very vague thing, but the aggregate demand is highly unstable, since in a recession period, the first expenses to be curtailed are those related to services, especially information.

The FRI revenues are stable: they vary only with the number of members. On the other hand, during the development period 1968-1971, expenses had steadily increased, exceeding revenues. The members were certainly getting a better service, at the cost of a deficit from the FRI, which they had to bridge. The FRI needed a financial cure: Professor Armstrong decided to leave his charge of president in order to facilitate a change in policy. When Richard Hamilton, the new Chief Executive Officer, took over in March 1970, he had the task to cut expenses below revenues, which he did accomplish: these drastic measures essentially consisted in reducing the staff for program development.



This experience is consistent with Bower's remark[1] that operating deficit is usually quite high in the first years of operation of a

---

[1] Richard S. Bower, "The computer service industry," Bell Journal of Economics, Autumn 1973.

computer service bureau and then disapperas, due to early expenditures made in order to get revenues later.

It should be noted that the services rendered in the early years of FRI could be classified as "research and development" whereas much emphasis is now placed on regular routine information services, e.g. portfolio evaluation. Present structures are therefore less subject to recession.

Another innovation made in 1970 was the unbundling of services. Partial membership was now offered, and the fee matched better the service and users' requirements.

1.2 Organizational Structure

The FRI is a very light, flexible, small organization, of around ten to twelve people: the director and his assistant, programmers and secretarial staff. A branch office in Toronto (two people) essentially provides assistance to local users.

The function of the organization consist into:

(i) design and writing of application programs for data maintenance and editing; initiation and coordination of industry applications.

(ii) data updating activity: new data entry, error checking.

(iii) program documentation; file directories.

(iv) user assistance: support of members' use, training.

There are two main influencers in the management of the FRI: the members and the McGill Computing Centre. The members sit on the Board of Directors and elect the Chief Executive Officer of the FRI. A particular user will often want the FRI to write a special program for him (e.g. portfolio managers complaining about their computer bills and wanting the portfolio program to be converted for the stocks they are interested in) and of course, the FRI will have to deal with that particular request.

The development of new programs is then made on a trial and error basis: at each stage, it is checked what the users are doing with them; if they use them, they are taken to the next stage. There is a delicate balance to be maintained between what will be written by the FRI (resource availability is limited because of the small size of the organization) and what will be passed on to the user, in terms of his own time or his computer bill.

The second party is the McGill Computing Centre. The small size of the FRI is due to this association with the Computing Centre, which performs all hardware-related operations; the FRI just buys the computer time required for the file update and the program debugging, as any other regular user, plus the disk storage space. Specialization also is a characteristic; the FRI acts as a middleman between his prospective data-bank users and the Computing Centre[1], although the relation between the end user and the Centre does not go through the FRI there is a special contract for this arrangement.

The bargaining power of the FRI vis a vis the Computing Centre (for such matters as rates and services) much depends on the number of active users the FRI brings to share the computer system cost, and on the possibility of alternative arrangements with other time-sharing vendors.

---

[1]Same thing between the telecommunication network and the end user for the connection of remote customers to the computer.

[2]McGill bills the customer. Assistance to the customer is the responsibility of FRI.

1.3 Membership

Present users[1] of the FRI files divide into commercial users and university users. Prospective commercial users are all the members of the financial community: banks and trusts, mutual funds, brokerage house, investment counsels, insurance companies, exchanges. University users include the four Montreal universities plus six others.

It is not sure that university users are more sophisticated than commercial users. Although a considerable amount of data exists on on-line storage (a differential advantage upon the Compustat tapes), it does not seem that it induced a proportional amount of research (at least not comparable to that engendered in the University of Chicago's Centre for Research on Security Prices). Each year a number of Master's research papers are written on financial subjects, but only a fraction actually refers to empirical research carried on the FRI data-bank.

Membership can be obtained by payment of a fixed fee proportionate to the extent of service desired: a segment of the data bank, or the whole file.

---

[1] See Appendix 1 for a complete list.

The annual membership fees are shown below:

Full membership	$15,000
Partial membership	
Financial Analysis*	9,000
Economic Analysis	2,800
Portfolio Management	5,800
Chartered Banks	2,500

Beside this, each member has an arrangement with the McGill Computing Centre by which he pays for the use of the system (basically, he is charged for the computer time, the connect time and the storage, if any; we will treat this specific topic in the next chapter).

As far as computer science is concerned, about eighty percent of the users of the FRI data-bank are very unsophisticated. Financial analysts in specialized departments of their respective company, they do not want to learn about computers or programming. However, as they get more involved, they can be assisted by a programmer to explain the data more fully. Until this moment though, one of FRI's goals is to design its application programs so that they are not imposing on the user the need to learn about computers. We already attended to this tradeoff earlier and we shall deal with it more extensively in Chapter 3.

*including the Chartered Banks.

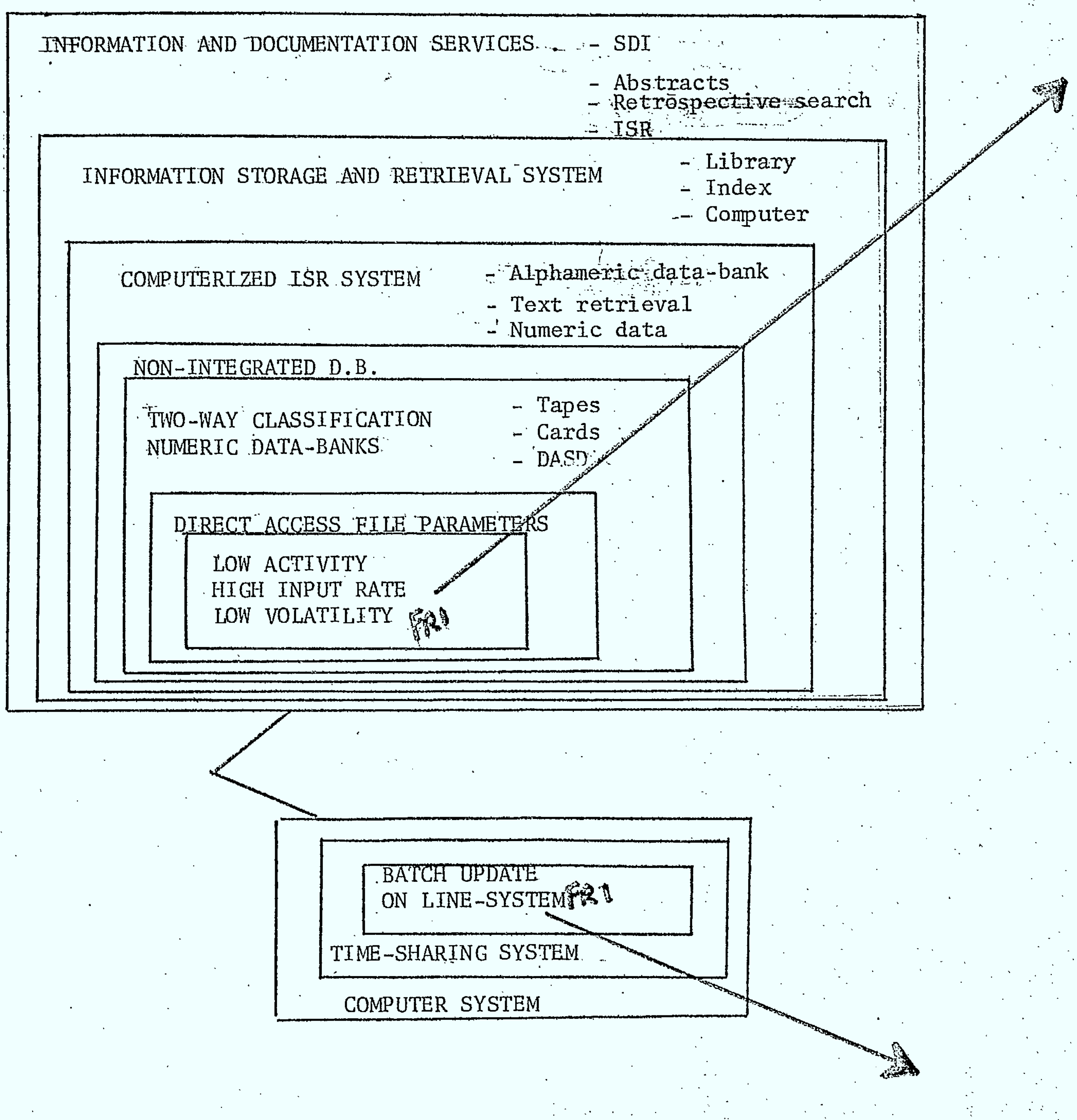
1.4 Future Plans

The plans for expansion head into three main directions: number of users, new data-bases, and availability of service across the country. Partial membership, making entry less costly, continue to attract four to five new members each year. A major project is now under way: the bond market data bank (by data-bank is meant both the data base and the access and manipulation programs). Due to the size of this data-base, the FRI is considering setting it up on another computer system. At the user end, the huge lists of data will require high speed printers instead of the usual terminals. This project may require as much of manpower resources as what has been done by the FRI so far, in terms of development.

Another objective of FRI is to make its services available across the country. Lately for instance, it was decided to install a Dataroute line between Vancouver and Montreal to serve a new member in Vancouver. This was made possible through the dramatic decrease in communication costs.

1.5 Relation to Other Similar Data-Banks

A data-bank today develops along two dimensions: one is the documentation science dimension, the other is the computer science dimension. Any data-bank can be described in terms of these two coordinates, and the FRI fit at the intersection of these fields according to the graph below:



Having located the FRI data-bank among the whole spectrum of information-related systems, and thus shown how specific it is, let us now compare it with other financial data-banks.

Perhaps the most famous of them is Investors Management Sciences, Compustat tapes. Investors Management Sciences is a subsidiary of Standard and Poor's and thus gets first-hand data. The Compustat tape[1] contains 20 years of annual data for 1,800 listed industrial companies, 850 over-the-counter companies, 154 utilities and 500 Canadian companies. Quarterly data are also available for the past 10 years. There also exists a Bank Compustat covering 110 banks. The tapes are regularly updated and available directly from IMS for a subscription fee or through various time-sharing systems. However, the distinction must be clear between the data-bases (= the files) maintained (e.g. formatted, updated, checked) by IMS or Data Resources Inc., another firm specialized in economic data, and the processing programs to access, manipulate and display information from these data bases, which are developed by other firms (most of the time, time-sharing vendors, like Interactive Data Corp.).

A data-bank is a combination of the data-bases and the processing programs. Such a data-bank is for instance, provided by Interactive Data Corporation which makes available:

---

[1] "Investment Analysis and Portfolio Management," Cohen, Zinberg and Zeikel.

- the Securities' Price Data Base: price and volume data for all New York and American Stock Exchange common stock.
- the Corporate Financial Data Base: similar to the Compustat tape (obtained from IMS)
- the Economic Data Base (obtained from another company)
- the Financial Return Data Base: returns on common stock.
- the On-line Ticker Data Base: data on trading throughout the day.
- the Bond Data Base: daily price and volume information.

Interactive Data Corp. also provides the following product program:
- Analystics: screening, ranking and statistical functions.
- XPort: portfolio management
- IAL: investment analysis language.

IDC is typical of a time-sharing vendor which specializes in the development of programs, while buying the data-bases elsewhere. The FRI also concentrates on the programs, but is not a computer-time vendor. This has an important implication, in that there is not such an incentive on the part of the FRI to have efficient programs on one hand, and to press the members to use them, on the other hand.

A question arises at this point, following these last considerations: to what extent the FRI is representative of a data-bank? We first noticed that the FRI is not an integrated operation: there is no data collection, FRI does not own a computer. However, we mentioned that most financial data-banks get their data-base from outside: in this respect then, FRI is no different. Second, FRI is not a computer-time vendor.

To ensure comparability then, from now one, we shall consider the FRI and the McGill Computing Centre[1] as one entity.

Other features of FRI, which will be dealt with at length later, also make it somewhat unique. The FRI data-bank serves a small number of intensive users. This has an impact on both the economics of the data-bank and its management:

- effort is more oriented to providing custom-tailored service than standard, efficient programs. The optimality of file structure and programming algorithms may be sacrificed to their simplicity.
- this tendency is encouraged by members which do not only have a right of control, but a right to ask for specific services (e.g. write a program which satisfies a particular need).
- the marketing and the public relations aspect overrides the cost consciousness. The structure of revenues (membership fee) makes it more interesting for FRI to recruit a new member[2] than to improve operation efficiency.
- in a technique of production analytical framework, the system formed by the FRI and the Computing Centre would then tend to substitute more marketing to research resources, e.g. more manpower time will be dedicated to promotion and assistance than to program development (a clear implication of this is that the FRI uses more uncontrollable inputs, if we can go as far as to say that program development time is controllable).

---

[1]At least, the segment of the McGill Computing Centre activity devoted to servicing the FRI and the FRI members.

[2]FRI's unit of operation is more the number of users than any other measure. The marginal revenue from a new customer is thus the fee he pays, while the marginal cost of servicing him is almost nil.

In conclusion, at the FRI, management factors are more critical than cost factors, and the small scale of operation[1] is a limitation of the study. At the scale of FRI operations, costs will reflect more policy decisions than environmental, design and operations parameters. It is certain that on another scale, factors which are now irrelevant to FRI can take a considerable importance. Keeping in mind these inherent limitations, let us now proceed to examine FRI operations more closely.

---

[1] In terms of volume of data, and volume of usage.

APPENDIX 1

LIST OF F.R.I. MEMBERS

FINANCIAL RESEARCH INSTITUTE
suite 910,
635 dorchester boulevard west,
montreal 101, quebec
telephone: 866-9558

January, 1974

LIST OF MEMBERS BY AREA

MONTREAL

Bank of Montreal
Banque Canadienne Nationale
Bell Canada
Bunting (Alfred) & Co. Limited
Caisse de Depôt et Placement du Québec
Canadian National Railway
Canadian Pacific Limited
DuPont of Canada Limited
Greenshields, Inc.
C.J. Hodgson & Company
Levesque, Beaubien Inc.
Montreal Trust Company
Nesbitt Thomson & Company Limited
O'Brien & Williams
Royal Bank of Canada
Royal Trust Company
Sun Life Assurance Company of Canada
Walters Securities Limited

OTTAWA

Bank of Canada
Government of Canada - Dept. of Communications

TORONTO

A.E. Ames & Company Limited
agf Management Limited
Bache & Co.
Burns Bros. & Denton Limited
Canada Life Assurance Co.
Draper Dobie

TORONTO - CONT'D

Canada Trust: Huron & Erie
Canadian Imperial Bank of Commerce
Canavest House
Dominion Securities Harris
Fry, Mills Spence Limited
Independent Order of Foresters
Midland-Doherty Securities
National Trust Company
North American Life Assurance Co.
Toronto-Dominion Bank
Wood Gundy Securities Limited

WATERLOO

Mutual Life Assurance Co. of Canada

WINNIPEG

Richardson Securities of Canada

REGINA

Houston Willoughby & Co.

UNIVERSITY MEMBERS

Ecole des Hautes Etudes Commerciales
McGill University
McMaster University
University of Manitoba
University of Quebec
University of Sherbrooke
University of Toronto
University of Waterloo
University of Western Ontario

CHAPTER 2

THE FRI: DESCRIPTION OF OPERATIONS

2. FRI: Description of Present Operations

As emphasized in the previous section, a data-bank includes both the data-base and the access programs. Following this distinction, we first present the types of files, then the product programs.

2.1 Data-Bases Maintained

2.1.1 The Economic Data

This data base contains 9,200 monthly, quarterly and annual time series of economic variables, both U.S. and Canadian. These data highlight a wide variety of economic indicators, organized around the National Income and Product Accounts, the Gross National Product Components by Industry, Retail and Wholesale Trade data, Population and Labour Force, etc. . .

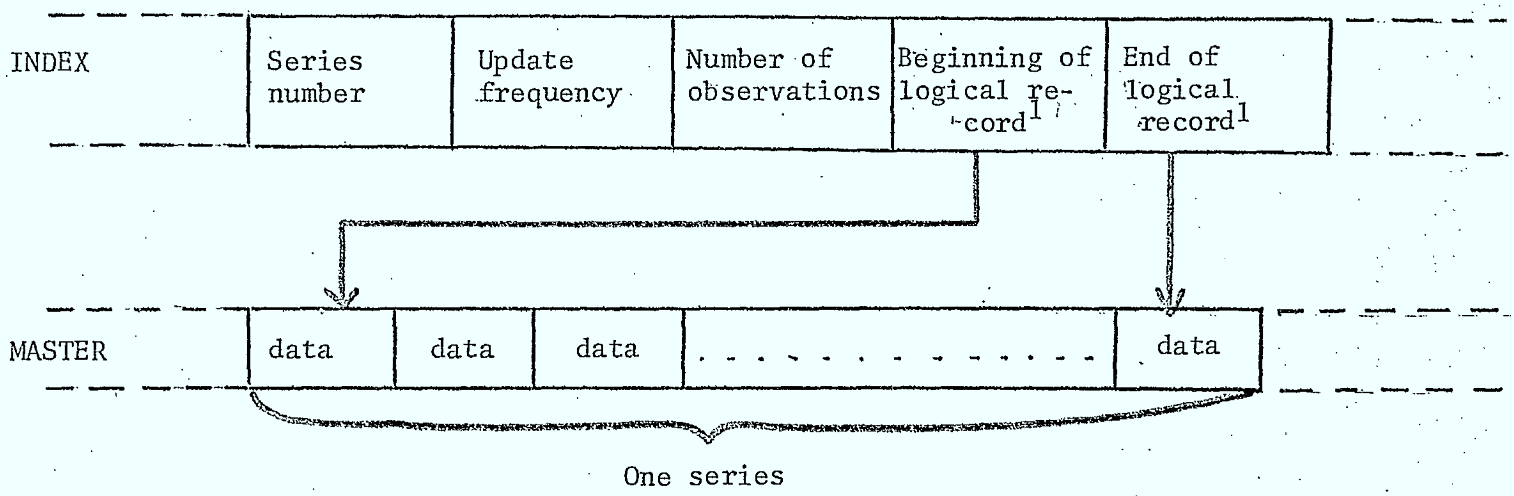
2.1.1.1 Source of Data

The data are obtained from Statistics Canada. A subscription contract entitles the FRI to receive a magnetic tape every month from Ottawa.

2.1.1.2 Data Elements, Record Layout, Data and Update Volume

Each economic series is referred to by a series number.[1] The file organization is indexed sequential (a homemade indexed sequential organization; FRI does not use any of the manufacturer-provided data-management packages). The index record layout is shown below:

[1] The user can find this series number in a series directory similar to that provided by Statistics Canada.



The 9,200 series require 5.8 million bytes. They are updated every month upon arrival of the Statistics Canada magnetic tape.[2] Moreover, it is found to be more economical to totally rewrite the whole file: no space is wasted waiting for forthcoming updates; the physical location of the records has thus to be upset every time.

The computer update time amounts to seven (7) minutes/month. The average number of observations being 360 (30 years of monthly data) per series, the update volume amounts to 0.28% of the total volume.

---

[1]Track number, word number.

[2]Which is mounted on a tape reader at the McGill Computing Centre.

2.1.2 The Financial Data

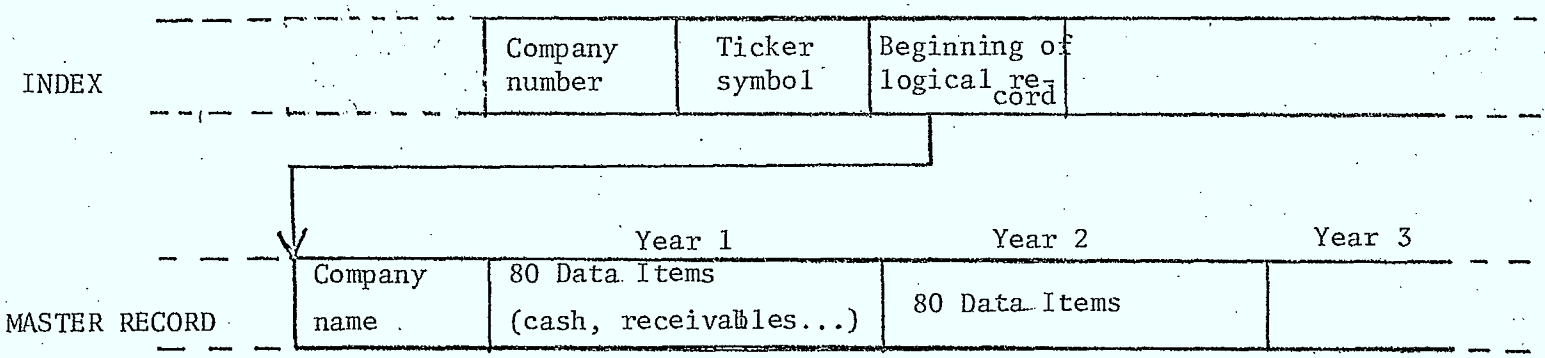
The U.S. and Canadian Financial Data files total 175,000 series.[1] There are 1,800 U.S. companies (annually and quarterly) and 150 utilities (annually) each with 60 data items of the balance sheet, income statement and market data; there are 250 Canadian companies and 80 utilities, each with 80 data items.

2.1.2.1 Source of Data

Data are acquired from the Financial Post, with which the FRI has a partnership deal for financial information. Data are entered on a tape in Toronto by the Financial Post services, and the tape is sent to the FRI in Montreal.

2.1.2.2 Data Elements, Record Layout, Data and Update Volume

The organization of the file is indexed sequential with fixed length records. A typical record is displayed below:



[1]So as to be consistent throughout, a series is defined as the time series: there will then be one series for the dividends of company X from 1953 to 1973: The time series is a concept, different than the record layout of the data.

This file is updated weekly and monthly. Each U.S. company is assigned half a disk track and this segment of the file is entirely rewritten at each update; each Canadian company is allocated one full track and new data are simply appended. The update time amounts to 28 minutes per month. The total file volume currently is 17.3 million bytes.

2.1.3 The Stock Price Data

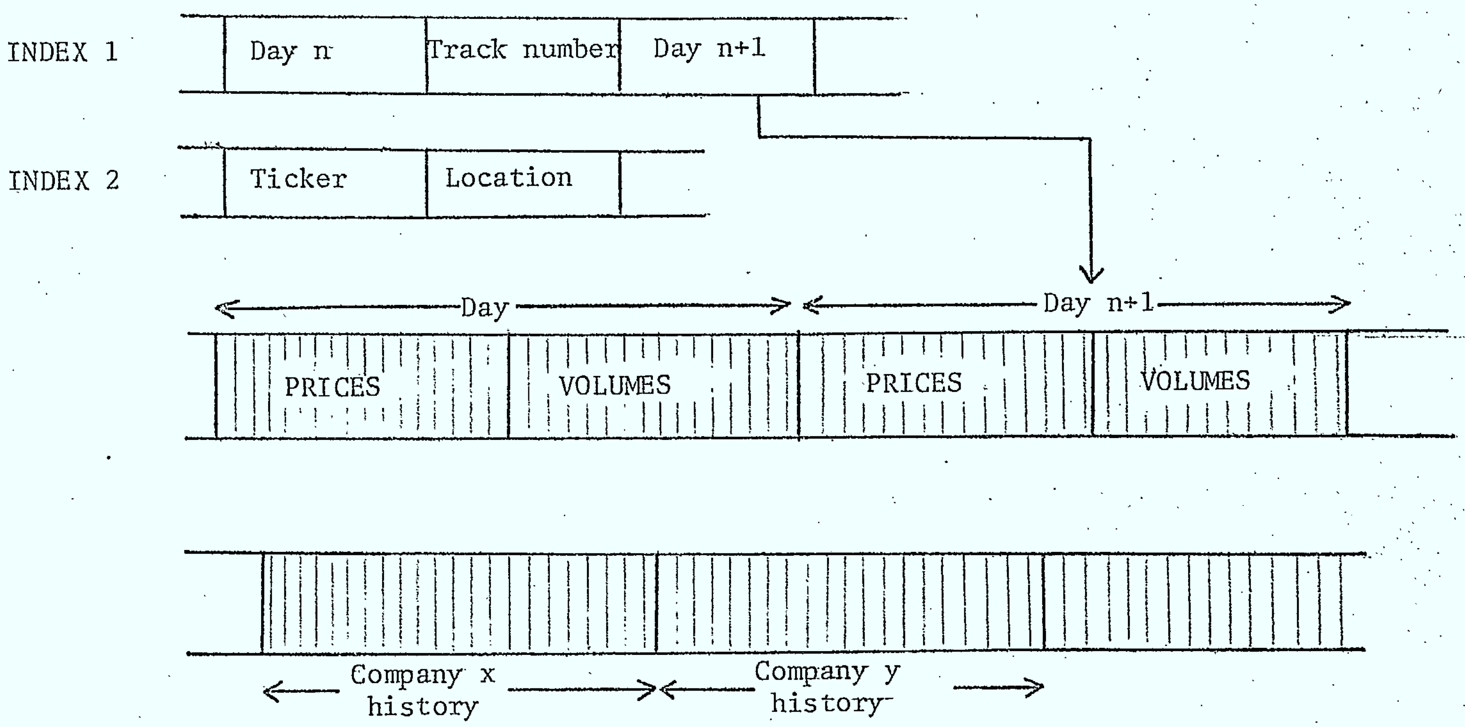
2.1.3.1 Source of Data

According to our definition of series, there are 2,000 series of daily data from the New York, American, Toronto and Montreal Stock Exchanges. The Montreal Stock Exchange sends paper tapes to the FRI, while the two U.S. and the Toronto Stock Exchanges data are sent over by lines.

2.1.3.2 Data Elements, Record Layout, Data and Update Volume

This file contains information about the daily price and traded volume for the stock of 2,000 listed companies together with the date and amount of dividends and splits. The file organization is indexed sequential, with variable length chains. To accommodate the needs of different users,[1] the same set of data is stored twice: once in day sequence, and once in company sequence. The record layout is shown below:

[1] A portfolio manager is interested in the prices of a number of stocks on a particular day; other analysts wish to have the history of prices of selected stock.



These series are updated daily and quarterly, and the update time is 100 minutes per month (CPU).

2.1.4 The Banks Data

2.1.4.1 Source of Data

This is one major area where the FRI does the data collection itself. Published information[1] concerning the Canadian chartered banks is taken and processed by the FRI personnel.[2]

2.1.4.2 Data and Update Volume

The file occupies 1.4 million bytes, and is updated weekly, monthly or quarterly depending on the particular information (monthly update time is one minute of CPU). This file is organized randomly with a computed record address.

---

[1]Consisting in quarterly and annual reports, and a monthly publication from the Treasury.

[2]Due to the format of the data, it amounts to typing it from the terminal into save files.

2.2 Services Offered

The FRI has developed and made available to users a number of data access, manipulation and reporting programs. It does not use a data base management software package; reasons that were given are: they are very expensive to operate, and they are not sufficiently oriented towards users.[1]

Most available programs are essentially data-related, i.e. they only access data in one file, which is by no means a drawback since in each of the areas - economics, financial data, stock market data, banks data - financial management activities are totally different. However, a set of manipulative programs (such as growth rate, internal rate of return, regression calculations) can be used for any type of data, through user-created files.

Accordingly, each of the files - economics, financial fundamentals, stock prices and banks data - is accessed through a set of programs, ranging[2] from:

(i) the basic data-retrieval subroutine, which is the elementary building-block in any of the other programs. Users which are able to write their own programs also make use of these routines. These routines take advantage of the McGill system routines for input-output: large blocks of data are read[3] into core and the desired information is then extracted, by direct table look-up.

---

[1] e.g. there would be a considerable time spent in training the user.

[2] it must be kept in mind that this is more a continuum than very distinct categories.

[3] Most of the time, the block consists in a whole track from the disk-pack storage. See MUSIC User Reference Manual, McGill University Computing Centre.

(ii) the elementary print out programs, which display the desired cross-sectional or time series, according to a format specified in command parameters. A typical command is:

PRINT, CYNAME, ITEM, YEAR1 TO YEAR2

where: CYNAME = the company name referred to by a ticker symbol.

ITEM = the series name, e.g. assets.

YEAR1 = Start of time series

YEAR2 = end of time series

(iii) the report-generating programs, which display data in a sophisticated format. They generally are an extension of the previous category.[1]

(iv) the file-related manipulation programs: these are the most complex and relate to specific financial management activities. Such programs are:

(a) the screening programs: according to a number of criteria,[2] a selection is made of the x stocks presenting performance regarding these criteria.

---

[1]However, the following category could also fall into the report generating programs, although they involve much more computation.

[2]This system recalls the Boolean logical data retrieval system by which a search is organized with commands such as:

(MAN) AND (OVER 30) OR (MAN) AND (LESS THAN 30) AND (COLLEGE)

(b) the portfolio management programs: for the setting up of the classification scheme of stocks, the creating and/or updating the portfolio with transactions (sell, buy . . .) and the valuation of the portfolio at a particular moment in time.

(v) the general-purpose manipulation programs, which have already been briefly alluded to, work upon data from the user files. They compute growth rates, internal rate of return, bond yield, regression planes, make plots and macroeconomic forecasts.

There are approximately one thousand of these programs,[1] requiring a total of thirty-million bytes.

This set of programs is completed by a thorough documentation concerning both the data files and the programs.

The FRI also engages in what is called the support of the FRI members' use: training of the members' staffs, maintaining contact with them, meeting with them to discuss new things or problem areas. Since more and more users do their own programming, a growing activity is to help them with the debugging of their programs.

This, of course, leads us to consider the costs incurred by the FRI in order to provide these services. In the following section, the emphasis will then be on descriptive figures and cost relationships specific to the FRI, leaving for later the implications for an integrated data-bank. It will necessarily be a very brief section, since the data-retrieval cost study will be included in the chapter bearing on the user statistics, and the discussion about the categorization of costs will be found in Chapter 4.

---

[1] These are not entirely distinct programs; there exist many versions of a basic program.

2.3 FRI Operating Costs

Categories and Percentage Breakdown

Costs at the FRI break down into:

	1972	1971	1970	1969
- Data acquisition costs	20%	21%	18%	21%
- Net Computer Costs	12-15%	17.5	19	19
- Salaries	35%	32	29.5	24
- Communications	12-15%	12	13.5	12
- Administrative and Overhead	15%	17.5	20	24

2.3.1 Acquisition Costs

In some cases, the contract between the FRI and the data source includes some services from the FRI (notably in the case of the Stock Exchanges) and the value of these services are considered as a cost. Otherwise, they are simply contracts for cash.

2.3.2 The Computer Costs

The computer costs fall into three categories:

(i) the storage cost, which is proportional to the storage space. The McGill Computing Centre normally charges 15¢ weekly for one "IBM 2314 DASD" track and 30¢[1] for a "3330 DASD" track (the IBM 2314 Direct Access Storage Device has a track capacity of 7294 bytes, while the 3330 DASD track houses 10,000 bytes).

---

[1] 20¢ if the user does not need his file backed up by a copy on tape.

The FRI because of its large and relatively constant demand for storage space, has exclusive use of two 3330 on-line disk packs. These are utilized for storage of data and production programs. For the data storage, the FRI has an arrangement with the Computing Centre and pays $665 per week.[1]

(ii) the update CPU time is normally proportional to the update volume, although when the file is entirely rewritten (the economic data, a part of the financial data), the update time becomes dependent upon the total file volume. When storage cost and update time cost are compared:[2]

	Economics	Stock market	Financial	Banks
STORAGE COST	$700/mo.	$9,000/mo.	$2,100/mo.	$168/mo.
UPDATE TIME COST	$ 70/mo.	$1,000/mo.	$ 280/mo.	$ 10/mo.

It is noticed that there is a ratio of 10:1 between storage and update costs. This is consistent with the fact that storage volume is a multiple of update volume in this type of application (time series).

The CPU charging rate depends on the period of day as shown below:

---

[1] Storage hardware will include the following components: block multiplexor channel, controller, and disk spindles. One controller supports up to eight spindles. Manufacturer's quotations for rental costs comprise approximately $600 per month for a 3330 spindle, $2400 per month for a controller and $500 per month for a block multiplexor channel. The arrangement between FRI and the McGill Computing Centre then comes closer to rental cost.

[2] These costs have been computed with the rates of $1.20/track/mo. and $10/min. CPU.

Hours		1 Min. CPU
8:00 to	9:30	$10
9:30	12:00	12
12:00	2:00	10
2:00	4:30	12
4:30	6:00	10
6:00	10:00	8

(iii) The programming and debugging involve both CPU time and connect time.[1] Accordingly, in addition to the CPU rates shown above, we have the connect time rates:

Hours		10-15 cps[2]	30 cps	60 cps	120 cps
8:00	9:30	$3.00	4.50	6.00	9.00
9:30	12:00	4.00	6.00	8.00	12.00
12:00	2:00	3.00	4.50	6.00	9.00
2:00	4:30	4.00	6.00	8.00	12.00
4:30	6:00	3.00	4.50	6.00	9.00
6:00	10:00	1.00	1.50	2.00	3.00

This last category is essentially research and development and maintenance.

The total computer costs at the FRI[3] have been drastically reduced since 1970, when the FRI had a much larger staff for development:

1970	$ 100,300
1971	57,900 (-42%)
1972	49,300 (-15% on 1971)

[1] While the update activity involved only a negligible connect time.

[2] characters per second; 1 cps ≅ 9 bits per sec. 10-15 cps is the speed of the usual TWX or IMB 2014 terminals.

[3] The appended figures are net of a 15% rebate to FRI on total computer usage costs of FRI and users combined in excess of $2,500 per week, and of a 25% rebate on combined usage costs in excess of $4,000 per week.

2.3.3 Salaries

The summary table at the end of the section gives a breakdown of manpower resources used for operations and capacity.[1] Capacity would also include the staff time for the documentation of programs and data. A last category that has been left out is the programming and debugging time, and the development of products and markets; all three are in the development category.

2.3.4 Communication Costs

The communication costs include:

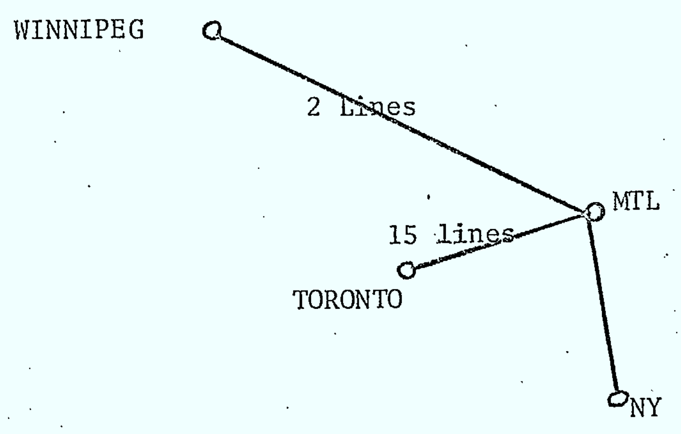
- 8 terminals at the FRI connected with the McGill Computing Centre.
- one tape receiver to be connected to a tape reader in New York (the NYSE and AMEX Stock Market data)
- a transmission line from New York to Montreal.

It is owned by a member who lends it to the FRI at night. It costs the FRI $8 monthly.

- 15 lines from Toronto to Montreal.
- 2 lines from Winnipeg to Montreal. The policy of FRI is to pay for the distance Montreal-Toronto in a "two basic point pricing" fashion, and to have the user pay the rest. The cost to FRI is thus two lines Toronto-Montreal. The network is thus:

---

[1] These two terms will be discussed in Part III, section 2. Capacity roughly defined includes all activities necessary for being ready.



2.3.5 Rent, Administrative and Overhead Costs

This cost component involves two office rents, one in Montreal and one in Toronto, the administrative expenses, postage, telephone, supplies, etc.

2.3.6 Development Costs

While we do not have estimates of its cost, development has consisted of programming effort, research on financial data comparability and format (notably the establishement of homogeneous classification systems for financial data). The FRI has also incurred major one-time costs in validating the data in the files.

SUMMARY TABLE OF OPERATING STATISTICS

	ECONOMICS	STOCKMARKET	FINANCIAL	BANKS
Acquisition Costs				
- contract for cash	$150/mo.	$600/mo. } NYSE AMEX	$3500/mo.	
- value of reciprocal services		$250/mo. } MSE TSE		
- transmission cost		$380/mo. } NYSE-MTL MSE		$8/mo.
Storage & Update				
- storage	5.8 M. bytes	75 M. bytes	17.8 M. bytes	1.4 M. bytes
- update time (CPU)	7 min/mo.	100 min./mo.	28 min./mo.	1 min./mo.
Staff Time (Operations) Man-Days/Mo.				
- data quality	0	20	1	1
- customer support	4	20	10	2
- data loading	1	10	1	1

2.3.7 A Naive Cost Model - Part I

While examining the costs incurred at FRI, it was possible to identify some determinants of costs, e.g. some factors, such as the file size and the update volume, which had a direct impact upon the cost figures.

The major part of costs are relatively independent of these quantitative factors, and are mainly function of the type of data, some contract passed with suppliers, etc. . . Moreover, it is difficult to relate the magnitude of the staff and other costs to the magnitude of the file size and update volume.

In summary, and this is the inherent limitation of a cost study, most costs are fixed with respect to output,[1] whether output is the use of the system or the number of series stored. Since, so far, we have just dealt with the FRI and not with the user subsystem, the discussion can be limited to the primary output of the FRI: the files and the programs.

However, even this is not so simple: is the FRI primary output the number of bytes stored, or the number of series stored? In the first case, storage is variable with the output, while in the latter case, it is variable only to the extent of changing the number of series stored. The same applies to the update time, which is fixed with respect to a given number of series.

---

[1] A more thorough discussion of the concept of output appears in Chapter 3 of Part II.

A naive approach is to disregard these problems by focussing the model on the capacity costs, where the main cost factor is the scale of operation. The model would then be able to predict the magnitude of period costs with respect to some parameters of interest. As we mentioned earlier, this cost model incorporates in a very crude way the data acquisition and salary costs: the only cost area where its accuracy is faultless in the computer costs.[1]

The main determinants of the storage volume are:

- the age of the series (in years)
- the update frequency (number of updates per year)
- the number of series (defined as the <u>time</u> series)
- the record length
- the index (if any)

The factors to take into account for the update volume are just:

(i) the number of series

(ii) the update frequency

(iii) the record length

(this would be the case for a simple appending of the new data; now, if the whole file is rewritten for each update, as it happens for the economic series for instance, then the update volume is the whole storage volume).

---

[1]Which are the "direct" costs of storage and update; the balance is considered as indirect and overhead costs.

Then, if X_1^i: update frequency for series i $1 \leq i \leq X_2$

X_2: number of series stored

X_3^i: age of series i

the update volume is: $\sum_{i=1}^{X_2} C_1^i X_1^i$ measured in bytes and the storage

volume is: $\sum_{i=1}^{X_2} C_2^i X_1^i X_3^i$ in bytes. If we assume:

1. an average update frequency for all series i: X_1
2. an average age for all series: X_3
3. the update volume is negligible in comparison with the storage volume
4. the update cost is proportional to the update volume,

the computer and other costs are the sum of the storage-related cost: $C_2X_2X_1X_3$ and the update-related cost: $C_1X_1X_3$; thus:
cost dependent on size of data-bank: $X_1X_3(C_1 + C_2X_2) + C_3$.[1]

The purpose of this model is to point to some parameters of cost, the most important being the file size and the update frequency. Its shortcomings appear when one considers the variety of file organizations and update strategies. However, a greater limitation lies in its concentration on a limited portion of the FRI's total costs.

---

[1] C_3 includes, among other things, the storage control unit cost.

2.4 Usage and Cost to the User

We mentioned earlier that the data-bank system should include the McGill Computing Centre which provides the usage facility. Since the overall system cost includes the usage cost, this section considers usage parameters.

2.4.0 User Needs and Characteristics

In order to place the FRI into perspective, it is necessary to turn to the user, see who he is, what are his activities, his information needs and sources. A more detailed survey is presented in Appendix 2. In this section, only the major findings will be outlined.

The first observation to be made is that data-banks and other specialized information services appear as the least important source of information for the financial community. Before coming to the explanations, it must be mentioned that the importance of data-banks is expected to increase in the next five years relatively to other sources.

The obstacles to a more rapid development of financial data-banks fall into three main categories. First, data-banks have to compete with traditional sources of information. More specifically, they are a duplication of other media. Data-banks are limited in that they offer something that already exists elsewhere. With the exception of in-house data banks, they do not include data for unlisted companies, in contrast to the expectations of financial analysts, who would prefer more breadth of information than depth.[1]

[1] From the data-bank management viewpoint, collecting data for unlisted companies would involve a considerable effort, to be offset by the (presumably) small incremental benefit, since unlisted company data are seldom looked for.

Data banks also have to compete with other non-computerized information services. If the criteria for a good information source are: ease and speed of access, timeliness, accuracy, convenience and completeness, this last quality does not seem to be applied to numeric data-banks, as notes and adjustments are not included. For instance, the users' personal files seem to be generally preferred as a source of past data. In some cases, however, computerized data-banks will favourably compete against other sources when a great deal of manipulation on large amounts of data (screening) is needed.

Another considerable obstacle to the greater usage of data-banks is that they are repositories of past data, which are not sought <u>per se</u>, but only to the extent that they can help to predict the future course of events, or derive a trend. Besides this, data are not used further back than 10 years. The usefulness of a data base is thus inherently limited, but it can be enhanced by more powerful forecasting tools.

Some additional information on the usage which is made of the FRI data is as follows: the price data are generally accessed to evaluate the portfolios, whereas the company data are used as a complement to other sources. The economic data are used mainly as inputs for econometric models and studies.

Several respondents' remarks point to the desirability of more selective services (to avoid duplication) and in particular to new applications such as Selective Dissemination of Information (SDI), an article retrieval service, and international economic data (IMF date, for instance).

2.4.1 Types of Users

Table 1 shows the FRI commercial users:

TABLE 1

USERS	NUMBER
- brokers and investment firms	17
- banks	7
- trusts	6
- insurance companies	5
- industrial companies	4
- government services	3

The FRI provides four main services: stock market data, financial data, banking data and economics, according to the broad classification of its series. Each category of user may be expected to show a bias towards a service, banks for banking data, industrial companies for economic data, etc.

A matrix of interdisciplinary bias was built where the coefficients are derived from the actual usage of the files by categories of users:[1]

---

[1] Source: FRI

TABLE 2

	Financial	Stockmarket	Banks	Economics	
Brokers & Investment firms	6%	76%	7%	11%	= 100%
Banks	20%	42%	21%	17%	= 100%
Trusts	5%	92%	0%	3%	= 100%
Insurance	5%	79%	3%	13%	= 100%
Industrial	55%	33%	0%	12%	= 100%

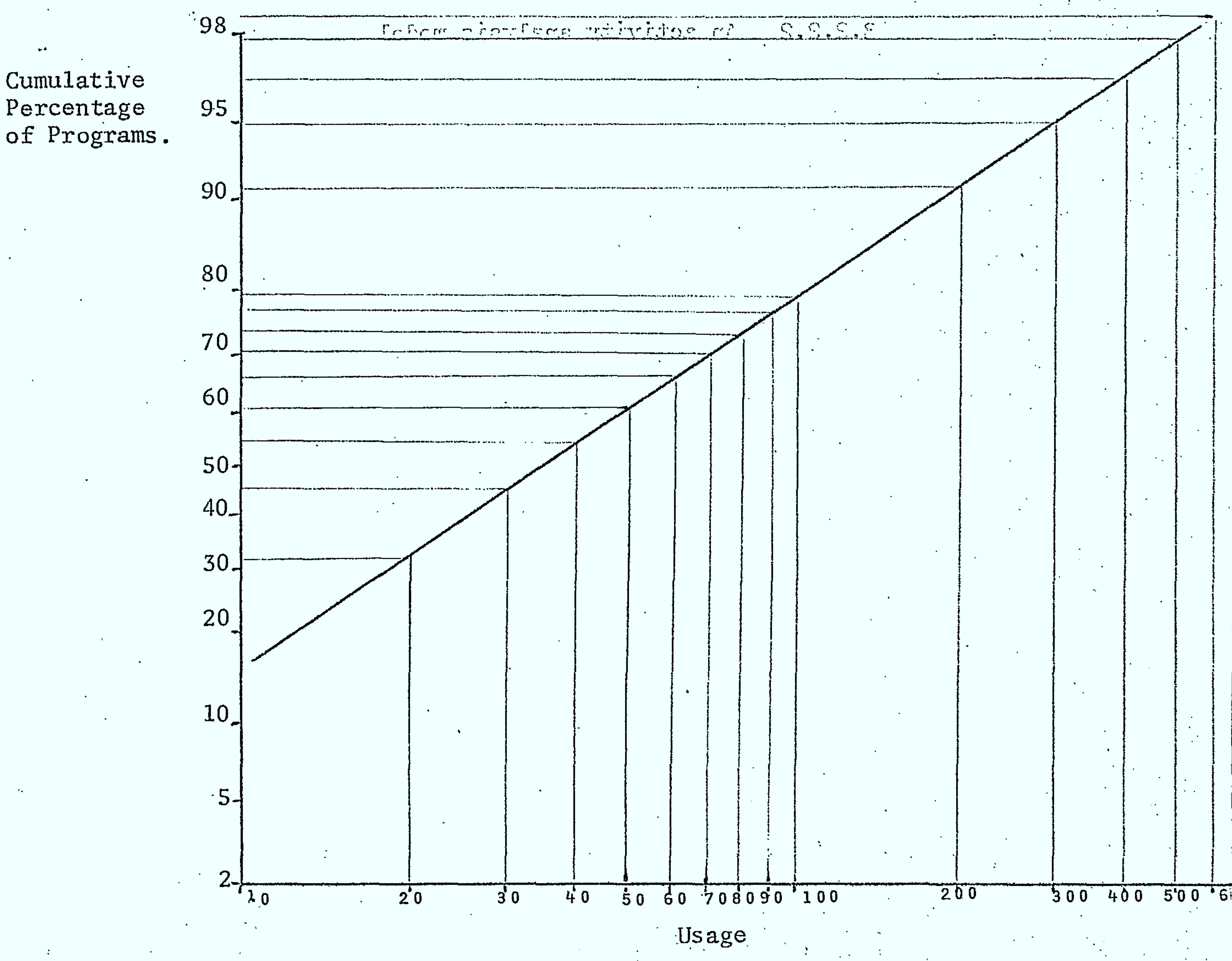
From this table, it can be seen that the trusts are the most biassed towards a particular file (presumably because of the portfolio management programs), while the banks are the most diversified users, followed in this by brokers, investment firms and insurances (which probably have the same financial management activities). The industrials notably do not follow the general pattern of bias towards the stockmarket data: on the contrary, they show greater interest for the financial reports.

We believe that a generalization of this matrix provides a good description of the pattern of usage among services. Moreover, it has an important role to play in the prediction of demand for particular services given a population of users categorized according to this scheme.

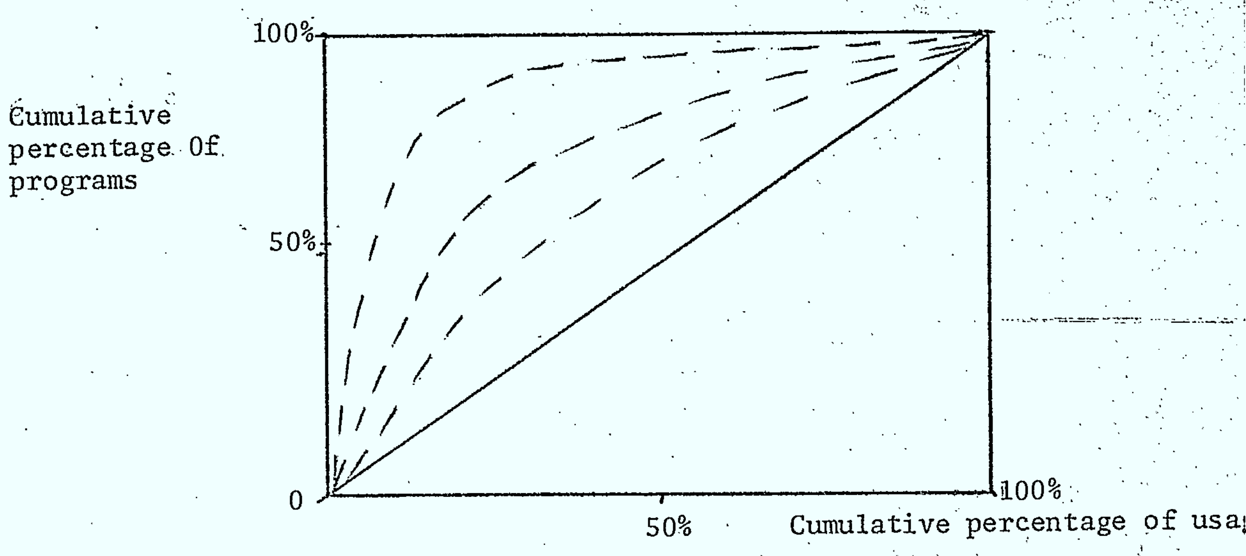
2.4.2 Usage Patterns

2.4.2.1 Among Programs and Files

Another interesting feature, revealed by the study of usage, is that usage follows some kind of Pareto's Law: that is, a relatively small percentage of programs contribute a high percentage of usage. More specifically, using log-normal paper, it was observed that the number of programs run is lognormally distributed, or equivalently, that the logarithm of the number of programs run is normally distributed.



If no precise explanation for this particular shape of the contribution-by-usage curve was found,[1] it must be remembered that Pareto's Law is a well-known phenomenon in the social sciences. Indeed, one could hardly expect usage to be uniformly distributed among programs; the uniform distribution would have been represented by the diagonal on the chart below:



The more distant the distribution curve is from this diagonal, the more dispersed (e.g. concentrated in some segments and light in others) the distribution is:

---

[1] Beside the fact that there cannot exist negative use of a program and that most of the programs are run between 0 and 10 times, e.g. the distribution is very skewed.

The implications of this observation are of two orders:

(i) it is of valuable interest to assess trade-offs such as cost of keeping and maintaining the program versus expected benefit from usage. Its importance lies precisely on the emphasis it places on marginal analysis: what impact will it have on usage to dispose of, say, 20% of the programs?

(ii) with this relationship, it becomes possible to project the usage of a particular program for various levels of total usage since the parameters of usage distribution can be expected to remain fairly constant.

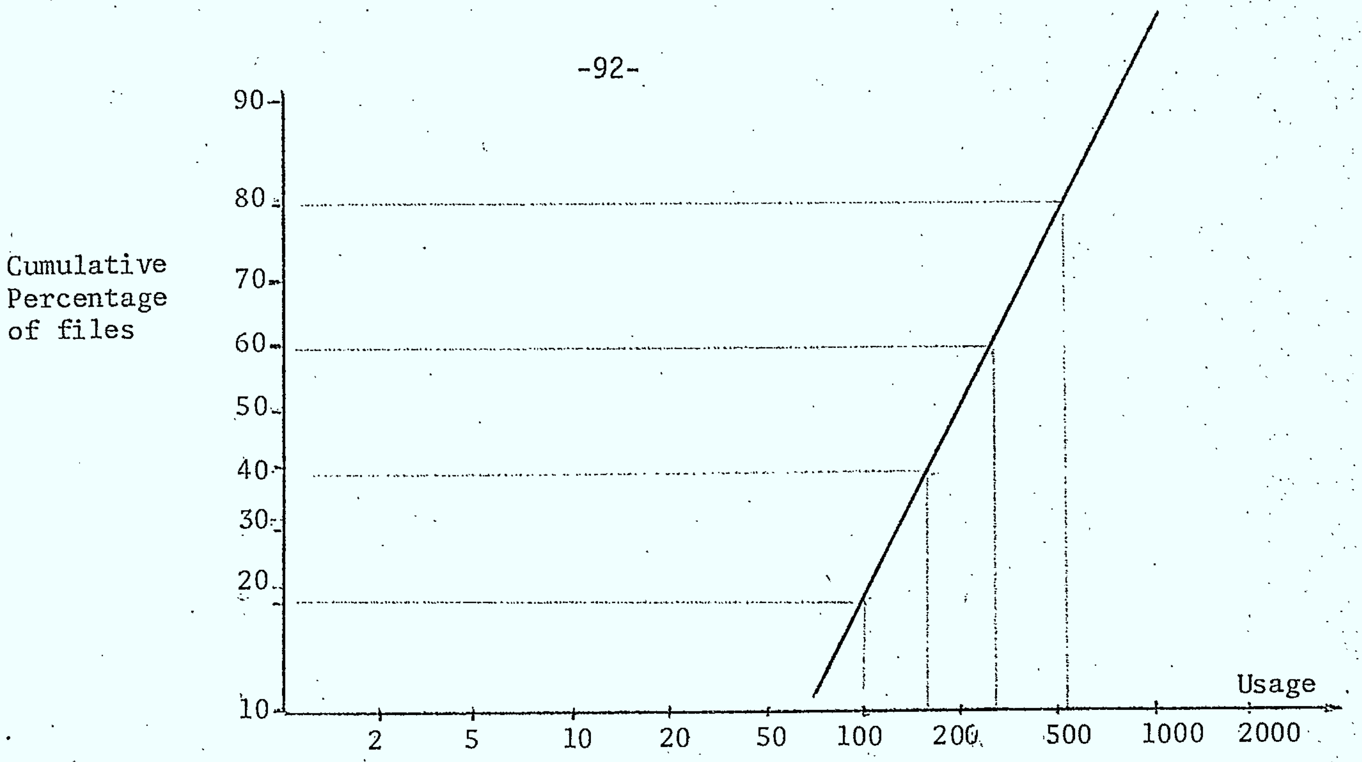
In fact, usage distribution among programs may not be such a critical matter, since once a program is in the system, there are relatively small costs associated with keeping it.[1] It becomes a key factor when dealing with files instead of programs. Files are expensive to maintain and store, and the trade-off is more relevant in this case.

The same analysis was carried upon the usage among the four files (a finer analysis was not possible[2]) and approximately a similar relationship showed:

---

[1]although this will be discussed in Chapter 3 of this part.

[2]Usage is only recorded by the FRI in terms of number of program runs; since one file is accessed by many programs, we only have a very rough idea of file usage.



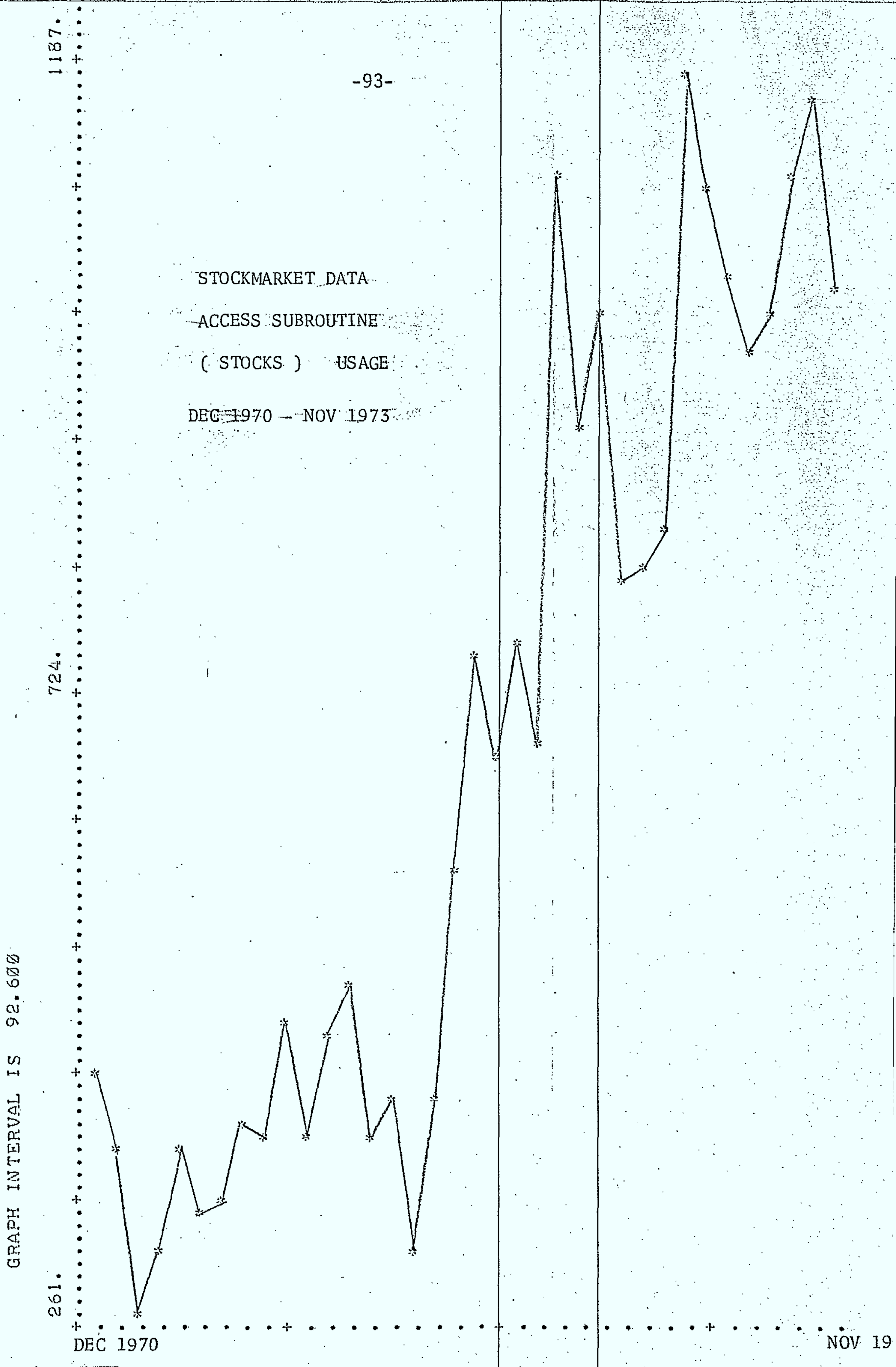
Although it was not possible to see whether Pareto's Law holds for usage against age of the data, we suspect that it would: usage would then concentrate on the most recent data items.

2.4.2.2 Usage Over Time

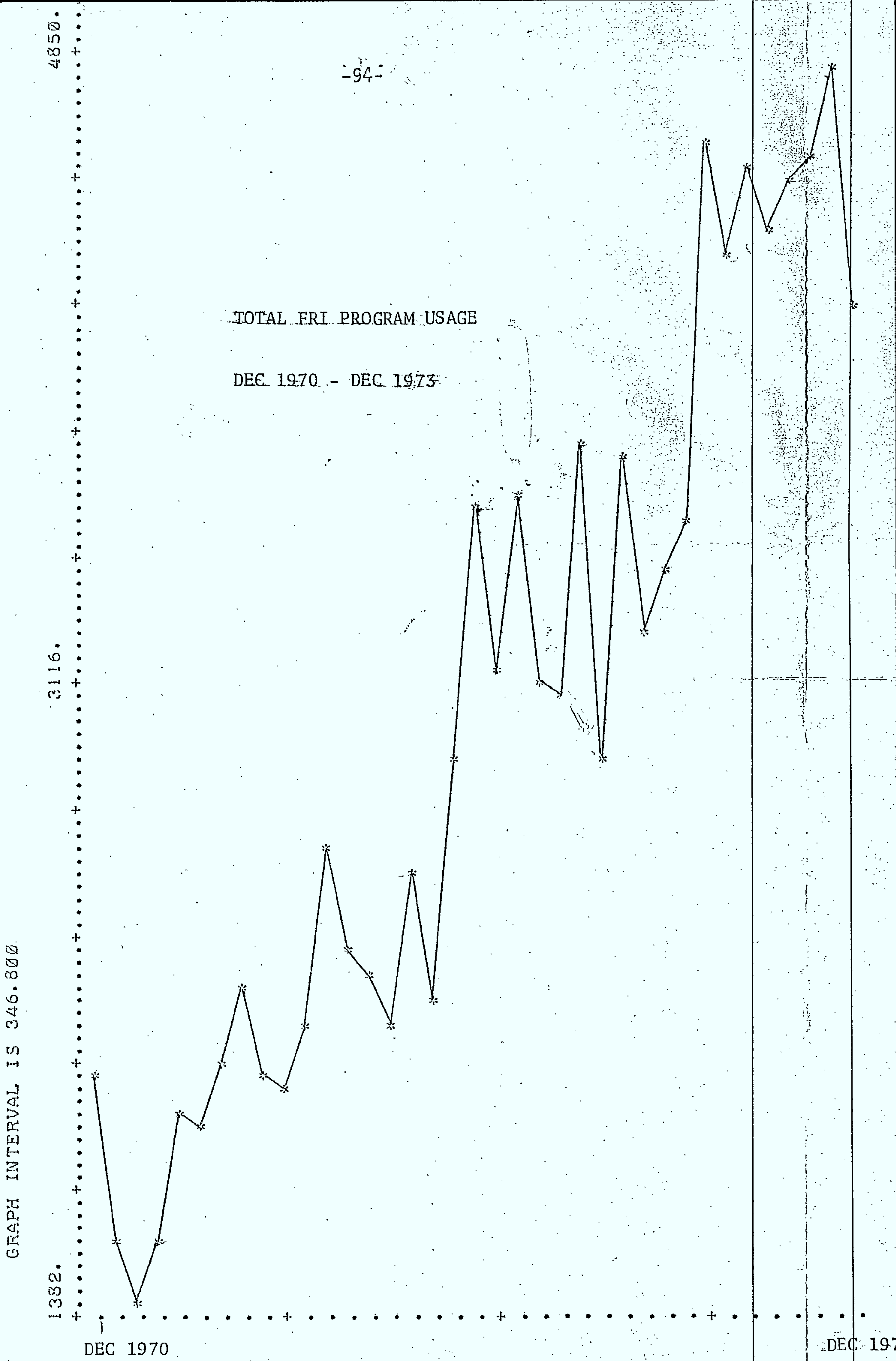
Usage, as measured by the number of program runs, has dramatically increased since the beginning of FRI operations. There has been a 50% increase of 1971 usage over 1970 and a 30% increase in 1972 over 1971. Although no seasonality can be observed, the year high is in November, with annual jumps in usage in the period March to June (annual company reports are published at this period of the year).

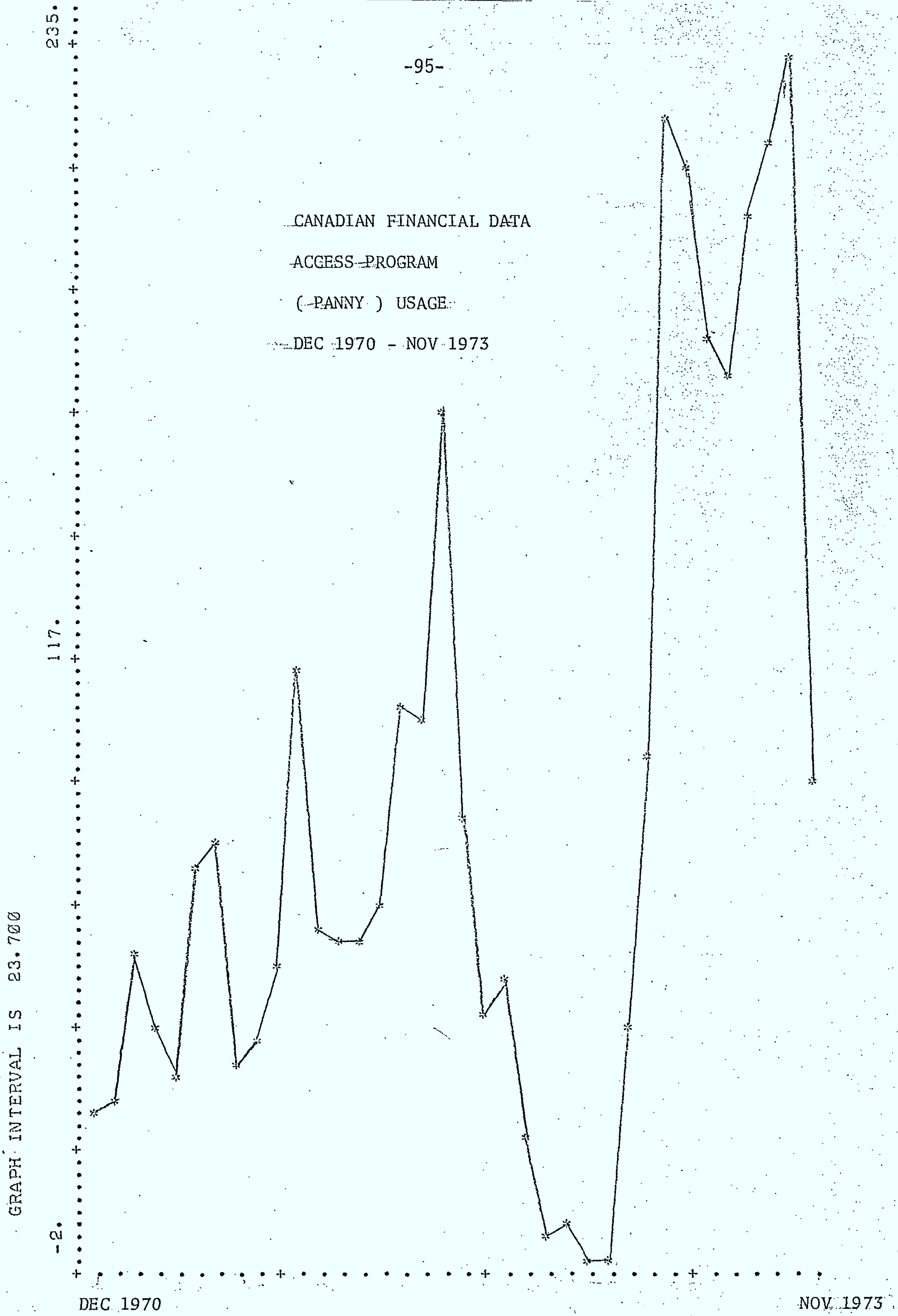
A time series analysis along the Box and Jenkins method was carried out on the FRI usage. The best model to fit the data was a moving-average process of the first order on the first differences:

$$Z_t = 58 + U_t - .70U_{t-1}$$



STOCKMARKET DATA
ACCESS SUBROUTINE
(STOCKS) USAGE
DEC 1970 — NOV 1973
1187.
724.
261.
GRAPH INTERVAL IS 92.600
DEC 1970
NOV 19


TOTAL FRI PROGRAM USAGE
DEC 1970 - DEC 1973
4850.
3116.
1382.
GRAPH INTERVAL IS 346.800
DEC 1970
DEC 197


CANADIAN FINANCIAL DATA
ACCESS PROGRAM
(PANNY) USAGE
DEC 1970 - NOV 1973
235.
117.
-2.
GRAPH INTERVAL IS 23.700
DEC 1970
NOV 1973

where Z_t are the first differences of the observed series. The random shock interpretation of this model suggests that the effect of a disturbance at time t (a sudden upsurge of activity by the users) carries over the period t+1 but with a negative weight: in other terms, a period of high usage will generally be followed by a period of lower usage, as if an equilibrium had to be achieved.

2.4.2.3 <u>Intensity of Usage</u>

Data concerning characteristics of terminal sessions and the usage of CPU by FRI members are presented and commented on in this section, in an effort to assess the load placed on the computer system by the very existence of the FRI data-bank.

The terminal session time is defined as the interval of time starting when the user signs on the system and ending when the terminal is disconnected from the system. The CPU time is simply a surrogate for the amount of work done during a session. For our purposes, a user is identified by his code; since an FRI member (a company) may be attributed a number of codes, there are more "users" than members. The sample period chosen was one week in November (the peak period for the FRI data-bank usage). It has been checked that the sample is representative from the point of view of the stability of the results. It must be recalled that terminal session time (or connect time) and CPU time are two of the three components of the computer bill.

During this week, for 1282 terminal sessions of the FRI data-bank users, the average terminal session time was $\mu_T = 23$ minutes (with a standard deviation $\sigma_T = 31.5$ minutes). The average CPU time per terminal session was $\mu_{CPU} = 0.267$ minute (with a standard deviation of $\sigma_{CPU} = 0.551$).

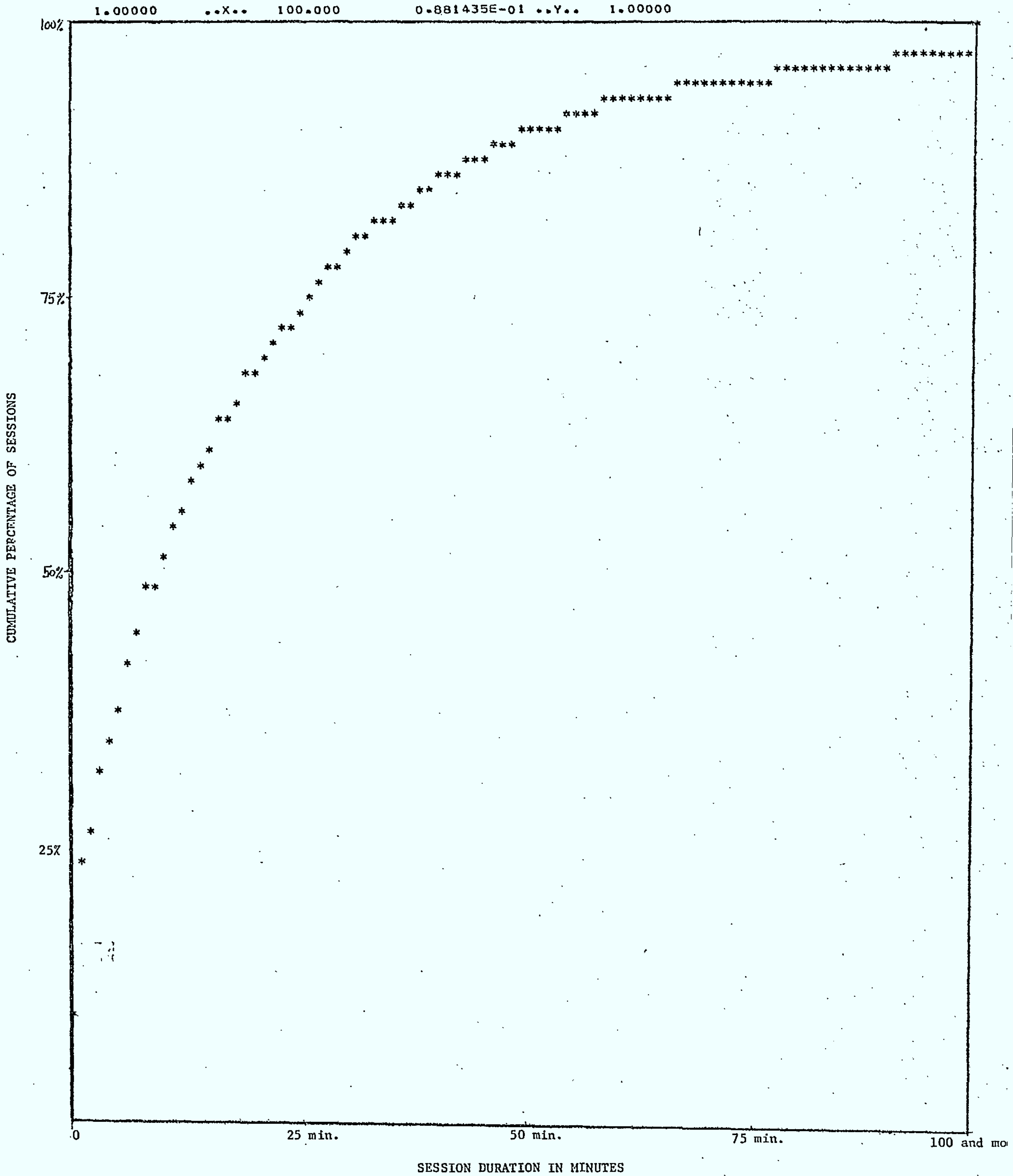


FIGURE 1: Session Duration Pattern

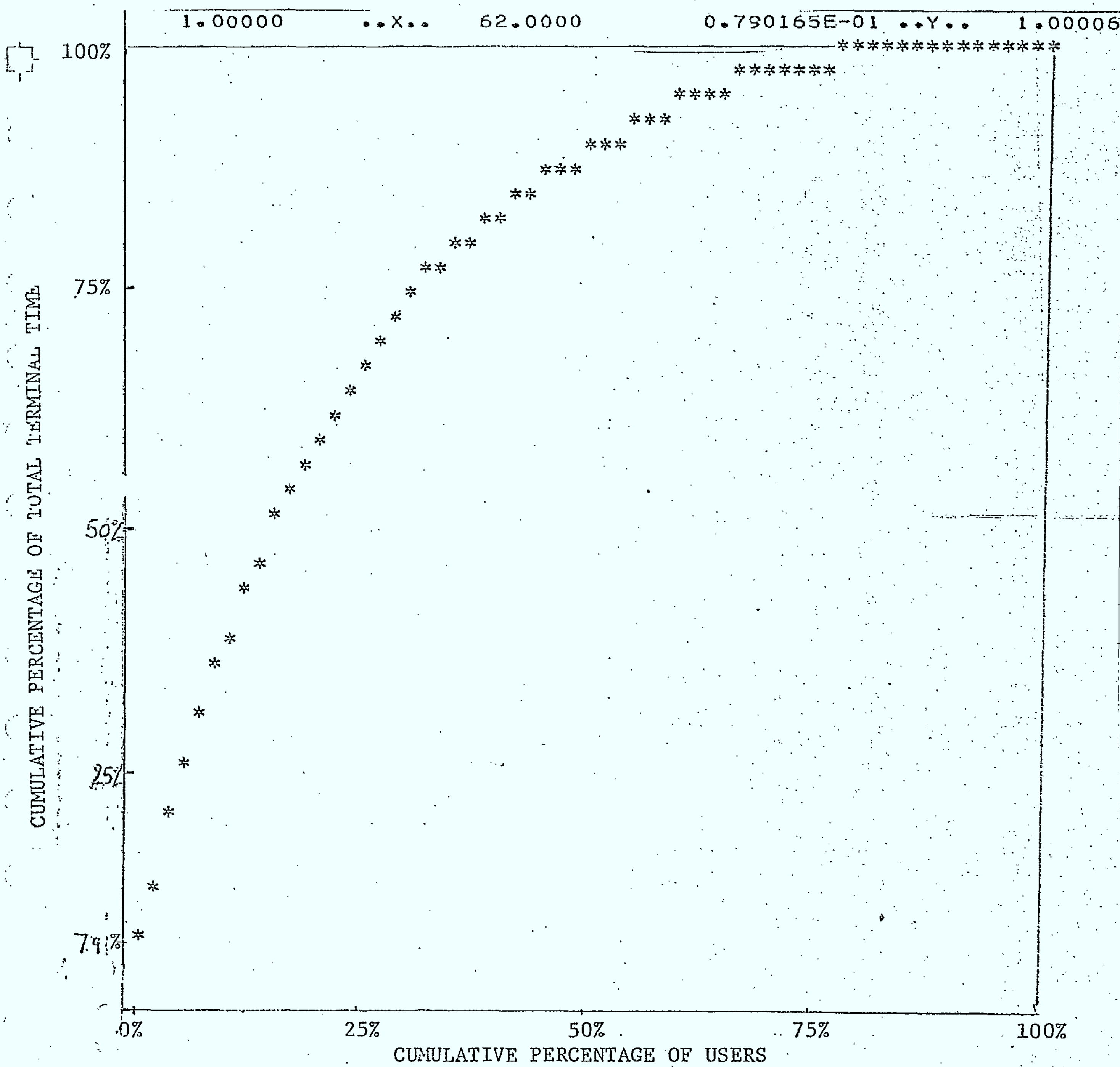


FIGURE 2: Distribution of Terminal Session Time

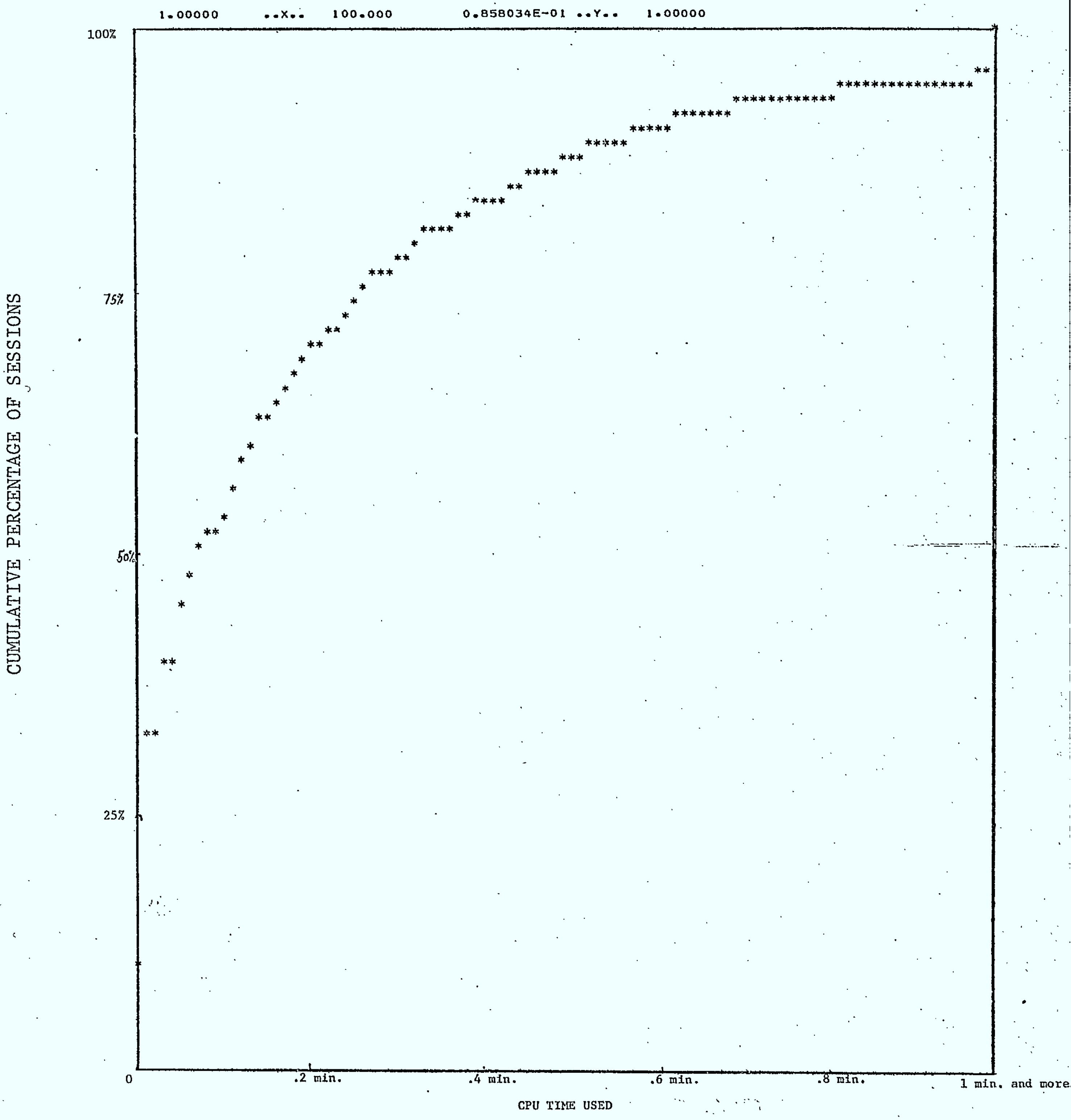


FIGURE 3: Session Work Pattern

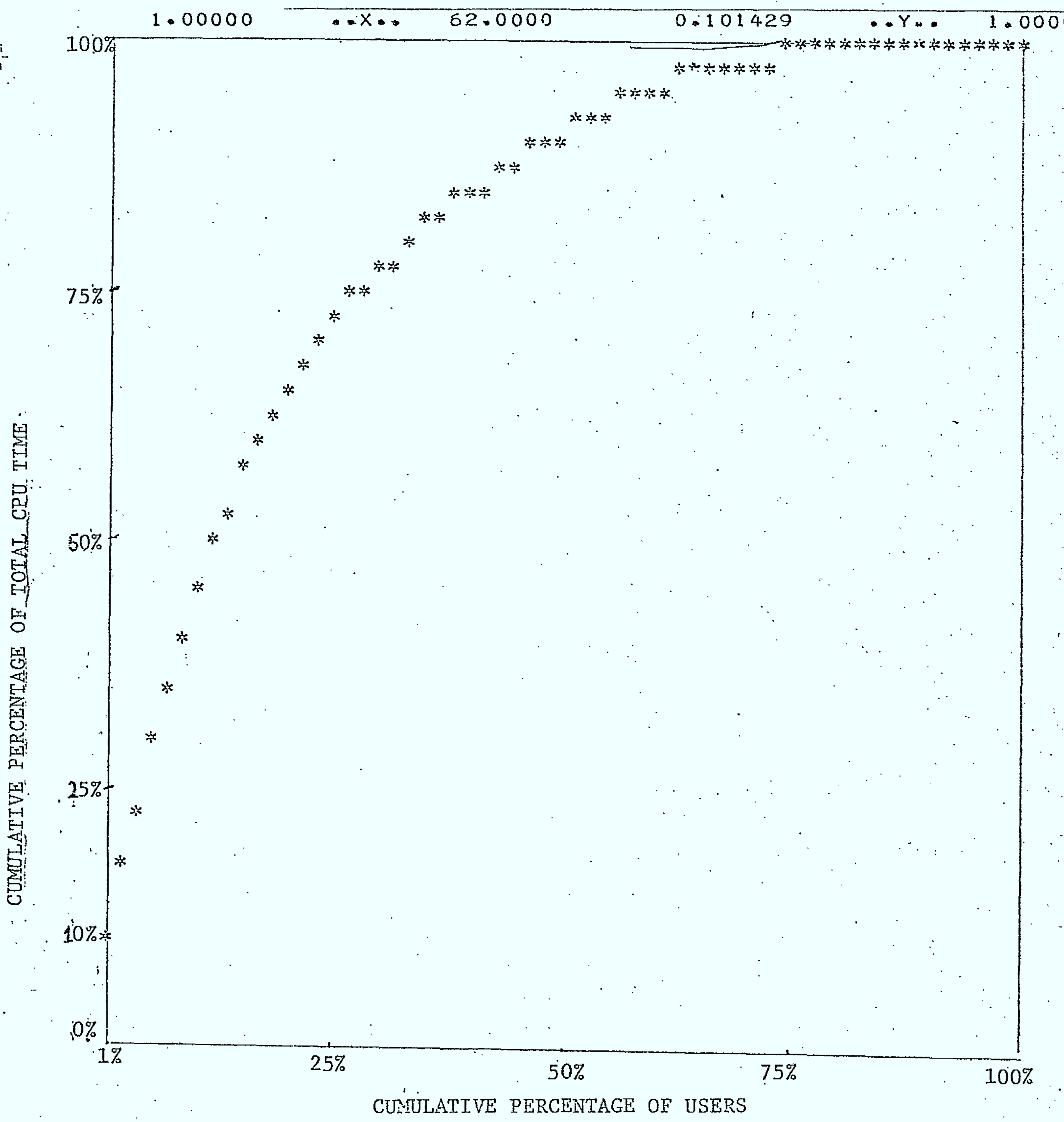


FIGURE 4: Distribution of CPU Usage

Figure 1 shows the plot of the cumulative percentage of terminal sessions ranked by session duration. Figure 2 shows the cumulative percentage of total terminal session time accounted for by users. As can be inferred from these graphs, a small percentage only of the session (25%) last more than 25 minutes, and there is a large number (45%) of short terminal sessions (less than 5 minutes). From Fig. 2, one sees that 25% of the users account for more than 75% of the total terminal session time, and the 50% most active users account for a little less than 90% of the total connect time.

Figure 3 is a plot of the cumulative percentage of terminal sessions ranked by CPU usage. Figure 4 shows the cumulative percentage of total CPU time accounted for by the users. These graphs depict similar relationships. Approximately 75% of the sessions use up less than 0.20 minute of CPU time (Fig. 3) and 75% of the users require less than 25% of the total CPU time used (Fig. 4). The conclusion is that there is a number of very active users (25% of the total) who put a heavy load on the system (consuming 75% of the recorded resource usage), whereas the light users (the remaining 75%) are responsible for only 25% of the resource usage.

These observations provide a basis for assessing the load placed by data-bank users on a time-sharing system, both in terms of processing time and communication sub-system requirements (terminals, lines, channel ports). They bear direct implications for the models exposed in part III (the terminal session time, or service duration, is one of the two basic parameters of the queuing problem arising with the contention for channels.

2.4.3 Cost as Related to Usage

Since we include the user-servicing subsystem (e.g. the McGill Computing Centre) in our data-bank system, the costs the users are bearing are inputs to the operation. In section 2.3, we examined inputs that were related to the file size and the update volume. In this section, we deal with inputs that are dependent on the usage of the data-bank, in other words, on the number of accesses to the data-bank.

2.4.3.1 User Computer Bill and Other Costs

Besides his FRI membership fee, the user pays for his use of the time-sharing computer service. In section 2.3, the rate structure for the computer cost was exposed: both CPU time, connect time and storage of user files are the basis for charging the user. Beside these computer costs, he pays Bell and/or General Electric for the transmission hardware (the terminal, the line, etc.) and service. But first, some observations about the computer bill must be made.

On one hand, data about user access to certain programs per period of time and on the other hand, the computer bill per member for the same period of time, were available. Eleven members, heavy users of the stockmarket data, were selected, and we used a regression analysis of their monthly computer bill on their corresponding monthly program usage. Two explanatory variables were used as surrogates for program usage.

S: stockmarket data access program runs

M: other program runs

TABLE 3 - Computer Bill vs. Usage

Member			R^2
1	Computer bill	= 7.5S + 3.7M + 1060	R^2 = .92
2	C	= -0.8S + 3.6M + 539	.12
3	C	= 0.5S + 3.3M + 2441	.22
4	C	= 8.5S + 9.7M - 373	.83
5	C	= 10S + 8.51M - 1227	.48
6	C	= 3.4S + 3.8M + 69	.72
7	C	= 1.8S + 13M + 180	.84
8	C	= 13S + 36	.63
9	C	= 3.9S + 29M + 992	.63
10	C	= -11S + 21M + 3430	.66
11	C	= 6.7S + 1.7M + 274	.60

More detailed independent variables could have been added, but since this would have required more observations than available (12 months) and would have resulted in the instability of the coefficients. Some coefficients obviously cannot be relied upon[1] (possibly because a number of outliers distorted the least-squares computation). If we assume

[1] A low R^2 or a negative coefficient casts some doubt about the validity of the relationship. The t-values are not shown; they give significant values to all coefficients but those of the rejected regressions 2, 3, 10.

some stability of the members' private files, then the regression intercept gives an estimate of the storage cost component.

The computer bill however includes the cost of many other things than just the FRI programs run: user-made programs on non-FRI files, notably. In the matrix below, the processing of FRI files with absolutely non-FRI programs is impossible, but because of the nature of the time-sharing system, the user can have computer activities which are not recorded by the FRI:

	FRI programs	non-FRI programs
FRI files	measured	impossible
User files	measured	not measured. ?

On an overall basis (e.g. all members together), there is a .78 correlation coefficient between the computer bill and the stock-market data access program usage, and only a .50 coefficient between the bill and the other programs usage.

The overall regression equation is:

$$\text{Computer bill} = 6.66S + 3.9M + 101 \qquad R^2 = .79$$

t-values (8) (2.23)

This means that on the average, a user spends $6.6 on a stockmarket program, and has a storage cost of a hundred dollars per month. However, the individual regression lines show wide variations, with an access cost between $1.8 and $13. The determinants of these variations will be explained in the next subsection.

The conclusion is that the number of program runs is a fairly good predictor of the computer bill, at both the individual and the aggregate level. We are thus faced with two different measures of data-bank usage: the computer bill and the number of program runs.[1] On one hand, the number of accesses is subject to criticism, because it does not specify what is being done during the run (how many portfolios handled, how many queries, etc.); in other words, with which intensity the data-bank is consulted. On the other hand, the computer bill includes such items as non-FRI program runs, storage cost, and connect time, which eventually bias the measure upward.

In summary:

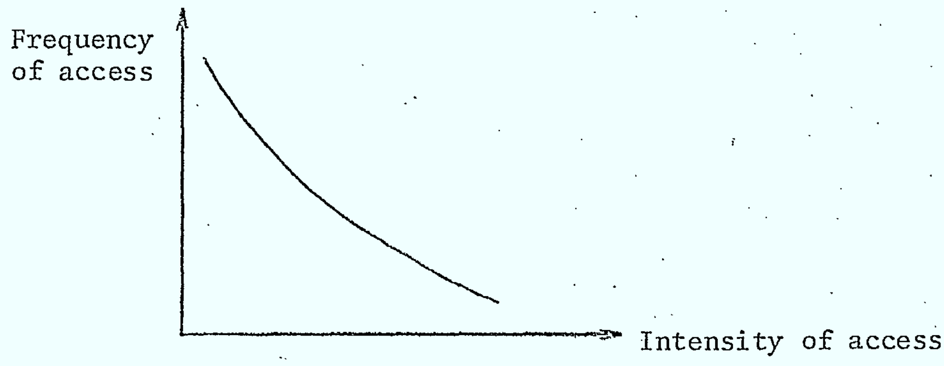
the computer bill is a surrogate for:	number of runs is a surrogate for:
- intensity (number of queries, or access to file)	
- frequency of access (access to programs)	- frequency of access
- storage cost	
- non-FRI program work	

The only common factor between the two measures is the frequency of accesses, but the intensity of access is not. If we assume that a user has a stable workload[2] (e.g. a certain number of

[1] or indifferently, the number of accesses to the program.

[2] This seems to be supported by our earlier remark (section 2.4.2.2) that there exists a tendency towards equilibrium over the periods.

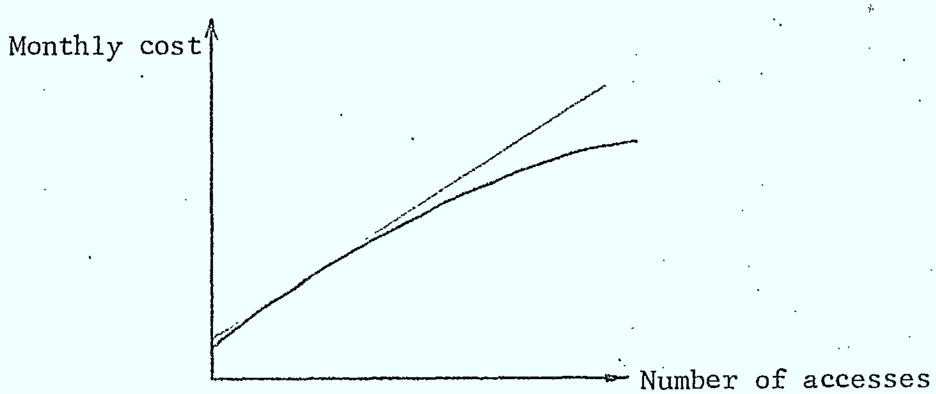
companies he looks after, a number of portfolios to evaluate), either he will access the data-bank a few times, but intensively, or many times, but less intensively.



The regression coefficient (\$6.6) is a surrogate for the intensity of the average program run. More precisely, the computer bill is such that:

\$bill = (frequency) x (intensity) x (proportionality constant)
# of accesses   # of queries in one access

If, as hypothesized, there is some inverse relationship between the number of access and intensity, then we should expect a computer bill less than proportional to the number of access:



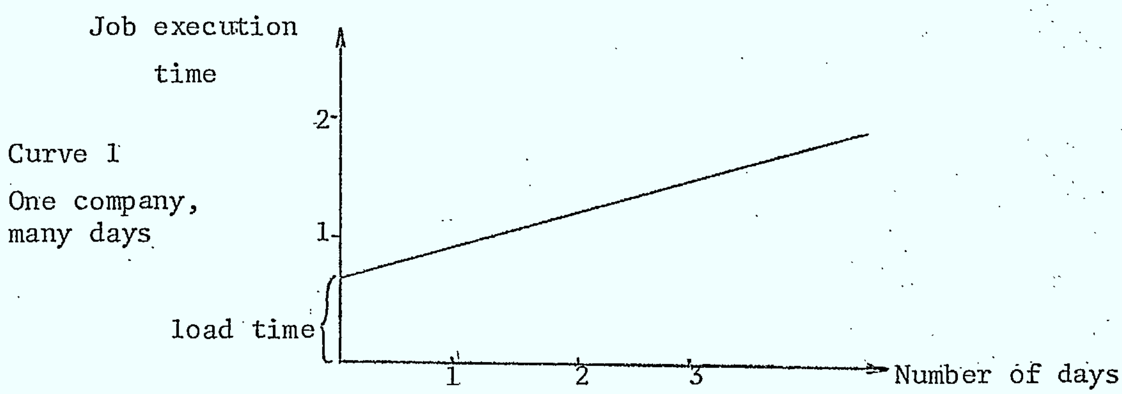
Some slight evidence was in effect found of a decreasing slope, but its statistical significance is low. The important feature thus remains in the linear relationship between usage and cost. The variations around this linear pattern are due to the variations in usage intensity, which will be the topic of the next section.

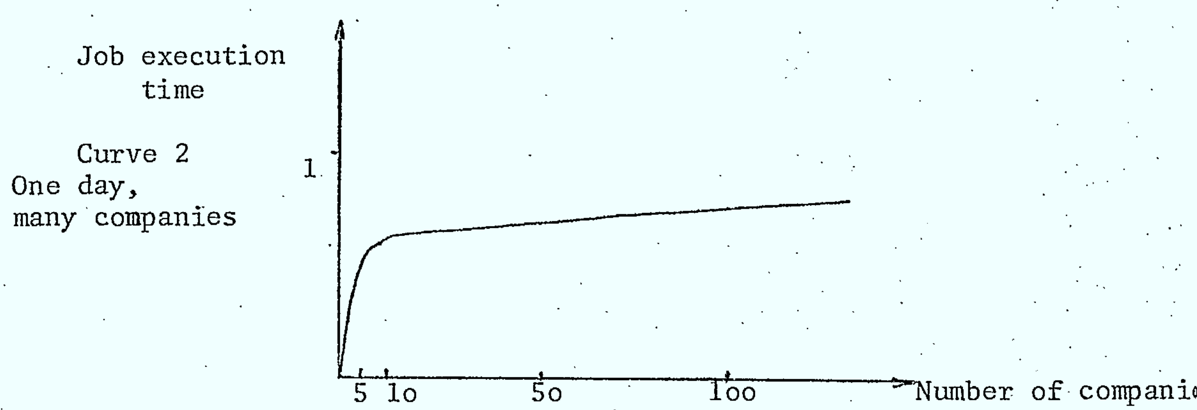
2.4.3.2 Unit Cost of One Program Run

An elementary program, PRICIT, was run. This program searches the stockmarket data file and prints out the price quotations for the desired stocks, for a specified day. Given a list of ticker symbols representing the stocks, the program first computes the relative location of these companies in the block of data that will be read in, then actually searches the block containing price quotations for all companies for the day requested, reads it in core as an array, selects the stocks and prints out the quotations.

The execute time thus depends on two parameters:

1. the number of days searched.
2. the number of companies searched.





Curve 2 flattens out very quickly: this is due to the fact that once the block corresponding to the right day is in core, it is almost immaterial whether 10 or 200 quotations are asked. Neglecting some non-linear terms in curve 2, a relation giving the job CPU time against the number of days (N_D) and the number of companies (N_C) was arrived at:

$$CPU_t = N_D(r+sN_C) + t + uN_D \qquad \text{with } r, s, t, u = \text{coefficients}$$

which reduces in: $CPU_t = N_D(u + r + sN_C) + t$[1]

From this formula, it is possible to predict access cost: assuming a CPU time charged at rate p_1 and a connect time charged at rate p_2, and a given proportional relationship v between the size of request ($N_C N_D$) and output time, the cost per access is:

$$C = [N_D(u + r + sN_C) + t]p_1 + vN_C N_D p_2$$ [2]

With X as a number of accesses per period of time, the cost to the user accessing PRICIT would be: COST = X.C+F (F being the fixed costs associated with the rent of terminal and lines).

We cannot, unfortunately, generalize too fast, because PRICIT is not by far the only program. However, the parameters N_D and N_C which define the search profile, can help to categorize users,

[1] two similar models can be found in recent articles of the Journal of Chemical Documentation: - "the development of a general model for estimating computer search time", Park, et al., Vol. 10, no. 4, 1970. - "Evaluation of search time for computerized IR systems", Ware et al, Vol. 12, no. 4, 1972.

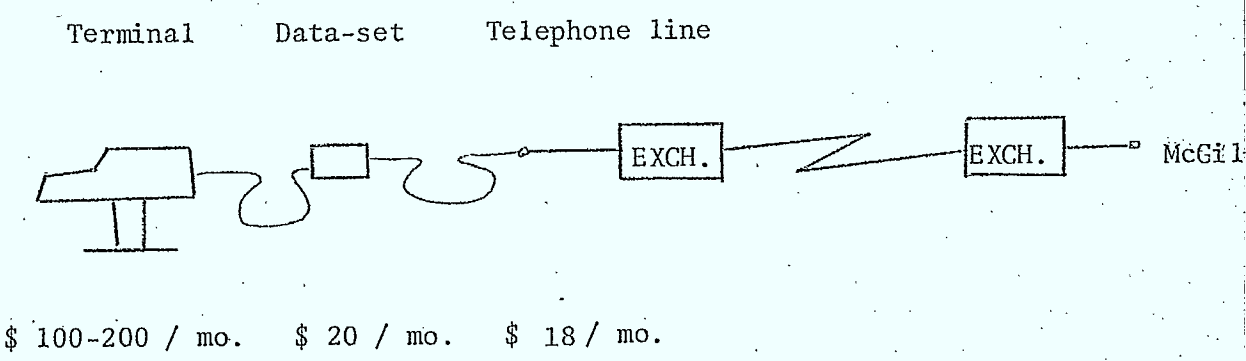
[2] Original estimates of u+r, s, and t for PRICIT were: u+r = 1.15 sec., s = .0053 sec., t = .80 sec.

a large N_C relative to N_D being typical of portfolio managers, and the converse being true for security analysts. Different classes of programs can be accommodated by different coefficients u+r and s.

This model can be useful for justifying a unit-pricing policy, according to which users would be charged on the basis of queries of "calls".

2.4.3.3 Additional Cost to the User

We mentioned in passing that there were some additional costs for terminal and line rent. The terminal monthly rental charge is between $100 and $200 (from Bell or General Electric). To each terminal must be attached a modem or data-set[1] (leased from Bell) with a monthly rental charge of $20. Eventually, the user has to pay for a business line, which amounts to $18/mo.:



[1] Its role is explained in section 3.1.2 below.

2.4.4 The Naive Cost Model — Part 2

In this subsection, we are concerned with costs that are dependent on the data-bank usage. Communication lines, data set, terminal rent are fixed inputs, but the computer costs of access are the variable inputs. These variable inputs can easily be related to:

(i) $X = X_4$: the number of accesses to the data-bank per period

(ii) $N_D = X_5$: the number of series searched and printed per access.

Then the cost of the user-servicing subsystem is:

$$COST = X_6 \{X_4[(C_4 + C_5) X_5 + C_6] + C_7\}^{1} + C_8$$

where X_6: number of users of the data-bank

C_4: CPU time cost per series searches: $(u+r+sN_C)p_1$

C_5: connect time per series searched and printed: $(vN_C)p_2$

C_6: program load time: (tp_1)

C_7: terminal, data set, line rent. C_8: communication controller.

This descriptive model is as crude as the part 1 naive model, although it might prove to be a rather simple device for predicting the overall usage-related system cost.

---

[1] King and Bryant in "Evaluation of information services and products" developed a similar model for retrospective search systems.

APPENDIX 2

USER NEEDS QUESTIONNAIRE

[CHAPTER 6]

THE METHODOLOGY OF THIS STUDY

6.1. Objectives of This Study

The information required in any user survey is dependent on the objectives. It is necessary therefore to repeat the objectives of this research report.

These objectives are the following,

(a) to describe the existing sources of information used by professional investors, and to ascertain the main differences between brokerage and institutional need.

(b) to formulate ideas of how professional investors think their future information needs will be met.

(c) to make recommendations as to how financial information can be improved, and future requirements met.

(d) to utilize the methodology, which has been developed in scientific and social science information need and use surveys, in the investment information field.

(e) to comment on the relevance of future computerized information systems.

As noted in the previous methodology chapter, King and Bryant distinguished two types of situation. In the first the purpose of the study is to add to the store of knowledge concerning information flows, man-machine interfaces etc.[1] The second evaluation situation arises when facts are needed about users of specific or contemplated systems.

This current information survey is largely of the first type in that the purpose of this study is to add to the store of knowledge concerning information flow. King and Bryant state that in this type of study, needed information is loosely structured, and often the structure of the study develops only after the investigation is under way. This has been the case for this study.

However, several members of the financial community are subscribers to existing information services and data-banks. Therefore, there is an element of the second type of situation in this study. In this second situation there is a sharper focus on specified questions.

6.2. Definition of This Study

In the previous chapter on methodology, 'user studies' were discussed. The distinction between use of, demand for, and need for information was made. This research report is largely a 'use of' type of user study, as the prime concerns of the questionnaires and interviews are the information sources utilized

---

[1]King & Bryant, op. cit., p. 240.

by analysts. The 'demand for' aspect is also brought out, though to a lesser extent, because some of the intended respondents of the study do make use of information services. Consequently, information demand can be assessed. The questionnaires and interview questions make little attempt to assess information needs, especially of unfelt or inarticulate needs.

A distinction was also made in the previous chapter of the different meanings of the word 'use' in information studies. It can mean the gathering stage of use, or it can mean the use to which information is put once it has been collected. This study is clearly of the former type, although Chapters 3 and 4, on the Canadian financial system and previous investment studies respectively, discussed the use to which information is put. For example, the research report of a company's prospects for the analysts and portfolio managers of the institutions.

In terms of the five key aspects of communications, discussed in the methodology chapter - source, message, channel, receiver, and response, the questionnaire is largely concerned with the source element. This source element is present in terms of - from the company under discussion, not from the company under discussion, from general sources, from broker analysts, from Government sources etc. The message aspect is present in terms of annual report data, supplementary statistics data, conversations with management, etc. The channel aspect is covered by the media distinction between, for example, accounting data from the annual reports generally made available and the same data but from a computerized data bank. The receiver aspect is present in terms of broker analyst, portfolio manager etc. The response aspect is, however, not covered (although the response of an analyst may be to recommend that a certain stock be bought, sold or held).

The interview questions are less concerned with the data source, but more concerned with the use to which the data is put.

6.3. Draft Questionnaire (Preliminaries)

A draft version of the questionnaire, which had been designed in conjunction with the Financial Research Institute on the basis of the previous Opinion Research Corporation survey, was used in pre-test interviews with members of the financial community. This draft questionnaire is shown in the Appendix.

The following people were interviewed - 2 managers of investment research of banks, an institutional sales representative of a brokerage company (who answered on the basis of his previous experience as a broker analyst), a research analyst of a trust company whose main area of interest was the banking industry, an investment manager of a utility company, and the investment editor of a financial newspaper.

It was obvious that they were many faults with the initial questionnaire. Firstly, the respondent was asked to rank the different items within each category. For example, under the heading General Communications the respondent was asked to rank annual reports, supplementary statistics, management reports and news, press releases. This proved relatively difficult, and many respondents thought that a grading scale would have been better.

The second fault was the applicability of the item to the respondent. For example, the direct contact category was taken differently by broker analysts and institution analysts. The broker analysts took it to be with the company under review, while the institution analyst took it to be with broker analysts. Other problems were encountered with the definition of the items.

The questionnaire was then revised with some of the wording made clearer, and the ranking system was changed to a grading scale. The revised questionnaire is discussed below in section 4.

The above analysts were also asked preliminary interview questions, which are discussed in section 5. These questions were of a probing, exploratory nature, and therefore it was necessary to ask these questions rather than put them in a questionnaire.

6.4. The Questionnaire

The revised questionnaire that was used in this study is given at the end of this section and in the Appendix.

6.4.1 General Questions

The first five questions are of a general nature requesting general information from the respondent. The person's name nor the name of his company were not asked to assure anonymity. No direct reference was made in the introduction of the questionnaire to the Financial Research Institute, which had assisted considerably in this study and the questionnaire designs, in order not to bias the responses.

Question 1.1 asks the nature of the firm's business - the main breakdown that is being explored here is the difference between broker and institution. Within institution there are many possibilities including banks, trust companies, insurance companies etc. The questionnaire can then be coded for analysis in as many ways as required, i.e. broker-institution, and then within institutions.

Question 1.2 asks the city location of the office. The previous study by Opinion Research Corporation, on which this study is partly based, distinguished between New York City and non-New York City information use as the vast majority of brokers are situated in that city (see Chapter 4). In this current study, information differences between Montreal and Toronto can be sought. Also, a breakdown of major city and smaller city information use is possible - for example between Toronto and London, Ontario.

Question 1.3 asks the respondent to state his position in the company - the major breakdowns we are requiring information on are the differences between analyst, representative and portfolio manager requirements, and the differences between brokers and the institutions.

Question 1.4 asks the respondent to state the activity that takes up most of his time. This is a direct result of the pre-test interviews with the analysts and brokers. Some respondents intimated that though they had a certain title they spent the majority of their time on a different activity. For example a broker representative could also be concerned with portfolio management for an institution, and an institution analyst could be largely concerned with economy, company or industry analysis.

Question 1.5 asks the main area of industry or other specialization of the respondent. Differences by industry specialization are being looked for here.

These five questions facilitate the analysis of the data. They enable the major breakdown of the data into analyst and non-analyst, and institution and broker, as well as other categories depending on the actual response of the survey.

The framework of the main part of the questionnaire has been borrowed from the previous study of ORC, which used the three classifications - General Communications, Special Communications and Non-Company Communications. This was solely on a criterion of information source. As mentioned in Chapter 2, this study was criticized for not including computerized data banks in its classification of data. The current research report has added this fourth category to their original classification, and has more rigorously defined the classification items.

The framework of the questionnaire is therefore the following - general communications made available by the company or agency under review; direct contact on the part of the analyst with the company under review for specific information; non-company communications from third party sources, and computerized date banks and information services.

For each section, the respondent is asked to state whether or not he uses the respective source, and if he does to rate that source on a five point scale based on a division of least useful, somewhat useful, useful, very useful and extremely useful. The replies can then be analysed differently depending on use and non-use, and on the extent of use.

6.4.2 General Communications (with the company under review)

The items under this category are annual reports, quarterly reports, management reports, supplementary statistics, news and press releases and Statistics Canada bulletins. These data items, and those of other classifications, were described in Chapter 2 on information sources.

The item Statistics Canada bulletins was included here for the consideration of an economist employed by a broker or institution in a capacity of examining Government policy. In his case the company or agency under review is the Government.

6.4.3 Direct Contact

The direct contact items are personal contact with company officials, group contact on an informal basis such as a management presentation to financial analysts, communications with company officials and tours and other special events.

6.4.4 Non-Company Communications

This item refers to communications either not specifically from the company under review, or not concerned with a company but with an industry or economy. In this category we include the written reports of analysts (these could be from brokers or institutions), conversations with colleagues and analysts, industry trade magazines, stock exchange information, Government affiliated sources such as Bank of Canada, Statistics Canada and Wire Services such as Dow-Jones.

6.4.5 Data Banks and Information Services

This is the classification which was not included in the 1967 ORC study. It comprises computerized data banks of an outside service organization, such as Financial Research Institute, computerized data banks of an in-house nature, data obtained in machine readable form, such as Statistics Canada and Cansim, and newspaper printed services such as the Financial Post cards.

6.4.6 The Sampling Frame

It was stated in the methodology chapter that a sampling frame should ensure that all types of data, types of work, types of information use, channels of communication and types of user behaviour are adequately represented.

This questionnaire is partly focused towards the research operations of the analysts of the brokers and the financial institutions, the relationship of their research to the sales representatives of the investment dealer-broker and to the portfolio managers of the brokers and institutions. As far as types of data are concerned, this survey does not emphasize current price and trading data, which would be of prime interest to the bond and stock trader of the investment dealer, but instead emphasizes this research process.

As noted above in the methodology chapter, any questionnaire sampling frame is always limited by the actual response. It is this reponse which determines the types of work, types of data, types of information use, channels of communication and types of user behaviour that are actually analysed.

It is expected that a broader survey of the Financial Analysts across Canada will soon be undertaken. This will broaden the size of the current sampling frame, and hopefully produce results which validate those of this research report.

FIGURE 6.1 THE QUESTIONNAIRE USED

Investment Community Information Source Survey

Information sources for the professional investment community can be classified under four main categories - general communications made available by the company or agency under review; direct contact on the part of the analyst with the company; non-company communications and computerized data banks and information services. This questionnaire follows this general classification.

First of all a few brief questions. It is not necessary to write your name, or the name of your company. Full anonymity can be assured.

1.1 The nature of your firm's business e.g. brokerage, brokerage-dealer, banking, trust etc.

____

1.2 The City location of your office.

____

1.3 Your position e.g. analyst, institutional sales, portfolio manager, etc.

____

1.4 The activities that take up most of your time. Please tick (several if necessary).

company analysis ( )
industry analysis ( )
economic analysis ( )
institutional sales ( )
retail sales ( )
portfolio management ( )
underwriting ( )
security trading ( )
other (please specify) ______________

1.5 Major area of specialization e.g. oils, banks, mining industry, pulp and paper etc.

____

For the following four sections, please indicate on the rating scales the strength of the value of the information sources to you by ticking the appropriate box.

Section 2. General Communications
(with the company or agency under review)

	not used	least useful	somewhat useful	useful	very useful	extremely useful.
2.1 Annual Reports	( )	( )	( )	( )	( )	( )
2.2 Quarterly reports, interim statements	( )	( )	( )	( )	( )	( )
2.3 Management reports, for example copies of presentations made by management, reports on annual meetings	( )	( )	( )	( )	( )	( )
2.4 Supplementary statistics, for analysts e.g. a breakdown of sales by division	( )	( )	( )	( )	( )	( )
2.5 News, press releases including bulletins, company advertising and company magazines	( )	( )	( )	( )	( )	( )
2.6 Statistics Canada bulletins, Bank of Canada reports, etc.	( )	( )	( )	( )	( )	( )

Section 3. Direct Contact
(with the company or agency under review)

	not used	least useful	somewhat useful	useful	very useful	extremely useful
3.1 Personal contact with company officials	( )	( )	( )	( )	( )	( )
3.2 Group contact on an informal basis	( )	( )	( )	( )	( )	( )
3.3 Communications with company officials	( )	( )	( )	( )	( )	( )
3.4 Tours and other special events arranged by management	( )	( )	( )	( )	( )	( )

Section 4. Non-Company Communications

	not used	least useful	somewhat useful	useful	very useful	extremely useful
4.1 Written reports of analysts	( )	( )	( )	( )	( )	( )
4.2 Conversations with colleagues and other analysts	( )	( )	( )	( )	( )	( )
4.3 Industry, trade magazines, including bulletins, business financial and professional publications	( )	( )	( )	( )	( )	( )
4.4 Information filed with the Stock Exchanges, Security Commissions	( )	( )	( )	( )	( )	( )
4.5 Government affiliated sources	( )	( )	( )	( )	( )	( )
4.6 Wire Services such as Dow-Jones	( )	( )	( )	( )	( )	( )

Section 5. Data Banks and Information Services

	not used	least useful	somewhat useful	useful	very useful	extremely useful
5.1 Computerized data banks of an outside service organization	( )	( )	( )	( )	( )	( )
5.2 Computerized data bank - in-house information service	( )	( )	( )	( )	( )	( )
5.3 Data obtained in a machine readable form e.g. Statistics Canada tapes	( )	( )	( )	( )	( )	( )
5.4 Newspaper, magazine printed service e.g. Financial Post Cards	( )	( )	( )	( )	( )	( )

Comments. Please feel free to further describe your information sources.

____

____

____

____

____

6. Summary

Could you now indicate an approximate weight percentage of the value of these major classifications to you (considering each classification as a whole). If possible, could you also give an estimate of the use to you of these data sources five years ahead.

	Indicate Weight Percent	
Major Classification	Now	In 5 Years
General Communications	___	___
Direct Contact	___	___
Non-Company Communications	___	___
Data Banks and Information Services	___	___
TOTAL	100%	100%

Comments: Please feel free to discuss your expectancy of future sources, or to discuss this questionaire.

____

____

____

____

____

Thank you for spending your valuable time completing this questionnaire.

Please return in the stamped addressed envelope to:-

John Meaker, Faculty of Management, McGill University, 1001 Sherbrooke W., Montreal, H3A 1C5, Que.

6.5. The Interview Questions

The following questions explore the reasons for using information sources and the data obtained from these sources. As these are concerned with ill-defined areas, they are much better suited to a personal question session, than to be included in a questionnaire where they could be misinterpreted.

Vickery, as noted in the methodology chapter, presented two lists of questions that could be asked in these interviews.[2] The first set is for assessing potential needs of a proposed information system; the second for gauging the reactions of users to an established information system. The questions of the first list were utilized in this study and are listed below:

Q.1 What is your main area of interest?

Q.2 Do you keep any personal information files? What is their content? How and when do you use them?

Q.3 The last time you needed job-related information, what was the first source you approached?

Q.4 What did you get from this source?

Q.5 How long did it take to get?

Q.6 Why did you use that source?

Q.7 How was the information used?

Q.8 What kind of information do you find difficult to obtain?

---

[2] Vickery, op. cit., p. 40.

Q.9 In what form do you prefer to receive information?

Q.10 What depth of information do you usually need (light, specific fact, detailed analysis)?

From his second list the following relevant questions were taken:

Q.11 How important are articles to you, e.g. Financial Analyst Journal, Journal of Finance?

Q.12 Do you follow up references in these journals? Were they of direct value?

Q.13 Do you receive bulletins from an information service? Are they of use to you? Does their coverage match your interests?

Q.14 Do you receive a Selective Dissemination of Information Service? (SDI was explained).

In addition the following questions were also asked to further determine information usage:

Q.15 How do you go about getting information?

Q.16 Is there one person that you go to generally for information?

Q.17 Do the companies send their annual reports to your analysts, do they keep them individually, do you have a central depository?

Q.18 Do you have an in-house librarian or information service?

Q.19 What use do you make of old information?

Q.20 What use do you make of libraries - e.g. University, Stock Exchange?

Q.21 What use do you make of surveys e.g. Financial Post Survey of Mines?

Q.22 Do you think too much information is available?

Q.23 How could you see the information improved?

Q.24 How important is the timeliness of information to you?

Q.25 How rational can you be in deciding the amount of historical data to be used in your analysis?

Q.26 What use do you make of computerized data banks? Do you have access to one in-house or outside?

Questions to users of data banks follow:

Q.27 Would you prefer more information on a computer data bank or less information but for a period of time?

Q.28 Would you prefer more packages available to you to improve your ability to assess data?

Q.29 Could you give an estimate of the value of a data bank to you?

[CHAPTER 7]

THE INTERVIEW AND QUESTIONNAIRE RESULTS

7.1. Preliminary Remarks

In total twelve members of the financial community in Montreal were interviewed and asked to complete the questionnaire. Two of these people passed on a copy of the questionnaire to other financial analysts, with the result that fifteen completed questionnaires were received. The author would have liked to have been able to interview more respondents, and to have received many more questionnaires. However, each interview took a long time to arrange as these are very busy people. It is anticipated that the questionnaire, which has been developed in this research paper, will be mailed to members of the Financial Analysts Federation across Canada in the near future, under the sponsorship of the Chapter Presidents. This mail survey should obtain a more representative and bigger response than this research paper. With this broader data base, it should be possible to validate the results and conclusions of the data presented here. The questionnaire to be used, the methodology to be employed, and the analysis of results to be undertaken, by this larger survey have been developed in this research paper.

The results of the questionnaire are analysed in section 7.2, and the interview responses discussed section 7.3. The comments on the questionnaire forms, further discussing these information sources and future information expectancies, are discussed in the latter section.

7.2. The Questionnaire Responses

The 15 completed questionnaires are analysed below. The actual responses and the complete raw data are given in the Appendix.

7.2.1 Replies to the General Questions

(a) Replies to Question 1.1: Table 7.1 below gives the nature of the business of the respondent's company. There were seven respondents from financial institutions (the banks, insurance companies and the trusts), six from broker-dealers and two other respondents.

Table 7.1

The Nature of the Business of the Firm

Nature of Business	Number of Responses
Broker-Dealer	6
Bank	2
Insurance	3
Trust	2
Newspaper	1
Utility Company	1

(b) Replies to Question 1.2: The city location of office replies were all Montreal as all the interviews were located there.

(c) Replies to Question 1.3: The positions of the respondent are given in Table 7.2. Two of the respondents, who had the positions of research managers in the financial institutions, have been classified here as analysts. The utility company financial analyst is also classified in this category.

Table 7.2

The Positions of the Respondents

Position	Number of Responses
Analyst	12
Investment Editor	1
Portfolio Manager (Broker)	1
Institutional Sales (Broker)	1

(d) Replies to Question 1.4: The replies to the question of the activities that took up most of the time are given below in Table 7.3. In general the brokerage analysts were concerned with institutional sales, as well as company and industry analysis, while the institutional analysts were concerned with economic, industry and company analysis in conjunction with portfolio management. The portfolio manager had a partial interest in security trading.

Table 7.3

The Activities Mentioned

Activity	No. of Times Mentioned
Company Analysis	10
Industry Analysis	13
Economic Analysis	6
Institutional Sales	4
Retail Sales	0
Portfolio Management	7
Underwriting	0
Security Trading	1
Other	0

(e) Replies to Question 1.5: The responses to the major area of specialization question are given in Table 7.4. Some of the respondents did not reply to this question, but repeated their job position such as portfolio manager, or institutional sales. The research managers are concerned with all the specializations of their research groups.

Table 7.4

Major Areas of Specialization

Specialization	Frequency
All (the research managers)	2
Construction	2
Pulp and paper	2
Banks	2
Beverages	1
Tobacco	1
Capital Goods	1
Chemicals	1
Textiles	1
Real Estate	1
Oils	1

(f) Commentary: Although fifteen respondents are barely sufficient for the making of valid conclusions, the responses are probably sufficient for the making of inferences as the data is fairly inclusive of the activities, positions and specializations of the Montreal financial community. There are divisions among the respondents between analysts and non-analysts, and between the institutions and the brokers. These divisions should be sufficient to allow inferences to be made of the information usage of these respective groups.

7.2.2 The Replies to the Questions on Information Sources

The data on information uses are analysed below in order to obtain rankings of the information sources of the different categories of respondents of the sample. As the questionnaire asked the respondent to tick the various classifications most applicable to him, it is first necessary to allocate a number scale to these classifications to be able to perform manipulative operations with the data. The arbitrary scale of 0 to 5 has been chosen, with 0 referring to not used and 5 referring to extremely useful. Any increasing number scale could have been chosen as the operations performed on the data are for relative purposes only. No absolute measure of information usage are used in this analysis. The raw data is included in the appendix with this coding employed -

0 = not used
1 = least useful
2 = somewhat useful
3 = useful
4 = very useful
5 = extremely useful

(a) Information by Source

The responses for each question on information usage have been summed across the rows (with the assistance of an electric calculator). These totals are given in the appendix. Table 7.5 below depicts the rankings obtained from these sums with rank 1 referring to the highest response sum, and rank 20 the lowest.

Table 7.5

Overall Ranking of Information Sources
(15 Respondents)

Questionnaire Number and Item Description		Ranking
2.1	Annual Reports	1
2.2	Quarterly Reports	2
2.4	Supplementary Statistics	3
4.1	Written reports of analysts	4
4.3	Industry Trade magazines	4
5.4	Newspaper, Printed Services	4
3.1	Personal Contact	7
4.4	Stock Exchange Information	7
4.6	Wire Services	9
4.2	Conversations	10
2.5	News, Press Releases	11
3.3	Communications with Company	12
2.3	Management Reports	13
4.5	Government Affiliated Sources	13
5.1	Data Banks (Outside)	13
3.2	Group Contact	16
2.6	Statistics Canada, etc.	17
5.2	Computer Services (in-house)	18
3.4	Tours	19
5.3	Machine Readable Data	20

The row summation and ranking procedure has been repeated for the analysts, the brokers and the institutions, and the rankings obtained incorporated in the table below. For example, for the analyst category the portfolio manager, the investment editor and the institutional salesman have been removed from the raw data and the row sums recalculated. The overall ranking is repeated in this table.

Table 7.6

Comparative Rankings of Information Source Usage

Questionnaire Number and Item Description		Overall 15 Respondents	Analysts 12 Respondents	Brokers 6 Respondents	Institutions 7 Respondents
2.1	Annaul Reports	1	1	2	2
2.2	Quarterly Reports	2	1	5	1
2.4	Supplementary Statistics	3	3	5	3
4.1	Written Reports of Analysts	4	8	11	4
4.3	Industry, Trade Magazines	4	6	2	8
5.4	Newspaper, Printed Service	4	7	10	4
3.1	Personal Contact	7	4	1	8
4.4	Stock Exchange Information	7	4	8	8
4.6	Wire Services	9	10	4	7
4.2	Conversations	10	10	5	13
2.5	News, Press Releases	11	13	12	8
3.3	Communication with Company	12	10	8	13
2.3	Management Reports	13	15	15	8
4.5	Government Sources	13	8	14	13
5.1	Data Banks (outside)	13	15	15	16
3.2	Group Contact	16	15	19	4
2.6	Statistics Canada	17	14	15	17
5.2	Computer Services (in-house)	18	18	13	18
3.4	Tours	19	19	18	19
5.3	Machine Readable Information	20	20	20	20

These rankings are referred to in the commentary below.

(b) Summary of Information Usage

In the final section of the questionnaire, the respondents were asked to weight the value of the major classifications (considered as a whole) to them now and in five years. These data are included in the appendix. The date for each classification have been added across all the individuals, and converted to percentage form for Table 7.7 below.

Table 7.7

The Respondent's Summary Classifications

Classification	Now	5 Years Time
General Communications	28.3%	23.9%
Direct Contact	28.9	26.1
Non-Company Communications	26.4	26.4
Data Banks	16.4	23.6
	100	100

These data have been analysed for Table 7.8 into institution, broker and analyst categories to obtain the indicated rankings by usage for the present and the future.

Table 7.8

Ranking by Classification and Type of Respondent

Classification	Analysts Now	Analysts 5 yrs.	Brokers Now	Brokers 5 yrs.	Institutions Now	Institutions 5 yrs.	Overall Now	Overall 5 yrs.
General Comm.	1	3	2	3	1	3	2	3
Direct Contact	2	2	1	1	2	2	1	2
Non-Company Comm.	3	1	3	2	3	1	3	1
Data Banks	4	4	4	4	4	4	4	4

Each individual's responses were then analysed to ascertain which of these categories were expected to rise, fall or remain constant in percentage use. This data is presented in Table 7.9. Note that only 14 respondents replied to this section.

Table 7.9

Changes of Classification Expected by the Respondents

Classification	Falling	Constant	Rising
General Communications	6	7	1
Direct Contact	6	7	1
Non-Company Communications	2	10	2
Data Banks	0	2	12

7.2.3 Commentary

The results obtained in this study can be compared between major categories, such as brokerage compared to institution, and can be compared with the ORC study of US financial community information source usage.

(a) Information by Source: Referring to Table 7.6, these results indicate that the brokers rank personal contact with the company under review, annual reports, industry trade magazines and wire services very highly. The institutions (as reported by this data) have ranked annual and quarterly reports, supplementary statistics, group contact and the written reports of analysts very highly. The analysts, as a group, rank these sources very much the same as the overall ranking (not surprisingly as analysts comprise 80% of the overall respondents), except that they make relatively less use of written reports and more of government sources.

Of note is that both government sources and computer data banks are ranked quite low in the table - government sources are ranked 13th overall, outside data banks 13th, and in-house data banks 18th. Consistent low ranking information sources are tours and data in a machine-readable form. For example, none of the companies polled used Cansim independently of the Financial Research Institute data bank services. However, several companies do obtain these data series from the FRI, and therefore its use is part of the overall outside data bank classification results.

The ORC study did not present results in this overall list form but merely within classification. The data of this current study have been restated from Table 7.6 to be directly comparable to the ORC rankings.

Table 7.10

Comparison with the ORC Study

Major Classification Item	ORC Ranking	This Study		
		Overall	Instits.	Brokers
General Communications				
Annual Reports	1	1	2	1
Quarterly Reports	2	2	1	2
Management Reports	3	5	4	5
Supplementary Stats.	4	3	3	3
Press Releases	5	4	5	4
Direct Contact				
Personal Contact	1	1	2	1
Group Contact	2	3	1	3
Communications	3	2	3	2
Tours	4	4	4	4
Non-Company Communications				
Written Reports of Analysts	1	1	1	5
Industry, Trade Magazines	2	2	2	1
Stock Exchange Information	3	3	4	4
Conversations	4	5	5	2
Wire Services	N/A	4	3	3
Government Sources	N/A	6	6	6
Data Banks				
News, Printed Services	N/A	1	1	1
Outside Data Banks	N/A	2	2	3
In-House Data Banks	N/A	3	3	2
Machine Readable Data	N/A	4	4	4

N/A: not applicable.

Table 7.10 indicates that the results of this study are fairly consistent with the ORC study (which had 1,758 total respondents). However, the sets of results may not be entirely comparable as the two studies used slightly different questionnaire procedures. One of the main differences is that US analysts rank management reports higher in the General Communications Category. The breakdown of non-company communications for brokers by ORC (this data is in the appendix) also places written reports of analysts as rank 5, with the same other rankings as this study. In the Direct Contact Classification of ORC for brokers, group contact is also ranked third.

(b) The Summary of Usage: The summary of sources (Table 7.7) indicates that the respondents expect the use of Data Banks to increase over the next 5 years at the expense of General Communications, and Direct Contact. Overall, it is expected that non-company communications will replace the direct contact category as the first ranking source, Data Banks will remain the fourth ranking source. Of note is that for the brokers Direct Contact will remain the prime source of information, while General Communications will lose ground to Non-Company Communications for all categories of respondents.

7.3. The Interview Results

Twelve separate interviews were conducted with the following people - four from brokerages, two of those analysts, one a portfolio manager and one an institutional salesman; two managers of investment research at banks; two investment analysts from an insurance company; two analysts from a trust company; the investment editor of a financial newspaper; and a financial analyst from a utility company. These were twelve of the fifteen respondents that completed the questionnaire.

The interviews were slightly disappointing in the results they produced. The main reason for this was that, as the respondents are very busy people, they could only grant thirty to forty-five minutes for the total interview session. As all the interview respondents also completed the questionnaires, the questionnaires and the discussion that was generated generally took fifteen to twenty minutes. This left between ten and twenty minutes for the interview questions. It was thus quite difficult to explore in depth the information requirements and sources of the respondents. When asked a specific question some respondents tended to speak at considerable length, particularly as some questions required development of the reply. Also, the respondents were experts in their field, and seemed to enjoy the opportunity of talking of their information sources.

Consequently, the reporting of the interview questions in the following section is somewhat impressionistic in nature. It is a result of the interviews and the comments of the respondents when completing the questionnaires. No one person was asked all these questions, though many respondents had time to answer several. The list of questions is that developed in the previous methodology chapter.

Q.1 What is your main area of interest?

This question is really redundant as it was asked in the questionnaires. Three financial analysts who were not interviewed, completed the questionnaire.

Q.2 Do you keep any personal information files? What is their content? How and when do you use them?

The respondents all kept some information files. All the analysts individually received financial reports from the companies they covered. To quote one person - "analysts are great hoarders of information". The financial newsletters and information services are also kept by most respondents. Several of the people interviewed had their own complete sets of journals and government monthly newsletters. Most of them received several newspapers daily - for example, the portfolio manager received six financial newspapers every day.

The content of these files is self explanatory - analysts store information on the companies and industries they cover, as well as government, journal and newsletter information.

The answers to how and when they used these sources were vague - for example, as necessary.

Q.3 The last time you needed job related information, what was the first source you approached?

Generally, the company annual report was quoted as the first source of information. However, it depended on the type of information required. For example: "if one was looking for the sales of a company for the last five years an index card, such as a Financial Post card, would be the easiest means of obtaining it." A brokerage analyst stated that if qualitative rather than quantitative information was required he might go straight to the company concerned.

Q.4 What did you get from that source?

The replies to this question depended on the data source - for example company information, economic data, industry trends. Management interviews were often used to "fill-in the gaps".

Q.5 How long did it take to get?

Often the analyst had the information in his own files. Information from Stock Exchanges can take about a week to arrive.

The New York Stock Exchange was quoted as the fastest followed by the Ontario Securities Commission.

Q.6 Why did you use that source?

The general replies were - easiest, fastest, most convenient. Past financial statements were used rather than index cards or computer services, to be able to adjust past data for such circumstances as changes in inventory evaluation. To quote one analyst - "a great deal of information on financial statements as notes or adjustments never gets on information cards or computers."

Computers were generally used when a great deal of information was required on several companies, or screening was required. The statistical reports for analysts and Stock Exchange prospectuses contained a great deal of detailed information.

Q.7 How was that information used?

Generally the reply was that the information was incorporated in a report, or used in an evaluation.

Q.8 What kind of information do you find difficult to obtain?

One reply was - "what will the future price of the stock be?" However, most information can be obtained. It is the timeliness of the information that is most important.

Q.9 In what form do you prefer to receive information?

The replies to this question were somewhat inconclusive. The research managers and the institutional salesman mentioned verbal form from analysts. The basic reply was that it depended on the information sought. Again timeliness was mentioned as more important than information form. Analysts did not mind if information came from computers or index cards as long as it was correct.

Q.10 What depth of information do you usually need (light, specific fact, detailed analysis)?

The replies stated that the depth depended on the information being sought. Industry, trade magazines are used for general background reading. Financial reports, Stock Exchange prospectuses, and supplementary statistics are used for specific facts. The reports of analysts are used for detailed analysis of a company.

Q.11 How important are articles to you - e.g. Financial Analyst Journal, Journal of Finance?

The Financial Analysts Journal was quoted as being much more important than the Journal of Finance. Several analysts spoke of skimming through the Financial Analyst Journal. Financial newspapers were mentioned by one person as being more important than journal articles.

Q.12 Do you follow up references in these journals? Were they of direct value?

Sometimes and rarely were the main replies. One financial institution is putting its own Key-Word-in-Context (KWIC) index on a computer based on journal article titles to provide an article retrieval service.

Q.13 Do you receive bulletins from an information service? Are they of use to you? Does their coverage match your interests?

Several respondents received information services such as the Financial Post services, Graphoscope and Investment Newsletter. Generally they were useful, though several respondents mentioned a great deal of duplication in these services, and that they were not very selective. Generally, the coverage matched their interests.

Q.14 Do you receive a Selective Dissemination of Information service? What do you think of the possibility of SDI with regards to financial information?

Selective Dissemination of Information was explained as (a) the user notifies the information service of his interests, (b) an information service supplied available information based on these interests, (c) the user subsequently informs the service of the usefulness of the information, and updates the profile.

The general reply was no, SDI was not received. One respondent spoke of a service such as this being available in the U.S.A. For example, if an analyst was interested in anything to do with gold, he could indicate this to an information service and receive everything concerned with this metal.

One respondent mentioned that he had his own SDI - he asks his secretary to scan the daily Dow-Jones ticker tape to indicate the items that would interest him, based on a list of his interests.

Several respondents mentioned that more selective information would be a great benefit, and save them considerable time. They would require more specific information on any recommended service, however, before being able to consider it properly.

Q.15 How do you go about getting information?

Most of the replies to this question are covered elsewhere. One useful comment was made about company sponsored tours - "tours can be extremely useful first time around to new analysts. Subsequently they are less useful."

A broker analyst noted the following further ways of obtaining information of a company - "its competitors, industry trade associations, its advertising agents, its suppliers, its sales agents, government agencies and by listening carefully in elevators."

Q.16 Is there one person that you go to generally for information?

The general reply was no. Sometimes a secretary is used as the keeper of information files, but basically analysts keep their own data. Specialists are consulted for specific information - for example one of the research managers mentioned that he talks to the appropriate analyst for industry and company information.

Q.17 Do the companies send their annual reports to your analysts, do they keep them individually or do you have a central depository?

Generally, analysts receive annual reports directly from the companies concerned. Otherwise a copy can always be obtained within short notice. Analysts keep their own copies, and have built up their own information files.

Q.18 Do you have an in-house librarian or information service?

In most companies a secretary functions as the custodian of certain information files. There were no in-house librarians for financial data. One company, as previously mentioned, is planning its own information service of journal article titles based on a KWIC index.

Q.19 What use do you make of old information?

One analyst replied - "the focus of all our work is towards the future, therefore the past is only useful if it can be used to help predict the future course of events." Past data is often used in trend analysis to help predict the future values of such items as earnings and dividends.

Q.20 What use do you make of libraries? e.g. University, Stock Exchange.

Generally very little, Stock Exchanges sometimes are used for detailed information not available to companies.

Q.21 What use do you make of surveys? For example, the Financial Post Survey of Industrials.

Some respondents thought that surveys were useful for specific information on the companies of an industry, especially those not covered on data banks and information services. Others thought surveys not useful at all.

Q.22 Do you think too much information is available?

The views on this question were conflicting. Some thought that there was too much, some thought not enough of the items they were concerned with. A representative conclusion might be that there is not too little information available.

Certain analysts thought that too much information was not a bad thing, - "if it was missed in one source, or if that source was late, it could be picked up in another." Another replied that too much information was received by the analyst, requiring him to be very selective in what he read. One analyst replied that he read only a very small part of the information he received, the rest was skimmed.

Coupled with the quantity of information is the timeliness of the data - if one information source had the information quicker than another then it was beneficial to the analyst.

Several analysts mentioned that there was much duplication, especially in investment newsletters.

Q.23 How could this information be improved?

Several mentioned avoid the data duplication. An institutional analyst replied that he would like to be able to receive in one source a summary of all the broker research analysts' projections of earnings and dividends for each major company. He would be then able to decide for himself from this list. When confronted with this possibility another analyst mentioned that this was available in the U.S.A.

Q.24 How important is the timeliness of information to you?

Several respondents stated extremely important. However this depends on the type of data and the purpose to which it is to be put. If it is for a past trend analysis then the absolute timeliness of the data is not critical. If, on the other hand, it is for stock selection or portfolio evaluation then timeliness of the data is very important.

One analyst, who was concerned with the relationship of the market to economic variables, criticized government statistical data for being very late. He mentioned that it can take two months for certain GNP time series to be updated. He noted that the U.S.A. is much faster in this respect. There, this data is available much earlier, correct in direction it not in absolute magnitude, and subject to adjustment to the correct absolute amount. This analyst prefers earlier, perhaps slightly inaccurate data, to later accurate data.

Q.25 How rational can you be in deciding the amount of historical data to be used in your analysis?

One respondent replied that he uses five years data for fundamental company statistical work, and up to 10 years for technical market data analysis. It depends on how the circumstances have changed since then. Often analysts use the data most readily available, - if it is on a data bank they may use it as far back as possible. The consensus was between 5 and 10 years of past data.

One broker mentioned that they generally forecast up to two years ahead for serious purposes, and up to four years ahead for general guidelines.

Q.26 What use do you make of computerized data banks? Do you have access to one - in-house or outside?

Several of the respondents had access to data banks in their companies, both in-house and to an outside service bureau. The in-house data banks were used for specific uses - for example one financial institution has developed its own in-house credit evaluation model to analyse companies that are not on the outside data banks. The raw data on these companies is obtained from information services. The company of one of the brokers has developed its own market price index similar to the Toronto Stock Exchange index.

The outside data banks were used for fundamental data retrieval, to obtain data for such purposes as portfolio evaluation, and to provide data for specialized studies. One company used the fundamental data only for the evaluation of bonds.

The companies using the price data generally accessed the data banks regularily - once a week, or every day, to evaluate the portfolios they managed. The accesses to the fundamental data series were often made on an ad hoc basis.

The economic time series data were used by some companies to provide data for econometric models to ascertain the relationship between selected companies and industries and the economy.

One analyst in a brokerage company not using computerized data banks for information retrieval mentioned that - "research analysts do not really need computers - they can build up their own information sources, which they supplement with first hand data from the companies."

Q.27 Would you prefer more information on a computer data bank or less information but for a longer period of time?

The unanimous reply to this question was that they would prefer information on more companies rather than longer time periods (as long as at least 5 years of data was available).

Q.28 Would you prefer more packages available to you to improve your ability to access and manipulate data?

One respondent would have preferred more packages available, but of course these would have been for his specific needs. One research manager stated that he would like to have the programs to manipulate the data written in-house to ensure that his specific requirements were met.

Q.29 Could you give an estimate of the value of a data bank to you?

Those respondents, whose companies had in-house computerized data banks, thought that these data banks were invaluable, especially as the programs were 'tailor-made'. Those companies using outside data banks found them quite useful for obtaining large quantities of past data for company analysis, and for obtaining current price data.

No respondent referred to the accuracy of the data on the data banks (Shaw and Archibald mentioned that several of their respondents criticized data banks for the accuracy of the data). One can infer from this admittedly small sample that the data accuracy has improved since 1970.

One respondent mentioned the problems of comparing data from the data banks between companies - there could be differences of interpretation of the standards for reporting accounting information, thus making comparisons difficult and sometimes necessitating a resort to the actual financial statements.

Q.30 How do you expect your information sources to change over the next five years? (This question is from the questionnaire).

One comment was - "you never project five years into the future in the investment business."

Another analyst predicted that there would be a smaller, more compact brokerage community.

A third comment was that there will be more information on the international scene, - "international economic information will become more important. Behaviour of a stock or bond market cannot be explained by domestic data alone. This will become increasingly more obvious and the need for foreign data more pressing."

A fourth comment was that there will be improved disclosure from companies in the future, and more general information available.

A fifth respondent predicted that individual direct contact may go down because of 'insider trading' scares. He mentioned further that several companies have become less willing to talk to analysts.

A final comment is that improved service from Statistics Canada, and better timeliness of data, is to be expected.

[CHAPTER 8]

CONCLUSIONS

This chapter makes inferences and draws conclusions based on the survey data presented in Chapter 7, and the previous descriptive chapters. The five main objectives of this study are considered in turn below, and conclusions related to these objectives stated.

Objective 1: To describe the existing sources of information used by professional investors and to ascertain the main differences between brokerage and institutional use.

The description of the existing sources of information has been covered quite well in Chapters 2, 3 and 4. There is therefore no need to reconsider this description here.

Table 7.5 presents the overall ranking of information sources that was developed from the results of this survey. Annual and quarterly reports, supplementary statistics, written reports of analysts and industry, trade magazines were given the highest rankings by the overall respondents of this survey. The five lowest ranking information sources, on the basis of this data, were group contact, government sources such as Statistics Canada, in-house computer services, tours and machine-readable data.

In Section 7.2.3 the data of this survey were compared with the rankings obtained from the ORC study. The rankings from this study are very similar to those obtained from the American study (see Table 7.10). Two inferences that can be drawn from the differences in rankings are that management reports are used by financial analysts slightly less in Canada than in U.S.A., and that the rankings of group contact and communications with the company are reversed in the two studies. These differences are quite marginal; therefore a conclusion could be that the rankings obtained from this limited Montreal-based survey are validated by those of the much larger U.S. survey.

The second part of this objective is to ascertain the main differences between broker and institutional data use. These differences can best be made by using Table 7.6 which presented comparative rankings of information source usage. In Table 8.1 below, the differences between the rankings of brokers and institutions have been made and ranked in order of largest difference. Only the 7 largest differences are given.

Table 8.1

Differences in Rankings Between Institutions and Brokers

Item Description	Institutional Ranking	Broker Ranking	Diff- erence	Rank
Group contact	4	19	15	1
Conversations	13	5	8	2
Management Reports	8	15	7	3
Personal Contact	8	1	7	3
Written Reports of Analysts	4	11	7	3
Industry, Trade Magazines	8	2	6	6
Newspaper, printed service	4	10	6	6

These differences are largely to be expected due to the nature of the different tasks of the two groups. For example, it is the nature of the brokerage research task to undertake original research on a company, and to be the first to present an analysis of new data on this company to the institutions. Hence, the very low ranking by the brokers for group contact. Group contact, based on the results of this survey, is largely for the information benefit of the institutional analysts. As a direct contrast, there is the high ranking by brokers for conversations with other analysts, and personal contact with the company under review. Their high ranking for industry, trade magazines can be explained by the fundamental research nature of brokerage activities. The industry magazines are used to obtain ideas of how future dividends and earnings will be affected by events in the industry.

The institutions rely more on written material, both from the company and from the broker-analysts, and less on personal contact with the company.

These differences are further brought out by the summary data presented in Table 7.8. The institutions, of this sample, rank general communications with the company under review first and direct contact second, while the brokers reverse these rankings.

Objective 2: To formulate ideas of how professional investors think their future information needs will be met.

This is covered by the summary section of the questionnaire, and by some of the interview questions. Table 7.9 shows the changes of classification expected by the respondents. Twelve of the fourteen respondents who answered this section thought that their use of data banks will increase. The majority of the respondents expected their use of non-company communications to remain fairly constant, while it was generally expected that general communications and direct contact will decline in relative use.

Table 7.7 presents an average of the respondents summary classifications. The ranking stated by the respondents, for the present, is the following - direct contact, general communications, non-company communications and data banks. The ranking expected by the same respondents for five years time is the following - non-company communications, direct contact, general communications and data banks. One of the conclusions of this research paper is that the use of non-company communications will become the first ranking information source (on the basis of these questionnaire results). General communications are expected to fall both in relative and absolute use, while data banks will increase in absolute usage but maintain their fourth ranking in this categorization of the four main information source types. Direct contact will maintain its second ranking.

The interview questions on this topic were not very conclusive, and do not really explain why the respondents expected data bank usage to increase, or general communications to fall. Six interesting points, made by the respondents, relevant to this question of future information sources, are presented below -

(a) there will be more international information available - "behaviour of a stock or bond market cannot be explained by domestic data alone."

(b) there will be improved disclosure of financial and operating data from companies.

(c) improved timeliness of data in general, and computerized data in particular, is to be expected in the future. Government source data is expected to improve in timeliness.

(d) there will be more general information on industries and companies available.

(e) several respondents expected to see a greater breadth of company data available on outside data banks in the future.

(f) one respondent (and the Shaw and Archibald study) think that the securities industry will undergo increased concentration and amalgamation. The result may be a smaller number of brokerage firms, and consequently less duplication of research material from the brokers to the institutions.

Objective 3: To make recommendations as to how financial information can be improved, and future requirements met.

The first recommendation is that the timeliness of information can be improved. Several of the respondents stated that they did not make considerable use of data banks because of the lateness of information. Government source information was especially criticized for lateness.

The second point is that there is a considerable amount of data duplication both to brokers and institutions, and from brokers to institutions. Several respondents thought that there should be a more selective information dissemination system to avoid data duplication and redundancy. The concept of a selective dissemination of information system for financial information appealed to some of the interviewees. However, there was insufficient time possible to explore this in further detail. A SDI service would differ from a scientific or technical information system in that analysts are less concerned with abstracts and references to journal articles, but, perhaps, more oriented to the future. The timeliness of the information would be very important to analysts.

A third point is that more complete financial disclosure by companies should improve the quality of information available.

As mentioned above, several of the respondents would like to have a larger breadth of data available on computerized data banks. However, without having looked into the costs, benefits and problems involved, it is difficult to make a recommendation that there should be increased coverage of companies.

Finally, the point has been already made that increased usage of international financial statistics in the work of analysts will be made to evaluate companies, industries and economies.

Objective 4: To utilize the methodology, which has been developed in scientific and social science information needs and use surveys, in the investment information field.

Chapter 5 reviewed the methodology and techniques of information surveys. The concepts of information use, demand and needs were explored; the problems of sampling noted; the psychological and environmental variables related to information needs and uses discussed; and a decision framework for user surveys presented.

Chapter 6 presented the methodology used by this survey. The information science methodology was useful for the design of the questionnaire, the preparation of the interview questions, and somewhat useful for the sample selection. The questionnaire has been a quite successful means of obtaining data on information sources in this study. The interview questions have not been so successful, firstly, because there was a lack of time for in-depth exploration in each interview, and secondly, perhaps, because of a difference in nature of financial analysis to scientific information work, thus making some of these questions unsuitable for discussing financial information needs.

<u>Objective 5</u>: To comment on the relevance of this study to future computerized information systems.

This study has concluded, on the basis of the results of the survey undertaken, that the use of computerized data banks is expected to increase. However, data banks are expected to remain as the fourth ranking source of the four classifications used in this study. Data banks should be seen as an integral part of the information sources available to analysts. These sources basically consist of information from the company under review, information obtained from the company, and more general information from third sources. The computer is basically a media for obtaining both general communication information, such as financial statement data, and third source material, such as economy data and stock trading data.

The following points pertaining to computerized data banks have already arisen, both from the survey itself and from the descriptive and analysis sections of this study.

(a) the timeliness of fundamental and economic data on computerized data banks could be improved.

(b) consideration should be given to the possibilities of placing relevant international financial statistics data on data banks.

(c) there seems to be a need for a greater breadth of company fundamental data on computers. However, any enlarging of this data depends on the costs and benefits involved.

(d) none of the respondents mentioned data inaccuracies with regards to data banks. In fact, one survey respondent, speaking of government source data, stated that he would prefer slightly less accurate but earlier data to accurate, later data.

To conclude, in this research paper, information needs and sources have been discussed, surveyed and analysed. The study has produced a ranking list of the information sources of professional investors, has presented the respondents' expectations of future information sources and has discussed the information requirements of the survey respondents.

BIBLIOGRAPHY

Aguilar, F., Scanning the Business Environment. New York: Macmillan, 1967.

Amey, L.R., & Egginton, Management Accounting: A Conceptual Approach. London: Longman, 1973, Chapter 1.

Biehal, G.J., "Computer Languages for Financial Analysts", MBA Thesis, McGill University, 1969.

Bond, E.P. and Shearer, The Canadian Financial System. Toronto: Prentice-Hall, 1972.

Boutell, W.S., Computer-Oriented Business Systems. Englewood Cliffs, N.J.: Prentice Hall, 1973.

Brittain, J.M., Information and its Uses, a Review. Bath: Bath University Press, 1970.

Clarkson, G.P.E., Portfolio Selection: A Simulation of Trust Investment. Englewood Cliffs, N.J.: Prentice Hall, 1962.

Cohen, J.B., Zinbarg & Zeikel, Investment Analysis and Portfolio Management. Homewood, Illinois: R. D. Irwin, 1973.

Cuardra, C.A., Editor: Volumes of the Annual Review of Information Science and Technology. Chicago: Encyclopaedia Brittanica, 1965 onwards.

Duesenberry, J.L., "Criteria for Judging the Performance of Capital Markets" in Elements of Investments. Edited by Wu, H.K., & Zakon. New York: Holt, Rinehart and Winston, 1965.

Fama, E.F., "Efficient Capital Markets", The Journal of Finance, May 1970, p. 383.

Finance and Financial Data References Sources: A Guide. The Management Library Reference Department, McGill University, 1973.

Flowers, B.H., "A Survey of Information Needs of Physicists and Chemists", Journal of Documentation, Vol. 21, No. 2, 1965, p. 83-113.

FRI, An Introduction to the Financial Research Institute, Montreal 1973.

Guthrie, A., A Survey of Canadian Middle Manager's Attitudes Towards Management Information Systems. Ottawa, Carleton University, 1972.

Jordon, J.R., "Let the Computer Select Your Reading List", Datamation, 16, Feb. 1970, p. 91-94.

Latané, H. & Tuttle, Security Analysis and Portfolio Management. New York: The Ronald Press, 1970.

Latané, H., Tuttle & Jones, "Quarterly data: E/P ratios versus changes in earnings in forecasting future price changes". Financial Analyst Journal, 25, Jan.-Feb. 1969.

Leroux "Computer Application to Financial Analysis", MBA Thesis, McGill University, 1967.

King, D.W., and Bryant, The Evaluation of Information Services and Products. Washington: Information Resource Press, 1971.

Martyn, J., and Vickery, "The Complexity of the Modelling of Information Systems", Journal of Documentation, Vol. 26, No. 3, 1970.

Meadow, C.T., The Analysis of Information Systems. New York: John Wiley, 1967.

Miller, D.C., Handbook of Research Design and Social Measurement. New York: David McKay, 1964.

Mintzberg, H., "The Nature of Managerial Work", Chapter 2 of Readings for Management Policy, McGill University, 1973.

Neufeld, E.P., The Financial System of Canada. Toronto, Macmillan, 1972.

Opinion Research Corporation, ORS Public Opinion Index Special Report. Princeton, N.J.: Opinion Research Corporation, 1967.

Radford, R.J., Information Systems in Management. Reston, Virginia: Reston Publishing, 1973.

Sharpe, W.F., Portfolio Theory and Capital Markets. New York: McGraw Hill, 1970.

Shaw, D.C., and Archibald, The Management of Change in the Canadian Securities Industry. A long term study by the University of Western Ontario and the Toronto Stock Exchange. Toronto: Toronto Stock Exchange, 1972.

______________, Study 1: "Canada's Capital Market".

______________, Study 2: "A Survey of the Research Operations of Brokerage Firms in Canada."

______________, Study 3: "Trust Companies in Canada".

______________, Study 4: "Life Insurance Companies in Canada".

Sparks, D.E., Chodrow, and Walsh, "A Methodology for the Analysis of Information Systems". Columbia University School of Library Science, In. 3, Vol. 1 & 2, 1971.

Stone, D.E., "Caveats in Computer-Aided Financial Analysis", Financial Analyst Journal, Jan.-Feb. 1968, p. 149.

Vickery, B., Information Systems. Hamden, Connecticut: Archon Books, 1973.

Williams, E.E. & Findlay, Investment Analysis. Englewood Cliffs, N.J.: Prentice Hall, 1974.

Wood, D., "Discovering the User and his Information Needs", Aslib Proceedings, 21, 1969, pp. 262-270.

CHAPTER 3

THE COST OF A DATA BANK

As indicated earlier, a data-bank is a mixture of computer technology and library science. This chapter thus takes a look at costs from these two points of views. In the computer cost section were included considerations related to the current communication technology[1], as we cannot think today of a computer system deprived of communication links with its users. In the information storage and retrieval system cost section, two models of interest can be found.

[1]With particular reference to the Bell Canada offering: Dataroute.

3.1 The Computer-Communication Technology

3.1.1 Computer costs

3.1.1.1 Introduction to the computer system elements

3.1.1.1.1 Storage problems

3.1.1.1.2 The time-sharing environment

3.1.1.2 The cost of a computer: the price for its services

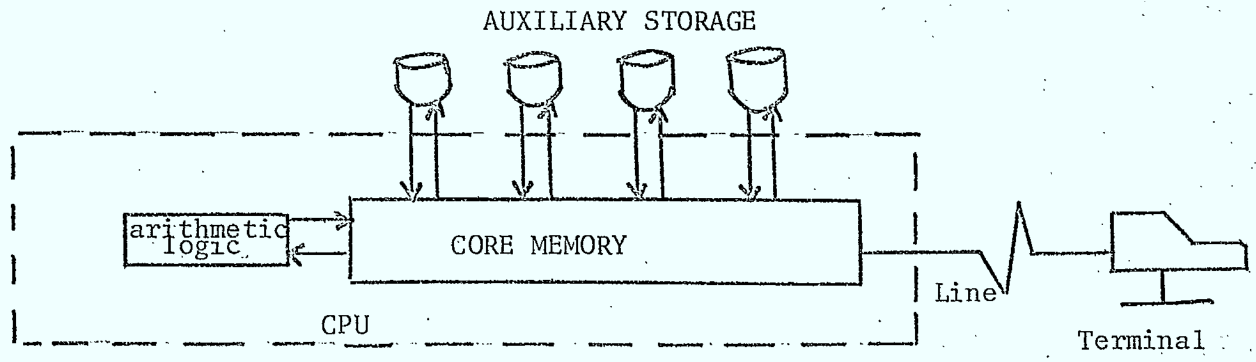
3.1.1.2.1 Costing computer services: the capacity problem

3.1.1.2.2 Time-sharing: the problem of shared resources

3.1.1 Computer Costs

3.1.1.1 Introduction to the Computer System Elements

In this section, we shall briefly summarize the elements of a computer system on which a data-bank relies. For our purposes, a computer system can be defined as comprising input-output devices, central processing unit (arithmetic-logic plus core storage), and auxiliary storage:



The CPU is the actual computer. Within the core memory is located an operating system which reduces the working storage for the users' programs.

Tapes, and more and more often, disks, are the components of auxiliary storage. Since it has a particular importance for a data-bank operation, a section will be devoted to the storage problem.[1]

Of the input-output devices, only terminals can be considered for our purposes, although one must be aware of the range of existing devices, from card readers to optical character recognition (OCR).

---

[1] See section 1.1 et seq. of Part IV, where file organization is examined in detail.

Between the storage devices and the CPU on one hand, and between the terminals and the CPU on the other hand, there exist hardware interfaces called controllers. The storage device controllers were already alluded to in section 2.3.2, page 77. The communication controller supports a certain number of terminals (up to 100 low-speed terminals in the case of the IBM 3705 communications controller[1]).

The computer system supporting a data-bank can thus be best depicted as comprising:

(i) the terminals

(ii) the modems

(iii) the local lines

(iv) the line control unit

(v) the CPU

(vi) the direct access storage devices

(vii) the storage control units.

---

[1] Its cost is approximately $3,000 per month.

3.1.1.1.1 Storage Problems

The storage cost is critical in a data-bank operation, the whole concept of data-base being based on the availability of data from files. It is the more critical the faster is the storage medium. For on-line applications, such as a data-bank, fast storage media are essential: the implications of this make storage organization a key factor in costs.

There are three cost-related aspects in storage:

(i) the storage volume: since cost is proportional to the space occupied, an obvious consideration is to save space, by suppressing redundancy, blanks, and useless information. Redundancy would normally happen because files are duplicated to allow for different search strategies. But why would one want to allow for the search strategy options? Because of the search cost.

(ii) the search cost: in reality, it is the search time, but it ultimately comes to cost. The search time is directly related to the search strategy, which itself is dependent on the file structure. The search strategy will be different according to whether or not the records in the file are ordered, direct access is possible and there is an index. However, the alternatives in file structure all have a cost, in terms of storage space again: setting up an index consumes memory space. Thus some alternatives save search time, but consume more space.

A tradeoff has to be worked out, with costs being a function of the file activity, e.g. the frequency of consultation of the file. However, file activity should not be the only parameter: file volatility and file growth are two other parameters, which will be briefly examined now.

(iii) the update cost: file maintenance requires that new data be substituted for old data, or simply added. But these operations themselves involve locating the item to be replaced (and we come back to the search cost above) or finding free space (we here have another type of search cost) or keeping the logical link with other records, in short, costs. File volatility refers more specifically to records being replaced by others (and in this case, file volatility entails almost the same costs as the file activity), while file growth refers to records being simply added (in this case costs are more difficult to evaluate).

As can be readily perceived, the storage volume, the search cost and the update cost represent the microcosmic counterparts of the file size, the usage cost and the update cost of data-bank. The main difference is that, in dealing with problems of optimal file organization, the design is much more technology-dependent.[1]

---

[1] An excellent introduction to the file organization problem is found in Lefkovitz, D. "File Structures for on-line systems." MacMillan and Co. (1969). A somewhat more advanced text is Knuth, D.E. in Chapter 2 of "The Art of Computer Programming" pp. 228-463, Vol. 1, Addison-Wesley (1969).

3.1.1.1.2 The Time-Sharing Environment

Time sharing is an extension of multiprogramming towards strongly user-related applications: interrogation, interaction, online real-time. Multiprogramming basically permits the handling of many independent programs concurrently by overlapping or interleaving their execution. Taking advantage of the speed hierarchy of the computer system -the CPU, the auxiliary storage, the input-output devices in decreasing order of performance level - the higher level elements can jump from one program to another, while the lower level elements remain allocated to one job only (example : the terminal) . This provides a better utilization of the higher-level elements (CPU) which would have been idle most of the time in a monoprogramming environment. There are various degrees in multiprogramming, from each job being executed until completion to fixed or variable-length time slices being allocated to the jobs (together with a scheduling algorithm for the sequencing of the programs) .

Against the better utilization of the higher-level, expensive components of the computer system through multiprogramming must be offset by its heavier executive control requirements (which are paid for in decreased problem program area in bytes, in reduced computer time available for useful work and in extra hardware) . The impact on the user cost, however, is not clear. Probably, and this is the most significant thing before the next section, multiprogramming also provides a better balance between user costs and user needs: a given monoprogramming configuration never will be optimal for a given job. The user pays for input-output devices he does not use (the typical case of CPU-bound jobs), for instance, or conversely for a CPU capacity which he wastes. In a multiprogramming environment, the user can expect to be charged more in relation to what he actually used, and has the performance, in size and speed, that he desires.[1]

[1]A tutorial text in multiprogramming (real-time oriented) computing is Mortin, J.J. "Design for real-time computer systems,: Prentice-Hall 1967.

3.1.1.2 The Cost of a Computer: The price for its Services

The pricing (external or internal) of computer services is a rather challenging matter. It is similar to the problem of pricing utilities such as communications, energy or transportation, mainly because a large amount of costs are incurred in order to be ready to meet the demand.[1] The difficulty is compounded by the specificity of today's computer operation, multiprogramming and multiprocessing. In order to isolate the issues, the first section will deal with the monoprogramming computer, while the second section will introduce the additional complexity brought about by running jobs simultaneously.

3.1.1.2.1 Costing Computer Services[2]

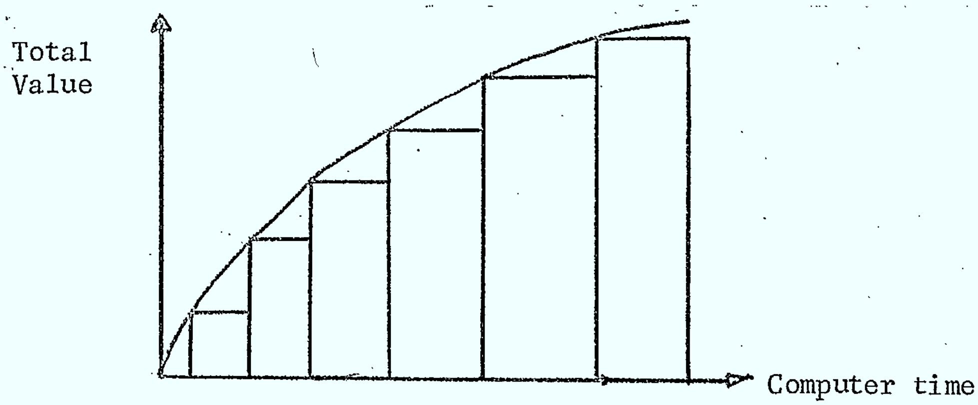
A computer facility must be paid for its services, which are provided to external users (to the organization) as well as to internal users. The problem arises of setting a price for the commodity represented by computation. The underlying philosophy here is that of the efficient allocation of scarce resources (in this case, computing power) to alternative uses (the user's jobs). In this context, where the organization seeks to maximize overall benefits, prices should be divorced from costs: if a charge is to be applied, it should essentially be based upon the strength of the demand at the time, not upon costs;

---

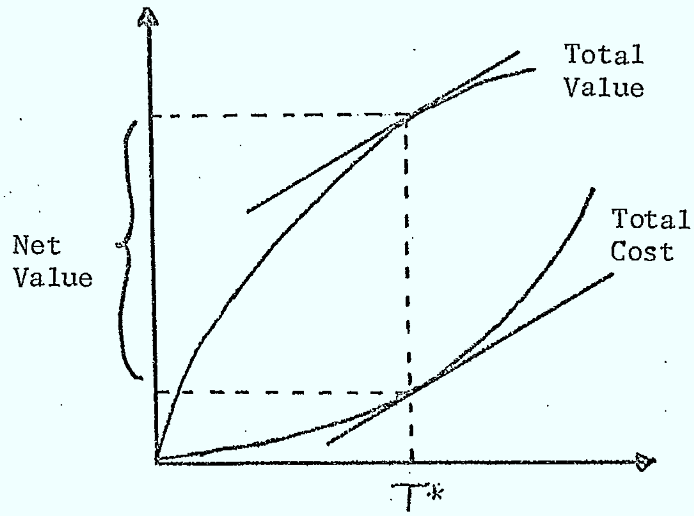
[1] e.g. the debate about marginal vs. peak-load pricing for electrical utilities.

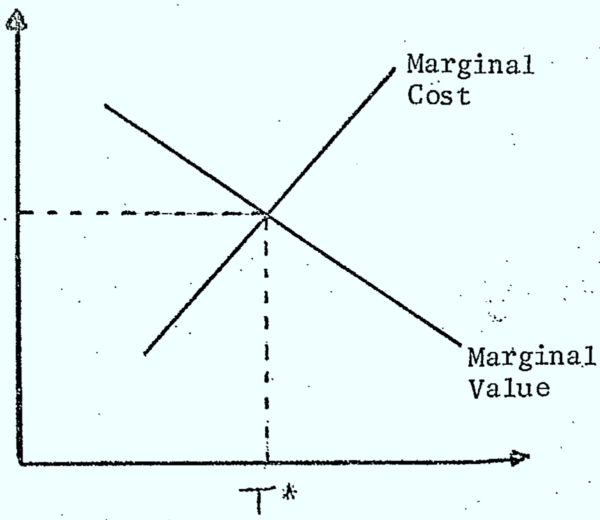
[2] Most of what follows in this section is inspired by "Economics of Computers," by W.F. Sharpe, Columbia University Press, 1969.

in other words, as much should be charged for the computer services as the traffic will bear. If we make the restrictive assumption that this commodity is homogeneous over time (i.e. the value of computation at one time during the day is the same as at another time), then we can rank the jobs by value only, obtaining a schedule of total value of computer time against required computer time.

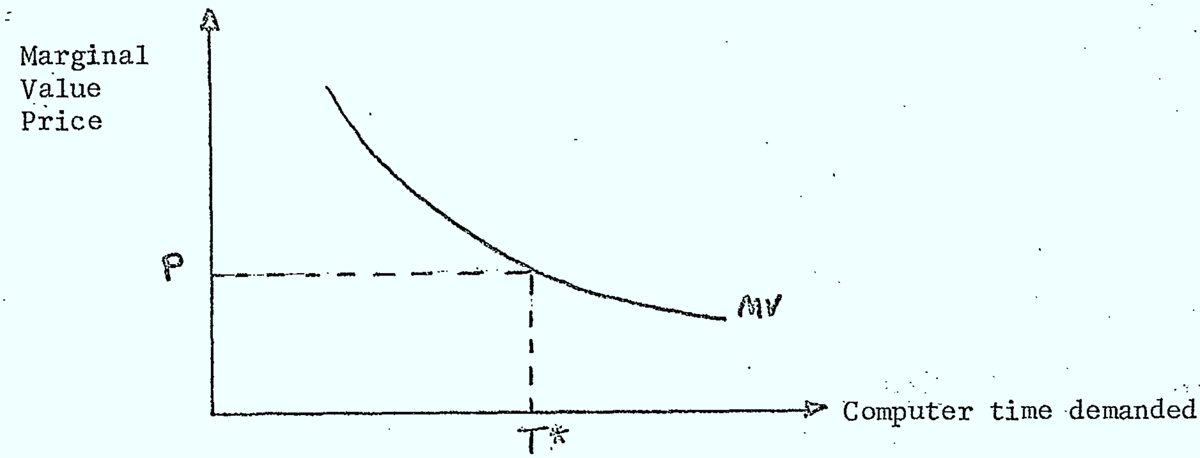


The marginal value curve is the demand curve for computation: the users, seeking individually to maximize net value (= total value - cost = total value - price charged) will choose the amount of computation (= the demand) which makes the difference between value and cost largest for the price charged:

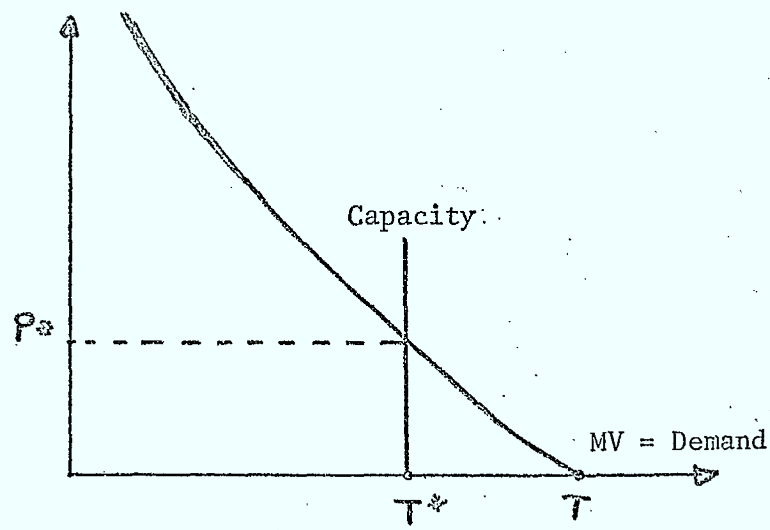




At any price, then, the slope of the total value curve (= the marginal value) represents the demand for computer time:



In the case where the availability of computer time is limited by contract or capacity, then obviously if the price of computer time is set to zero, the quantity of computation demanded (T) may exceed (this depends on the shape of the demand curve) the capacity T*. If a price is charged, the amount demanded will gradually decrease until it matches the capacity: the price P* at which this happens is called an equilibrium price:



This result can also be achieved through mathematical programming:

$$\text{maximize} \quad \sum_i f(x_i)$$

$$\text{s.t.} \quad \sum_i g(x_i) \leq T^*$$

$$x_i \geq 0 \text{ for all } i$$

where x_i is the proportion of job i run, f(.) the value function and g(.) a production function which can be homogeneous of degree one. If the value of job i is proportional to the time t_i it is run, we have:

$$\text{maximize} \quad \sum_i t_i \frac{V_i}{T_i}$$

$$\text{s.t.} \quad \sum_i t_i \leq T^*$$

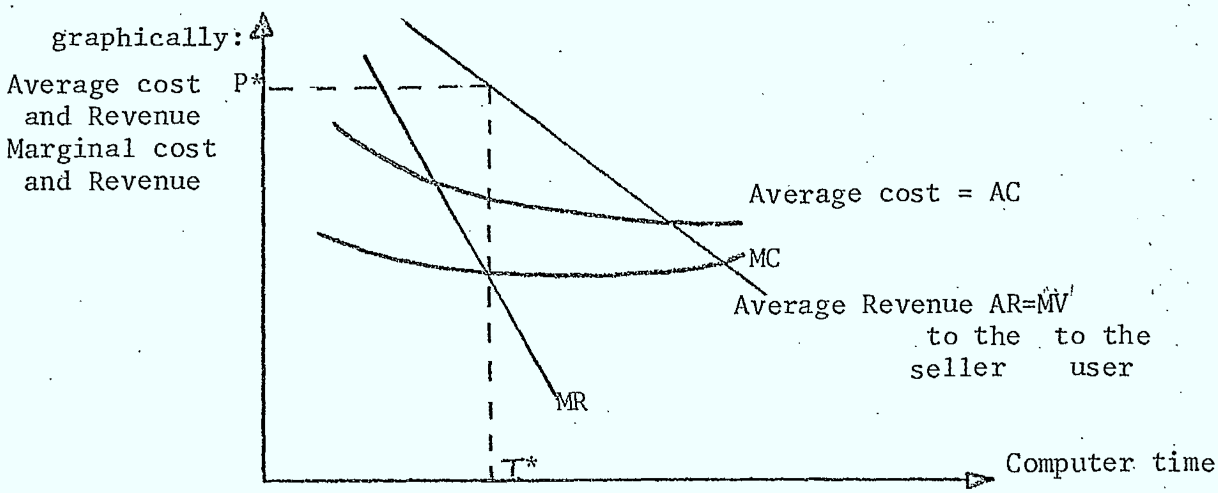
$$0 \leq t_i \leq T_i \text{ for all } i$$

where V_i and T_i are, respectively, the value and the required computer time for job i. In these two programs the solution may include jobs which will not be run to completion; for instance, one third only of job 2 is to be run; this can be interpreted as job 2 being run every three periods.

Beside the solution to the program which gives the optimal utilization of computer time, the dual variables associated with the constraints will yield shadow prices which are precisely the equilibrium rationing prices obtained graphically above.

So far, the discussion has centered on internal (transfer) prices for computer services. When external users are to be charged, or in general in the case of a profit-maximizing organization, a monetary yardstick of value is used: revenue. When dealing with value, the guide for the optimal utilization of the computer was to equate marginal value with marginal cost; similarly, the optimal utilization (in a profit-maximizing sense) of the computer will require equating marginal revenue with marginal cost.

The difference is that the organization now faces a demand curve which is the average revenue. The optimal price will then be that price for which marginal revenue is equal to marginal cost; graphically:



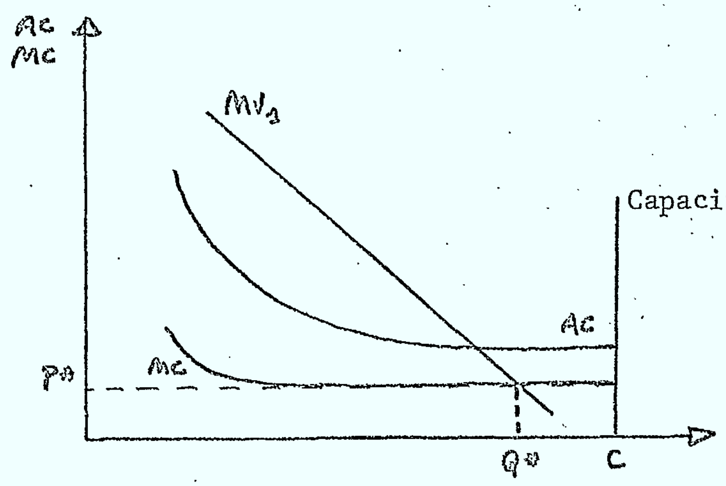
With the problem of joint resources, which we shall treat later, the problem of capacity constraints is the most challenging for purposes of pricing.[1] In the above graph, we did not indicate a capacity

[1] "The Theory of Public Utility Pricing and its Applications" by R.H. Coase; "A Mathematical Formulation of the Peak-load Pricing Problem" by Israel Pressman; "Peak-load Pricing of Telephone Calls", Littlechild; all in The Bell Journal of Economics and Management Science, Spring 1970.

constraint. Reverting to our earlier graphs (value) in order to simplify the analysis, let us recall that the rule was to equate marginal value and marginal cost, and charge the corresponding price: this is marginal cost pricing.

If the capacity is fixed, and total cost over the period does not vary much with quantity of computation up to capacity, then average cost $\frac{TC}{Q}$ is an hyperbola (= economies of scale) and marginal cost is always below average cost:[1]

CASE 1



In case 1, where the intersection between marginal cost and marginal value occurs for a utilization less than capacity, the marginal-cost pricing rule leads to a loss (since costs, average costs, will not be covered) and the computer is not used to capacity.

---

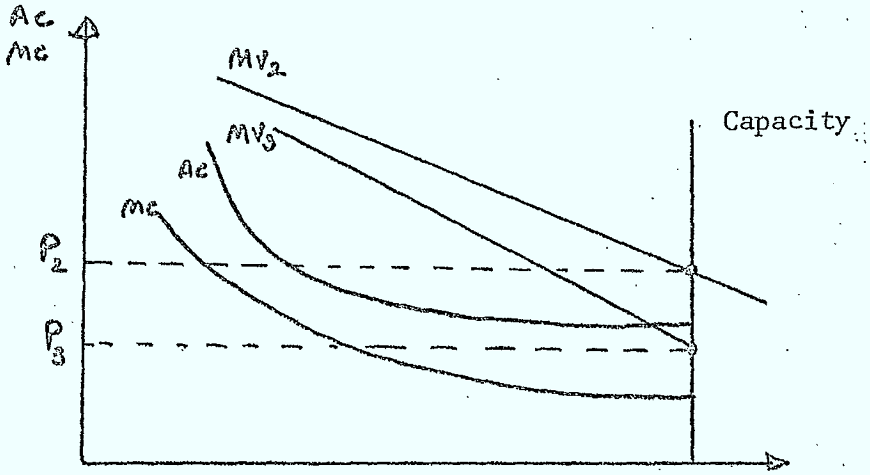
[1]Since the slope of average cost curve is: $\frac{d(\frac{TC}{Q})}{dQ} = \frac{MC.Q-TC}{Q^2} = \frac{MC}{Q} - \frac{AC}{Q}$

and since the slope of average cost curve is negative (hyperbola):

$\frac{MC}{Q} - \frac{AC}{Q} < 0$ or $MC < AC$.

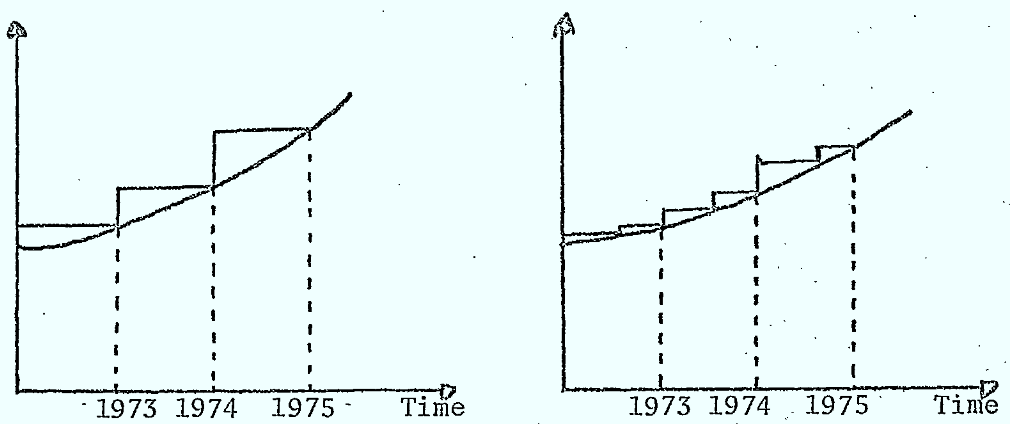
In the cases 2 and 3, the marginal value curve simply does not intersect the marginal cost curve, so that the marginal cost pricing rule must be replaced by a price set in order to ration the available capacity (intersection of MV curve and binding capacity constraint: optimization at the boundary of the domain).

CASES 2 & 3

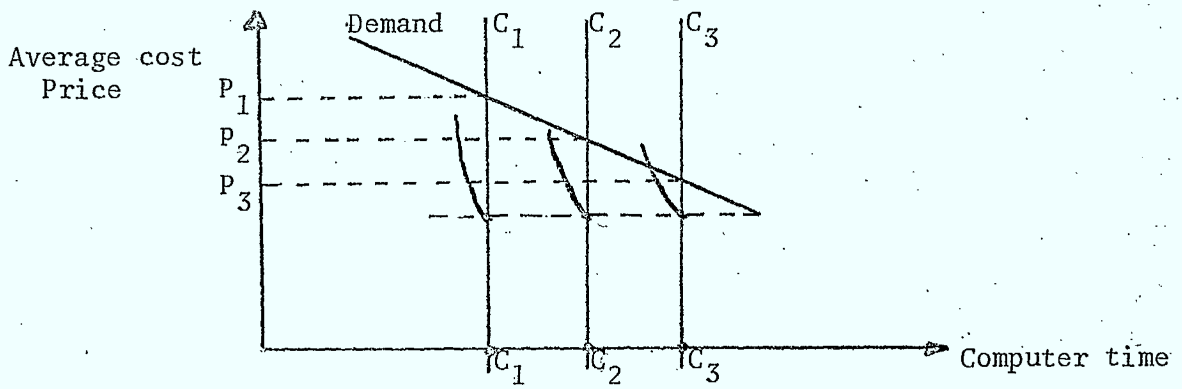


But the price can as well be above the average cost as below, depending on the shape of the demand curve.

Over the long run, however, capacity is not a constraint, since additional computing power can be put in operation. The adjustment to demand still cannot be in perfect balance, because of the indivisibilities of equipment; smoothness and balance can, however, be achieved by adding smaller discrete increments:



There is a trade-off between the opportunity cost of idle capacity created by large increments and the outlay cost of frequent conversions: a larger than necessary capacity at a moment in time will entail a deficit as shown by the graph below:



If the cost of equipment is proportional to capacity, then the average cost for each of the capacity utilizations C_1, C_2, C_3 is the same: k, the proportionality constant between cost and capacity. Yet the equilibrium prices are decreasing with increasing capacity. It can easily be seen that the excess of revenues over costs is decreasing with increasing capacity: this is the opportunity cost of idle capacity. It is an increasing function of the idle capacity, which itself is a function of the frequency of capacity changes: the more frequent the changes, the smaller the increment of capacity at each change. This opportunity cost is to be balanced against the conversion costs, which are essentially dependent on the number of changes during the period. The trade-off can be best recognized in the diagram below:

Two computers should be used at capacity, and the price for their utilization should be set at the intersection of the marginal value curve with the capacity constraint (cases 2 and 3 above), which may happen to be greater or less than average cost.

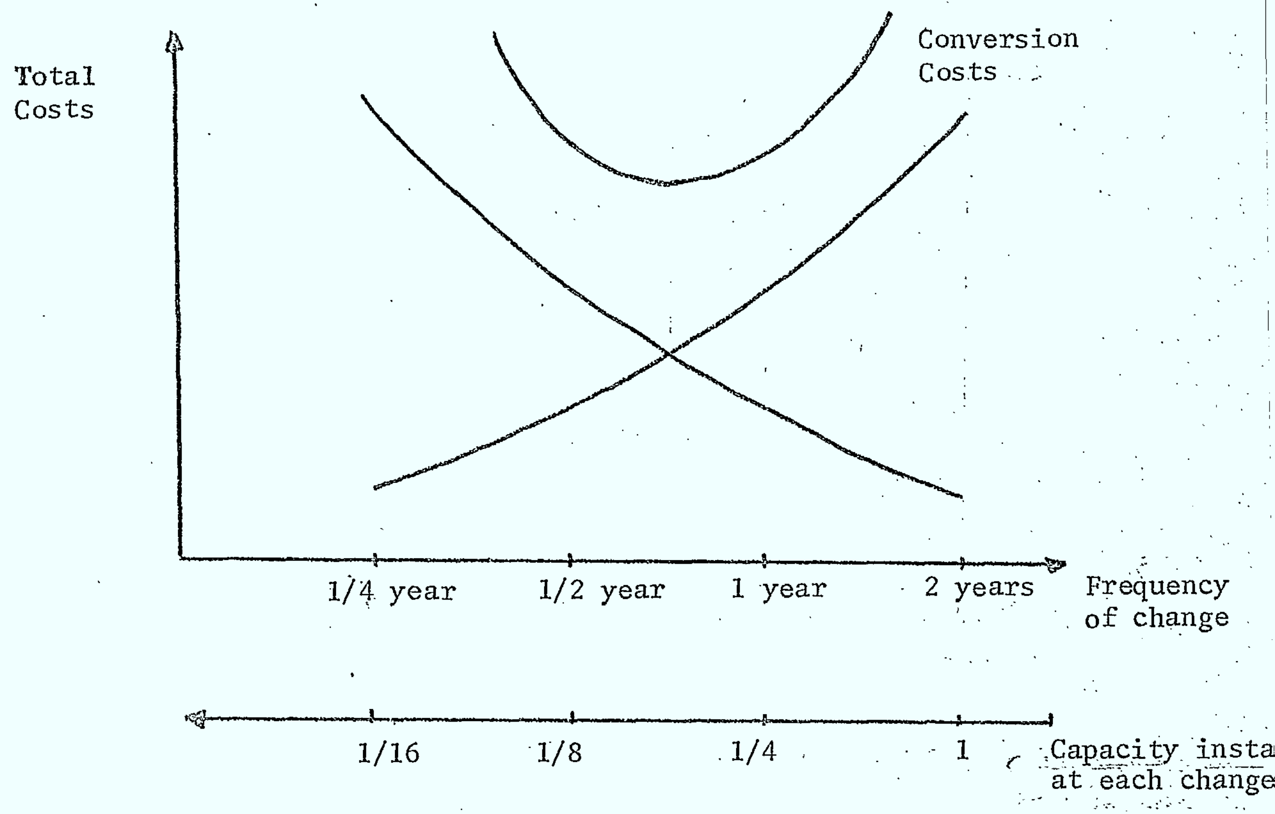
Note that when costs vary with utilization within capacity limits, the marginal cost is not zero, and that optimal utilization will lead to some unused capacity. A second point of interest is that the cost of an underutilized system is greater than the average cost of a fully utilized computer when economies of scale are present: an implication is that users will tend to merge their operations if all the above conditions hold (mainly fixity of costs, and monotonic decreasing average cost).

When the computer system comes close to capacity, however, some complications appear. Queuing theory, for instance, tells us that the turnaround time (time spent in the queue, plus time of service) will rapidly and progressively increase as capacity is approached. If arrivals of jobs follow a Poisson distribution (with parameter λ: average number of jobs submitted per unit of time) and they are run at a rate of μ per unit of time, then the proportion of time the computer will be in use is: $P = \frac{\lambda}{\mu}$

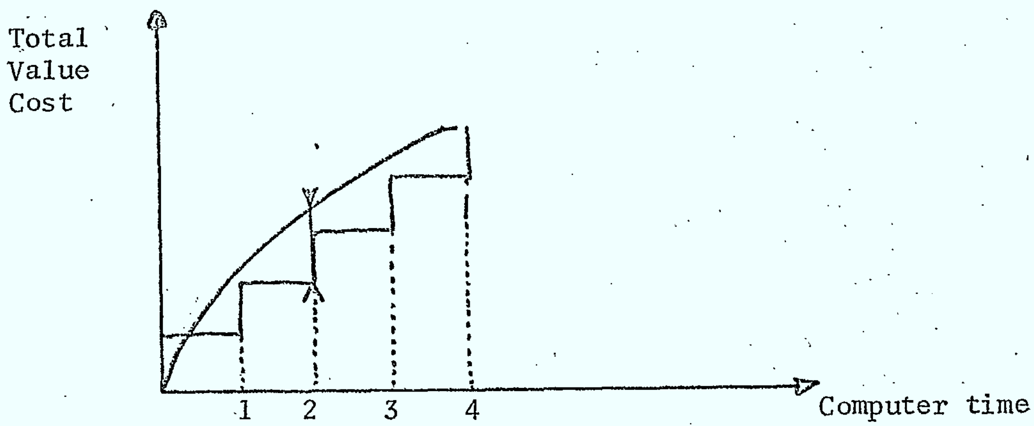
and waiting time (excluding service time) will be:[1]

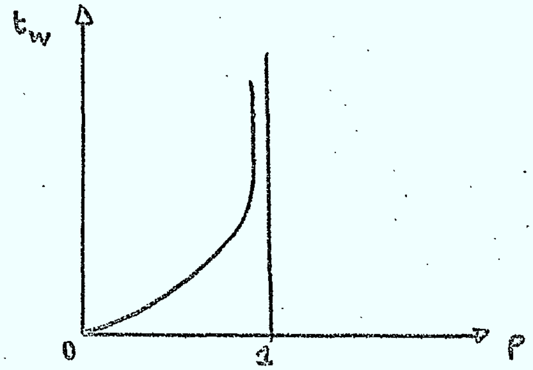
---

[1] James Martin, "Design of Real-Time Computer Systems," p. 383 et seq.



Whatever the incremental capacity, the long run total cost function is a step function (with flat steps if the costs do not vary with utilization within capacity limits). Marginal costs will be zero for all utilization levels and provide no guides for pricing. However, one can determine the optimal investment from the graph below:





$$t_w = \left(\frac{P}{1-P}\right)\frac{1}{\mu} = \left(\frac{P}{1-P}\right)S$$

$$(S = \frac{1}{\mu} = \text{service time})$$

Thus, there is a non-monetary penalty associated with the increased utilization of a system. This penalty will be at work to restrain utilization to a reasonable level, since jobs of low value will not be run at all, or a price can be set for priority (the introduction of priority is equivalent to a multiple service station queuing problem, or to the addition of new "capacity constraints" to the mathematical program for allocation of resources), in which case a low price is tagged to otherwise undesirable turnaround time and a high price for low turnaround.

In fact, computer time is not the homogeneous commodity we have assumed so far, and its value depends very much on the time of day, since most users will submit their jobs during a preferred time interval, which brings us back to the turnaround time considered before. The waiting-time will serve as a rationing device here also, since at peak hours a high turnaround will be associated with the desirable time of day, and users who are willing to accept a low turnaround will be confined to less desirable hours.

The turnaround time, and its variant, the peak-hour, is thus an essential consideration in the pricing of computer services.[1]

---

[1] "Cost-benefit evaluation of scientific computing services", D.N. Streeter, IBM Systems Journal, Vol. 11, No. 3.

It is even more critical when dealing with a data-bank, since response time is twofold: it applies both to the update frequency and the retrieval time. Computer applications will cover the whole spectrum of response times, from real-time (for airline reservation systems, stock price quotation services, industrial processes) and conversational (interactive education) to remote batch (various business applications), according to the aggregation level of information, the prediction span and the dynamics of the environment. But before going further into this, we shall examine how the economic analysis is modified by introducing the multiprogramming, multiprocessing and time-sharing modes of operation.

3.1.1.2.2 Time-sharing

The previous section dealt with the problems of costing computer services when costs were fixed relative to utilization. This section will discuss more specifically the issues posed by the joint and shared costs involved in multiprogramming. Costing then means allocating costs across joint activities. In the preceding section, we could safely take the elapsed computer time as a measure of service and of value, because only one job was running at a time; there were no overheads which benefited all users, only direct costs which benefited only one activity. The price of the computer time was considered as a rate per unit of time, and the product of rate and elapsed time gave the final charge.

When many jobs are run and share the resources during a period of time, then the computer cannot any longer be considered as an entity, but must be broken into a number of "resources," like central processing unit (CPU), storage, I/O channels and so on. One must keep track of every "resource" used by each job. This calls for an accounting of as many resources as there exist in the system, to the smallest element (a byte in core, a track on disk storage, et. . .). The finer the accounting, the more equitable the charging will be. One extreme would be to charge for the use of the tiniest component, but clearly this would be undesirable because the cost of getting the information would by far exceed the envisaged benefits. This is essentially a technical feasibility problem.

Other problems are more conceptual in nature: those which arise in connection with the charging for joint resources. Staff is such an example. They fall into what is commonly called indirect or overhead costs, which is another way of saying that they will be allocated in a completely arbitrary way. An operator is associated with the computer as a whole, but it is impossible to allocate his salary among the basic activities or resources (CPU, storage, I/O). The allocation percentage can reflect anything from the computing centre manager's will to shift usage from one component to another (e.g. from disk storage to tape) to an appearance of rationality (allocate more overhead to equipment categories which require human intervention: card puncher and readers, tape readers . . .). Falling back on economic concepts, the allocation scheme should be worked out not in relation to costs, but to demand. The major portion of overhead should be charged to the inelastic demand segments of users or usage, without this becoming a matter of ethical concern. Indeed, since the price of the computer services should be such that marginal costs equal marginal revenue, and marginal cost does not include the overhead portion, it should be set independently of the overhead. The implication is clear: any cost-based overhead allocation is artificial, and the demand elasticity provides the only guide from which an allocation is to be worked out. This also is the basis for differential pricing, the benefits of which should be fully reaped in the pricing policy of computing centres.

Because of the difficulties of estimating demand elasticities, computing centers rely more on equitable charging algorithms than on efficient equilibrium prices.[1] Such a charging scheme is to charge according to the CPU time, adding a share of all other resources as overhead. The fallacy of such a charging algorithm is easily seen when one considers the variety of resource mixes: some jobs will use much input-output and little CPU time; does this mean that they should be charged less than CPU-bound jobs? (which is the implication of this scheme). Also the equity and the realism of CPU time-charging

[1] See "The great cost allocation debate" by J. Gallop, Canadian Data Systems, November 1972, for a behavioral, pragmatic approach to alternative pricing schemes.

is in question since it does not take into account the core storage usage (a job which requires more core storage makes it unavailable for other jobs). A better solution would then be to combine CPU time and core storage requirements in a single measure.

Further along the lines of charging equitably for computer services, two recent articles[1] deal with the pricing of the rest of the computer resources. These papers are very similar in their conclusions: costing for jobs should proceed from the application of a costing rate (standard) to the resources used or tied up for a period of time. These resources fall into a number of categories: CPU and core memory, I/O devices (printer, card reader and puncher, tape I/O, disc I/O), and storage (on-line disc storage, off-line disc storage), channels. The costing rate is obtained by dividing the monthly cost of each resource by the total actual monthly usage. However, at this point, the opinions differ. For Wiorkowski a critical stage is to obtain the actual monthly usage, which is only a percentage of the theoretical usage; because of preventive maintenance, down time (failure), rerun time, no resource is used 100 percent of the time, and this is why it is necessary to correct the ideal usage figure. This step may have grievous consequences: if the percentage of utilization is too large then the denominator of the rate formula will be too large and the charging rate will make that cost exceed

---

[1] IBM Systems Journal: "Accounting control of data processing" by Rettus and Smith.

[2] Datamation, August 1973: "A cost allocation model" by Wiorkowski.

revenues. For Wiorkowski, then, the rate formula is:

$$R = \frac{C}{PTU}$$

where R = rate per unit of time per unit of resource

C = total cost of the resources in a category (including an overhead share)

P = percent usage

T = time per month (theoretical)

U = units of resource

For Rettus and Smith (hereafter R & S), who using accounting terminology, the standard rate SR is the ratio of the monthly budget cost (including direct costs of the component category, and a share of indirect costs and overhead costs) and the productive time PT (PT is, unlike Wiorkowski's actual usage, defined by the difference between total monthly time [720 hours] and preventive maintenance, down time, rerun time and multiprogramming degradation): $SR = \frac{BC}{PT}$.

The standard rate for each component in the category of n components is then $\frac{SR}{n}$.

One of the main concepts developed by R & S is the multiprogramming degradation. Multiprogramming is desirable because the degradation affecting certain components is balanced by the increased utilization of other components (mainly the CPU). The elapsed time for each job in a multiprogramming environment exceeds the time it would have required in a monoprogramming set-up. The multiprogramming degradation (MP) is precisely this excess time. More specifically, MP is measured by the total time a job is in the wait-ready status, which is the period during which the job is tying up core and I/O, is ready to continue executing, but is waiting to obtain the CPU, which works on another job. Productive time PT[1] is obtained by subtracting the wait-ready time (or MP) from the elapsed time: PT = ET - MP.

However, the operating system does not record the wait-ready time but it can be obtained indirectly by the theoretical resource utilization (TRU) time[2] according to the relationship, MP = ET - TRU time. As a consequence, the productive time for each component category is:

main and auxiliary storage = PT=ET-MP=ET-ET+TRU time=TRU time.

CPU utilization: CPU time

Channel utilization: number of reads and writes (EXCP's X physical record length X transfer rate.)

So far, the rate calculation has been discussed. For the charging of a job, the standard rate (R & S)[3] or the rate (Wiorkowski) will be applied to the utilization (R & S) or the number of units (Wiorkowski), according to the tables shown (tables 1 and 2).

---

[1] The productive time correction only applies to core and I/O since the multiprogramming degradation only affects these components.

[2] TRU time is provided by the Systems Management Facility of OS/360 and OS/370.

[3] Empirical estimates of the multiprogramming degradation are available in "Servicing priority job requests on a computer", R.E. Carr, Canadian Datasystems, January 1973.

TABLE 1: Derivation of Rate

For R & S

(1) Component Group	(2) Number of Units	(3) Total Time	(4) Productive Time	(5) Standard Budget	(6) Component Group Standard Rate	(7) Unit Standard Rate
Main storage	400 Kbytes	720	610	xxxxx	xxxxx/610	(6) / 400
•						
•						
•						

For Wiorkowski

Cost-charge Category	Cost/Month	Percent Usage CPU	Percent Usage Core	Unit of Time	Units of Resources	Rate
		CPU	Core	Hours	1,000 bytes	
Kilobyte Hour	$50,000	30	40	720		$.9645061
On-line Printer	5,000	40		Minutes 43,200	Lines 1,700	.0002292
Card-Reader	2,000	5		Minutes 43,200	Cards 2,000	.0004629
Disc I/O	17,000	60		Seconds 2,592,000	Channels 2.38	.0003215
Tape I/O	19,000	50		Seconds 2,592,000	Channels 2	.0003665
Off-line Tape storage	6,000	80		Month 1	Tapes 4,000	1.8750000

TABLE 2: Determination of Charge

For R & S

Component Group	Utilization	Number of Components	Standard Rate for Group	Total Cost
Tape	40*	4	$ .30	$ 48.00
Disk	40*	200**	.01	80.00
Printer	40*	1	.20	8.00
Main Storage	40*	150***	.01	60.00
CPU	10	1	2.90	29.00
Multiplexor channel	2	1	.40	.80
Selector channel	5	1	.50	2.50
				$228.30

For Wiorkowski

Category	Rate	Unit of Time	Unit of Resource	Charge
Kilobyte Hour (CPU + Core)	.0645	.01485	96K	1.4090
On-Line Printer	.0002		2006.14	.4598
Card Reader	.0004		299.51	.1386
Disk I/O	.0003		5008.71	1.6103
Tape I/O	.0003		4830.86	1.7705
Off-Line Tape	1.8750	1 month	1.00	1.8750
				$7.2622

* Productive time
** Tracks
*** Kilobytes

The McGill time-sharing system (MUSIC), which charges the user as per CPU time, connect time and storage space[1] can thus be viewed as going into this direction of charging for resource utilization. Yet the sophistication of the approaches advocated by R & S and Wiorkowski lies a bit too far ahead of current practice, mainly because of information and accounting cost.

The other large component of a computer system cost is the communication cost. This is the topic of the next subsection.

---

[1] rates were shown in chapter 2 of this Part.

3.1.2 Communication Costs

3.1.2.1 Introduction to Data Communications

3.1.2.1.1 the importance of communication links

3.1.2.1.2 DC vs. AC transmission - telegraphy
- voice-band
- modem

3.1.2.1.3 types of communication links - dial-up or dedicated
- half of full duplex
- synchronous vs. asynchronous

3.1.2.1.4 cost considerations in relation to network design

3.1.2.2 The Advent of Dataroute

3.1.2.2.1 description

3.1.2.2.2 elements of equipment

3.1.2.2.3 cost

3.1.2.1 Introduction to Data Communications

3.1.2.1.1 The Importance of Communication Links

The components of a computerized system as listed in the preceding section (CPU, storage devices, terminals) are separate units connected by communications channels. Decentralized data collection or bulk data transfers, remote job entry or time-sharing, inquiry/response or information and text retrieval are many activities which call for a sophisticated communications network. The more numerous the transactions between the above components (as is the case in any type of on-line, real-time, man-machine interaction), the more critical the communication system becomes.

Some idea of the communications system cost is given by a recent Diebold Group Survey.[1] It shows that the user of a data communications system spends about 50 percent of his dollars on lines, 35 percent on hardware (modems, multiplexors and communications processors) and 15 percent on salaries and overhead, with data communications cost representing 8 percent of the total data processing cost.

In the Bell System[2] in the U.S., telecommunications costs break down as follows:

---

[1] "Trends in Data Communications," Lynn Hopewell, in Datamation, August 1973.

[2] Ibid.

- long-distance transmission 17%
- switching 45%
- local loops 15%
- terminals[1] 23%

In spite of the critical economic role played by communications, the complexity of the rate structure is such that charges must be worked out on an individual basis with the common carriers. The charge is determined by considerations of the type of terminal equipment needed, the volume of data transmitted, the grade of service, the line capacity, the distance of transmission, and the mode of operation. Since a minimum of understanding of the elements involved is needed, an attempt will be made to cover them briefly in the following section.

Communications costs are insignificant below a distance of 10 miles[2] as a rule. When the distance is greater, much care has to be brought to the network design, since communications costs may well dominate the cost equation.

---

[1]The literature often refers to the terminals as being part of the communication equipment, while for our purposes we consider terminal costs as making up the computer system cost.

[2]At least, local loops cannot benefit from network design studies.

3.1.2.1.2 Digital vs. Modulated Transmission

With the recent introduction of Dataroute by TCTS and Infodat by CP-CN, both digital data transmission services, there exist now two main types of data-transmission. Digital Signals[1] can be sent as they are, without modification - digital transmission technique - or they can be carried by a higher frequency waveform - analogue technique. Until recently, only the analogue technique could be used because of the distortion which affected the digital transmission as soon as the distance exceeded a few miles. Although in the sequel we shall concentrate on Dataroute, it is useful to examine the characteristics of analogue networks, for the simple reason that in the present state of the art there is not a pure digital network but a combination of the two.

In an analogue (or modulated) transmission system, we find all the elements of a communication system: the source (the computer), the encoder (the modem), the communication channel (the line), the decoder (another modem) and the receiver (the terminal). The binary coded signals of the computer are handled by a modem (contraction of modulator-demodulator) which convert them into analogue signals before they are sent over the line. The converse is realized at the other end of the line, the analogue signal being converted back to digital by another modem.

[1] "Telecommunications and the computer," James Martin, Chapter 14, Pulse techniques.

The waveform which carries the signal can be a voice-band grade (on the telephone line) or broad-band grades (joint telephone lines with greater capacity). Typical such lines are TCTS's Multicom and Dataline and CP-CN's Broadband.

The modem thus is an essential building-block in the analogue transmission system. Another one is the multiplexor, which permits a certain number of low-speed lines to be derived from one voice-band grade line. Most often, multiplexing is economically attractive, since one voice-grade gand line is cheaper than the twelve or twenty low-speed lines of equivalent total capacity, and this the longer the distance. Multiplexing can be achieved in many ways, either by rapidly switching one low-speed channel at a time (time-division multiplexing), or by buffering the input at low speed, or by sending many frequencies on the line (frequency-division multiplexing).

In any transmission system, the signal gradually fades away as it travels. In an analogue system, it is periodically amplified, but with each amplification, noise is introduced and thus data errors. In a digital system,[1] the signal may also be amplified, but the essential is that it triggers a regenerator which reproduces a new identical signal without noise. The signal is thus periodically cleaned up: this technique reduces transmission errors.

---

[1] op.cit.

Another advantage of a digital system is that no conversion interface is needed between the digital computer equipment and the transmission network since both use the same signal. Translated into economic terms, it means that the overall system cost per terminal is no longer charged with the modem price, and the per terminal cost is likely to show a fast downward sloping curve tending to more reasonable levels.

At this point it must be noted that in an analogue network advantage can be taken of the statistical nature of the messages, e.g. instead of providing for the peak load, the line only provides for a multiple of the average load (the adequate multiple is computed by a Poisson distribution table, given a desired grade of service or average response time). The line can thus be more efficiently used.[1]

---

[1]Currently, only multiplexors are used in digital networks: multiplexors, although lowering the per-channel cost, still allocate a fixed channel to a terminal during the transmission, even if the channel is inefficiently used. By contrast, concentrators (which were used in analogue networks) can effectively connect one channel to many terminals. The experience of Dataroute is that no customer has yet proposed to provide concentrators, although the capability exists.

3.1.2.1.3 Types of Communication Links

There are two main transmission modes, depending on the directional flow of data: half-duplex, when transmission can be in both directions, but not at the same time; and full-duplex, when transmission can be made in both directions at the same time. Another distinction, which we alluded to, is the grade of the channel (its capacity or speed, in terms of bits per secons bps). Depending on the type of multiplexing, the required channel capacity may either be a direct function of the number of low-speed terminals or take advantage of the statistical nature of data traffic parameters.

Of relevance also is a technical aspect of transmission: synchronous or asynchronous. Asynchronous transmission is the slower mode, signal elements being transmitted to indicate the start and stop of each character. Synchronous transmission is used for high-speed and eliminates the need for start-stop bits, since the receiver is bit by bit synchronized with the sender.

A fourth consideration is the type of service the system asks for. A dedicated line (or private line P/L) remains connected between its terminal points for the duration of the lease. A switched line (dial-up service), on the contrary, is allocated on a pay-as-you-use basis. Payment varies according to mileage, connect time and time of day. Choice between the two alternatives (if available) is based on the proportion of time the line is in use, and distance.

3.1.2.1.4 <u>Cost Considerations in relationto network design</u>

The problem of communications network design is a real-time, on-line computer system can be stated fairly simply. There is a central computer and n terminal locations. Among all possible designs, choose the one that minimizes the overall network cost.

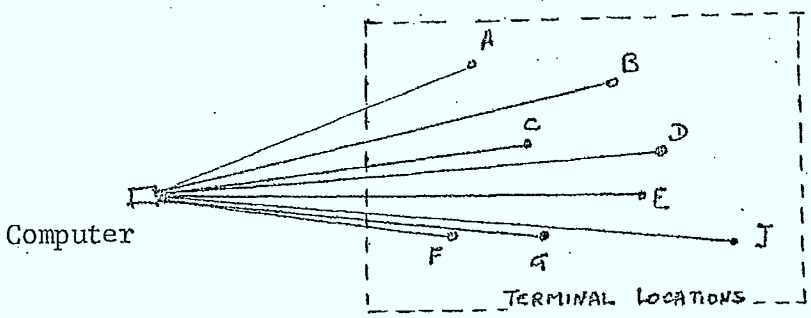


Fig. 1

At one extreme, each terminal can be connected to the computers by a dedicated line. Of course, each line will be used intermittently and inefficiently, since the terminals will not be receiving or sending constantly and are not able to use the line capacity fully.

There exist a number of devices[1] that allow for a reduction in costs:

(i) the use of line switching (or dial-up access),

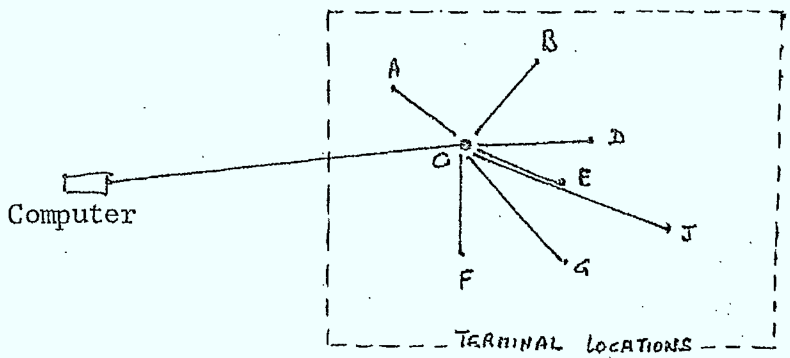


Fig. 2

[1] James Martin, "Teleprocessing Network Organization," Chapter 7 Prentice-Hall, 1970.

which allows one terminal to be on-line with the computer (more than one if there are more lines from the exchange to the computer: the main consideration here is the waiting time for connection). With uniform and longitudinal distribution, line mileage is reduced by a factor of 1/n.

(ii) multidrop lines: here, only one terminal at a time can transmit or receive; however, the terminal is not waiting for

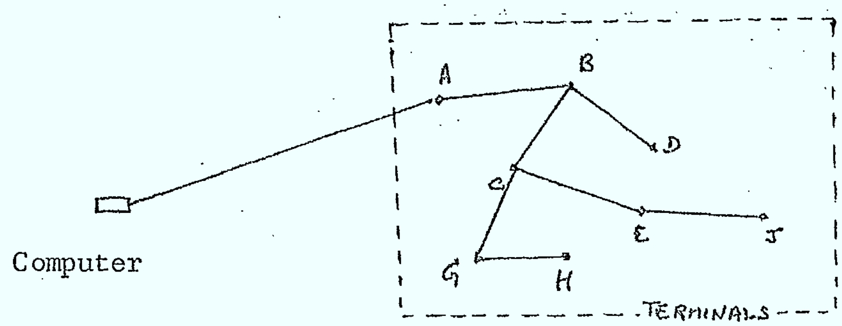


Fig.

connection, but for message transmission only, i.e. the waiting time has not the same parameters. Note that the line mileage again is lower than in Figure 1.[1]

(iii) multiplexing: this refers to simultaneous transmission. In the above design, only one terminal at a time had control over the line; multiplexing allows for many terminals using the line at the same time, by dividing the line into as many channels (the ways in which the dividing-up can be achieved have already been alluded to). The design with multiplexing is shown in figure 4.[2]

[1]The additional equipment cost in this case is a polling arrangement.

[2]The additional equipment cost here is that of a multiplexor in C and one at the computer.

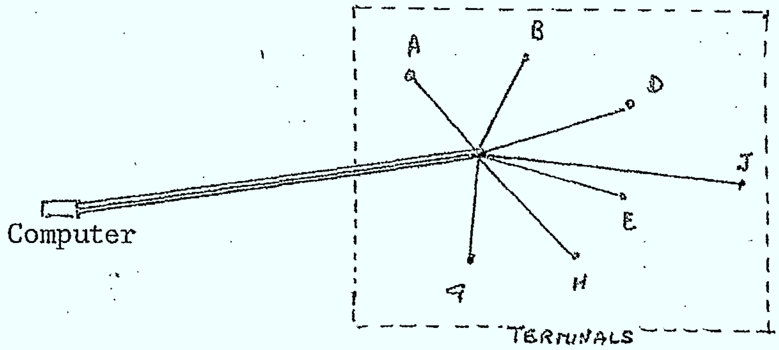


Fig. 4

The difference with Figure 1 is that we have here a high-speed line between C and the computer.

(iv) Buffers and concentrators are devices which further enhance the efficiency of the network. Terminals are man operated, and thus are themselves not used at optimum speed (consequently inefficiently using the line speed). Buffers store blocks of data and then send them in bursts at maximum line speed, as soon as the block is full. Concentrators are more complicated buffers, at a higher speed level, but they basically have the same function.

3.1.2.2. <u>The Advent of Dataroute and Infodat</u>

These two digital data transmission services offer the user cheaper lines and improved error performance. Using pulse code modulation[1] (PCM) together with Time Division Multiplexing (TDM), the digital networks will be the basis of our study, since they are the latest telecommunication technology.

3.1.2.2.1 <u>Description</u>[2]

This section will deal with TCTS's Dataroute offering more specifically. Dataroute is a private line digital data transmission service, in full duplex, over a broad range of asynchronous (lower speed range) and synchronous speeds (110 to 50000 bps).

The Dataroute network connects the major cities of Canada (called for this purpose the Dataroute Service Areas: DSA's) with the possibility of:

(a) end-to-end private line digital data service for customers within DSA's (on-net locations).

(b) Dataroute service combined with analogue private or dial access lines for customers within or outside a DSA (off-network location).

(c) point-to-point service, multipoint service (more than two points), multidrop service (a number of points within a DSA connected via a single channel).

---

1 Martin, op.cit., chapter 14

2 Dataroute Customer Reference Manual

3.1.2.2.2 Elements of equipment and network design

We shall first have a look at the basic features of the Dataroute inter-DSA line, then consider the related devices, and eventually discuss the alternative accesses to the Dataroute line. Among the services offered, the point to point is the simplest: it involves two ends (a terminal in Toronto and the computer in Montreal, for instance). When more than two Dataroute Service Areas are linked, we have a multipoint network, with any one DSA being connected to the n-1 others (such a design would be suitable for a distributed data-bank, where requests and updates are made across data bases). When more than one terminal point is needed in a given DSA, we have a multidrop design. These are the fundamental linkings between and within DSA's; combinations of multidrop and multipoint arrangements are possible with the constraints set by channel capacity. The combinations satisfy almost any service requirements, as long as the terminal points belong to a DSA. Outside DSA's, the combination of other data services (Multicom and Dataline) with Dataroute provides a complete coverage.

Two types of multiplexors make up the interface between a DSA and the inter-city Dataroute line. The synchronous multiplexor concentrates the input speeds into the 56,000 bps speed of the Dataroute common data stream and decodes it on arrival. The asynchronous multiplexor (or characterplexor) converts and combines lower-speed, asynchronous channels into synchronous, 2400 to 50,000 bps channels, which will then be handled by the synchronous multiplexors.

Medium speeds thus enter the synchronous multiplexor directly, while low speeds pass first through the characterplexor. This hierarchy of multiplexors thus provides maximum flexibility, as will be seen later (Low-Speed Deriving Service).

Within a DSA, there exists a number of possibilities for access to Dataroute (Dataroute Access Arrangements, DAA) besides the normal private line on-network access. These are the dial access (which uses the analogue data services between the terminals and the asynchronous multiplexor) and the lower speed deriving service. Through dial access, one user only has Data route on line at a time. With the lower speed deriving service, a user leasing a 2400 bps. or higher Dataroute service can obtain many lower speed services (110-600 bps) through the characterplexor, according to the table shown below:

low speed channel rate \ leased channel speed	2400	4800	9600	19,200
110 bps	25	50	100	127
134.5	16	33	66	127
150 bps	16	33	66	127
300 bps	8	16	33	66
600 bps	4	8	16	33

Number of Low Speed Channels

3.1.2.2.3 Elements of Cost

Regarding costs, Dataroute offers a pricing scheme which is exposed below. Dataroute is a private line service offering and the monthly cost has two main components:

- the inter-city line, which cost is a function of both distance and capacity in bits per second; the service period, day, night or 24-hour, is also taken into account. The city-to-city rates are shown by matrices of line capacity ranging from 0 bps to 50,000 bps.
- the local arrangement (multidrop, lower speed deriving service, dial-up access for on-net locations; dial-up access, private line in combination with analogue services for off-net location).

Since the local arrangement is more complex, it is useful to distinguish the alternatives. From here on, we shall be dealing with a point-to-point Dataroute line (say, Toronto to Montreal) with on-net (e.g. within DSA) users. The simplest design would occur if there were one terminal in Toronto linked to a Montreal computer. With two or more terminals, each either has:

(i) the dial-up access through the analogue network (terminals operating at speeds lower than 1200 bps). Rates are given by a Dataroute Access Arrangement table per terminal:

speed	monthly rate
110 → 300	$ 40.00
301 → 600	50.00
601 → 1200	75.00

In this case, the Dataroute line will not exceed the 1,200 bps.

(ii) the straight private line digital access (also called multi-drop), in which each terminal has, one at a time, the full capacity Dataroute service on-line. Rates are given by another Dataroute Access Arrangement table per terminal:

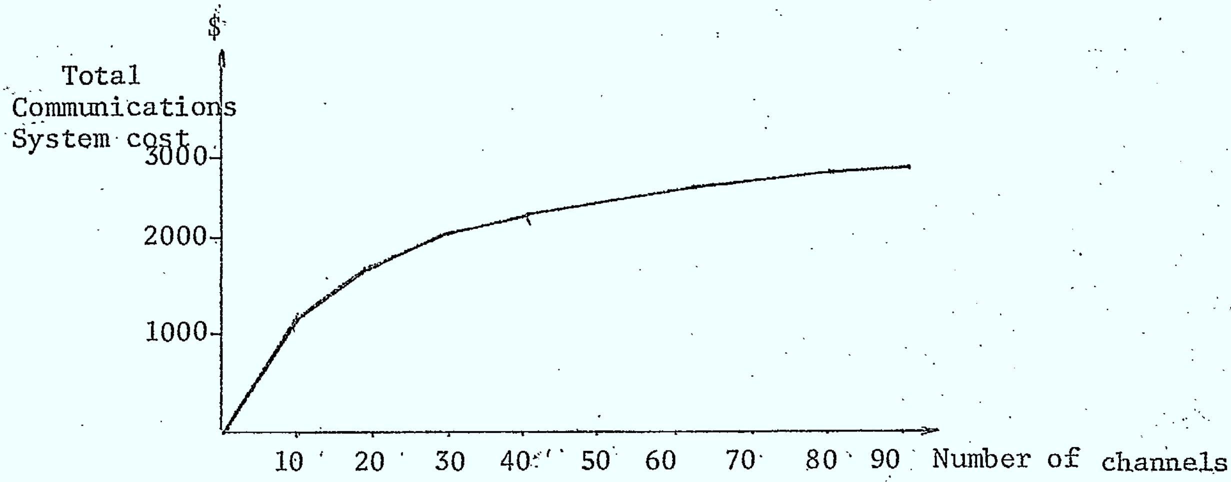
speed	monthly rate
110 → 300	$ 40.00
301 → 600	50.00
601 → 1200	75.00
2400	125.00
4800	225.00
etc.	

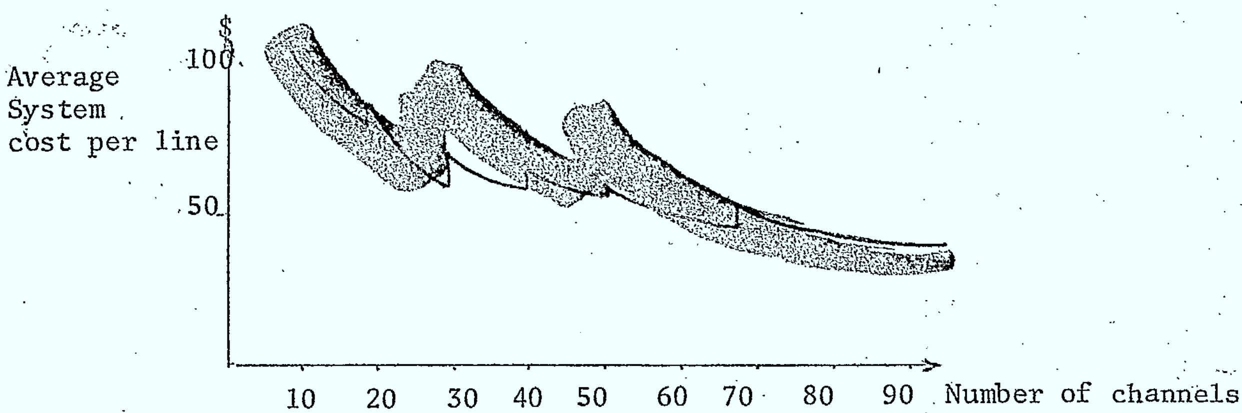
(iii) the multiplexing of low speed local (private or switched) lines into a high-speed Dataroute line. With private line access, each terminal has permanent access to a low-speed Dataroute channel. With switched lines, terminals have dial-up access to low-speed Dataroute channels, through the analogue network. Rates are given by lower-speed deriving charges according to the table below:

	monthly rate
first four terminals	$115.00 ea.
next twelve	30.00 ea.
next ten	25.00 ea.
over twenty-six	20.00 ea.

and are additional to the Dataroute Access Arrangements shown above, for the high-speed lines.

In total, if both line costs and local arrangement costs are summed, the communication system cost curves appear as under:





Moreover, calculations shown in the appendix demonstrate that for a sufficient number of terminals in each DSA, the cost per channel is relatively insensitive to both distance and channel capacity. Thus, distance truly drops from the communications cost equation.

Beside this, if an adequate number of users share the communication facility, the communications cost itself becomes negligible. For instance, a community of 10 users will pay $200 for each channel, whereas an organization of a hundred users will be charged $60 per channel. This conclusion holds true mainly in reason of the specificity of the offering: Dataroute as a private line means its cost is a fixed cost which has to be spread over as many users as possible. A bulk leasing of Dataroute lines is thus the more economical the larger the organization.

3.1.3 Synthesis

3.1.3.1 One-time vs. recurring cost

3.1.3.2 Maintenance vs. development

3.1.3.3 Salaries vs. hardware

3.1.3.4 Internal system cost vs. user-to-system cost

3.1.3 Synthesis: alternative categories of costs

So far, a certain number of data-bank components have been reviewed: notably computer system and the communication system, in order to arrive at a cost model. Model building is but one step in an integrated methodological process aimed at designing an optimal system. In the small, model building should aim at a representation which identifies the decision variables. In one word, a cost model should include the relevant costs of a class of decisions.

The specification of a class of decisions limits the scope of the model. For instance, a model which helps to decide between alternative policies concerning the retention of series, necessarily entails considering the cost associated with a particular series: what is the cost of keeping this series in the data-base? What is the cost of not keeping it? A model which is valid for all types of decisions is infeasible. The alternative is to present a range of "models", each displaying a facet of the complex reality.

Viewing a model as a way to describe situations, a model should provide two main pieces of information: the magnitude of the total costs and the breakdown of this figure into its components. It is necessary to emphasize the importnace of the breakdown: the characteristic of a technology like that of computers is its pace of change. Each year sees a number of technical breakthroughs, and the relative costs of inputs change dramatically.[1] The technical design

---

[1] Such an example is the relative share of hardware and software in the total cost of an installation.

of a system can always be translated into costs at a point in time, but perhaps more important is how this cost will behave over a period of time.

Three models directed to this problem are presented first, to give some perspective to the costing problem. A fourth model, drawn from a paper by D.N. Streeter, exposes some useful concepts.

3.1.3.1 <u>One-time vs. recurring cost</u>

This breakdown points to the balance between the initial investment and the annual cost. The big effort can be made at the start on spread over the planning horizon. This decision depends technically on the modularity of the system (e.g. how small the increments in capacity can be), but also on the user demands and on the funding method. The total cost over the planning horizon $t = 0$ to $t = n$ is:

$$TC = I + \sum_{t=0}^{n} \frac{C_t}{(1+r)^t}$$

where the C_t are the annual cash outlays and r the cost of funds. The higher is r, all other things being equal, the more the designer would want to spread the costs and reduce the initial outlay. Consequently, some implementation decisions can be viewed in this framework:

(i) On-line vs. batch processing: On-line involves terminal and communication costs, additional control hardware, on-line bulk storage, large terminal and job supervisory programs. Although, through the manufacturers' pricing policy, some or all of these additional costs can be spread over a period of time, there still remains a heavy set-up cost.

(ii) Time-shared vs. dedicated system: In a time-shared system, someone else takes the risk of a high set-up cost, while the user spreads his cost. Furthermore, by the very nature of time-sharing, users' costs more closely match users' needs because the system tailors its resources to the particular job, besides all the data processing technical support and back-up capabilities of which the users can take advantage.

(iii) Lease vs. buy: The relevance of the formula

$$TC = I + \sum_{t=0}^{n} \frac{C_t}{(1+r)^t}$$

is here obvious, since the evaluation of the appropriate policy will be directly made through the application of discounted cash flows techniques.

(iv) Integrated vs. individual applications:

The set-up and conversion costs of integrated applications are huge.

Most computer system managements prefer to build up the facility gradually, application by application. As can be seen, the funding subsystem is critical: more than the users' demands, the amount and timing of the subsidies will ultimately determine the time path of development for the data-bank.

3.1.3.2 Maintenance vs. Development

This view stresses the bottleneck in development, e.g. the bulk of the resources are devoted to the maintenance of the programs and the data-base. The cost of operating and maintaining existing applications limits the part of the annual budget available for new application areas. Often, this breakdown is a consequence of the initial set-up vs. recurring cost decision. What is development actually? In the case of the FRI, it would certainly be appropriate to include marketing costs to the extent that they are aimed at finding new customers , developing new usages.[1] Customer support cost then itself breaks down into operating support (e.g. completing the written documentation) and user education, which truly is development.

Perhaps a more useful cost distribution is that proposed by Richard Bower,[2] between development, capacity and operations. For him, development goes in three directions: system development, new customers, new usages. Capacity is the fixed cost (in equipment as well as in personnel) that must be incurred to service when requested (this means a significant reserve capacity must be maintained); capacity costs can be assimilated to the period costs. Operations are the direct costs of servicing users[3] (they also include, as our maintenance does, the operating support). The correspondence between our distribution and Bower's is to be viewed as:

---

[1]Developing new usage is the deepening of the usage of existing data and programs by promotion and user education.

[2]"The Computer services industry" in The Bell Journal of Economics. September 1973.

[3]Or variable operating costs.

Bower's	Development	Capacity	Operations

Ours	Development	Maintenance

Another allocation is quoted by Bower in his article, referring to a Diebold Survey (1971), between new systems development, ongoing applications and conversion.[1] The Diebold estimates give:

New Systems Development:	30.4% of EDP resources
Ongoing Applications:	59%
Conversion:	10.6%

which Bower redistributes according to his scheme:

Development:	25.4%
Capacity:	49.7%
Operations:	24.9%

Bower quotes another study by Adapso (1971) and his own survey of computer services with these results:[2]

	Adapso	Bower
Development	22%	12%
Capacity	48%	58%
Operations	30%	30%

[1] It seems that this last distribution best applies to industrial and manufacturing firms, which the study did in fact survey.

[2] In the Diebold and Adapso Surveys, percentages are Bower's own. In Bower's survey, percentages are ours (average of Bower's results).

Variations can be explained by different interpretations and allocation of systems programming staff costs between development and capacity.

3.1.3.3 Salaries vs. Hardware

The Diebold, Adapso and Bower surveys give data about cost distribution between inputs:

		Adapso	Diebold	Bower	FRI alone	FRI + Comp. Centre
	Hardware	30%	35.4%	30%	30%	65%[1]
salaries	Systems personnel Operating personnel	42	25.4	50% (approx)	35%[2]	17%
	Supplies	4	6.8		20%	11%
	External services	2	3.8	20%		
	Other	22	0		15%	7%
		100%	100%	100%	100%	100%

Overall, close to 60% of the cost is tied to staff and personnel and only 30% to computer hardware. If hardware costs were to drop to zero, there would not be a proportional incentive to develop things at a faster pace.

Some figures are distorted due to the necessary redistribution:

[1] This high percentage is the result of taking the overall user computer bill as hardware cost. But the computer bill is the user-allocated computing centre costs which would include 50% staff costs. A correction would then be to take out 35% of hardware costs and put it in salaries.

[2] The low salary percentage is balanced by a high external services percentage, due to the data acquisition costs. Entry of data on machine-readable media is made by the outside suppliers.

For an automated system, personnel cost, either systems and programming staff or data preparation staff, breaks down this way:

1. planning
2. actual design
3. coding
4. testing-debugging
5. file conversion
6. actual data preparation and correction
7. program maintenance

while equipment costs are:

1. computer costs:
   (a) compile time
   (b) test-debug time
   (c) run time
   (d) maintenance
2. additional:
   (a) terminal
   (b) storage
   (c) communications
3. material costs, if any.

More automation means more hardware cost and less manpower cost. But again, what falls into these two categories? Consider, for example, acquiring tapes in which some manpower has been inputted, if it is not considered as only manpower. When the system design decision has to be made, relative costs of hardware and manpower have to be assessed, together with their predictable evolution, since relative price changes may drastically modify the economic desirability of a given design.

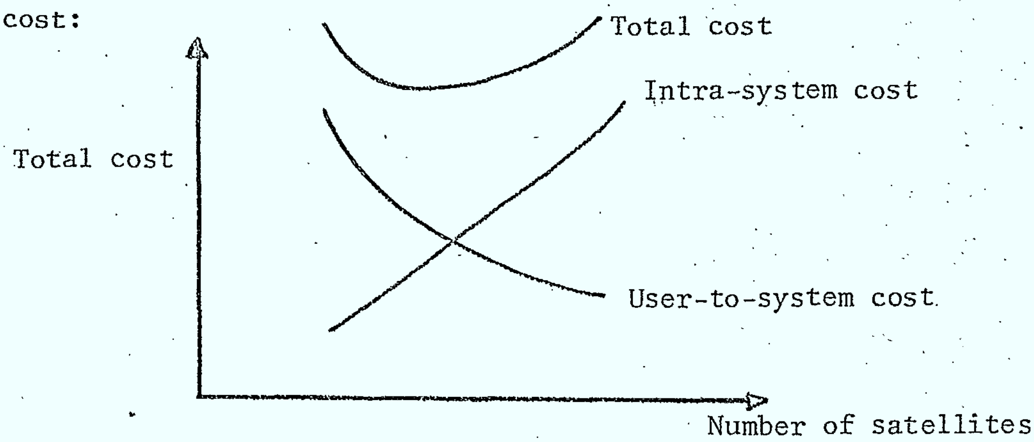
3.1.3.4 Internal System Cost vs. User-to-System Cost

In a recent article,[1] D.N. Streeter proposes a model for evaluating the optimum dispersion of computer facilities. We shall retain from his model a distinction between the internal system cost and the user-to-system cost:

(i) the internal system cost essentially is computing and storage costs plus intra-system communication costs (e.g. the costs of transferring data between the satellites of the system: updates or requests for specialized services).

(ii) the user-to-system cost is the communication network cost applying to the local loops from users to the next data-bank.

All other things being equal, increasing the number of satellites also increases the intra-system costs; the total cost, however, may or may not increase due to savings in the user-to-system cost:



---

[1] IBM Systems Journal, Vol. 12, No. 3, 1973: "Centralization or dispersion of computing facilities."

The trade-off eventually leads to an optimum of total cost.

The distribution of costs is an essential consideration in view of dispersion problems, for which the only relevant costs are storage and communications; computing costs (servicing the queries) can be assumed to be fairly constant across different computing systems, and thus it makes no difference which data bank the user accesses. Administrative costs may be subject to economies of scale but we cannot be far from reality by assuming them proportional to the number of users serviced; they are then independent of the design decision, and the only relevant costs remain storage and communications.

It is important to remember this conclusion as it will be a key assumption in the dispersion model of Chapter 2, Part III.

3.2 The information storage and Retrieval viewpoint

3.2.1 Costing problems

3.2.1.1 The output measure

3.2.1.2 Accounting difficulties

3.2.2 Useful cost categories

3.2.2.1 The cost of a series

3.2.2.2 An activity analysis model

3.2 Costs of a databank : An Information Storage and Retrieval Point of View

There is a striking analogy between databank operations and library management : they both accumulate material over time ; the information inventory is in never-ending growth ; the inventory does not depreciate, indeed it appreciates. There is an even greater analogy between the FRI operations and a library of serials : they both go through the process of acquiring material on a subscription basis, storing and retrieving it by its issuing date. Many a problem in the management of a library of serials finds its counterpart in the context of the FRI, where the commitment of subscribing is but one aspect : the retention period for a journal is equivalent to the problem of fixing the lenght of the series ;the number of journals decision is similar to determining the number of series ;the decision of whether to own a journal or to borrow it[1] is made on the same grounds as a decentralization-centralization problem for the files : does the volume of request justify the cost of keeping the journal (or the file) versus the cost of borrowing (or transmitting) from the central library (or the central file location).

Because of this fruitful parallel, the first section of this chapter will discuss the recurring problems of

---

[1]From another library, when a request is made.

accounting for information systems ;the second section will expose two planning and decision models which hopefully are in accordance with the concepts reviewed.

3.2.1 Costing problems : output definition, shared inputs, cost allocation, accounting procedures

3.2.1.1 The output measure

Cost models have been established for industrial organizations where value is added to a flow of input material until the final input. Libraries and databanks depart so much fron this basic model[1] that it is necessary to contrive new concepts, if not in economics, at least for accounting purposes.

One possible approach would consist in considering an input flow of information, basically disseminated, unstructured, non-standardized, thus devoid of informational content, and with almost no value to the data in terms of accuracy, timeliness, completeness and structure. In short, the system reduces the entropy of information. Some terms need comments : what is the value of information, to begin with. Theoretically, in the Bayesian probabilistic framework, the value of information is the incremental effect it has on achieving some desired benefit, and the analysis requires the prior probability distribution of each possible outcome, the accuracy of new information, the posterior probability distribution and a payoff function of benefits associated with each outcome[2].

---

[1] As do hospitals, educational organizations and public organizations in general, by the way. Note that cost-benefit, cost effectiveness analysis methods have been devised to deal with the evaluation of these systems.

[2] An excellent rapid presentation is given in James C. Emery "Fundamentals of computer-based Management Information Systems" (Addison-Wesley 1973).

Clearly, this is not a viable solution to the problem of determining the value of information.

It is often easier to directly measure benefits by referring to an objective monetary value[1] of information. Benefits from an information system occur through a reduction in operating costs or an increase in revenues of activities using the information system. Obvious examples are lower inventory carrying costs resulting from an inventory control system, reduction of cash balances not bearing interest from faster updating of bank accounts, suppression of routine paper work in an investment firm (e.g. ratio analysis of income statements and balance sheets).

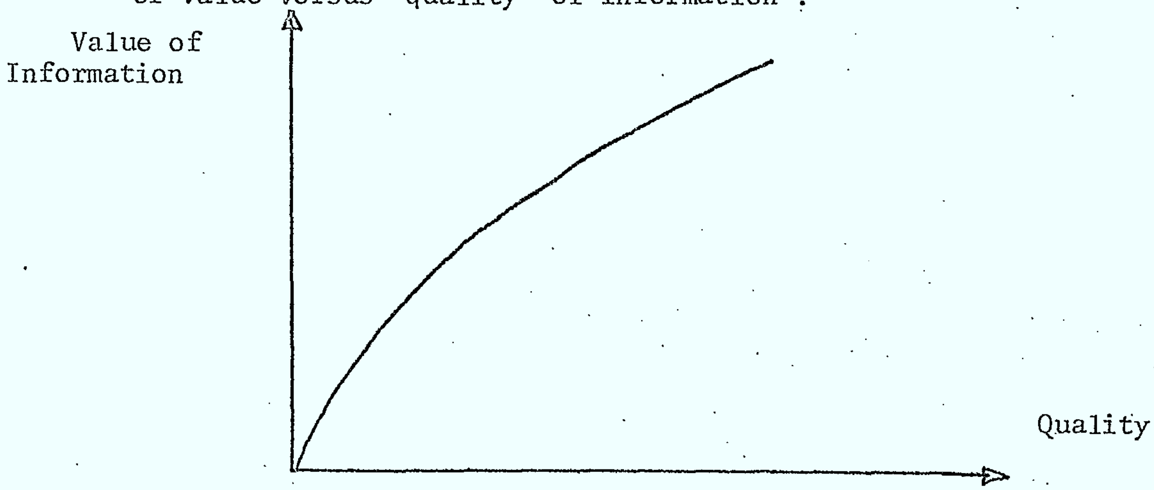
Although benefits of this sort can be assessed fairly accurately when alternatives are stated (e.g. the current method of operation vs. the proposed system) and cases are simple, it does not help in finding the optimum, and thus estimating the improvements due to the change.

Other benefits can be found in improved services, and in better decision-making . However, estimates here become increasingly difficult to make. How, for instance, can one evaluate the capital gains made by investors through a more practical investment service (e.g. the FRI) ?

We can thus say that an information system enhances the value of information. But how ? What are the characteristics of the information system which improve the "quality" of the information ? We mentioned them in passing : accuracy, timeliness, completeness, structure, ease of accessing. The problem is to find the mix of

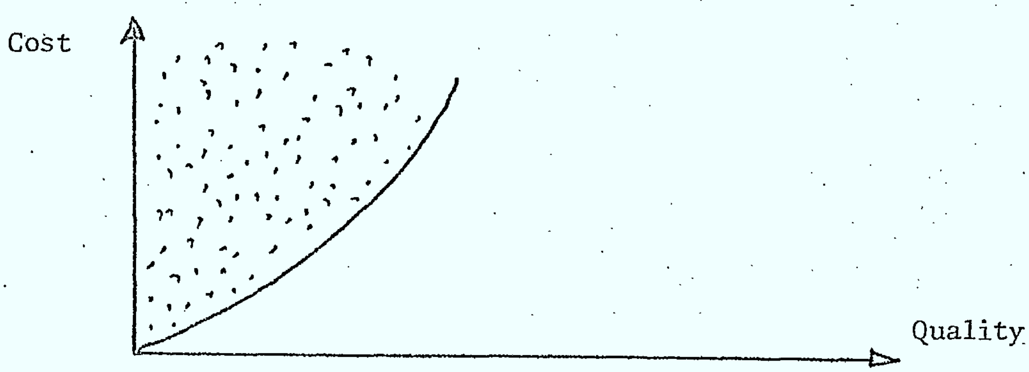
[1]Chapter 3, "Theoretical Analysis of Information Systems", Borje Langefors (Barnes and Nobel, 1970).

characteristics which will provide the best benefit/cost measure. Some relative measures of these "qualities" of information can be imagined, and a system be entirely defined. Accordingly, we can draw a schedule of value versus "quality" of information .



Usually the curve will show diminishing marginal value of quality.

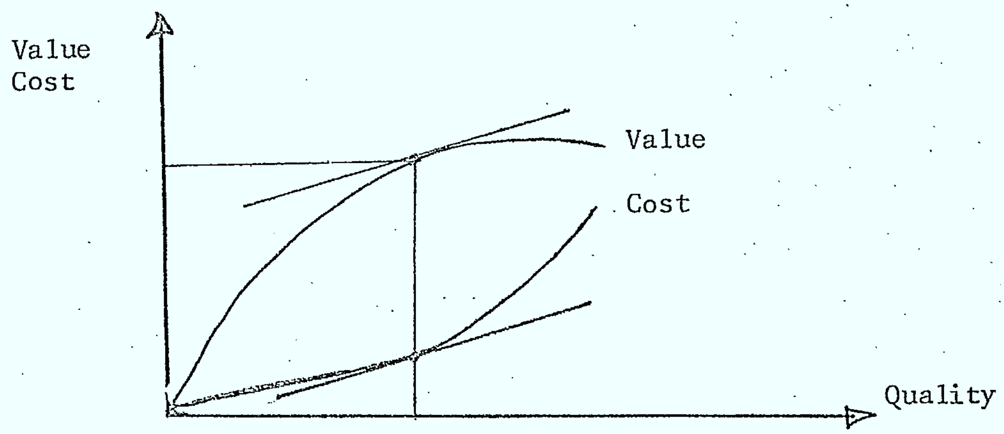
Let us now look at the other side of the equation - costs. Each system defined by its qualities has a cost. There exists a finite number of possible systems[1]. However, a rational decision-maker would choose a system with higher qualities among all those of the same costs and the cheapest system among all of the same quality. If we concentrate upon only one quality, the set of efficient systems is thus given by the graph below .



1This is cost/effectiveness analysis;it is inspired from a succint exposition by W. Sharpe in "Economics of the Computers" (Columbia Univ. Press, 1969).

The analysis cannot stop at this point. We have the set of dominant systems, but which one should be selected? This choice is achieved by indifference curves connecting quality-cost combinations which are equally desirable. Thus a choice can be made without direct reference to the value of information. However, this is particularly difficult to implement, mainly by reason of the indifference curve determination.

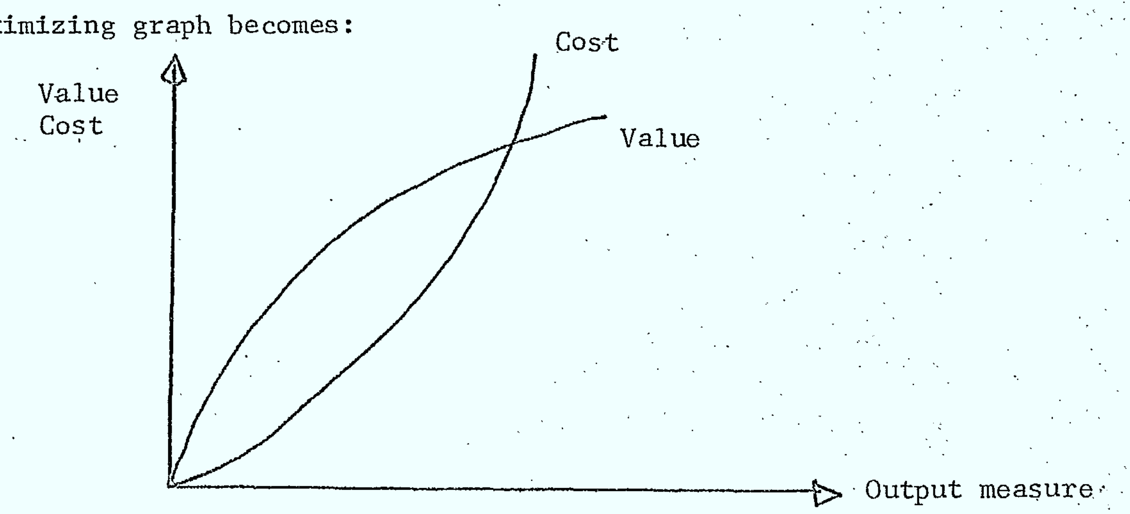
Including the value of information in the framework allows for greater tractability. Indeed, if we graph both the cost and value curves, it is fairly easy to see how the choice can be made: by maximizing the net value (= difference between value and cost), or equivalently by equating marginal value with marginal cost.



The value of information is thus the index of desirability attributed to the quality of the system. We explored some of the ways it can be computed, but we did not indicate one of the pitfalls. There is some indeterminacy concerning the level at which the value should be computed[1]: the value of an information system can (theoretically

[1] An interesting discussion of these aspects appears in "Library objectives and performance measures and their use in decision-making," in Swanson & Bookstein: O.R. Implications for Libraries (Chicago University Press).

at least) be attributed to the attainment of community or societal objectives. This is the ultimate and definitive level at which it should be judged. But it is clearly impracticable. A more modest level would be the impact of the system on its potential users: on their sophistication, ability, creativity. These benefits also are exceedingly difficult to estimate. A more reasonable approach to value is through quality measures (accuracy, timeliness, . . .) as we saw before. However, the multiplicity of measures and, within each measure, the variety of yardsticks, makes it difficult to associate a unique measure vector with a particular system. The simplest performance measure is the output measure, and the net value maximizing graph becomes:



The cost curve then can be interpreted as the aggregate price of input quantities in relation to the level of output. For instance, input efforts in a library can be listed: selection, acquisition, cataloguing, abstracting, shelving. All these inputs are in terms of staff time and dollars, and they result in an output level, which might be the circulation volume or the number of accesses per dollar input or the exposure count times the exposure time.

The main decision variable in a cost model is the output level. However, an information system cannot, as a business enterprise can, categorize its output with precision or certainty: it cannot predict the use that will be made of its material. And of course, because of the variety of candidate measures, it cannot easily assess the product and service levels or even the use intensity. For the FRI, will it be the number of users, the time spent on the terminal, the number of queries, the number of program runs? The definition of an output measure is not simply of academic interest: the pricing of services has to be made on a convenient and objective ground.

In order to define the output measure, there is no need to spell out the physical output: only the operational notion is useful, e.g. the measure. We mentioned earlier the difficulty of finding the appropriate scope for the measure (societal objectives, knowledge, and exposure to information). Another difficulty lies in the inherent limitation of the measure itself. There are two aspects in a data-bank operation: a passive reactive aspect of satisfying the requests (measures of this service are, for instance, the proportion of user demands satisfied, or the response time to an average demand), and an active aspect, which involves creating the demand and multiplying the use (measures which intrinsically include this aspect are exposure counts, the exposure time, or a combined measure, item consulted times the duration of consultation). It is important briefly to review these measures because they all miss some critical dimension of usage. The exposure count (= the number of program runs in the FRI case) does not discriminate between the different exposure types: enquiry (simple display of data) or manipulation (data analysis and synthesis). Adding all exposures thus leads to assembly of dissimilar things, notably, the cost of providing these services might be very variable. Focussing more on the time dimension, the exposure duration time (= the session or connect time at the FRI) provides a better measure of overall output[1] but still does not distinguish between enquiry - and manipulation time. The item consulted times the enquiry - time[2] is a measure which integrates the usage over items as well as over depth of access. However, this measure is more difficult to get and requires implementing a special measurement procedure.

3.2.1.2 Accounting Difficulties

One would think that there is no particular reason why conventional accounting procedures should not be applicable to information services. It seems, however, that their implementation raises more problems than it solves. The main purpose of accounting is to match revenues with costs, and output with costs. Here, one must distinguish between costs and expenditures. Conventional accounting statements

---

[1] In fact, some time-sharing systems charge users according to their connect time only.

[2] A measure similar to passenger-mile, in the case of transport.

show the expenditures, not the costs. Costs, by their economic definition, are those inputs associated with the present output; in contrast, expenditures made in a given period are not necessarily designed to produce usage or exposure in that period. Present exposure is a result of past expenditures, and present expenditures are often designed to produce future exposure (the acquisition and update "costs" are in fact expenditures, inputs that will be used and useful for many periods).

At this point, a parenthesis should be added. Costing a product or a service usually means finding its average cost. For what purposes? Budgeting (requests for funds), allocation of resources and planning, control and pricing. This is in contradiction to the discussion in Section 3.1.1.2 where it was found that only the marginal, or more generally avoidable, cost should serve as a basis for decision-making in general, and pricing in particular. This conclusion must be justified, however: the pricing policy essentially depends on the goal the information centre wants to achieve. Profit maximization, for instance, will entail a pricing scheme which has nothing to do with fixed costs; yet, a self-imposed constraint of budget equilibrium (short-run costs equal short-run revenues) is the only case where fixed cost is relevant for pricing. The latter is the case for a number of organizations, and certainly for information centres in practice. The implication is that some consideration should be given to short-run average cost (or unit cost) which economic theory hardly mentions.

The unit cost computation (total ÷ output) has to bridge the gap between expenditures and costs. This is accomplished by the measurements of costs on an accrual basis, with depreciation and other adjustments for prepayments as main devices. The rationale for depreciation is the loss in value of an asset investment; the user is then charged for this loss in value, which can occur in either of two ways: the investment loses value by the simple passage of time (some physical deterioration will occur without use, and loss of value as a result of technological change), and the depreciation applies to one period of time; or the investment loses value by usage, and the depreciation applies to each use. Most of the time, these two causes of loss of value are mixed, each unit of output bearing a share of the depreciation of the asset during the period (a year). Is this rationale still valid for a document collection? Not quite, since the holdings of libraries or information centres may not depreciate, but on the contrary, appreciate. Then we cannot justify the practice of depreciating a document collection in the usual manner. And the rule for depreciating, e.g. allocating a portion of the asset value to each period, does not make sens, since one cannot measure the useful "life" of the collection (unlike machinery, or buildings, it does not wear out). Allocation of asset investment expenditures to periods is thus, at best, an arbitrary procedure for information-related production. Conventional cost accounting principles, when applied to information services, raise more problems than they solve. The conventional unit cost computation is straightforward: the direct material costs, through some inventory valuation rule (LIFO, FIFO, weighted average cost), are added to the direct labour costs; a portion of indirect, overhead costs is allocated to the units produced during the period, and these total input costs are divided by the number of units produced.

For information services, the first of the problems becomes apparent. Since the production process is not clearly defined, the number of units "produced" is a fuzzy notion. In order to get a better idea of where the difficulties lies, two fundamental modes of operation of an information service centre can be identified.

(i) A proactive mode of operation: this mode is typical of subscription services: current awareness service, recurring bibliographies. Its main feature is that once a user has subscribed, the output of the information service can only be assessed through a quality measure and that, within a range of quality, the centre is relatively free to operate. The main determinant of the activity is the incoming flow of data, which greatly affect the volume of information processed. The effectiveness with which the centre covers the field and filters the data is what is meant by quality. But what concerns us is that the amount of environmental data triggers the level of operations. Since one can expect some regularity in this amount of data, the information centre knows the level of operation it will reach in the next period. Most of its budget goes into fixed costs for capacity. Since, as already mentioned, production units in the short run are a fuzzy concept, and large amounts of costs are joint, shared costs, the costing of services is difficult. In the long run, its success depends on the number of customers and on quality. This situation can be likened to that of a newspaper which has only subscribers, or an academic journal.

(ii) A passive mode of operation: this mode is typical of the responsive information service, which is set in motion by requests only. Although the content of the next request is highly unpredictable (inventory-like information storage is useless), this mode of operation allows for a better, faster quality adjustment of output to the request, through frequent feedback passes.[1] Since most costs are direct costs, cost allocation is made easier.

At one end of the spectrum, then, we have the pure initiative service, whereas at the other extreme, there is the pure "search on request" type of information service. Most information storage and retrieval systems are midway between the two because of their product mix (some "on special requests", some "freely available") and they have to focus on two different things:

a) deal with as many environmental data as possible, e.g. comply with the requirement of completeness.

b) service effectively as many requests as possible.[2]

---

[1]See Kochen and Deutsch, "Decentralization by function and space," Management Science, Vol. 19, no. 8, April 1973, pp. 841-856.

[2]The balance between these two objectives is thoroughly examined in R.R. Wiederkehr, "A net benefit model for evaluating elementary document retrieval systems," in "Evaluation of document retrieval systems: Literature, perspectives, measurements, technical papers," Westat Research Inc. (1968).

This has important implications for cost models: all the mixed initiative - responsive modes of operation require a distinction to be made between acquisition and input processing of the data <u>and</u> the usage, whereas the two extreme pure modes of operation do not require it, since their operating costs are either uniquely related to the incoming flow of data (initiative mode) or to the demand flow (responsive mode).

Fig. 1 . The initiative mode

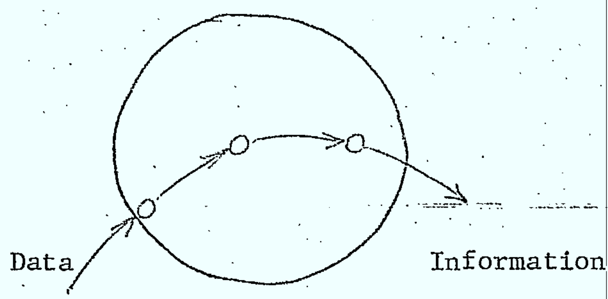


Fig. 2 . The mixed mode

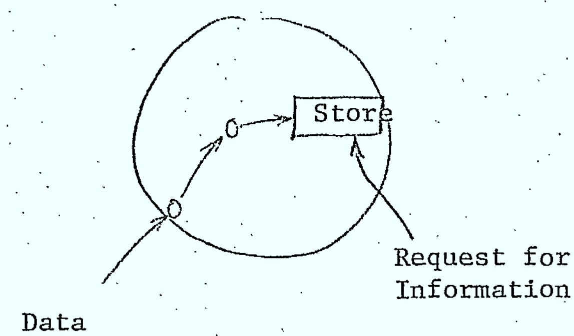
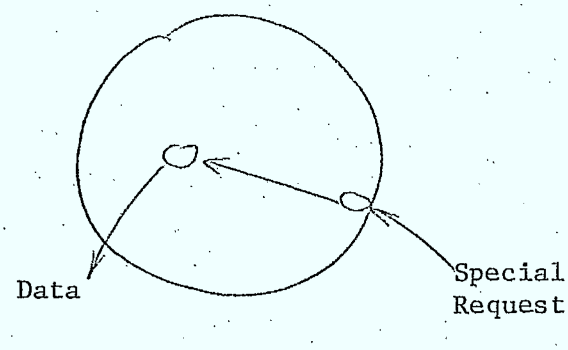


Fig. 3 . The request mode



A second and related topic of accounting for information services is the question whether the acquisition and input (update and indexing) expenditures are to be capitalized or expensed, i.e. are these expenditures an investment or an operating cost? This issue[1] has not been resolved yet, although the capitalization approach is more consistent with the issues raised in the preceding paragraph.

The dilemma of unit cost computation is clear. Insofar as libraries, information centres and data-banks are concerned, a relatively large capital outlay must be made to collect, store, and document information regardless of its use, but documents and information cannot be submitted to the same depreciation rules as machinery and equipment. In other words, there is no fair way of charging the end product with a portion of the capital outlay, and no way of computing a unit cost which would allow the recovery of all costs. It is thus tempting to calculate acquisition and input costs independently of costs related to usage:[2] then, for instance, the cost per exposure (= per access) would simply be the total usage costs divided by the number of accesses. To make up for the depreciation of the

---

[1] Cf. King and Bryant: "The evaluation of information services and products," ibid., and Swanson and Bookstein: "O.R.: implications for libraries," ibid.

[2] Although this might be contrary to real life experience, where initial acquisition and input costs weigh heavily on subsequent usage cost (more effort from FRI to improve file accessibility would result in faster access).

capital outlay, a cost-plus pricing method could be applied or alternatively a two-part tariff: a subscription charge could be levied regardless of usage (this is the case at the FRI: subscription fee plus charge for every access).

3.2.2 Some Useful Cost Categories

Two categorizations of cost are investigated in this section, and are shown to lead to decision models.

3.2.2.1 The Cost of a Series or a Program

The cost associated with keeping a series or a program is a useful decision concept. If we extrapolate the scale of the FRI operations in both the volume of stored data and the volume of usage, then a number of situations may occur where one has to ask what is the cost of keeping a series for, say, 10 years. If the number of series increases up to unacceptable levels, then some series obviously will have to be deleted: the add-delete decision will require some ground rules. Similarly, after 20 years of data, some decisions concerning the retention period will have to be made.

One can think of the cost of keeping a series for a period of t years as being formed of:

(i) the initial cost: the cost of modifying the index, of making the subscription arrangement, etc.

(ii) the annual recurring cost: the acquisition and input cost, a_t+i_t which applies only to new items, and which depends neither on the number of years the series has been stored nor on usage; the storage cost, s_t, which applies to the whole series, and which depends on the number of years of storage (through the series length); eventually the usage cost, u_t, which concerns the whole series and depends on demand.

These two components make up C_t, the annual cost:

$$C_t = \delta a_o + (a_t + i_t) + (s_t + u_t)$$

The total cost of keeping the series in file for n years is:

$$C(n) = \sum_{t=0}^{n} C_t$$

If, as can be expected, the acquisition and input costs are reasonably constant every year, and the storage cost is proportional to the series length (which increases in proportion with time), e.g.,

$$s_t = ns$$

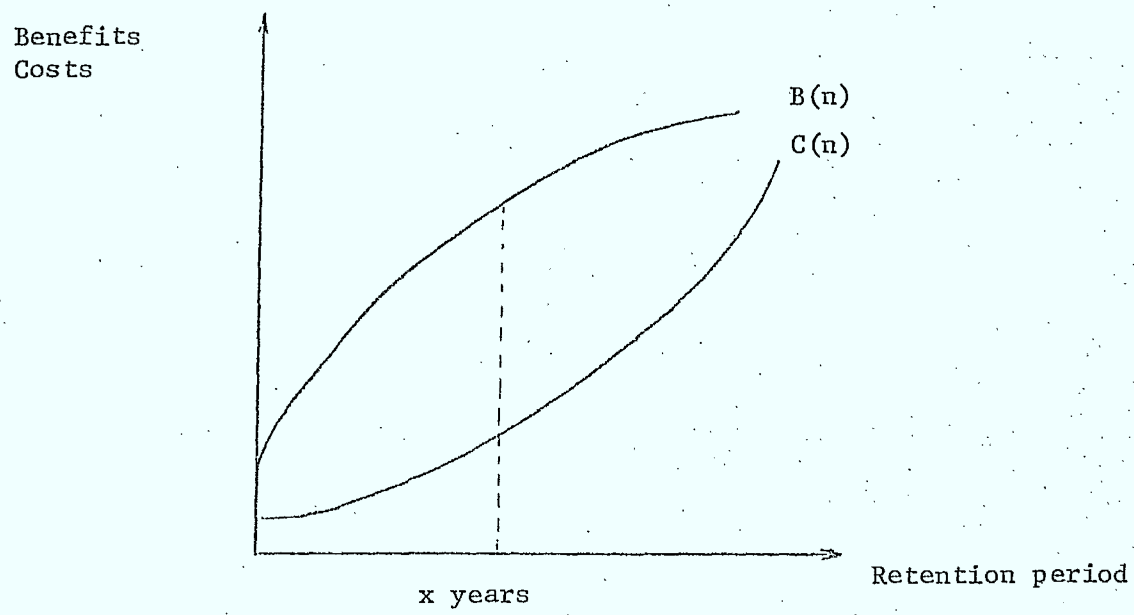
then some terms are removed from the summation:

$$C(n) = a_o + (a_t + i_t)n + (ns)n + \sum_{t=0}^{n} u_t$$

$$= C_o + C_1 n + C_2 n^2 + \sum_{t=0}^{n} u_t$$

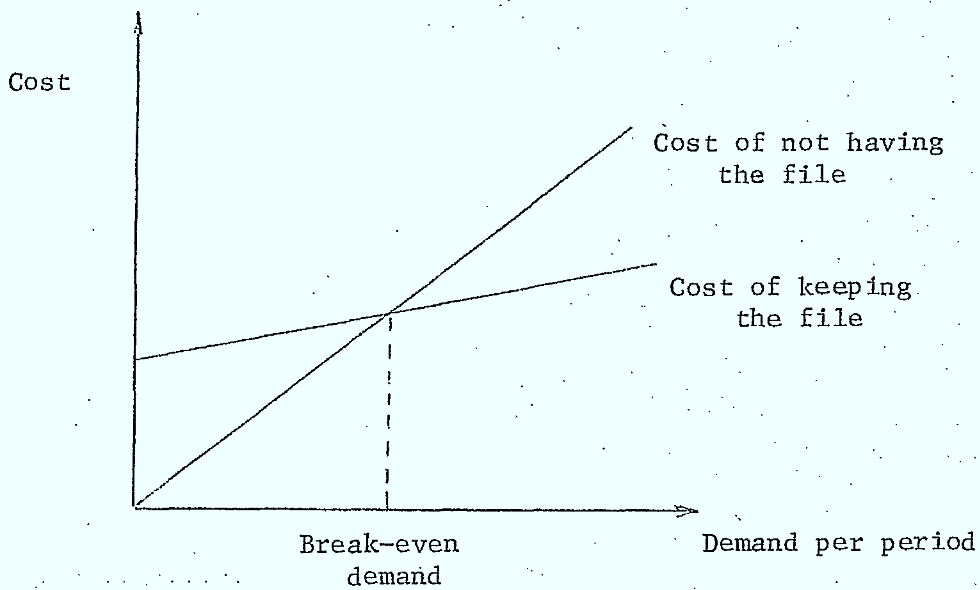
This equation is startling at first, because it is hard to visualize the cost associated with so small a unit of operation. In fact, if we remain consistent in our definition of a series, namely a time sequence of numeric data, a time series, then the series may or may not be small. For instance, the whole FRI stockmarket file is organized day by day, in one series; in contrast, the FRI price history file, organized company by company, constitutes many series. However, it must be recognized that the concept of a series is not a very useful or realistic one, since a decision will never bear on a single series, but more likely on a package of series, like the Vancouver stock exchange file. Accordingly, an alternative is to take one file as the unit, and compute the cost associated with a whole homogeneous file. One more reason to take one file as the unit is that usage cost is not easily assigned to a particular series, whereas it is simpler to allocate usage cost to an aggregate of series, all of which are accessed through the same set of programs. The usage cost for a file is then the sum of the computer costs of running the file access programs.

This cost model can easily accommodate a wide range of decision problems concerning the file. The retention period, for instance, is the maximum length of the file kept available on direct access storage: it can be 5 years, 10 years. The problem can be formalized graphically:



If some pattern of information obsolescence can be postulated,[1] then a curve showing the cumulative value (or benefits) of information against time (or, equivalently, value of information against retention period) will be monotomically increasing at a decreasing rate. The maximization of net value will yeild an optimal retention period.[2]

Another set of decision concerns the subscription to a new information service or the discontinuance of the subscription. Some measure of the cost of not subscribing (cost to the user or outlay cost for the data-bank which has manually to find the data) must be set, and this cost is weighed against the cost of subscribing and keeping the file for a certain horizon period, at different levels of demand:



[1]For instance, the value of current awareness information sharply dim-ishes after one year; the value of the daily stock exchange information becomes negligible after one month.

[2]This illustration is drawn from F.F. Leimkuhler and M.D. Cooper: "Analytical models for library Planning," Journal of the A.S.I.S., November-December 1971.

The choice between storage media, method and frequency of updating, can be facilitated by this type of cost model, which focusses on a manageable decision unit, the file. It is not, however, a global cost model of a data-bank. In the following section, an overall model is proposed, which tries to integrate all the aspects of a data-bank operation.

3.2.2.2 An Activity Analysis Model

In Section 2.2.1, a categorization of operating costs was proposed in which costs were associated with four activities:

(i) acquisition

(ii) input

(iii) storage

(iv) output or usage

This breakdown was applied to a particular measurable unit. In this section, this generic cost distribution is extended to the whole information service. It focuses on the physical processes of information transformation, and takes each step of the processing as a cost centre. It is a mix of cost accounting principles and linear algebra.

A data-bank essentially transforms raw data in various forms and categories into user-oriented information: what is needed, therefor, is a study of the transformational process.

For our purposes, the incoming flow of data is characterized by:

(i) The categories of data: for some kinds of information services the relevant categorization is the source, for others, it is the physical form (journals, reviews, books, microfilms) or the discipline-subject (engineering, medicine, economics). The categorization may not limit itself to one of these dimensions and may present itself under the form of a 2- or 3-dimensional matrix. This attempt to categorize the incoming flow of data is similar to the accountant's effort to distinguish between the different input materials entering the manufacturing of a product. The number of categories is n.

(ii) Their rate of flow: each of the above categories of data is also distinguished by its rate, e.g. its volume per period of time. For instance, economic data are available with certain frequencies (stock transaction every minute or so, stock closing price every day, economic indicators every month, financial reports every year, etc.) which greatly affect their volume. The incoming flow of data can be represented in R^n by $[x_1, x_2, \ldots x_N]$.

The analysis of the transformational processes is facilitated by the above characterizations, because each process can be viewed as the transformation of a vector in a category space into a vector in another category space. For instance, n different categories of raw data (classified according to their source) are turned by the FRI into m categories of data (classified according to their purpose).

The categorization of the resulting data flows is a useful representation of what happens technically or organizationally in the transformational process. For instance, if the resulting data flows were categorized according to the type of machine-readable form (microfilms, tapes, cards), the data flow rates would give a rather accurate idea of the volume of data-processing taking place.

When dealing with the raw data flows, there may be many candidate criteria of classification (we mentioned source, form, subject), whereas within the system only two classifications at most seem relevant: by physical support medium, and by subject.

Thus each process, in successive order, acquisition, input, storage, takes a vector of its incoming data flows and transforms it in another vector. Although this direction was not followed, a measure of the transformational effort could be found through the eigenvalue theorems applying to the linear transformation.

The last process, output, takes a slightly different form, however. The distinction between initiative and responsive modes of operation was noted in the preceding section. Since this model has to accommodate the intermediate range of mixed initiative-responsive types of information service centres, it must recognize the demand for data as the determinant of the output activity. Accordingly, the last process, output, takes a vector of demand flows and transforms it into a vector of demands for stored data. A summary follows:

<u>Acquisition Process</u>

[ED] [TRANSF1] = [SO]

1 or 2 x n-dimensional
ED: environmental data flow rate matrix
1 or 2 x m-dimensional
SO: system-oriented data flow rate matrix

TRANSF1: transformation matrix

Input Process

[SO] [TRANSF 2]= [ΔSS]

SO: same as before
ΔSS: storage growth rate matrix
TRANSF2: transformation matrix

Output Process

[D][TRANSF 3] = [SD]

D: demand rate matrix
SD: stored data demand rate matrix
TRANSF 3: transformation matrix

Thus, as a summary, a vector or a matrix (depending on whether the categorization of environmental data is made on one or two criteria) of data flow rates is transformed into a vector of system-oriented data flow rates through the acquisition process (represented by the transformation matrix TRANSF 1). The meaning of this transformation is that a new categorization is applied to the incoming data, together with a new distribution of flow rates (due to synthesis or analysis of the data).

The vector of system-oriented data flow rates is then transformed into a vector of increments of the collection [ΔSS].

As for the output process, the demands from users, represented by a demand flow rate vector, categorized by service (offered by the information centre), are translated into rates of search for stored data.

In the following an attempt will be made to describe the FRI operations with these concepts. It will be useful to think of the flow of data incoming to the FRI as categorized by source:

(i) the New-York (NYSE), American (AMEX), Montreal (MSE) and Toronto (TSE) Stock Exchanges.

(ii) Statistics Canada (STATCAN)

(iii) Financial Post (FP)

(iv) the chartered banks (CB)

These flows can be represented by a 7-dimensional vector (with units in bytes):

(NYSE, AMEX, MSE, TSE, STATCAN, FP, CB)

Through the acquisition process, this vector is transformed into a vector of system-oriented inputs for which an appropriate categorization (in regard to the final services) is by information classes:

(i) stockmarket data (SP)

(ii) stock price history (TA: for technical analysis)

(iii) montly stock prices (MP)

(iv) high-low price range (HILO)

(v) financial data (FF)

(vi) economic data (ECON)

(vii) banks data (BK)

The vector is thus 7-dimensional (with units in bytes):

(SP, TA, MP, HILO, FF, ECON, BK)

The input processing consists in taking these data and storing them on the files. Since a useful categorization is by storage medium and there is only one storage medium (magnetic disks), the resulting vector is 4-dimensional (with units in bytes): (S). A word of warning is necessary:

(i) the units do not necessarily have to be the same throughout the transformations.

(ii) because some activities which are necessary in transforming the real-world data into machine-readable form are performed by other organizations, the distinction between acquisition and input processing may not exactly match the activities of FRI proper.

One could argue, for instance, that the operations described under the heading "acquisition" would better fit under input process, because of the particular computerized work flow existing at the FRI (that is, the merger of the NYSE, AMEX, TSE and MSE tapes is made during the input processing). We feel, however, that the breakdown which is proposed here is more typical of information service centres which have to convert the environmental data into system-oriented inputs.

The purpose of this representation is to give some physical content to the production processes involved in the transformation of data into information, and to divide the whole operation into meaningful segments of activity, or cost centre. A key factor in doing this is to be able clearly to identify the output of a production process and this is precisely what has been attempted here. The remainder of this subsection considers how costs can be related to output.[1]

Having identified the outputs of four processes, it now remains to relate them to resources or costs involved. Unfortunately, available cost data do not map into the four processes, and the following discussion can be viewed as an attempt to realize this mapping. The main tool used will consist in successive allocation of physical resources or costs to activities.

A first allocation will translate the income statement data into more meaningful terms. This has to be achieved by inside investigation of the accounting records. The resulting distribution may appear as follows:

	CPU	I/O	DK	COM	DOC	SUP	PROG	A	I	S	O
X_1: salaries											
X_2: computer costs											
X_3: communications				A				B			
X_4: administration, etc.											

[1]Part of the impetus of it comes from the seminal work of Yuji Ijiri, in "Management Goals and Accounting for Control," North-Holland Publishing Co. (1965) and "The foundations of Accounting Measurement," Prentice-Hall (1967).

The symbols used will appear throughout the following discussion and they are explained here:

CPU: central processing unit computer time

I/O: Input-output computer time

DK: disk storage

COM: communication

DOC: documentation (written material)

SUP: support (user service, answering questions, etc. . .)

PROG: programming

A: acquisition

I: input processing

S: storage

O: output processing

Four requirements for this distribution must be followed:

(i) costs X_i can be allocated to an activity j without ambiguity.

(ii) since it is practically impossible to list all possible activities, only a few classes of activities are chosen, mainly based upon the organizational structure and/or the use and conversion of economic resources.

(iii) two sets of activities are distinguished: one set is the four main processes defined earlier: A, I, S and O; the other set of activities will serve as a medium for allocating the resource uses into the four processes.

(iv) each activity level is measured by a unit (minutes of CPU for central processing unit, man-hours for programming, etc. . .). This measure provides a means of checking whether the resource which is a priori associated with an activity in fact bears a relationship to the level of this activity.

The resulting matrix will be called [A⋮B], where B stands for the submatrix of main processes. If we sum the columns of the submatrix A we obtain a vector q which represents the resources[1] Q_1, Q_2, . . . Q_m used per period of time. The vector q is thus:[2]

	CPU	I/O	DK	DOC	SUP	PROG
q	200	1000	10,000	300	50	300
units in	Min.CPU	Hrs.	number of tracks	manhours	manhours	manhours

The real problem starts here: we want to allocate each of these resource uses to the other activities (recall each of these resource uses emanates from one activity) and to the main processes. For instance, CPU will be distributed over the programming activity and the input and output processes. The amount of arbitrariness in this allocation is limited by the fact that there will usually be a one-to-one relationship between allocated resource and activity or process (this has been the basis for selecting the activities).

---

[1]When expressed in the unit of the activity level; if they are expressed in dollars it is the cost per period of time.

[2]Only the columns of [A] are summed, since those of [B] are "direct" resource inputs.

CPU	I/O	DK	DOC	SUP	PROG	A	I	S	O
200	1,000	10,000	300	50	300				
-200					20	10	20		150
	-1,000				60	80			860
		-10,000			500			9500	
			-300						300
				-50					50
					-300				280

The elements of this matrix are based on multiple physical units, however, if we consider the following symbolic representation:

$$q = (Q_1, Q_2, \ldots Q_m)$$

$$\left[\begin{array}{ccc|ccc} -Q_1 & & q_{km} & d_{11} & & d_{1s} \\ & -Q_2 & & & & \\ & q_{kj} & & & d_{k\ell} & \\ q_{m1} & & -Q_m & d_{m1} & & d_{ms} \end{array}\right]$$

[C] [D]

The following relationship holds, line by line:

$$-Q_k + \sum_j q_{kj} + \sum_\ell d_{k\ell} = 0 \text{ for all } k. \qquad (1)$$

This allocation operation can be viewed more formally as a matrix multiplication, if the resources q_{kj} exchanged by an activity k to an activity j are proportional[1] to the column activity j, e.g. $q_{kj} = c_{kj}Q_j$ in column j. The c_{kj} can thus be viewed as normalized values of the q_{kj}. We then have, instead of (1):

$$\sum_j c_{kj}Q_j + \sum_{\ell} d_{k\ell} = 0$$

or $\boxed{Cq' + g = 0}$

where g is the vector:[2]

$$\begin{bmatrix} \sum_1 d_{1\ell} \\ \cdot \\ \cdot \\ \cdot \\ \sum_1 d_{k\ell} \\ \sum_1 d_{m\ell} \end{bmatrix},$$

$q' = [Q_1, Q_2, \ldots Q_m]$ and

$$C = \begin{bmatrix} -1 & & & c_{1m} \\ & -1 & & \\ & c_{kj} & \ddots & \\ c_{m1} & & & -1 \end{bmatrix}$$

[1] This assumption is critical; we shall come back to it later. This useful simplification is suggested in Henon: "L'econometrie au service de l'entreprise", Gauthier-Villars, Paris, 1964.

[2] Whose elements are the row sums of the matrix [D].

The same normalization procedure can be applied to the elements of matrix [D], but the column elements are now divided by the output levels of the processes, which are the data flows in bytes. The resulting matrix is a technology matrix

$$[T] = \begin{bmatrix} t_{11} & & t_{1s} \\ & t_{k\ell} & \\ t_{m1} & & t_{ms} \end{bmatrix}$$

The output levels (or data flows) are represented by the vector $h = (h_A, \ldots, h_1, h_s, h_o)$. This normalization is expressed by the identity:

$$G_k = \sum_{\ell} t_{k\ell} h_\ell \text{ for all } k,$$

or in matrix notation $\boxed{g = Th'}$

The matrix [T] has elements of dimension: resource/process activity level. For instance, in the case of the storage resource, the corresponding row has elements expressed in: tracks/bytes of storage.

In this first stage, we carried out the allocation of the resource uses to the main processes, through the intermediate step of recording the resource uses by activity account. Now, the relationship can be worked out backwards from process activity level to resource usage:

$$g = Th'$$

$$Cg' + g = 0$$

That is, $Th' + Cg' = 0$

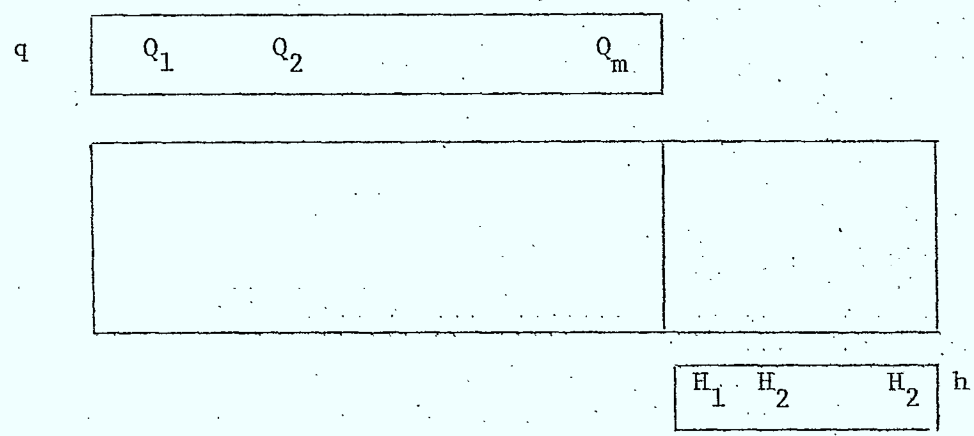
or $$q' = -C^{-1}Th'$$

This is the production function for indirect resources, yielding the input use as a function of the output level (= the data flows).

The crucial assumptions in this model are that stability in the coefficients c_{kj} and $t_{k\ell}$, and proportionality[1] between resources and activity be observed. Hence, the limitations of this approach; it also relies heavily on the practical assumption that it is possible to find some relevant allocation information. Yet, it has the advantage of being applicable to any kind of information science with minor modifications: it stresses the basic production processes that take place in an information service system.

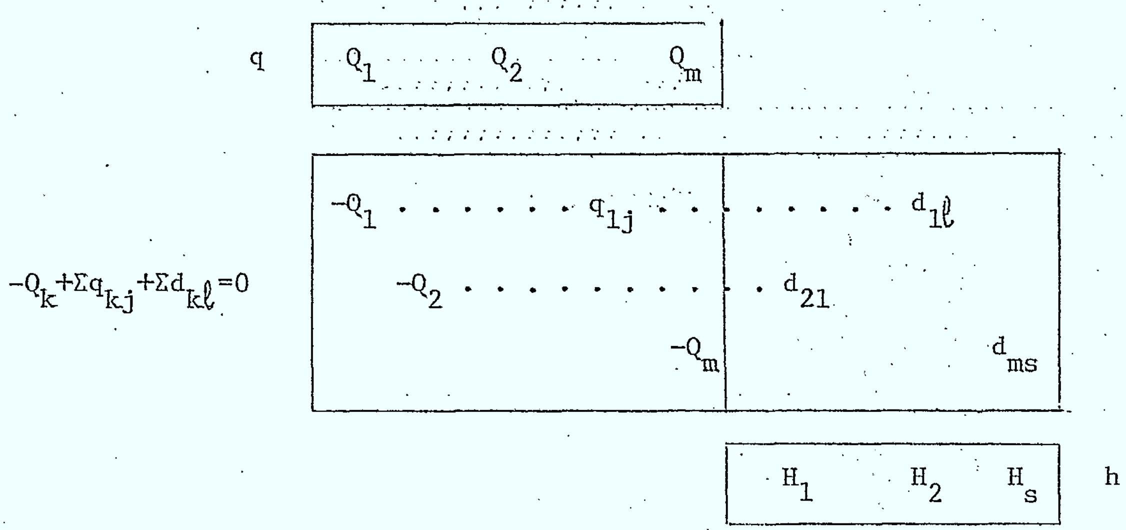
A summary of the approach is given below.

Step 1: available information: q and h.

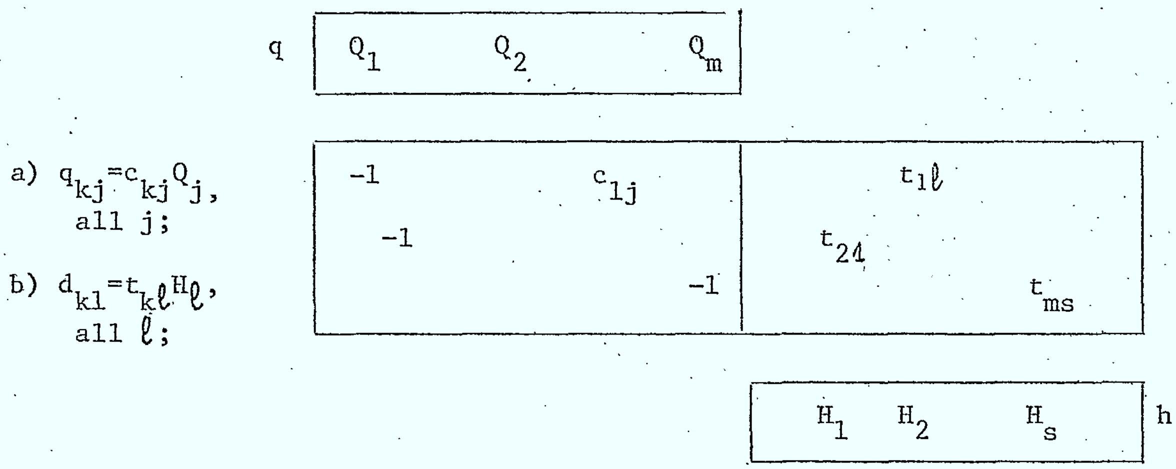


[1]Strict proportionality; fixed factors are not recognized in this analysis.

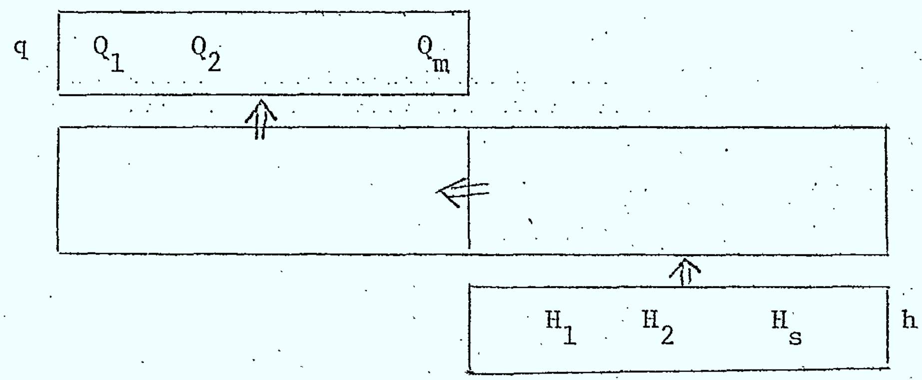
Step 2: allocation of resources to activities and processes



Step 3: normalization of matrix elements



Step 4: backward induction



BIBLIOGRAPHY

A.C.M. (Sponsors), National Colloquium on Information Retrieval, Philadelphia, 1965 –

Becker, J. and Hayes, R.M. Information Storage and Retrieval: Tools, Elements, Theories. New York: Wiley, 1963.

Brown, G.W. and Miller, J.G. (eds.). EDUNET, A Study on Information Networks. New York: Wiley, 1967.

CODASYL Committee: various publications.

Cohen, Burton J. Cost-effective Information Systems. New York: American Management Association, 1971

Cuadra, Carlos A. (ed.). Annual Review of Information Science and Technology. New York: International Publishers, 1966 – .

Elias, Arthur W. (ed.). Parameters of Information Science. Washington: American Documentation Institute, 1964.

Fisher, Gene H. Cost Considerations in Systems Analysis. New York: American Elsevier, 1971.

Flores, Ivan. Data Structure and Management. Prentice Hall, 1970.

Havelock, Ronald G. (ed.). Planning for Innovation Through Dissemination of Knowledge. Centre for Research on Scientific Knowledge. Ann Arbor: University of Michigan, 1971.

Henley, John P. Computer-Based Libraries and Information Systems. New York: American Elsevier, 1970.

Lamberton, Donald M. (ed.). Economics of Information and Knowledge. Penguin Books, 1971.

Little, Arthur D. Centralization and Documentation. Final Report to the National Science Foundation, 1963.

Martin, James T. Teleprocessing Network Organization. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

__________ Design of Real-Time Computer Systems. Prentice-Hall, 1967.

Meise, Norman. Conceptual Design of an Automated National Library System. Metchen, N.J.: Scarecrow Press, 1969.

Morse, Philip M. Library Effectiveness: A Systems Approach. Cambridge, Mass.: MIT Press, 1968.

Newman, Simon M. Information Systems Compatibility. Washington, Spartan Books, 1965.

Overhage, Carl F.J. and Harman, R.J. Planning Conference on Information Transfer Experiments. Cambridge, Mass.: MIT Press, 1965.

Sharpe, William. The Economics of Computers. New York: Columbia University Press, 1969.

Stimler, S. Real-Time Data Processing Systems. New York: McGraw-Hill, 1969.

Vickery, B.C. On Retrieval Systems Theory. London: Butterworths, 1965.

Whittenburg, John A. and Schumacher, Anne W. Guidelines for Planning a Task-Oriented Information System. U.S. Department of Commerce, 1969.

Williams, William F. Principles of Automated Information Retrieval. Elmhurst, Ill.: Business Press, 1965.