A THEORY OF
DATA SYSTEMS FOR ECONOMIC DECISIONS

by
PAUL B. HENDERSON, JR.

B.S.I.E.  Washington University
B.S.M.E.  Washington University
  (1948)
M.S.B.A.  Washington University
  (1950)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF
PHILOSOPHY
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1960

Signature of Author
Department of Economics and Social Science, May 9, 1960

Certified by __________________________________________ Thesis Supervisor

Accepted by __________________________________________ Chairman, Departmental Committee on Graduate Studies
ABSTRACT

Title: "A Theory of Data Systems for Economic Decisions"

Author: Paul D. Henderson, Jr.

Submitted to the Department of Economics and Social Science on May 9, 1960 in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Industrial Economics.

This thesis is a presentation of a general theory of data systems; of the empirical content of the theory in the context of a data system for economic decisions of the firm; and of the formal properties of the resulting mathematical model.

The foundation for this theory is a concept of data elements, each consisting of an identifying label and a piece of substantive content. The content of such an element can range from a single item of quantitative data to the full array of data in a complex report. Although formulation of a data system model by this approach carries the disadvantage of a significant increase in the number of variables that must be considered, there is a compensating advantage in that all of the data elements can be consistently identified and precisely manipulated. The associated system model is shown to be specifically applicable to data systems in support of the economic decisions of the firm; to be susceptible to formulation as an integer linear programming problem; and to possess a theoretical solution for the optimum system design.

The setting of the problem, the current status of research and practice in related fields and disciplines, and an outline of the research program and general structure of this thesis are presented in Chapter I.

Chapter II presents the theory. It includes an explanation of the relation to information theory, the nature of the data elements, the requirements and sources of data, the manner of aggregating the elements into a complete model, and the means of expressing the model as a connected graph and as an integrated set of matrices.

The theory is expressed in the context of a data system for the economic decisions of an "industrial firm" in Chapter III. The resulting model permits all management data to be consistently and precisely located in a five-dimensional matrix and a completely
integrated data system to be generated from such a matrix in successive stages.

In Chapter IV, it is demonstrated that the management data model can be expressed as an integer linear programming problem incorporating fixed costs, time and capacity constraints. The existence of an optimum solution of such a formulation is cited as theoretical justification for the elemental data approach.

Finally, the model is evaluated with particular reference to the implications on organization and accounting and to the relationship of the data system model to certain machine scheduling and sequencing problems and computer applications.
Since men first fix goals, they have been beset by the members of the associations and intentions. Thus questions: What information and, how to convey it to.

The patterning of the signal, was initiated by the content and the communications were either small gatherings, or, if large and simple structure of ends a building. Given a common: well be conveyed orally, by drums.

Despite the relative very early development of and institutions attests to practice of cooperative ends. West and the East there is and more abstract languages pictures and ideographs to progress.
With the advent of intricate layering of production and distribution organizations, of large-scale industrial enterprises, and of complex products and processes, came accelerated development of a series of institutional practices intended to regularize certain aspects of the association in order to enhance technical efficiency and, not incidentally, to reduce the burdens of those who assumed responsibility for the administration of the association. Organization planning and accounting were the most significant of those structural developments, encompassing the most critical aspects of the form and the results of the association. However imperfect and static they may have been, they were indispensable to the development of our present industrial institutions.

Recent advances, notably in the development of mechanical and electronic equipment to assist in data processing and transmission and of mathematical tools to assist in analysis, have been concentrated on the other side of the basic problem. Linear programming, marginal analysis, symbolic logic, matrix algebra and other techniques can be applied to a wide range of problems. Collection, processing and transmission equipment will accommodate all types of data -- from theoretical abstractions to nonsense data that masquerades as "pertinent fact" about reality.

It is in this setting that much of the recent enthusiasm for "sophisticated" computers finds its most critical test. There is undoubtedly merit in finding a faster way to do that which is already being done -- but in the quest for perfection there is also the possibility that the preferable solution lies in developing a completely new pattern. It is not altogether trite to hold that the
search for a faster or better way to make a buggy-whip may be misspent effort.

In contrast to early times, there is now greater emphasis and correspondingly greater progress in the science and art of conveying signals. The development of languages continues, with special emphasis on languages that machines can "understand" and there is a growing realization that between the chosen language and the transmission medium is a broad area of choice of the information to be recorded and communicated.

This area between is both a land for development of new concepts and for extension of the old. It is the province of theory, model building, and application in which the structures need to be viewed in new ways in relation to the tools with the hope that each can be better understood and better used.

Direct attack upon "correct" message content is complicated because the message is not justifiable in and of itself while a language or transmission medium must, first of all, meet the rather obvious test of self-consistency.

The traditional approach to improving message content is based on the review of several common data aggregates, notably, the form and the report. It typically involves a detailed study of the necessity for each form and report and of the number of copies and frequency of submission thereof. In all but the simplest situations and with all but the most perceptive analysts this approach is self-inhibiting. It provides a direct means to criticism of existing practices, thereby creating a negative attitude without compensating strength as a basis for a new design. Each analyst is isolated by
lack of a common reference framework and after correcting the more or less obvious discrepancies faces an ever growing network of individuals who's raison d'être is challenged by demands to justify their data requests. Details of format, number of copies, and frequency may be negotiable points but there are few who will view objectively a suggestion that some of their functions are unnecessary.

The ideal alternative to the traditional approach would be the development of the data system as one aspect of an overall optimum design. Such an objective is not yet attainable; the present state of knowledge is inadequate for practical application although there has been some progress toward understanding of the problem in general systems theory and cybernetics.

A reasonable alternative would be the design of the data system given a fixed, or nearly fixed, environment. The essential criteria for such a design would be a unified internal structure and a consistent relationship among the other components of the firm, specifically organization and accounting.

This thesis is concerned with such an alternative approach. It involves the presentation of a theory of data systems rather than a "cookbook" of principles, practices and techniques. It provides a basis for systematic identification and manipulation of data; a demonstration that it has meaningful empirical content in the context of the data system problems of the industrial firm; a model reflecting the theory which is both a tool of analysis and a formulation amenable to treatment by other mathematical and computational tools; and an argument that the approach is justified because it can lead to a theoretical optimum system design.

viii
The manner of presentation of the theory and discussion of its formal and empirical implications presumes a familiarity with linear programming and related tools; an interest in the theoretical and practical significance of this subject; and its relation to other problems in economics, operations and management research, and quantitative analysis.

For my introduction to the problems and promises of data system development and analysis I am indebted to many personnel of the United States Navy, Bureau of Ordnance. To Rear Admiral John Quinn, USN, Rear Admiral E. B. Hooper, USN, and to Captain E. A. Ruckner, USN: for their patience and their support; to Captain B. I. Lubelsky, USN: for a liberal education in the complexity and yet essential simplicity of the data needed for management decisions; and to Captain T. A. Brown, (SC) USN: a special debt of gratitude for his guidance and for his assistance in resolving the conflicting interests of those whom the system was to serve.

Of those who assisted during the period when I attempted to draw together the vague empirical notions into a coherent proposition, special thanks are due to Professor Robert M. Solow, who has supervised my research; to Professor R. B. Maffei for his encouragement and his insistence that my results be understandable; and to Professor M. J. Gordon for his assistance in interpreting the essential relationships between the accounting system and the data system.

Finally to my wife, I express my gratitude for her cheerful support and skillful typing.

Cambridge 39, Massachusetts
May 9, 1960

ix
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>THE DATA SYSTEM PROBLEM</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2. The State-of-the-Art of Data</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>System Design and Analysis</td>
<td>19</td>
</tr>
</tbody>
</table>

| II. | A THEORY OF DATA SYSTEMS | 24 |
|     | 1. From the Real to the Abstract | 24 |
|     | 2. Information or Data | 29 |
|     | 3. Data Elements and Networks | 35 |
|     | 4. Requirements and Sources | 40 |
|     | 5. Derivation and Aggregation | 44 |
|     | 6. The Data Matrix | 46 |
|     | 7. A Theory of Data Systems | 47 |

| III. | A MODEL OF DATA SYSTEMS FOR ECONOMIC DECISIONS | 52 |
|      | 1. From the Abstract toward the Real: | 52 |
|      | The Scope of Management Data | 55 |
|      | 2. Matrix Representation of Management Data | 59 |
|      | 3. The Program Dimension | 62 |
|      | 4. The Natural Structure | 63 |
|      | 5. The Management Structure | 65 |
|      | 6. The Content Dimension | 69 |
|      | 7. The Time Dimension | 70 |
|      | 8. The Management Data Matrix | 78 |
|      | 9. The Expression of Data Needs | 80 |
|      | 10. The Determination of Data Inputs | 81 |
|      | 11. The Selection of the Solution | 81 |
IV FORMULATION OF THE MODEL FOR THE SOLUTION FOR THE THEORETICAL OPTIMUM SYSTEM

1. The Components of a Data System 83
2. Mathematical Formulation of the Model 88
3. The Objective Function 93
4. The Output Constraints 94
5. The Time Constraints 100
6. The Capacity Constraints 104
7. The Nature of the Solution 107

V IMPLICATIONS OF THE THEORY AND THE MODEL 109

1. Summary 109
2. Conclusions 117

BIBLIOGRAPHY 119

BIOGRAPHICAL NOTE 122
LIST OF ILLUSTRATIONS

Figure
1. Graphical Representation of a Data Element ... 39
2. Data Element Relationships ............... 39
3. The Data System Model ................. 49
4. The Data System Block Diagram ....... 50
5. The Data System Simplified Network .... 50
6. The Data System Network ............... 50
7. Matrix Representation of a Data System ... 51
8. Matrix Representation of Data System Design
   Alternatives .......................... 51
9. Matrix Representation of the Classification
   Dimensions ........................... 75
10. The Data Matrix for a "Production Planning"
    Program .............................. 76
11. The Components of the Data System Model ... 86
12. The Sequencing Network ............... 116
CHAPTER I

THE DATA SYSTEM PROBLEM

1. Introduction

The decision maker occupies a role in many social action theories that reflects his importance in real situations. The general theoretical construct of the decision process which, under certain sets of assumptions, leads to a unique decision criteria intentionally captures the essence and not the whole reality of the problem.

The problem of present concern is that of the provision of the information which is needed. The background of the problem includes the theory of the decision making process, specifically in the realm of the economic decisions of the firm; the concepts and theories of information, organization, and accounting; and the tools and techniques of communication, computation, and analysis.

The impact of these subjects on the data system which does, or at least should, link them together is not necessarily unidirectional. Each may affect the others. If individually "optimized" in isolation there is a possibility that attainment of higher goals may be partially prevented. Until it becomes practicable to optimize the overall system, the continuation of the next best practice -- individual sub-optimization -- is in order.
In order to permit an orderly development of a theory of data systems for economic decision, the influences of the essential factors exogenous to the data system are stipulated by the assumptions that:

a). There is given an "association" which pursues some "function" to the end of attaining chosen "objectives".

b). There is a "decision maker", or set of decision makers, in a "relationship" which establishes a structural and functional pattern of the decision authorities and all "function performing" components.

c). The "activities" of the association are observable and possess an "accountable relation" to each other, to the association and to the "information" which the decision maker requires.

d). The decision maker uses the information in terms of the chosen objectives to decide upon "control variables" which regulate the association.

e). The purpose of the "data system" is to fulfill the information requirements in a manner consistent with the chosen objectives.

Some of the most significant consequences of relaxation of these assumptions will be considered following the development of the theoretical model of a data system, the demonstration of its applicability to the setting of the "industrial firm", and the analysis of the implications of the mathematical formulation of the model.
The view of the management process given in the economic theory of the firm is orderly and explicit. The firm is conceived to be a well-defined entity, under the direction of an autonomous decision maker who controls certain decision variables in accordance with a decision criteria which has been selected on the basis that it will yield "best" results in terms of a specified objective. There is a growing body of knowledge on the extension of the several components of the basic theory toward a general formulation more in keeping with the variety of circumstances that actually appear.

These developments are of more immediate concern for their effect on the data system problem than are the now well developed information requirements of the basic economic theory.

For example, the differences in firms, beyond those treated in the theories of competition, monopoly and oligopoly, have found a more meaningful expression in the context of the theory of games.¹ The consideration of the effect of multiple decision makers, beginning with the debate on the possibility of decentralized decision making, as in a socialist economy, has been extended to a study of decentralized decision making within a firm ²

---

¹ The application of game theory to the firm is discussed in general terms by R. Duncan Luce and Howard Raiffa in Games and Decisions (New York: Wiley, 1950) and in some detail by Martin Shubik in Strategy and Market Structure (New York: Wiley, 1959).

² Thomas Marschak, "Centralization and Decentralization in Economic Organizations" Technical Report No. 42, Department of Economics, Stanford University, 1957 includes a review of classical decentralized decision theory and a bibliography of related studies on decision-making in the firm.
To the extent that those results provide a more realistic view of the firm, they offer a better understanding of the proper content of the data system output.

To the practicing decision-maker the specification of data needs is itself a difficult decision. And to those who try to fulfill the needs there are further complications.

Hidden in the apparent definitiveness of the decision criteria are a host of difficulties. The determination of some data would require a vast experiment that can well be conceived but not performed. Apart from any objection to the static nature of the assumptions involved, there is no assurance that it is possible to identify the desired information in the data on the host of relationships that are observable.

Further, there is in the economic concept of the firm an assumption that the decisions necessary to the attainment of an efficient point of the production function will be properly made. This problem alone has fostered extensive research in industrial dynamics, game theory, and mathematical programming.

Finally, the assumptions as to the objectives of the firm need to be tempered by the inclusion of conditions that will assure that secondary goals are not neglected. Since it is now an accepted proposition that optimization is limited to a single objective, there is perhaps less worry about the means to assurance that some minimum level is attained in the other goals.

The firm faces many obviously interrelated problems which have been treated separately, often by different disciplines. It is certain that optimization of each of the constituent aspects of the
firm can yield a less-than-optimum solution in terms of the overall objective. It is conceivable that a more comprehensive approach would open a new domain of activities by its effect on individual incentives and attitudes.\(^3\)

Although these theoretical developments are not incompatible, they are not yet integrated into an overall theory or model. The rapid growth in complexity that attends a relatively minor extension of the theory suggests that economic theorists now are facing the problem that has always plagued the decision maker -- how to accommodate a great number of very different elements in an analysis without becoming overwhelmed by the sheer mass of data.

Just as organization planning and accounting are significant because they offer a means for creating and maintaining a uniform pattern for part of the whole problem, a similar discipline is needed in data systems. In each instance there is need for a cohesive theory that can be adapted to the particular variations that are found.

Although not complete, there is now extant a meaningful body of theory of organization and accounting. In many respects the present theories are the outgrowth of past practice. Certain constituent

\(^3\)The Scanlon Plan for employer-employee cooperation is illustrative of attempts to reap some of the advantages of this situation. See, F. G. Lesieur, Editor, *The Scanlon Plan: A Frontier in Labor-Management Cooperation* (Cambridge: Technology Press, MIT, 1953).

The difficulties may stem from the same phenomena reported by C. N. Parkinson in Parkinson's Law, and other Studies in Administration (Boston: Houghton Mifflin, 1957). The conditions favorable to increased efficiency cannot readily be established without a basic change in techniques or attitudes.

\(^4\)See, for example, the game theoretic formulations in Shubik, *Op Cit.*
aspects of both organization and accounting found separate expression
before the field itself was well defined. There should be little
surprise in the recognition that mathematics, the language of science,
was the forerunner in the developments that initiated the transition
from an art toward a science.

In the field of data system design there is a similar pattern
of successful practice and well developed theory on some constituent
elements -- of which the mathematical theory of information is
particularly noteworthy.

There is however, an apparent need for better guidance of the
efforts to apply the more powerful analytical and computational tools
to the increasingly complex data requirements associated with the
economic decisions of the firm. The successful application of large-
scale digital computers in many firms and to a variety of military
problems indicates that particular solutions can be found.5 The
recurring comment that a complete restructuring of the data system
is frequently a prerequisite to the use of the computer and that such
a change often gives an increment of benefit greater than that offered
by the computer suggests that there is a possibility of substantial
benefit, financially and operationally, to those companies who have
not mounted a massive assault on the data system problem under the
impetus of an appreciable financial commitment for a computer and

5An excellent summary is given by Roger Nett and Stanley A.
Hetzler, in An Introduction to Electronic Data Processing (Glencoe:
that the patterns which are available to guide the tailoring of the system leave much to be desired.  

If system redesign must await commitment to a computer program it is inevitable that some of the commitments will prove to be inappropriate as the redesign effort demonstrates that a computer is not necessary or even desirable.

The pervasive character of data in all aspects of the firm and the wide variety in the composition and proportions of the data mix intensify the importance of the fundamental choice of a concept to guide the design and analysis efforts.

It has been suggested that there is a meaningful analogy to the data systems problem in the recent and continuing concern with weapon systems development.  Unfortunately, there is a further similarity in the sense that the weapon system approach is well discussed, but poorly practiced. Solutions by crash-program management and brute-force engineering find their counterpart in the forced redesign of data systems under the shadow of a computer commitment.

This thesis is an argument for the systems approach on a theoretical rather than trial and error basis. It consists of an abstract representation of a data system as an ordered set of data

---


elements; a multidimensional matrix in which to manipulate the elements; an assertion that management data systems can be formulated in a reasonably simple integrated pattern, and a demonstration that such an approach is justified in that it permits a formulation that is known to possess a determinate optimum solution.

2. *The State-of-the-Art of Data System Design and Analysis*

There is extant a considerable body of literature on the aggregative approach to data system development and on the many contributions of related disciplines. The design or systems approach is of more recent origin, at least as a separately recognizable effort, initiated by the force of necessity for better data services and encouraged by the advances in equipment and analytical techniques.

The review of forms and reports is based on techniques similar to those used in the analysis of work motions and performance times for shop jobs. Each data aggregate is reviewed to determine the necessity for its content, distribution, and frequency. There is an implicit assumption that the data requirements will be adequately represented by those who have need for the information thereby permitting the analyst to concentrate on reducing and simplifying the number, frequency and content of the communications.

There are several readily apparent deficiencies in this approach. It is centered on the means rather than the objectives of the data system; it relies on the hope that improvement of the parts will improve the whole system; and it is dependent upon the abilities of the individual analysts. Of these, the most critical deficiency is the isolation of the analysts.
It can well be argued that any technique adequate to a little job can be scaled up to a big job. If so, the analysis could be conducted on a level where overall optimization would replace piece-wise suboptimization. To be successful in developing an overall system it is essential that all the review evidence be brought together for analysis. The traditional approach does not provide any adequate means to deal with the same type of data in different forms and reports; in most cases, identical data is not recognized except by fortuitous circumstance. Thus, the analysts find themselves with the same problem they are trying to solve for someone else -- how to comprehend the scope and variety of data that is involved in the decision?

Despite these limitations, the review technique has been a useful tool. It has been highly developed in large organizations, notably in the many agencies and bureaus of the federal government. It was, until recently, the way to try to improve data systems.

The deficiencies of the review approach were recognized long ago and some positive steps have been taken to minimize their impact. With each new advance in equipment there have been attempts to locate the tasks which could be restructured to take advantage of the new capabilities. From continued experience there has been distilled a set of principles intended to provide general guidance in the treatment of the areas which by their scope and complexity resisted the best direct efforts of the reviewers.

Two principles merit specific mention. The "exception principle" because it has long been followed and is frequently misleading,
and the "integrated-data principle" because it is a recent development intended to improve data services by the application of new equipment.

The exception principle is a specific application of the more general principle of "management by exception". As defined by Taylor, it provided that:

... the manager should receive only the condensed, summarized, and invariably comparative reports, covering, however, all of the elements entering into the management, and even those summaries should be carefully gone over by an assistant before they reach the manager, and have all of the exceptions to the past averages or to the standards pointed out, both the especially good and especially bad exceptions, thus giving him in a few minutes a full view of progress that is being made.  

It involves the implicit assumptions that problem-solving is the central function of management and that each organizational level should regularly monitor the data submitted to superior levels. It is an example of suboptimization by decentralized decision-makers without a common objective function. As with other "principles" it permits such wide discretion that any data system could be "justified" by an appropriate revision of the statements of managerial problems.

If carried too far it leads to excessive layering of administrative levels and a concommitant loss of background and analytical data. Although Taylor was careful to call for "all the elements entering into the management" it is difficult to determine what elements are appropriate and easy to misinterpret the concept of "management by exception" as a concept of "data service by exception".

---

A somewhat exaggerated analogy would be a police chief who cannot
determine from police reports whether he is trying to solve the crime
problem of a very evil small town or a most peaceful metropolis.

Parkinson's studies of administration suggest that the problems
are very real whatever their origin and Nett and Hetzler seem to have
found many "police chiefs" in their comment that:

...there is an even chance that people in the bureaus of
large organizations really do not understand at funda-
mental levels all that they are doing, although they
might think that some other bureau does this for them. 9

The "integrated-data principle" seems to have originated in
the recognition that machines, unlike people, would not tolerate an
illstructured operation. From the experience in applying the new
data processing and transmission equipment came a set of guiding
principles with obvious potentiality for improving data services:

1. Original data are recorded at their point
   of origin in a mechanical form ...

2. From then on ... data are processed exclu-
   sively in a mechanical manner ...

3. All processing of data is integrated so
   that original data in mechanical form serve all sub-
   sequent applications. 10

Using these principles, and variations of them, several major
industrial firms have developed substantially better data systems,
usually if not always involving large scale computers. None have

10 American Management Association. Establishing an Integrated
Data Processing System. Special Report No. 11.
professed to have solved the problem of design and maintenance of a comprehensive system.\textsuperscript{11}

Researchers in other disciplines have made substantial contributions to the understanding and partial solution of some aspects of the data system problem. Their efforts can conveniently be classified as:

a) The specification of data needs,

b) The collection and processing of data, and

c) The tools and methodology.

Since there are few fields where information is not of primary concern or where further research does not contribute to greater understanding of the data problem, there is an abundance of theoretical and empirical research on the specification of data needs. Those of immediate interest are the economic theory of the firm, game, decision and organization theory, and accounting.

The data requirements which arise in the theory of the firm are quite explicit but very difficult to fulfill. Much of the desired information on factors, processes, products and markets is not directly measurable and the data from which it might be estimated are hidden among many other relationships and disguised by the effects of stochastic elements. The research of present interest is that

\textsuperscript{11} In addition to the general summary of computer applications in Nett and Hetzler. Op Cit, Chapters V and VI, there are numerous publications by the American Management Association on specific installations. Special Reports No. 9, "Pioneering in Electronic Data Processing" and No. 11, "Establishing an Integrated Data Processing System" are representative.
devoted to the extension of the bounds of the theory to encompass a more realistic set of assumptions as to the firms' environment and objectives.

Since the appearance of the original definitive work on the applicability of game theory to economic analysis in 1944, the expectation has been that it would provide a useful tool for analysis of the complex world of economics that lies between monopoly and competition. A part of this promise is fulfilled in Shubik's work on the application of game theory to the market strategy of the firm, but it is apparent that the information requirements are to be increased rather than diminished by this approach. This is, perhaps, not surprising when the information sets for a normalized game are considered in the context of the firm.

Uncertainty about the actions of others and the risks which arise from stochastic processes introduce elements that are not amenable to solution by the conventional theory of the firm. Probability and game theory can extend the range of application of the fundamental concept of marginal analysis but each introduces additional information needs. Both require a broadened time span of data and more careful attention to the evidence in order to discern the state of nature or the opponents' strategy.

---


13 Shubik, *Op Cit.*

14 Luce and Raiffa, *Op Cit.*
When account is taken of the ill-structured nature of the actual decision environment the precise character of the proper choice is no longer obvious. Decision-making becomes a matter of separate theoretical interest. Just as better data can be expected to contribute to better decisions, a greater knowledge of the use of data can lead to better stipulation of data needs.

In the field of organization theory there is a continued concentration on the relations between the patterns of duties and responsibilities and the data provided. The central importance of information is readily apparent in Forrester's work on industrial dynamics, in Haberstroh's work on control and in the many studies of the process of organization in a purposive association.

In accounting there is increasing dissatisfaction with many time honored concepts. The fundamental purposes of accounting are being reconsidered and in some instances, revised to reflect the result of research in many of the same fields that have influenced the development of data systems. Decentralization is accepted as a

---

15 An excellent reference bibliography on decision-theory for the period 1945 to Sept. 1957 has been provided by Paul Wasserman and Fred S. Silander, Decision-Making -- An Annotated Bibliography, (Ithaca, New York: Cornell University, Graduate School of Business and Public Administration. 1958). More recent work is reported in the professional journals of the several disciplines and in those devoted more generally to operations research and management science.


a necessary phenomena in control and new concepts provided as the
basis for accounting data. Statistical inference, and experimental
designs are replacing the notion that exact measurement was not only
possible, but almost a professional point of honor.\textsuperscript{19}

Accounting, by its relation to the legal and financial struc-
ture, is one of the strongest influences on the conception of data
needs as well as a prime determinant of the acceptability of new data
processing techniques. It is certain that the changes in accounting
concepts and procedures will induce corresponding changes in the de-
mands for data services.\textsuperscript{20}

There is some suggestion of possible contention between the
fields of accounting and data processing. Churchman and Ackoff,
writing in The Journal of Accountancy, suggest that there is need for
a new "operational" accounting system to supplement the "traditional"
one, in order to handle this sort of problem: "If we were studying an
important executive problem in the year 1965, what data would we wish
we had collected systematically in the period 1955-65 to assist us in
making our decision?"\textsuperscript{21}

\textsuperscript{19}See, for example, Experimental Designs in Industry. Edited by

\textsuperscript{20}Several trends are now discernable. There is a growing aware-
ness that sampling techniques are applicable to certain accounting
functions and that the accounting system must be designed to serve all
levels of decision-making. In the development of the latter point
there is some promise that the use of transfer costs between organi-
zational groups will not only improve managerial efforts but may also
encourage changes in organization concepts that will add to the advan-
tages.

\textsuperscript{21}C. West Churchman and Russell L. Ackoff, "Operational Account-
ing and Operations Research", The Journal of Accountancy, Feb. 1955,
p. 38.
The understanding of the collection and processing of data, apart from the concern with procedure and technique, has been advanced by the work on information and cataloging theory.

Information theory has been of greatest interest to those concerned with the transmission of data. It is equally, but not so obviously, applicable to those who wish to send or receive useful messages.

The distinction between information and data and between plain languages and reference codes is of direct concern and is therefore discussed in detail in Chapter II.

Cataloging theory is concerned with the problem of retrieving data that has been placed in storage. The rather obvious part of the solution is represented by the card catalog of a library. The more meaningful part is concerned with the desire to locate data which relates to the subject of interest but is not so classified in the usual coverage of title, author, and subject headings. The search of technical articles and books is a specific example of great current interest.

The present approach to this problem is to develop general classification and coding schemes that identify the most significant works or passages of the document and use these extracts as a basis for scanning the whole library.

The data system problem is similar in some respects, but it is to be expected that the contents of data file must be more carefully organized and more precisely identified than the documents in a general library. For this purpose, cataloging theory and the related "library search" problem, offer two points of interest. First, the
recovery of small amounts of information from a larger library is essentially dependent upon the classification scheme in relation to both the material being stored and the type of requests to be honored. Second, the analysis and classification of the material in terms of the selected scheme is best performed by the authors when the material is generated.

Perhaps the most active area of advance relative to data systems is that of the development of data collection, storage and processing equipment, or analytical and computational procedures and of accounting theory and techniques.

The advance in equipment has been induced by technological progress and technical demands in other fields. For many years the fields of scientific computation have supported the development of data processing equipment at a pace faster than could be absorbed in managerial and business applications.

In similar fashion, it appears that the use of proven techniques of analysis and computation are restricted in their applications to economic decisions by the inadequacies of existing data systems and the near uselessness of much of the mass of data already collected.

There have been a few attempts, all of recent origin, to attack directly the problem of original design of data systems, to accomodate data in a detailed analysis where it could be precisely identified, and to develop a guiding theory for integrated-data systems.

The most fundamental approach is by way of the very concept of control. Cybernetics has become familiar as the title for an equally familiar problem -- the study of self-regulating systems.
It may, in time, provide a far more sophisticated approach than is now proposed.\(^{22}\)

Other approaches have also given some useful insight into the problems usually considered for solution by application of advanced computers. Some of the results appear to be useful in the field of data systems.\(^{23}\)

Despite the variety of research in closely related fields there is little in the literature to indicate that a direct assault on the data system problem has been mounted except in conjunction with the application of large-scale computers.

This thesis is concerned with the same objectives but a somewhat different approach than that proposed by Jim Rosenzweig.\(^{24}\) There are several points of similarity with the ideas presented by Irving Lieberman in "A Mathematical Model for Integrated Business Systems".\(^{25}\) The matrix representation which Lieberman used is especially pertinent and, therefore, discussed in detail in Chapter IV.

---


\(^{24}\) Rosenzweig, Op Cit.

3. Plan of the Study

The concept of data systems presented herein is the result of two separate study phases. The first was in connection with the actual design and installation of managerial and analytical data systems, the second was in connection with this thesis. It involved the development of a data system model which could meet three fundamental criteria:

a) **Consistency.** The constituent portions of the model must be such as to permit their consistent identification and precise manipulation.

b) **Validity of Empirical Content.** The model must be demonstrably capable of meaningful interpretation in the context of the economic decisions of an "industrial firm".

c) **Significance of Formal Content.** The approach must permit a formulation in which it can be shown to possess a solution consistent with the optimizing objective.

The final result, although expressed in theoretical and mathematical terms, is, of course, not unrelated to the partial successes and partial failures, the bits of information and the voluminous files which originally highlighted the need for a theoretical guide in data system design and analysis.

The contribution of the earlier phase to the purposes of the later research and to the final results consisted of (1) a recognition that something better than trial and error and repeated review of systems that, like Topsy, just grew, was essential to the solution
of the data system problem and to the exploitation of the potentialities of large-scale computers and (2) a conviction that the answer was not to be found in the aggregative approach centered on "forms and reports".

The research conducted in connection with this thesis logically involved three sub-problems related to each of the three criteria.

Consistency was the first and most fundamental problem. A way to accommodate the great variety of data in an all-encompassing framework was an essential prerequisite to both of the other criteria. Since the aggregative approach is based on inherently imprecise elements the search was concentrated on various concepts of irreducible bits of data -- a micro-data rather than macro-data approach.

In view of the great increase in the number of elements expected to result from the adoption of a micro-data basis the original research efforts were devoted to attempts to formulate a model which excluded all "artificial" additions of detail. Such a formulation seemed to offer the advantage that all data could be processed in plain language rather than in code with attendant advantages to the system users. Its fatal defect is that some portions of the system could not be isolated for study with any assurance that they can be returned to the correct place in the model without loss or alteration.

The alternative was to provide some never changing "handle" by which to grasp the data. The model chosen is based on the addition of an "identification" label to each and every entity of substantive data. This labeling is not an unmixed blessing, but the disadvantages of a roughly two-fold increase in the size of the problem are at least
partially offset since the resulting formulation can be segmented, reconstructed, and even generated by analysts working separately but within the overall framework. Further, the model can be readily and precisely expressed in either a mathematical matrix or connected graph with attendant benefits in manipulation and analysis.

The second problem was related to the question of empirical content. Whatever the properties of the data matrix it was deemed essential that the interpretation of the model in the context of the "industrial firm" give a meaningful result. The essential need for order and routine in management suggests the necessity for managers to relinquish part of the concept of a perfectly representative data system in return for a carefully structured system that provides quick orientation and meaningful information.

Using the data model based on elemental entities the first phase of the study of empirical content was readily completed. It provided a simple classification scheme -- each "label" identified the subject to which the substantive information applied. In this there was no significant reduction in the overall complexity of the pattern.

Further research indicated that a multi-dimensional matrix would provide a simple but comprehensive framework which was directly related to the managerial environment and readily adapted to a mathematical formulation.

Thus the model is here expressed in the form of a five dimensional data matrix.

The first dimension was chosen to reflect the functional distinctions that arise in any large association. The division is
usually, but not always, closely related to organizations. It differs in that the functional similarities, as in accounting or personnel administration, are rigorously maintained even though they cross organizational lines.

The second dimension recognizes the natural structure imposed by the physical characteristics of products or processes and the administrative structure dictated by superior authority. Although not always suitable as a basis for control because they lack regularity such natural structures are significant because they limit the scope of action of both inferior and superior levels.

The managerial perogative to create its own structure is recognized in the third dimension. Here there is freedom to define a consistent structure that can greatly reduce the number of variations that must be accommodated in the planning and review processes.

The fourth dimension recognized the substantive content of the data elements. Here, for example, there is recognition of the distinction between costs, quantities, and descriptions and between estimates, targets and actual values.

The final dimension is time. In some cases several different classifications of time -- the time when an estimate was made, the period to which it applied, and finally the time at which it was recorded -- each significant to some aspect of management.

Having developed a suitable model and defined its empirical content it remained to show that the Formal Content of the model was justified in the sense that it could be formulated as a problem to which there was a demonstrable optimum solution.
This criterion was fulfilled by the development of a linear programming formulation with a linear objective function. With the requirement that the solution must be in integer values, it was possible to incorporate fixed costs, time and capacity constraints. The result, although it fulfills the solution objective, is of limited practical application because of the great number of equations required. Given an efficient algorithm for the solution of large matrices composed of many zero and otherwise integer values this problem, and the analogous machine assignment and scheduling problem, will be subject to practical solution.

The presentation of the theory follows generally the line of its development. The theory is first stated in general terms in Chapter II, then re-expressed in the context of management data problems in Chapter III and finally shown as a linear programming formulation in Chapter IV. The thesis is concluded with an analysis of the implications among organization, accounting and the data system and an evaluation of the potential application of the theory to data system design and analysis and related problems.
CHAPTER II

A THEORY OF DATA SYSTEMS

1. From the Real to the Abstract

There is, in the consideration of data system design and analysis, a complex semantic problem. The complexity arises from the attempt to communicate an unfamiliar concept about communication itself with the expectation that the final result will justify the initial excursion into new meanings for old words.

In this endeavor, no new words have been coined, and several familiar, but not precisely definable, concepts are used. The intention behind the definitions is, perhaps, best conveyed by a brief preview of the argument to be presented and the background for the choice of approach.

There may be, in time, industrial firms without a regular staff of personnel; the more conventional firm is of present concern. The intrinsic need for responsive operating personnel and for an industrial environment in which people find satisfaction in their participation in economic processes are interrelated.

So long as people provide a purpositive influence, all systems—certainly including data systems—are best constrained to be suitable working associates. Evolution to a high order of sophistication will be feasible— if those parts of the data system which are closely associated with the operating personnel do not demand too much, too soon.

- 24 -
Thus, it is to be expected that the system will be redesigned periodically, as more advanced design concepts are developed and as the level of acceptance rises. A once and for all time "perfect" system is not envisioned -- it is doubtful that "perfection" is even a meaningful concept in such a design problem.

In the initial stages of development the scope of a particular advance will be limited. At later stages, there will be greater significance in integration across previously distinct lines and between increasingly diverse areas. The difficulties of system analysis will be correspondingly increased. With sufficient time and effort it is likely, if not certain, that continued advances will be made.

The discontent with this view of the future course of design efforts stems from the recognition that (1) each redesign of a system will require virtually complete review of the old design and (2) the experience gained in one area cannot readily be extended to the same type of problem in different settings because the similarities cannot yet be recognized without detailed study of both situations.

The central reason for this isolation is the lack of a generally applicable means of characterizing a data problem and its present solution. There are some common prototypes -- payroll systems are becoming commonplace -- but even in them there is substantial diversity.

The four-point solution here proposed is (1) that a common means be provided by which to identify all "pieces" of data from the smallest entity to the largest aggregate, (2) that each data problem be considered in terms of the purposes which the span of operational
events are intended to serve rather than the methods presently used i.e. to concentrate upon the ends and not upon the means, (3) that recurring problem types -- which will become recognizable by the common characterization -- be solved in a variety of ways appropriate to different levels of sophistication, and (4) that new problems first be characterized in order to determine if any prototype systems are applicable.

The preceding sequence can be illustrated by an inventory control problem.

Assume that the problem is the maintenance of dealer's inventories. Point (1) suggests that every item in the sales line be given a unique identity number, that every type of data about the items and all operating events be classified and precisely identified.

Four "data elements" which would result from such a characterization are of particular interest:

a) The identification of an item.
b) The stock level of an item at a particular time.
c) The desired stock level of an item.d) The sales of an item in a particular time period.

Point (2) suggests that the desire to maintain a suitable stock be taken as the system purpose in preference to considering ways and means to improve the present methods (which may be assumed to involve a monthly decision on the reorder quantity for each item based on stocks on hand and desired).

Point (3), in this simple example, would demand that the possible ways for maintaining the desired stock level be catalogued. At the very least it should be recognized that the reorder quantity
assuming no unfilled previous orders) can be determined in different ways — any one of which may provide the means for triggering the reorder. The note above suggested that the reorder was presently computed as [Intended Stock Level] - [Current Stock Level]. It is also equal to [Accumulated Sales] + [Change in Intended Stock Level] and to [Accumulated Sales] + [Sales Forecast] - [Average Sales].

The fourth point involves not only characterization of a problem to aid in its internal solution but also to provide for transfer from and to other systems that contain or need the same data.

There is a real possibility that the second or third computation may be the basis for an improved system. If sales records are accumulated daily; sales forecasts or stock levels are prepared as a linear projection of past sales or levels — subject to managerial review —; and a simple order rule is provided (e.g. ship when a full truck load of merchandise has been sold or when a controlled maximum number of days have passed since the last shipment), the reorder decision can be reduced to a monitored routine.

There is some reason to expect that the data for the cumulative sales approach may be more readily available and more adaptable to other uses. Production needs can be anticipated from the accumulation of sales and a consistent body of data is provided for sales analysis.

One point to be emphasized is that past records were seldom maintained in such a systematic way that any rigorous analysis can be conducted or any general conclusions drawn on whether or why
certain computation patterns are to be preferred. For example, there may be a valid theorem in the idea that more responsive systems must be based on more "original" data (original in the sense that the [Current Stock Level] is obtained as a consequence of prior [Accumulated Sales] in the example above).

If such analyses are to be conducted there must be provided some way to identify and manipulate the data which is or becomes available. It is not enough to call forth simple elements of data in an example -- the characterization must be suitable to accommodate any and all data that is, in fact, extant in the setting of industrial firms.

The present approach to this goal is conducted in three steps:

First, a completely general theory is offered to establish the necessary definitions and to provide accommodation for any and all data. This is the subject of the remainder of this chapter.

Second, because the fundamental interest in a reasonable approach to actual data problems cannot be satisfied by generalities, the theory is interpreted in the context of a data system to support the economic decisions of an industrial firm. This argument, presented in Chapter III, is not a "proof" of the sufficiency of the empirical content of the data system model. It does provide a clearly recognizable image of a real system and leads to several specific conclusions that are susceptible of test and verification; these practical implications are considered in Chapter V.
Third, and finally, because the merit of an approach is suspect unless it can be shown to lead to a desirable objective, the model is then formulated as a problem in integer linear programming in order to show that the data element theory encompasses a meaningful concept of an optimum solution and by that fact a measure of justification for verification of its empirical content.

2. **Information or Data**

There is a readily apparent difference in the meaning conveyed by the substantive content of communications. Although the differences are rarely precisely defined, it is not uncommon to distinguish between the mass and the essence. In military contracting, for example, the standard clauses are known as "boilerplate"; they are accorded little attention in relation to the clauses that convey the particular objectives, schedules and terms of the contractual agreement.

The full scope and variety of the obtainable data which exists in the environment of each decision maker is, to some degree, pertinent to the decisions to be taken. But the measure of contribution of the data actually recorded and provided to support a decision does not lie in the mere number or variety of "facts" obtained and considered. The essential part is its specific contribution to awareness of reality -- the degree to which it differs from other possible data and the degree to which it reflects the existence of one particular situation out of a number of possible situations. Variety and choice
are vital if a communication is to have meaning; there is no meaning conveyed by a null or unchanging signal.

The information content of data, which is one approach to meaning, has been precisely defined and extensively studied under the names "information theory" and "communication theory". The "amount of information" has been given a particular meaning and a standard unit of measurement by Shannon, drawing upon previous work on entrophy, quantum mechanics, and particle physics.\footnote{1} His definition of information, as a measure of the freedom of choice of content in a message, has contributed much to the theory and the practice of data transmission and is of substantial interest in the study of languages and communications.\footnote{2}

The amount of information ($H$) in a two-choice signal is expressed in "binary digits" and in multiple-choice signals in "equivalent binary digits" according to the equation $H = -K \sum P_n \log_2 P_n$ where $n$ is the number of possible choices,

\begin{align*}
P_n & \quad \text{is the probability of the } n\text{th choice}, \\
\sum P_n & = 1, \text{ and} \\
K & \quad \text{is a constant, here equal to one, which indicates the choice of the measure system.}\footnote{3}
\end{align*}

\footnote{1} Claude E. Shannon and Warren Weaver, \textit{The Mathematical Theory of Communication}. (Urbana: Univ. of Ill. Press, 1949).


\footnote{3} \textit{Ibid.} p. 4. The "bit" was defined by Shannon, on the basis of a suggestion by J. W. Tukey, as the measure of information content of a binary digit. The "bunit" is a more general unit of measure for "equivalent binary digits" and therefore applicable to any signal.
Thus the amount of information that can be contained in a signal is dependent upon the number of distinguishable forms of the signal and the probability distribution of the forms. If the probabilities of the several alternatives are not equal there is a decrease in the maximum attainable information content.

An ideal on-off or zero-one signal can convey at most one binit (bit) of information, a decimal signal about 3.32 binit, and an alpha-numeric signal about 5.26 binit.

One other concept of information is especially pertinent to data systems. It is not necessary that the message contain all of the meaning to be communicated. As commonly expressed, the message may be either in "plain language" or in code. Since, in the final analysis all communication is in a code, it may appear that the distinction is not significant. On the contrary, there are two conditions that make it critical. First, the common languages of the world today are not efficient information carriers, and second, the transmittal of coded references can greatly increase the effective content of a message.

---

4 The decimal digit is sometimes defined as the unit of measure for a decimal signal in lieu of the binit and called a "Hartley" in honor of R. V. L. Hartley, one of the pioneers of communication theory.

5 Shannon, Op Cit. P. 100.

6 Pierce, Op Cit. p. 262 estimates that each English language symbol conveys about 1 binit of information. Shannon, Op Cit. p. 104 expresses the same inefficiency as redundancy and estimates that the English language is about 50% redundant.
The substantial degree of redundancy in ordinary English prose may even be exceeded in "catalog" listings where common titles may require detailed classification to attain unique identification. Lengthy names are frequently required in parts lists:

- Gear Shift Control Shaft Housing
- Gear Shift Control Shaft Housing Cover
- Gear Shift Control Shaft Housing Cover Gasket

In such cases the use of parts numbers is quite understandable — apart from any benefits to be gained by the preference of data processing machines for numeric codes of consistent length. Assuming that management information was no more redundant than common English prose it would contain about one binit per character — about 21% of the theoretical maximum.\(^7\)

To identify one among 10,000 things would require, approximately:

- a 1.4 character plain-language word,
- a 4 character numeric word, and
- a 3 character alpha-numeric word.

Unique identification of each of the 180 million persons in the United States would require:

- a 5 character alpha-numeric word,
- a 9 character numeric word, and
- a 38 character plain-language or name work — if in fact there were no persons with identical names.

\(^7\) The English alphabet plus the space gives a choice of twenty-seven forms and a theoretical maximum of about 4.75 binit.
Where unique identity is not provided by a single attribute, it may be attained by introducing other characteristics. For example, the population identification might be made unique by the addition of a birthdate or birthplace data.

It should be noted that it is difficult to assure uniqueness in plain language identification. Variable attributes such as residence or occupation are not suitable unless carefully classified and dated and there is no a priori means of knowing when enough characteristics have been included.

The second point is closely related to the first. Reference information provides a means of reducing the amount of information which must be transmitted rather than a means of increasing the capacity to convey information. In the preceding discussion the objective was to get maximum information from each character. Briefly, the way to attain that objective is to use the maximum number of distinguishable forms and to arrange that each form has nearly equal probability of use in all contexts. In tabling or reference coding the objective is simply to indicate which one of a specified list of things is intended. Thus a simple sequence of three numerals can indicate the selection of one thing from 1000. Discription of any one of the 1000 things may require many thousands of bits of information.

An extreme example would be the comparison of transmitting the contents of a book:

a) word-for-word,
b) by a language that uses efficiently all the available choice in the characters.
c) by citing the reference.
The obvious potential course, be discounted by that under which understand the form of the book itself.

Transmission efficiency meaning conveyed. A language 17,576 word vocabulary. We mission, such a language is and habits of people. It is to "noise" since any distort a highly redundant language distorted and is accordingly meaning despite the ineffic.

A perfectly efficient nothing but information, is sent or near future reality. Information system is actually A distinction must be made as in a message (which by definition choices at the source), the decision-maker, and the amount of

The first measure of it may be difficult to find a distribution. The second measure recipient may already possess the probability distribution. Knowledge, a small amount of sufficient.
By the initial assumptions, the purpose of the data or information system is to satisfy data needs. Since the amount of information needed for that purpose is not uniquely determinable there can be no uniquely determinable pure information system.

For the purposes of this study, information is conceived in terms of its mathematical definition. Data is used as a generic term for all records and messages. Thus, information is provided through the provision of data.

There is no essential basis for concluding that either "data system" or "information system" is a logically correct title. As a matter of convention, therefore, the discussion will be based on data and data systems rather than in terms of the information which they contain. Where information is used in the succeeding material it will refer to the mathematic notion of "amount of information".

3. **Data Elements and Networks**

A data element is defined as a quantum of substantive knowledge about a single subject together with the precise identification of that subject. Thus, the age of a specified person in terms of a particular unit of measure (years, decades, months, days, etc.) is an element of data. There are, then, as many data elements about a person's age as there are measure systems. Implicit in the definition

---

8 Neither is there a proper conclusion that one, or the other, excludes redundant information. Data processing equipment typically uses redundant information for validity checking and so long as there is any element of imprecision in decision-making there cannot be a unique, perfectly non-redundant, system to provide the information.
is a time reference for the element and a different data element for each different point in each frame of reference.

A data element so defined, is not identical with or even closely related to a binit of information. If the element of concern is one of many, its identification may contain an appreciable amount of information and, of more significance, may be a very meaningful reference since the substantive content referenced by the element could be a description of considerable length.\footnote{The initial impression that a data element is always a very simple thing is soon to be dispelled. Although the initial concept, like Bohr's model of the atom, is simple, the final result is sometimes quite intricate.}

It is for these reasons that a clear distinction is made between the identification of the subject of a data element and the content of the message about the subject.\footnote{An interesting illustration of this point is given in "A Data-Processing Dilemma: To Repunch a Large Volume of Data or to Process with an Inefficient Program", by Richard Smith, Larry Stember, Benjamin Schwartz of the Battelle Memorial Institute, presented to the Thirteenth National Meeting of the Operations Research Society of America, Boston, Massachusetts, May 1958.} With such a precise distinction each data element can exist in isolation without loss of any essential attributes. As with the conflicting descriptions of the elephant by the unseeing wise men, it is essential that there be no failure to distinguish between different, but similar, things or to detect the similarities.

One final essential property is needed - the means whereby any element can be taken from its place in relation to other elements and
consistently be returned to its rightful location. This structural linkage is required by and applicable to every element. It is provided by including in the identification portion of each data element two other pieces of knowledge -- the identification of one other element to which there is a relationship and the specification of the nature of the relationship.

For some elements there may need be no relationship other than to the identified subject, such elements are termed non-structural. This property may be used to simplify the representation of the model. In practice it is possible, even likely, that much of the detail would not require explicit specification. Existing conventions as to organization and accounting would make it possible to reduce detail and the manner of expression may, in itself, define the basis. For example, it will not usually be necessary to specify the language used in plain language descriptions, or to define the mathematical properties of the number system in reporting quantities.

Thus, the identification portion (I) of the data element is composed of three members:

a) The subject identification \( I_s \) to which the element refers.\(^{11}\)

b) The reference identification \( I_i \) to another element.

\(^{11}\) The specification of \( I_s \) is essentially a coding problem. The multi-dimensional vector that provides the code structure is not essential to the present discussion and is deferred until Chapter III when the data elements are interpreted in the context of a management data system.
c) The relationship \((I_r)\) to the referenced element.

The substance \((S)\) of the data element is a dual member. One part discloses the type of content \((S_t)\), the other may conveniently be considered as having one of two possible forms of content \((S_c)\):

a) A narrative or descriptive content \((S_n)\) or

b) A quantitative content \((S_q)\).\(^{12}\)

It is, of course, essential to specify in either case the form being used and the measure system involved.

The purpose of this type of data element is to make it possible to develop a complete data system and conversely, given a data system, to partition it into data elements so that it can be disassembled, analyzed, and correctly reconstructed, as desired.\(^{13}\)

This view of a data system and its constituent elements can be depicted as a connected graph by considering all non-structural elements of each subject, combined into a group, as nodes and the structural elements as links. As defined, every link will be bidirectional in the sense that every relation has a unique reciprocal.

A data element \((A)\) can be graphically represented in relationship to another element \((B)\), as shown in Figure 1.

\(^{12}\) The identification - quantification composition of data has been used before, although not in such detail as here employed. See, for example, Lieberman, Op Cit.

\(^{13}\) There is no intention to show, at this point, that a data system so constructed has empirical meaning. The model and its expression in matrix form will later be viewed in the context of a data system for support of the economic decisions of an "industrial firm."
**FIGURE 1**

A =

<table>
<thead>
<tr>
<th>$I_s$</th>
<th>$I_i$</th>
<th>$I_r$</th>
<th>$S_i$</th>
<th>$S_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_s(A)$</td>
<td>$I_s(B)$</td>
<td>$I_r(AB)$</td>
<td>$S_i(A)$</td>
<td>$S_c(A)$</td>
</tr>
</tbody>
</table>

Graphical Representation of a Data Element

**FIGURE 2**

<table>
<thead>
<tr>
<th>$I_s$</th>
<th>$I_i$</th>
<th>$I_r$</th>
<th>$S_i$</th>
<th>$S_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_s(A)$</td>
<td>$I_s(B)$</td>
<td>$I_r(AB)$</td>
<td>$S_i(A)$</td>
<td>$S_c(A)$</td>
</tr>
<tr>
<td>$I_s(A)$</td>
<td>$I_s(C)$</td>
<td>$I_r(AC)$</td>
<td>$S_i(A)$</td>
<td>$S_c(A)$</td>
</tr>
<tr>
<td>$I_s(A)$</td>
<td>$I_s(D)$</td>
<td>$I_r(AD)$</td>
<td>$S_i(A)$</td>
<td>$S_c(A)$</td>
</tr>
<tr>
<td>$I_s(A)$</td>
<td>$I_s(E)$</td>
<td>$I_r(AE)$</td>
<td>$S_i(A)$</td>
<td>$S_c(A)$</td>
</tr>
</tbody>
</table>

Data Element Relationships
Relationship to several other elements requires that separate elements be defined -- one for each relationship as shown in Figure 2. For the moment there is no specified beginning or end to the network. It is now necessary to define the boundaries and discuss their characteristics.

4. Requirements and Sources

The boundaries of the data system network are the decision-makers for whom the system is provided and the sources of the data.

The process by which the decision-maker determines his data requirements are not of concern. The initial assumptions provide that the determination be made and that the appropriate data needs expressed. For purposes of the construction of the theoretical model of the data system, the expression of data needs is subject to these assumptions:

a) Content. The data needs are expressed, or can be unambiguously interpreted, in the context of data elements. Thus, every subject about which data is needed and the exact type of data needed on every subject is specified.

Requests for "all pertinent information" and for "perfect information" are not operational and are, accordingly, excluded. It is further assumed that such requests will be clarified and made explicit.

In an organizational hierarchy the superior levels may well specify that certain "directive" data must be accepted by subordinate levels. This too is considered as a specification of need.
b) **Time.** The need for every data element which is
time constrained will be expressed in relation to a time
index. If any elements are listed for submission on a
"when requested" basis the permissible time span between
the request and the submission is provided. Time is assumed
to be measurable in units of such dimension that every time
reference is expressible in integer units. By convention,
the origin is chosen such that all events are associated
with a positive index and all measures of elapsed time are
taken as positive.

c) **Value.** There is implicit in the previous assump-
tion, that same "objective" will be sought, an expectation
that with every data element requested, either singly or as
a member of a set of elements, there will be associated a
value (cost) of its provision in the units of measure of
the objective. If, for example, the chosen objective was
to minimize the time between receipt of the element from the
source and submission of the element to the decision-maker
the "value" would be expressed in expended time units. Since
the objective of the data system is not necessarily identical
with the overall objective of the association, it is now
assumed that the decision-maker will express an additional
value index in terms of a premium or discount for the data
requested if such a condition is necessary to assure that the
objective function observed in the data system is compatible
with the overall objective.
d) **Content and Time vs. Value.** The objectives discussed above involve an implicit assumption that there is required only one possible set of data elements. The more general problem includes the possibility that several alternative sets are acceptable, each with different time constraints, and that one of the alternative sets must be provided. If, for example, the overall objective was profit maximization and the data system objective was cost minimization the decision-maker must assign such value indices to each of the alternate report sets as will assure that the objectives are compatible.

In considering the sources of data there is no assurance that the data elements in the output are identical with the data elements in the input. Quantitative data elements, especially, can be manipulated mathematically to obtain various sums, differences, ratios and products: great numbers of elements can be represented by averages or other statistical measures; and even narrative elements can be abstracted or consolidated to form new elements.

Discussion of the internal sources of data elements will be deferred until the next section since they are dependent upon both the final output and original input and related to the intermediary data elements that do not appear in the final output.

Whatever the data elements determined to be needed as inputs, their ultimate sources can be classified broadly as the decision-makers, the events within the association, and the environment in which it operates.
The decision-makers are the source of two classes of data. The most fundamental is that which establishes and modifies the structure of the association, its internal relationships, objectives, and functions. The second class is comprised of the control variables established by operational decisions. All such decision data are potential inputs to the data system.

The events within the firm constitute a completely different kind of source. "Event" data is intended to provide a representation of the activities of the association. These activities can be observed in many ways and thereby give rise to many alternative data elements. For the moment, all alternative observations are potential sources of the desired element.\textsuperscript{14}

Environmental data includes both event and decision information. Its distinguishing characteristic is that it is neither an event nor decision within the established context of the association.

The data elements and their interrelationships defined in the previous section and the requirements and sources of data discussed above provide all of the necessary components for the representation of a data system model except for the means to represent the alternative sources of each output element.

The model is now extended to accommodate the derivation process whereby an element of data can be created from other elements and the aggregation process whereby several elements can be combined into one without losing their separate identities.

\textsuperscript{14} The possibility of selecting the "best" representation and then defining it as a data need is considered in Chapter V.
5. **Derivation and Aggregation**

The data system must include the means for linking the data elements which appear as output to those which are to be collected from the several sources.

The input data elements requirements are determined from the output elements. Excluding those output elements which are unobtainable because there is no conceivable method of observation, generation or derivation of the element, every feasible output element can be obtained:

   a) directly, by observation of an event, from a decision-maker or from external sources as environmental data.

   b) indirectly, by derivation from two or more other elements in a definable process.

Analysis of the required feasible outputs may thus give a series of alternative choices of inputs.

By combining the possible sources and the derivation choices all data elements (E) can be classified as:

   a) Generated Elements (Eg). Those elements which originate with the decision-makers.

   b) Observed Elements (Eo). All elements obtained by observation within the association or from environmental data outside the association.

---

15 By the assumption on the decision-makers' expression of content, all infeasible requests are to be redefined until a feasible input source is found.
c) Derived Elements ($E_d$). All elements obtained from a definable combination of other elements. For example, a logical choice operation would provide a derived element; an estimate which involves a decision would be classified as a generated element.

The aggregation of data elements into special groups -- themselves data elements -- offers the possibility of simplifying the model in much the same way that palletization improves materials handling. If the exact composition of every aggregate is defined the structure of the model remains intact in terms of data elements and the aggregation process will not destroy the precision of identification in the model.

Three data aggregates are defined: the Group, the Form and the Report. Each is identified as a new subject and each element of its composition related to it by the reference identification. Thus the aggregates are themselves derived data elements and possess all the attributes of the other elements.

A Group is defined as any set of two or more elements having common subject identification ($I_s$). A complete group is the set of all data elements in the model on a given subject.

A Form is any specified set of data elements.

A Report is a form submitted as final output to a decision-maker.
6. The Data Matrix

The previous sections have provided the necessary components of a data system. The present section is concerned with the provision of a means for consistent manipulation of all of the components.

Matrix notation can readily be adapted to this task. It is only necessary to establish a consistent subject/content index to provide the necessary row/column reference. The set of all subjects \( I_s \) that are present in the data elements are assigned an index \( j = 1, 2, 3, \ldots, J \).

The set of all different types of substantive content \( S \) are assigned an index \( k = 1, 2, 3, \ldots, K \). If the feasible data elements required by the decision-makers are not infinite in number there is by the definitions on derivation at least one denumerable set of data elements that will permit preparation of the output set. Of course, all elements may be expressed in the matrix and any which are found to be infeasible can be isolated and removed from the output requirements. It is assumed that neither the output set nor the number of alternative input sets are infinite, and that infeasible outputs have been excluded.16

---

16 A distinction is made here between elements which cannot be obtained and are therefore excluded from the permissable output set and those which can be obtained only at great or "infinite" cost (in terms of the objective function). The final implications are not identical since in the first case there is no conceivable method to obtain the element while in the second instance better techniques or equipment may yet reduce the objective cost and make the solution feasible.
As convention, an extra column index \( K = 0 \) is established as an identification column. If the element exists the intersection \((j,0)\) will have a unit coefficient, if the element cannot be generated, observed, or derived its \((j,0)\) coefficient is zero.

Thus every element is representable by an identification matrix (actually a row vector) with a single non-zero coefficient -- a unit coefficient at \((j,0)\). Every generated or observed element is representable by a unit coefficient at \((j,k)\).

Every derived element is representable by a derivation matrix with a unit coefficient for each element required in the derivation. The exact process of derivation is not stipulated but, unless it exists, the process is termed decision-making rather than derivation and the result is a generated rather than a derived element.

The derivation matrix is obtainable as the matrix sum of the identification matrices of its constituent elements and replaceable by its own identification matrix.

7. A Theory of Data Systems

The necessary components of the proposed data system theory are provided in the initial assumptions as to the setting of the problem and in the definitions of the data elements and their relationships.

These components are now expressed as the data system model which is to be examined for its empirical and formal implications in the following chapters.

The setting for the data system includes four components:

a) The decision-makers who:
(1) specify the form, functions, and objectives of the association.

(2) request data

(3) issue decisions on control variables

b) The form and functions of the firm as established and modified by the decision-makers.

c) The productive processes which the association performs.

d) The external environment.

This Data System Model is shown in Figure 3.

Since the immediate concern is with the data system, assuming all other components as given, the model is simplified by limiting it to the data system (D), the input from the decision-makers (I_g), the input from events and the environment (I_o), and the output to the decision-makers (R).

It can be represented either in block form as in Figure 4 or in network form as in Figure 5.

Within the data system, the aggregation into data groups and forms (which are simplifying conventions for certain derived elements) may appear at any point. The only new component is the derived element (I_d) which is essential to the provision of elements not directly obtainable either as generated or observed elements.

Thus the final network diagram of the data system model is quite simple. It involves only two inputs I_g and I_o; one internal process-derivation-which yields I_d; and a single output R, as shown in Figure 6.
FIGURE 3

**DECISION-MAKERS**
(all levels)

Establish the association, its form and functions and direct its activities

<table>
<thead>
<tr>
<th>Issue directives and decisions on control variables, request reports</th>
<th>Provide data</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Provide decision data</th>
<th>Provide event data</th>
</tr>
</thead>
</table>

**THE ASSOCIATION AND ITS ACTIVITIES**

**THE DATA SYSTEM**

Provide exogenous data

**THE ENVIRONMENT**

The Data System Model
FIGURE 4

Issue directives and decisions on control variables; request reports

Provide observations of events

Provide Exogenous data

The Data System Block Diagram

FIGURE 5

The Data System Simplified Network

FIGURE 6

The Data System Network
In matrix form the complete system is represented as an identification matrix which is equivalent to the matrix sum of all reports.

**FIGURE 7**

\[
(D) = \Sigma(R)
\]

\[
\begin{pmatrix}
  0 & 1 & \cdots & K \\
  1 & 0 & \cdots & 0 \\
  0 & 0 & \cdots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  1 & 0 & \cdots & 0 \\
  0 & 0 & \cdots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  1 & 0 & \cdots & 0 \\
  0 & 0 & \cdots & 0 \\
\end{pmatrix}
\]

Matrix Representation of a Data system

Each report is in turn equivalent to the matrix sum of the identification and derivation matrices of its constituent elements.

Continuing this sequence, until every derived element has been reduced to either generated or observed elements, the system is then equivalent to the matrix sum of each of the combinations that provide exactly one source for each element.

**FIGURE 8**

\[
\begin{pmatrix}
  1 & 0 & \cdots & 0 \\
  0 & 0 & \cdots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & \cdots & 0 \\
\end{pmatrix}
= \begin{pmatrix}
  0 & 0 & \cdots & 0 \\
  \vdots & \ddots & \vdots & \vdots \\
  0 & 0 & \cdots & 0 \\
\end{pmatrix}
\]

or...or

\[
\begin{pmatrix}
  0 & 0 & \cdots & 0 \\
  \vdots & \ddots & \vdots & \vdots \\
  0 & 0 & \cdots & 0 \\
\end{pmatrix}
\]

Matrix Representation of Data System

Design Alternatives
CHAPTER III

A MODEL OF DATA SYSTEMS FOR ECONOMIC DECISIONS

1. From the Abstract Toward the Real:
The Scope of Management Data

The expression of a theory in the form of an abstract model, whether verbal, schematic or mathematical, conveys little practical meaning. A first test of its applicability is its interpretation in the context of some recognizable problem which offers a significant challenge in a familiar frame of reference.

The problem of interest is that of a data system intended to serve the needs of the economic decisions of a firm.

The concept of the firm and the scope of its activities which are to be considered as a part of the problem are defined in relation to the initial assumptions as follows:

a) The Association is a firm in the sense usually conveyed in economic theory.

b) The Function of the firm is to contribute some economic benefit involving the use and consumption of economic resources and the production and sale of products and services. In this sense it is called an industrial firm.

c) The Objective of the firm is to maximize profits, subject to the constraints imposed by administrative,
social, technological and economic conditions and attitudes.

d) The Decision-Makers of the firm are managers who establish and maintain its form and functions and control its activities.

e) The Relationship established between the managers and the other aspects of the firm is known as the organization. It involves both structural and functional relations.

f) The Function Performing components are the operators assigned to perform the activities of the firm.

g) The Activities encompass all direct and indirect, administrative, production and service events implicit in the concept of an industrial firm and the economic goods and services which are employed therein.

h) The Accountable Relation among the managers, the operators and the events are established by an accounting system.

i) The Information needed by the managers is the data which are to be provided in support of economic decisions.

j) The Control Variables are the policies and directives by which managers exert their influence on the firm.

k) The Data System is expected to provide the data needs (reports) at minimum cost.

Several assumptions are implicit in the above definitions. Two are of appreciable significance.
First, non-economic decisions which are inherent in the concept of an industrial firm are assumed to be made as necessary to the conduct of the firm. Whether or not the theory can be meaningfully interpreted in the context of technical decisions, such an interpretation is here excluded in the same sense that the technical choices implicit in the production function are not treated as matters for economic decisions in the theory of the firm.

Second, it is assumed that the managers determine their data needs in conformity with the profit-maximizing objectives and that cost-minimization in the data system is a compatible objective. Specifically, any effect of the provision of data on the income or other costs of the firm must be reflected in the value index of each report.

It should be noted that this definition of the data system does not encompass all communication. It is conceivable that some if not all of the managers are also operators and influence directly the activities and other operators of the firm and observe directly internal events and external environmental data.

Individuals and functions may perform several roles. The accounting function, in particular, and possibly other activities may involve both decisions and data derivations. Such components are accordingly viewed as managers and operators for their decision-making and operating functions and as an integral part of the data system for their data collection, processing and derivation functions.

The interpretation of the data system model in the desired context involves four problems:
a) A structuring and coding of the data matrix to permit ready accommodation of all necessary data elements,
b) The translation of data needs into matrix terms,
c) The determination of data sources and
d) The selection of a solution from the available alternatives.

The data matrix will be discussed in the next six sections, followed by consideration of the remaining problems.

2. Matrix Representation of Management Data

The data matrix discussed in Chapter II was necessarily very general. No specific dimensional measurement is suggested but the number of possible subjects is very large and the number of different types of content equally extensive. Although the full matrix can logically embrace any desired problem a much more specific relationship is necessary to the argument that there is meaningful empirical content to the theory when in the desired context.

It is now shown that the data matrix can be structured in such a manner as to make it readily suited to the accommodation of "management" data.

The criterion for successful matrix representation is provision of a specific address for all data such that identical data will be given the same address and similar data will be so classified that it can be readily located for analysis.

Such a criterion cannot be met by a single "linear" code in which the subjects are listed serially in any given order. A one-dimensional code is well suited to identification but not to data
retrieval or analysis. It is essential that a compact representation be provided and that the labeling of each subject provide a classification that serves well the needs for identification and retrieval.

A crucial problem is introduced by the necessarily temporal setting. It is relatively easy to develop an analysis plan for one point in time; what is needed is the capacity to accommodate a long sequence of events in a manageable, consistent pattern.

The design of the classification plan is essentially a solution to a "library search" problem. As data elements (books) are placed in the matrix (library), each must be classified and its character and location recorded in order that it may be found for future reference. Just as books are classified by code (e.g. Dewey decimal system) by title, by author and by subject (for ease of reference), data must be identified by its relation to several different classification groups.

The basic distinction between subject and content must be maintained throughout the classification scheme; there is both mathematical and practical advantage in the concept of a two-dimensional matrix. The deficiencies of a linear classification plan necessitate consideration of additional "dimensions"; the two-dimensional basis is maintained by aligning the classification dimension with the basic subject-content distinction.

The number of classification dimensions necessary for precise identification and meaningful association depends upon the type of data to be accommodated and the intended uses. One dimension is needed
for each category which is to be related to all elements. The categories that relate to some of the subjects are treated as sub-dimensions of the major ones.

The accommodation of data for the economic decisions of an industrial firm, as defined, requires a minimum of five major and six minor dimensions:

a) Program
b) Natural Structure
c) Management Structure
d) Content
   (1) Kind
   (2) Status
   (3) Unit of Measure
e) Time
   (1) of Generation
   (2) of Occurrence
   (3) of Record

It will be recalled that a data element was defined to be comprised of five parts:

1. The Subject-Identification $I_s$
2. The Reference-Identification $I_d$
3. The Relation to the Reference $I_r$
4. The Type of Substantive Content $S_d$
5. The Substance $S_c$

Of these, the Reference Identification ($I_d$) is a citation of some other Subject-Identification or a repetition of $I_s$ and may therefore be excluded.
The Substance, whether narrative or quantitative, is not involved in the matrix address. It is expressed in strict accordance with the type code and unit of measure provided by $S_1$ and is not of concern in the design of the system. The installation of a data system must, of course, include consideration of the necessary length and form of the Substance and the technical and administrative problems of its handling.

There are then three parts of the data element of specific concern in the matrix -- $I_s$, $I_r$, and $S_i$.

The Subject-Identification ($I_s$) and the Relation ($I_r$) are treated together in the first three classifications dimensions as the row index of the matrix. Since the essential logical order of these dimensions is not readily apparent they are treated in a somewhat arbitrary order of:

1. Program -- a general functional classification closely related to both of the structural dimensions.

2. Natural structure -- a consideration of the levels or echelons of the subjects as established outside the control of the manager either by the natural structure of the activities of the firm or imposed by administrative, legal, political, or social conventions.

3. Management Structure -- the levels and echelons which are under the control of the managers and can therefore be established and modified as appropriate by management decisions.

The Type of Substantive Content ($S_1$) is accommodated in two additional classification dimensions as the column index:
4. Content -- the specification of the character of the Substance \((S_u \text{ or } S_g)\) and the units of measure employed.

5. Time -- a reference to one or more of three possible time indices -- the time of generation (as for an estimate or decision), the time of observation (as in date-stamping mail), and the time of occurrence (as the point or span of time related to a level or flow).

These five classification dimensions are now considered in order both as to their internal composition and their interrelationships.

3. The Program Dimension

A Program is defined as any convenient "functional" classification of the activities of an industrial firm. Since the overall scope of the firm can be conceived, if not precisely specified, the classification could be made in any desired manner if a "Miscellaneous" program is provided to assure that the generality of the classification is not compromised. The criteria for convenience and the manner of preparing the classification can readily be developed from consideration of the nature of the problem and the reasons for the classification by program.

The classification is appropriate because:

1. The activities of an industrial firm are diverse and fundamentally different. Comprehension of any one program does not assure comprehension of all.
b) The organizational structure and functions create and maintain substantial distinctions between the operating components. The managers establish a concentration of interest in specific programs.

c) There are significant differences in internal program structure. Program identification will, therefore, contribute to structural identification.

The program dimension is a recognition of the operational and analytical merits of the common practice of subdividing work and specializing effort. Apart from any relation to natural programs such as sales, production and plant maintenance that are conceived to exist in an industrial firm, there is the additional advantage that sub-division of the overall problem into programs on any rational basis will aid in the comprehension of the data problem and in its analysis and solution.

The specification of the programs which will be recognized in a particular application is a matter of choice. There is no known basis for a claim that there is a single "correct" partition. The guide lines for choosing can be inferred from current industrial practice and from consideration of the intended uses of the program dimension.

The functional divisions of the existing organization or the sample organizations from any recognized industrial management text are a first approximation to the program list, since the induced interest of the managers may be expected to follow such a division and since there are developed structures within each division.
The improvement of the list of programs is dependent upon the uses intended for the classification.

Since the fundamental purpose of the program list is to provide a logical structure for all subjects in order to permit understandable analysis and manipulation of the data elements, it is apparent that the practical range of program lists is necessarily limited.

The internal structure of the program which is to be developed in the second and third dimensions provides a practical basis for revising the program list. The recognition of any natural structures and the necessity to provide an understandable subject pattern by the establishment of management structures will bound the number and scope of programs. The requirement that every program, except "Miscellaneous" be supportable by a logical segment of management structure will serve to provide an upper limit. The distinct structural pattern of present industrial organization will assure that there are at least a few meaningful divisions and thereby a useful lower limit.¹

¹ The most common organizational functions, administration, sales, production, maintenance, accounting, etc., provide the basis for enough programs, by whatever name, to be substantially adequate to reduce the lack of structure in the overall list of subjects. It is likely that the list need not exceed twenty to thirty programs, but, if in a complex firm, this estimate were low by an order of magnitude no essential compromise of the concept of the program dimension would result.
4. **The Natural Structure**

The Natural Structure of a program is defined as any pattern of the subjects which results either from external imposition or from the inherent relationships between the activities and objects which the subjects represent.

This dimension is included because:

a) Any structure imposed from outside the firm is necessarily of interest to the firm. Perhaps the most pervasive examples are the systems imposed by governmental agencies. If the firm must provide any data to such external agencies it must be prepared to report in conformity with the imposed structure.

b) Any structure which is inherent in the subjects is of potential interest because the managers and operators must be prepared to deal with the reality of activities and events. For example, the physical structure of a product and the necessary technological sequences of its manufacture provide a natural frame of reference.

c) It is possible, although not certain, that natural structures, whether imposed or inherent, may provide a useful point of departure for the development of management structures.

This dimension is thus a recognition that the data system must be prepared to accommodate reality as it is viewed by the managers and reflected in their data requests and as it exists in the events that constitute one of the primary sources of input elements.
Except for the problem of interpreting correctly the meaning of those who established the imposed structure and of identifying the inherent structures, this dimension does not require a creative classification. The extent and complexity of the structure are therefore not precisely bounded; some reliance can be placed on the limits to which external agencies are held by the difficulty of defining and maintaining a complex structure and the limits on the capacity to resolve the intricacies of physical relationships.  

5. The Management Structure

The Management Structure is defined as a pattern of levels or echelons established by the managers to provide an orderly frame of reference for the control of the firm.

This dimension is needed to:

a) Provide a means for bringing some of the diverse subjects of the natural structures of several programs into a simpler pattern. It might, for example, provide that all individual work assignments be given a common name and be identified within a single code system.

b) Provide a structure with a limited number of levels to represent a complex natural structure. The

---

2 Long established institutions do develop intricate structures but it is not likely that there will be more than fifteen to twenty structural levels or echelons in any program pertinent to an industrial firm. As in the case of the number of programs the number is not critical, but the usefulness of the classification is diminished if there are a great many subdivisions.
necessary comprehension of data is more difficult if the natural structure is intricate or if the distinction between levels is not readily discernible.

c) Provide a means to relate a program to the organization. The organization reflects both the functions to be performed and the personnel available and may be based on geographic or other non-functional segments. Thus, it is possible that the program and the organization are not co-extensive and that a common basis of reference need be provided.

The essential purpose of the management structure is to provide a means to accommodate the necessity for managers to view their operations in a relatively simple, consistent pattern. The organization and the accounting system are both reflections of this same need and both are intimately associated with the management structure.

The purpose of the management structure is best served if the number of levels is held to the minimum and if the management structures of similar programs are closely related. For example, in an organization which includes a product development program separate from the manufacturing program the use of management structures with the same number of levels and similar internal definitions for both programs will provide a more consistent pattern and may well enhance inter-program continuity.

Since the number of levels can be adjusted to suit the convenience of the managers, an intricate program can be recast in a simpler mold. If, for example, there are several products, each
with a different structure, there will be no common basis for comparison in the natural structure. Although a management structure may not fit the natural structure of any product exactly, the imposition of a common pattern can provide the uniformity needed for comprehension and a point of departure from which the differences can be recognized.

The circumstances that tend to limit the number of persons who can be directly supervised and the number of levels that can be maintained in a responsive organization also limits the number of levels in the management structure. A structure with more than about ten levels for a program is not desirable and perhaps not necessary. In contrast to the relative indifference with the number of programs and the number of levels in the natural structure, the utility of the management structure will be severely restricted if a simple pattern cannot be maintained.

6. The Content Dimension

The Content Dimension provides a classification of all of the different kinds of data about the subjects.

The primary basis of the classification is the common one of distinguishing among data on the basis of their fundamental substance such as cost, schedule, quantity, and description.

Each element is further classified by an indication of "status". This division permits concurrent treatment of those cases where there appear to be several conflicting values of a single measure. A common example is that of a product for which there are several different measures of "total cost". One may represent the engineer-
ing estimate of cost, one the burden of authorization. Each of these is based on the premise that the matrix must be suitable for a particular problem. This approach may be taken as a matter of course, but it is important to stipulate the possibility of confusion.

The final basis for classification is based on the substance of every element. Substance of every element can be determined by a process of measurement. In the instance of "plain language" may be taken as a matter of obvious units, as in the case of objects, but it is important to stipulate the nature of the proper place for each.

The necessity for this classification methods to the need to distinguish between different kinds of recognition and characteristics of a particular kind.

The classification of content in exclusive categories and an exhaustive listing of the matrix elements permit unambiguous and accurate classification of the proper place for each element.

If all else fails it is possible to provide a position scheme by extending it whenever an item is positioned. This approach may be considered a rigorous approach can be based on a short list of data which are likely to be requested.

Data requests can be broadly classified into:

a) Decision Data to receive support decisions to modify plan
b) **Background Data**

c) **Analysis Data**, study of which may yield either background or status information.

The need for decision data suggests that the data must provide the basis for some type of comparison. The basis for comparison may be standards established by the managers, past experience, or standards provided by trade or industry, associations or the government. In addition to direct comparison to standard targets, different categories of data can be converted to a non-dimensional basis (e.g., percentages) and thus compared.

Background data will include both descriptive and measure data to make specific the activities and events of interest to the managers.

Analysis may involve further study of both decision and background data and also data which is not available or not useful for regular decision making.

The second major basis of classification -- "status" -- specifically recognizes that many subjects in an economic setting are intimately related to and conveniently classified in terms of the conversion of resources into products.

The reasons for and the scheme of this classification can be illustrated in terms of a specific task that is necessary to some activity of the firm. The description data for the task, perhaps if subject to change, can be readily classified. The treatment of quantitative data introduces certain complications. The data system must accommodate three different figures -- the estimated requirements to perform the task (in dollars, man hours, machine hours,
parts quantities, etc.); the plant, inventory, etc.; and resources.

A simpler example is to refurbish a room. The costs substantially different from one another and different from her budget are resolved by a change in her available funds (by borrowing), or a revision of the three. In any event, the final values will be the same. The allocations, the "resources" converted to the final "allocation".

Thus the "status" problem is solved by repeating the selected computational headings of requirements/sat allocations.

The final major basis of the system. There is obvious means of reducing costs to a minimum and in doing so, the criteria of measurement. The decision is not on the ability to measure the result. Costs are frequently underestimated. They were collected in detail because managers can discriminate such costs without first conceiving the allocation, scaled to the user may involve...
hundreds, thousands or even millions of dollars.

7. The Time Dimension

Most of the data in an industrial firm are closely associated with a time reference.

The most direct reference is to the time of occurrence and the duration of an event. Estimates of future events introduce the need for a time index to permit reference to the time when the estimate was made. Finally, there is often need to record the time of entry of the data into the system.

These three indices permit precise identification of the time references. The same type of classification plan is desirable for each. Three items are needed. First the nature of the time index, i.e., whether a span of time or a single point in time is being cited. Second, the units of measure which may well extend from microseconds to decades. Third, the report of the measure in the chosen units.

The choice of a time reference is likely to be arbitrary although conventional practices provide a substantial guide. It is to be expected that the pattern of time spans will involve four phases. Past events summarized in relatively long periods; current events in greatest relative detail; the short range plans in periods of intermediate length; and long range plans in periods equal to or greater than used in the historical record.

The temporal problem is intimately associated with the individual manager's concept of "real time". For a "computer man" real time carries the implication that the processing of data representing
events is to occur "simultaneously" with the events themselves.\(^3\)

In a broader sense, each manager faces a time reference controlled not by a clock, but by the types of decision which must be made. As the manager is removed from proximity to physical operations his decision time is usually lengthened, so that where one hour may be an excessive time for a foreman, a week or month could be appropriate to the chief executive. Thus, to each decision-maker there are several imprecise, but recognizable "real time" references and it is this temporal framework to which the data systems must be responsible.

3. The Management Data Matrix

The generation of the management data matrix in terms of the five principal classification dimensions can now be related to the address parts of the data elements.

Subject-identification (I\(_S\)) is provided by the program and management structure. The natural structures, which are fundamental to the technical\(^4\) effort in each program, are not used in the development of the system except as they influence the management structures.

---

\(^3\) The technical implications of "simultaneity of events" are sometimes avoided by the use of concepts of micro-taneous and macro-taneous events with the implication of distinctions in time which are respectively undiscernable and apparent.

\(^4\) Technical is here used in the sense of the program specialty. Accounting is treated as a technical effort of the financial program in the same relation that science or engineering hold in a research and development program.
Careful design of the management structures will provide the necessary Relation \((I_r)\) as an intrinsic part of the system. It is necessary to define the intended meaning of the management structure but, once established, the basic structural relation can be assumed to hold unless specifically excepted. Relationships which involve derivation or generation of new data elements will require explicit specification of \(I_r\) in the data elements involved.

The Type of Substantive Content \((S_i)\) is accommodated by the major content and time dimensions.

The use of the matrix in the synthesis of a management data system can be considered in relation to a single program. The extension to a complete system follows if:

a) Any new elements which appear can be accommodated in the same matrix, and

b) The formulation for one or two programs can be consistently extended to all others without compromising the integrity of the system.

The question of adequacy of scope implied in the first point is answered by the form of the classification plan. Any and all data can be classified. The equally important question of the effectiveness of the classification code depends not only on the dimensional plan but also on the flexibility of the coding scheme and of the procedures used in the analysis.

The second point is fundamental to the attainment of an integrated system. It is known that many subsystems, e.g. payroll, inventory control, and personnel accounting, can be individually organized on an integrated basis -- if the integration can be extended
from one system to the next a fully integrated system is feasible of attainment.

The coding plan for the major and minor dimensions of the matrix must provide a compact representation while assuring the flexibility to accommodate any new identification or substance classifications that appear in the course of the study.

The code should enable an analyst to consistently and quickly classify any data element; permit the data elements to be conveniently sorted on any combination of dimensions; and provide a pattern that could be readily learned and remembered.

The principal features of the multi-dimensional matrix and its relation to current industrial concepts can be illustrated by a simplified model of a "production planning program".

Assume that the products require the manufacture and assembly of components -- some of which can be purchased. Only budgetary, schedule, and inventory control will be treated explicitly; personnel skills, plant capability and capacity and other resources are also involved, but are not essential to the example.

The dimensions of the data matrix could then be developed as follows:

a) **Program.** In addition to the initial "Production Planning" program, four others are needed for reference. A "Sales" program to provide customers; a "Financial" program to provide and administer authorized accounts for dollar resources; an "Inventory" program to maintain and to provide a source of status information on materials, assemblies, and finished goods; and an "Operating" program to purchase, manu-
facture, assemble and move the items. Each of these auxiliary programs will have an internal structure and may, in fact, be subdivided into several programs. Such details are not essential to the argument since the "Production Planning" program is assumed to be the initial basis for integration.

b) Natural Structure. It is assumed that the engineering effort necessary to the product design has been completed and that the required parts and processes are known. There is no assumption that the natural structures are identical for all products or that parts and processes are unique to a single product. The natural structures of the auxiliary programs are not of concern.

c) Management Structure. A four level segment of the management structure is assumed to apply to the Production Planning program. The top level representing the total production program; the second level the individual products that are sold; the third level the major subdivisions of effort required in the production of the product; and the fourth and lowest level the individual work orders that are issued to effect accomplishment of all necessary jobs. These levels are given standard names of program, project, subproject and task respectively. In this structure all orders enter at level two as projects and all work directives are issued at level four as tasks.
d) **Content.** Each subject in the natural and management structure will be represented by descriptive (engineering) data including information on the inter-relationship of parts and processes and the quantities and amounts of materials and components needed in each product.

Cost and quantity data and the "status" classification is required on each level. For simplicity it is assumed that costs are measured in whole dollars and quantities in units.

The schedule is represented by control events that indicate the significant milestones in the operations on each order. The "status" classification need not include resources since, in this case, the resource is time and need not be explicitly represented.

e) **Time.** The event times for the schedule milestones will provide the means for progress review as well as for determination of delivery estimates and for later analysis.

It will be necessary to record the time at which certain estimates are made and, if the operational sequence is rapid, the time reference of key data movements.

By concentrating on the management structure and data content, the five classification dimensions of the matrix can be represented in the desired two dimensional plan, as shown in Figure 9.

The representation of the "Production Planning" program example is shown in Figure 10.
The diagram outlined in Figure 9 and detailed in Figure 10 can be interpreted in several ways.

The physical flows occur in the Resources/Products columns. For example, dollars flow in on the basis of a sales order, are held in some account, then flow out for operations costs of payroll and purchases. Materials flow in to operations, thence to inventory, back to operations for manufacture and assembly, to finished goods inventory and finally to the customer.

Physical transformations occur in the rows under Resources/Products. The Sales program converts quantities into dollars; operations converts dollars (and time) into quantities. The Production Planning program does not deal with any resources -- its functions are limited to planning and control.
**FIGURE 10**

<table>
<thead>
<tr>
<th>Description</th>
<th>Requirements/Costs</th>
<th>Satisfactions/Quantity</th>
<th>Allocations/Challenge</th>
<th>Resources/Products/Schedule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Project</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>* * *</td>
</tr>
<tr>
<td>Subproject</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>* * *</td>
</tr>
<tr>
<td>Task</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>* * *</td>
</tr>
</tbody>
</table>

**SALES**

|                     | *                  | *                     | *                     | *                            |

**FINANCIAL**

|                     | *                  |                      | *                     | *                            |

**INVENTORY**

|                     | *                  |                      | *                     | *                            |

**OPERATIONS**

|                     | *                  | *                     | *                     | * * *                        |

<table>
<thead>
<tr>
<th>Planning</th>
<th>Control</th>
<th>Operations</th>
</tr>
</thead>
</table>

* The asterisk indicates the cells of interest to the production planning program.

The *Data Matrix* for a "Production Planning" Program
The differences between the requirements developed in connection with planning activities and the resources (wherever they are held) that are to be transformed are resolved through control allocations. Discrepencies between plans and controls are readily recognized when the requirements to do a project or task -- as planned -- exceed the control allocation. This discrepancy and related discrepancies between resources and allocations must be finally resolved. The recognition that the three values can and will differ permits corrective action to proceed without the necessity to deny or ignore the differences. There is, after all, a real problem involved in replanning a project to accomodate new dollar, time or quantity controls. The allocation data can be changed more quickly than the engineering and planning which support requirement estimates or the physical stocks of resources. All three measures are essential if there is to be a meaningful signal on which to base corrective action.

If desired, the complete story of the data representation of the events involved in the life sequence of a sales order can be traced through the diagram. The intra- and inter-program transfers; the sequence of planning, control and operations efforts; and the representation of physical movements of goods and dollars can be visualized in relation to conventional industrial processes.

The significance of this interpretation of the production planning and control process can be seen by reference to some current text on industrial management.\(^5\) Although the problem has been

\(^5\) A representative list of twenty-one duties of production
greatly simplified for purposes of illustration, it can readily be extended to encompass other duties.

The incorporation of other programs in the integrated system is seen to involve either additional treatment of some of the same data or inclusion of new data. The integrity of the system can be maintained by assuring that each addition to the system is compatible with all previous elements.  

9. The Expression of Data Needs

The interpretation of statements of data needs from all levels of management can readily be related to certain portions of the data matrix. The major programs, many natural structures, and a substantial pattern of content and time references are likely to be apparent. Management structures, if they exist, may not be well defined or sufficiently developed on a program basis across organizational lines.

Thus, the practical interpretation of data needs in the management data matrix involves three problems:

a) The completion of the pattern for subject identification with particular emphasis on the management structure.

control is provided by Claude S. George, Jr., Management In Industry, (Englewood Cliffs, N. J., Prentice-Hall, 1959), p. 457. Each can be recognized in the context of the example.

6 The program chosen as the initial basis should be the one that best reflects the central purpose of the firm. In this way the broadest possible base for integration will be provided.
b) The determination of the environmental data to be incorporated.

c) The development of a complete plan for content and time classification.

There are difficulties in conceiving an adequate overall program list and the associated management structures and practical complications in any later restructuring of the data system. These suggest that the programs on the tentative list be finalized one or two at a time as new segments of the management structure are developed and integrated with those previously established.

The accommodation of exogenous environmental data is not so involved. The subjects chosen can later be revised with the expectation that historical gaps can usually be "backfilled". This is in sharp contrast to internal data which, once ignored, is difficult or impossible to reconstruct.

The development of an adequate and servicable content and time classification is directly related to the approach used in establishing the programs and management structures. The essential difficulty occurs in the minor dimension: kind of content. The common compartmentation into description, quantity, quality, cost and schedule may not be adequate but is certainly significant of a large part of the needed classes.

Having developed the means for subject-identification each manager can be queried on his data needs -- first to identify the subjects of interest and second, to determine what data is needed on each subject. In like fashion, existing systems can be studied by recording and classifying each flow of data. Using a common
classification code, and assuring that no changes or additions are made without consideration of their effect on the integrity of the system, several analysts can work together to identify all the data needs in a program or overall system.

10. **The Determination of Data Inputs**

Each data element to be delivered as output must be linked to one or more input elements. The problem is to find at least one way in which the output element can be obtained.

Several different cases can arise, some of which cannot be resolved until selection is finally made among the available alternatives:

a) There is no conceivable way of obtaining the desired element. Such an infeasibility can only be resolved by redefinitions of the needs.

b) The acquisition may require that an intermediate element be generated (by a decision) rather than derived or observed as a function of the data system. This case can be resolved either by a modification of the data request or the inclusion of the decision requirement in the set of some other manager.

c) There may be a host of alternative sequences. The dilemma here is that ingenious new techniques must be permitted and that an infinite number cannot be tolerated. The practical solution would appear to lie in excluding those techniques that are dominated by others -- generally
retaining one alternative of each of the major types which are conceived to be available.

The goal of this task is to develop the smallest possible set of alternatives that contain the best solution. From the final set a choice must be made and the system selected.

11. The Selection of the Solution

In the development of the management data matrix; in the interpretation of data needs in terms of data elements; and in the statement of alternative ways to acquire the wanted data there was a clear possibility and apparent advantage in proceeding a program at a time.

These steps provide a comprehensive, integrated design -- not a unique combination as the basis for an installation. Any combination of the available alternatives can be the basis for a fully integrated installation -- but the selection of the best combination cannot be made on a step-by-step basis unless there are no procedures or equipment that could be advantageously linked together and if there are no economies of scale. Such an assumption is inappropriate in view of current knowledge of the proven capabilities of large-scale data processing equipment.

Evaluation of the combinations that, for some reason, seem most promising and selection of the "best" ones as determined by the chosen objectives is the most readily apparent practical approach. Given the integrated design, it will provide a relatively good installation and the basis for further improvement since the management data matrix is not voided by a less-than-optimum installation.
Judicious examination of the system may, in fact, provide sufficient opportunities for elimination of dominated choices and for separate solution of small segments of the overall system as to make the trial-and-error (simulation) approach a useful approximation to the theoretical optimum.

The necessity for the attempts to find a good solution by heuristic methods lies in the fact that specific solution of the equation set provided by the model is not yet practical. The justification for use of the model lies in the fact that there exists an optimum solution.
CHAPTER IV

FORMULATION OF THE MODEL FOR THE SOLUTION
FOR THE THEORETICAL
OPTIMUM SYSTEM

1. The Components of a Data System

The justification for the data element approach to the design and analysis of data systems, in addition to its empirical significance when interpreted in the context of the firm, lies in the implications of its suitability for formulation as a linear programming problem. The existence of a basis solution and of a theoretical optimum solution to such a formulation, and, therefore, to the model are of particular importance.

The purpose of the formulation of the data system model as a problem in integer linear programming is not to suggest that system design can be reduced to a simple routine. There are several real difficulties in selecting the alternatives which are to be considered and in determining the pertinent costs, or other objective values, to be minimized. Beyond this the formulation is unwieldy -- it requires more equations and more variables than can be accommodated in any known solution algorithm with existing computational devices. It is reminiscent of the scope of the centrally planned production
problem with its "millions of equations in millions of unknowns"\textsuperscript{1} -- without any internal pricing concept to permit decentralized treatment and solution.

The objective of the mathematical formulation is rather to show that there is a reasonable basis for suggesting that one concept is preferable to another. In evaluating the relative merits of different concepts the ease with which the several approaches can be applied to practical problems is certainly not a matter of indifference. As between concepts that offer a reasonable \textit{modus operandi}, the important question is whether there is any systematic approach to the best solution attainable.

Thus, the present goal is to show that the data element theory encompasses a meaningful concept of a theoretical optimum solution that is attainable by systematic application of the model in the analysis and redesign of data systems.

The demonstration of the linear programming formulation is based on a combination of the features of the theory and the principal system components as defined in the previous consideration of the empirical significance of the model to the industrial firm. The data system, so viewed, involves six major groups of components:

a) A report set which includes one or more subsets for each manager to whom data is to be provided. Each subset may include a selection of reports, if desired, of which only one is to be prepared.

\textsuperscript{1} Marschak, \textit{Op. Cit.}, P. 2.
b) A set of preparation functions which receive data elements, combine them into reports, and issue the reports to the managers.

c) A set of storage (memory) functions that receive data elements, retain them for varying periods of time and issue them to the duplicating functions.

d) A set of duplicating functions which receive data elements from the storage function, produce a fixed number of copies of the elements and issue the data either to the preparation functions, to the computation functions, or to a disposal function which eliminates the excess copies.

e) A set of derivation (computation) functions which receive data elements from the duplicators, produce new elements and issue the derived elements to the storage functions.

f) A set of data collection functions that observe events, receive data from the managers, and issue the observed and generated elements to the storage functions.

These components, together with the managers, the events and the sources of external environmental data provide a complete closed system.

In Figure 11 the formulation of the model as a linear programming problem consists of the specification of an objective function and such constraints as are appropriate to a reasonable abstraction of actual systems problems.
The Components of the Data System Model

The operation of the system is considered for a single complete cycle. The system cycle time is defined on the most infrequent certain report demand. Since the time at which a report is due is incorporated into the identification of the data element that represents that report there is no chance for confusion by an idea that the "same" report is submitted periodically. Reports which are established to be provided "when requested" are not considered in defining the cycle time, but are incorporated in terms of the probability that they will be demanded during the cycle.
Thus, if the cycle time of the most infrequent report was one year all specific report requests that are to be met within that span are pertinent to the system design problem. Each certain report is required exactly once—each "when requested" report is considered in terms of the probability that it will be requested once during the complete cycle.

The objective is the minimization of the cost of providing a report set which meets the managers' stated needs. By the previous assumption that the managers' provide a suitable report value index, and excluding any inter-report profit effects, the cost minimization objective is compatible with a profit maximization objective for the overall firm.

The costs to be minimized are associated with each of the data system components. The fixed costs are interpreted as the cost of availability of a specific function for one complete system cycle. Variable costs are associated with the intensity of use of each function.

The constraints can be considered in three groups. First, the system must include a selection of functions and sources that can provide the desired output set. This is called the output constraint. Second, any time limits imposed on the reports must be met. This is the time constraint. Third, and finally, the several functions may be subject to limitations on the number of elements that can be accommodated in any time period. This is the capacity constraint.
2. Mathematical Formulation of the Model

The mathematical notation and conventions used in the development of the objective function and the output, time and capacity constraints are as follows:

\[ D \quad (D_{jk}) \]

The overall data system and the complete data element matrix. The \( j \) and \( k \) indices established in \( (D) \) are used in all other data element matrices.

The values of the elements in \( (D_{jk}) \) and all other matrices in the model appear as one of three alternatives:

"0", if the element is not available for processing by that matrix or, in the case of reports, not requested.

"1", if the appearance of an element in a particular matrix is certain. The unit value signifies that the \( j^{th} \), \( k^{th} \) element appears in the function - it is equivalent to a "yes" or "on" condition in a binary system -- fractional values have no useful meaning.

"D_{jk}"', if the determination of the value is a part of the solution. In most instances the permissible values are 0 or 1, the exception being the element values in
the disposal functions. All values are, however, to be integer in the final solution.

\[ E_{jk} \] The event and external environment set and the data element source matrix. All elements in this matrix are, by definition, observed elements.

\[ G_{jk} \] The data generating set provided by the managers and the subset of data element matrices. All elements in these subsets are, by definition, generated elements. The lower case \( g \), \( g = 1, 2, 3, \ldots \), also represents the individual managers.

The values of all elements in \( G \) are assumed to be specified. Thus, each manager stipulates the elements that he generates which are to be included in the data system either as standards for reference or as the authorized source of that element. Any conflicts within the set \( G \) must be resolved before proceeding with the formulation.

\[ C_{jk} \] The collection set and the subset of collection functions. Each subset, \( c = 1, 2, 3, \ldots \), represents a specific data collection scheme.

The matrix elements are zero for all data elements that the particular collector cannot acquire by observation or from a manager. The other matrix elements are to assume values of zero or unity as appropriate to the solution.
The identification element, $C_{j,0}$, must equal unity if any other matrix elements are non-zero and be itself zero otherwise.

This conventional requirement on the possible values of the matrix elements and the identification element applies to all other function matrices, unless specifically excepted.

$M_{(m_{jk})}$ The storage (memory) set and the subset of memory functions. Each subset $m$, $m = 1, 2, 3, \ldots$, represents a specific storage function.

$X_{(X_{m_{jk}})}$ The reproduction set and the subset of reproduction functions. Each subset is associated with a specific storage function and receives data elements only from that function. The reproduced elements, "L" in number, are distributed to the single associated disposal function or to any of the computation and preparation functions.

The matrix elements in $X$ conform to the general convention on permissible values. $L$ is an arbitrary large number such that it assures sufficient duplicate elements to meet all computation and reporting requirements. Any variable costs incurred in duplicating
unneeded data elements are recovered by
the disposal function.

Y \quad (Y_{jk}) \quad \text{The derivation set and the subset of}
derivation (computation) functions. Each
subset is associated with a specific storage
function. Data elements are receivable from
all reproduction functions but the derived
elements can only be issued to the parent
storage function.

Each derivation function is able to
perform one or more different derivations.
The input elements to a derivation "d", are
symbolized as \( Y_{d} \); the output elements,
"d^*", as \( Y_{d^*} \). If all necessary inputs
are available the identification element
of the particular derivation set (i.e.
\( Y_{d,0} \)) is to assume a unit value as a pre-
requisite condition to the availability of the
derived set \( d^* \). Further, none of the input
elements reappear as output.

Z \quad (Z_{jk}) \quad \text{The disposal set and the subset of disposal}
functions. Each subset is associated with a
specific memory function. Data elements are
received only from the associated duplication
function. The matrix elements values in Z
conform to the general convention except that
they may assume any integer value from zero to L.
P \ (P_p_{jk}) \ The \ report \ preparation \ set \ and \ the \ subset 
of \ preparation \ functions. \ Each \ subset \ p, \ p = 
1, 2, 3, \ldots, \ represents \ a \ specific \ preparation 
function \ that \ receives \ data \ elements \ from \ any 
reproducer \ and \ prepares \ reports \ for \ any \ manager.

R \ (R_{jk}) \ The \ report \ set \ and \ the \ subsets \ of \ report 
outlines \ defined \ by \ the \ managers. \ Each \ subset 
is \ associated \ with \ a \ particular \ manager \ "g" 
and \ includes \ "r" \ report \ outlines, \ one \ of \ which 
is \ to \ be \ provided. \ The \ number \ of \ elements \ in 
given \ report \ is \ signified \ as \ "h".

If, \ and \ only \ if, \ all \ of \ the \ stipulated 
elements \ in \ the \ set \ h \ are \ provided, \ does \ the 
identification \ element \ for \ a \ specific \ report 
in \ a \ specific \ preparation \ function \ assume \ a 
non-zero \ value - specifically \ a \ value \ of \ unity. 

The \ transfer \ of \ a \ data \ element \ from \ one \ matrix \ to \ another \ is 
symbolized \ by \ showing \ the \ receiving \ matrix \ code \ as \ a \ superscript \ to 
the \ issuing \ matrix \ code \ (e.g. \ the \ transfer \ of \ the \ j^{th}, \ k^{th} \ elements 
from \ (X_m) \ to \ (Y_m) \ would \ be \ (X_{mj}^{Ym})_k). 

Fixed \ costs \ are \ associated \ with \ a \ function \ by \ reference \ to \ the 
identification \ element \ (j,0) \ and \ shown \ as \ superscript \ to \ "F" \ (e.g. 
the \ fixed \ cost \ of \ collection \ function \ Cc \ would \ be \ F^{Cc}, \ the \ matrix 
element \ being \ always \ understood \ to \ be \ (j,0)). 

Variable \ costs \ are \ associated \ with \ the \ matrix \ element \ and \ the 
function \ and \ shown \ as \ superscript \ to \ "V" \ (e.g. \ the \ cost \ of \ the \ j^{th}, 
\ k^{th} \ element, \ if \ collected \ by \ Cc, \ would \ be \ V_{jk}^{Cc}).
The additional symbols used are:

W  The value index of a report
T  Time constraints
t  Time variables
A  Capacity constraints
a  Capacity variables.

Each of these symbols is related to the appropriate function
or matrix element by superscripts and subscripts.

3. The Objective Function

The objective function is defined as the minimization of
cost. It involves the fixed and variable costs of the functions
and the elements and the value (negative cost) of the reports.
Thus, the objective is to:

\[ \text{Min: } F + V - W \]  

[1]

The fixed costs are to be charged only if the function is
included in the system. The identification element must be con-
strained to assume a unit value if the function is used and zero
otherwise. Thus:

\[ F = \sum_j C_{c, j, 0} \cdot F^{C_c} + \sum_m [M_{m, j, 0} \cdot F^{M_m} + \\
Y_{m, j, 0} \cdot F^{Y_m} + Z_{m, j, 0} \cdot F^{Z_m}] + \sum_p P_{p, j, 0} \cdot F^{P_p} + \sum_{r, j} R_{r, j, 0} \cdot F^{R_{r, j}} \]  

[2]

The variable costs are defined to be zero for the identifi-
cation elements \((j, 0)\) and are otherwise related to the function and
the element. Since the matrix elements must be constrained to be
zero if the element is not processed by the function, and integer values otherwise, the total variable cost is:

\[
V = \sum_j \sum_k \left( \sum_c c_{jk} \cdot v_{jk}^{Cc} + \sum_m [v_{jk}^{Mm} \cdot x_{jk}^{Mm} + x_{jk}^{Mm}] + \sum_p y_{jk}^{Pp} \cdot r_{jk}^{Pp} + \sum_f z_{jk}^{Fz} \cdot r_{jk}^{Fz} \right)
\]

[3]

The report value indices are given as negative costs associated with each report outline. The total cost premium or penalty is, then:

\[
W = \sum_g \sum_r R_{gj} \cdot w_{jk}^{Rgj}
\]

[4]

### 4. The Output Constraints

The output constraints are the fundamental basis of the formulation. Beginning with the requirement that a single report be provided from each set of report outlines, the data elements are traced, by the available alternate routes, to their ultimate sources.

The logical necessity for an integer solution is accommodated in the output constraints. The time and capacity constraints are defined in relation to the precedent output constraints. By this process the desired integer solution for the time and capacity variables can be assured by modifying the continuous solution only for the output variables.²

---

² For reference to the theory and application see, for example,
Report Requirements -- The first condition is that one report must be provided from each subset of report outlines:

\[ \sum R_{gr_j,0} = 1 \quad \text{for every } g. \quad [5] \]

The identification element, \( R_{gr_j,0} \), which represents a single report outline must be zero unless the report is included in the solution, and, if included, must be supported by the "\( h[R_{gr}] \)" data elements that are specified as comprising the report "\( R_{gr} \)."

Thus:

\[ R_{gr_j,0} = 0 \text{ or } 1 \quad [6] \]

\[ h[R_{gr}] \cdot R_{gr_j,0} = \sum_{j'k} R_{gr_jk}, \quad [7] \]

for each \( r \) and each \( g \), where the sum is taken over those \( R_{gr_jk} \) of the set "\( h \)."

\[ R_{gr_jk} \leq 1 \quad [8] \]

These constraints assure that each manager will be provided one and only one report from each subset of report outlines and that the report will be composed of the specific output data elements which were requested.

Preparation Requirements -- All the output data elements in a given report must be acquired from the same preparation function.

\[ \sum_{j} \sum_{k} P_{jk}^{Rgr} = h[Rgr] \cdot P_{j*,0} \] \[9\]

\[ \sum_{p} P_{j*,0} = 1, \text{ for each } g \text{ and } r, \text{ where } j^* \text{ is the } j \text{ of } Rgr_{j,0} \] \[10\]

The identification element of each P must assume a value of unity if the function is used:

\[ P_{j,0} \geq P_{j*,0}, \text{ for every } j^* \] \[11\]

\[ P_{j,0} \leq 1 \] \[12\]

All elements transferred from a given function must be available in that function:

\[ \sum_{g} \sum_{r} P_{jk}^{Rgr} = P_{jk}, \text{ for every element } \] \[13\]

Equations [13] assure that all elements used in the reports were available from the preparation functions:

\[ \sum_{p} P_{jk} = \sum_{g} \sum_{r} Rgr_{jk} \] \[14\]

The only elements that must be separately constrained: to integer values are the \( P_{j*,0} \):

\[ P_{j*,0} = 0 \text{ or } 1, \text{ for all } p \text{ and all } j^* \text{ of all } Rgr \] \[15\]

Reproduction Requirements -- Next, the needed elements must be acquired from the reproduction functions.
\[ \sum_{P} P_{jk} = \sum_{m} X_{mjk}, \text{ for all elements.} \] \[16\]

All elements in each reproducer must be transferred either to report preparation, to computation or to disposal functions:

\[ \sum_{P} \left[ \sum_{jk} P_{jk} \right] + \sum_{m} \sum_{d} X_{mjk} = L \cdot X_{mjk}, \text{ for every } j, k \text{ and each } X_{m} \] \[17\]

The \( X_{mjk} \) are in turn provided by the storage functions:

\[ X_{mjk} = M_{mjk} \] \[18\]

The identification element for each reproducer must assume a value of unity if any \( X_{mjk} \) are non-zero:

\[ X_{mjk} \neq X_{mjk} \] \[19\]

\[ X_{mjk} \leq 1 \] \[20\]

The required integer values necessitate that:

\[ X_{mjk} = 0 \text{ or } 1 \] \[21\]

**Computation Requirements** -- The computation functions transform a set "d" of elements received from the reproducers into a new set "d*" to be issued to the memory functions.

\[ \sum_{Xm} \sum_{Ym} X_{mjk} = \sum_{m} \sum_{d} Y_{mjk} \] \[22\]

\[ Y_{mjk} = \sum_{d} Y_{mjk}, \text{ for all derivation sets "d"} \] \[23\]
The set "d" is identified as d[j**] and the j** identification element is used as the transformation control:

\[ d[j**] \cdot Y_{m_{j**}0} = \sum_j \sum_k Y_{m_{jk}} \] for the j, k

in the set d[j**] and each Y_{m}.

[24]

The new set "d*" is identified as d*[j**]:

\[ d*[j**] \cdot Y_{m_{j**}0} = \sum_j \sum_k Yd_{m_{jk}} \] for the j, k

in the set d*[j**].

[25]

The identification elements for each computation function must be equal to unity if any derivation is performed by the function, i.e. if any of the Y_{m_{j**}0} are non-zero:

\[ Y_{m_{j0}} = Y_{m_{j**}0} \text{ for each } m \]

[26]

\[ Y_{m_{j0}} \leq 1 \]

[27]

The integer requirement can be met by controls on the identification elements:

\[ Y_{m_{j**}0} = 0 \text{ or } 1 \]

[28]

**Storage Requirements** The output of the storage functions is defined in equations [18]. The input is obtained from the collection and computation functions or from the managers.

\[ M_{m_{jk}} = \sum C_{c_{jk}}^{M_{m}} + Yd_{m_{jk}} + \sum C_{s_{jk}}^{M_{m}} \], for each m

[29]

The value of the identification element is dependent upon the use of any of the elements:
\[ M_{j,0} \geq M_{m, jk} \text{, for all } m \]  
[30]  

\[ M_{m, j,0} \leq 1 \]  
[31]  

The integer constraints are:

\[ M_{m, jk} = 0 \text{ or } 1 \]  
[32]  

**Collection Requirements** The storage requirements in equations [29] generate the need for collection functions. The final output constraint is that the collection functions must contain only integer elements, all of which are transferred to the storage functions:

\[ C_{c, jk} = 0 \text{ or } 1 \]  
[33]  

\[ C_{c, jk} = \sum_{m} C_{c, jm} \]  
[34]  

\[ G_{g, jk} = 0 \text{ or } 1 \]  
[35]  

\[ G_{g, jk} = \sum_{m} G_{g, jm} \]  
[36]  

**Summary of the Output Constraint Formulation** The preceding equations and the objective function define a formulation constrained only in terms of the elements necessary to provide the desired output set.

Given the specifications of the report outlines; the elements which can be accommodated by the several functions; and rules for derivation; the formulation can lead to several alternative results:

a) An infeasible solution if there is no source for some required elements or if the solution includes
elements which were expressed as being available at
infinite cost. Such infeasibilities might be eliminated by
redefining the data needs or by establishing some manager
as the source of the needed elements.

b) An optimum solution or solutions neglecting
all time and capacity constraints.

The imposition of the rule that all generated
elements must be accommodated in the data system can lead
to a situation in which the elements are processed through
the reproducer and then only to the disposal function.
Such a situation would suggest that a less costly system
could be provided by eliminating the decision data which
is unnecessary.

5. The Time Constraints

The essential concept of the formulation of the time con-
straints is to define a new time variable "t_o", that will reflect
the holding time (in integer units) during which a data element
remains within a function. The elements which do not appear in the
solution are to be treated in such a manner that their holding time
index is zero. The summation of the t_o times then gives a direct
index of the final availability, which can be constrained to a time
limit for each report. Where a distinction is necessary, the nota-
tion t_1 and t_2 is used to represent entry and exit times respectively.

The constraint set for most elements and most functions is
simply a product of a given processing time "t" and the value of
the matrix element. The derivation and report preparation processes
require consideration of the fact that all constituent elements
must be available before the new element or the report can be
produced. This requirement is accommodated through the identifi-
cation elements.

**Initial Availabilities** For each element in C there is a
specified time index $T_{jk}$: after which the element is available, and:

\[ t_0[C_{jk}] \leq T[C_{jk}] \cdot C_{jk} \quad [37] \]

\[ t_1[C_{jk}] \geq T_{jk} \cdot C_{jk} \quad [38] \]

Each element to be generated by a decision is available at
some time "$T_g$" after the supporting report is issued:

\[ t_0[G_{jk}] \geq T[g] \text{ for each } g, \quad [39] \]

\[ t_1[G_{jk}] \geq \frac{1}{2} t_2[R_{jk}] \quad [40] \]

**Processing Times** The time spent in the storage and repro-
duction, functions are:

\[ t_0[M_{jk}] \leq T[M_{jk}] \cdot M_{jk} \quad [41] \]

\[ t_0[X_{jk}] \leq T[X_{jk}] \cdot X_{jk} \quad [42] \]

**Derivation Time** The time required for processing in the
derivation function is defined on the availability of the latest
element used plus the computation time "$T[d\ast]$".

\[ t_2[Y_{jk}] \geq t_2[Y_{jk\ast\ast,0}] + T[d\ast] \cdot Y_{jk} \quad [43] \]

\[ t_2[Y_{jk\ast\ast,0}] = t_1[Y_{jk\ast\ast,0}] = t_1[Y_{jk}] \quad [44] \]

for each $j, k$ in the set $d[j\ast\ast]$
\[ t_1[Y_{dm_{jk}}] = t_0[G_g_{jk}] + t_1[G_g_{jk}] + t_0[C_c_{jk}] + \]
\[ t_1[C_c_{jk}] + t_2[X_{d^*m_{jk}}] + t_0[M_{m_{jk}}] + \]
\[ t_0[X_{m_{jk}}] \]

Equations \([43]\) define the availability time of the derived elements in terms compatible with the other source times, and provide the basis for the determination of the period during which the several elements are held in the function.

The holding times for the elements used in the derivation are:

\[ t_0[Y_{dm_{jk}}] = t_1[Y_{dm_{**}j{k}}] - t_1[Y_{dm_{jk}}], \text{ for the} \]
\[ j, k \text{ elements in the set } d[j**]. \]

The holding times for the new derived elements are:

\[ t_0[Y_{m_{jk}}] \geq T[d^*] \cdot Y_{d^*m_{jk}}, \text{ for the } j, k \]
\[ \text{elements in the set } d^*[j**]. \]

Report Preparation Time. The preparation of a report can begin when all necessary elements are available:

\[ t_1[P_{P_{jk}}] = \sum \bar{t}_2[C_c_{jk}] + \bar{g} \cdot t_2[G_g_{jk}] + \sum \bar{t}_2[X_{d^*m_{jk}}] + \]
\[ \sum \bar{t}_0[M_{m_{jk}}] + \sum \bar{t}_0[X_{m_{jk}}] \]

\[ t_1[P_{P_{j*,0}}] \geq t_1[P_{P_{jk}}] \text{ for the } j, k \text{ elements} \]
in the set \( h[R_{gr}] \) where, as before, \( j^* \) is the \( j \) of \( R_{gr_{j,0}} \).

\[ [49] \]
T[PP] time units later, the preparation function will be ready to prepare the specific report at:

\[ t_1[Rgr] \geq t_1[PP_{j*},0] + T[PP] \cdot PP_{j*},0 \] \hspace{1cm} [50]

And:

\[ t_1[Rg] \geq \sum_i t_1[Rgr] + T[Rgr] \cdot Rgr_{j,0} \] \hspace{1cm} [51]

Finally:

\[ t_1[Rg] \leq T[Rg], \text{ where } T[Rg] \text{ is the imposed time limit.} \] \hspace{1cm} [52]

Reports established to be provided on a "when requested" basis can be treated in several ways. The most rigorous constraint is to require that the allowable reporting time not be exceeded by the full processing time, i.e. that the system is able to provide the report, on time, if so requested at \( t = 0 \).

A more reasonable formulation would require that the necessary data be held in storage until requested. Then the sum of storage access, reproduction, and report preparation times must be equal to or less than the allowed reporting time.

**Summary of the Output and Time Constraint Formulation**  
The inclusion of the time constraints in the formulation introduces additional possibilities for infeasible solutions, and necessitates a substantial increase in the number of constraint equations in the formulation.

In the event that all of the reports in a subset are assigned infeasible time limits it would be necessary to obtain a relaxation of the time bounds or to introduce new sources for the critical
elements, perhaps by providing for acceptance of estimates rather than waiting until the actual events can be observed.

6. **The Capacity Constraints**

The limitations on capacity are assumed to be expressed in terms of (1) the maximum number of elements that can be accommodated by the collection, storage, memory, and duplication functions and (2) the maximum number of reports that can be accommodated by a preparation function, in a given period of time.

There is no defined limit on the capacity of \( G_c \) since there is no option as to the source of decision results.

The formulation of the capacity constraints involves the specification of the time index at which the controlled elements and reports enter and leave the functions and the definition of a set of equations that assume the value zero before entry and after exit and a value of unity during the holding period.

**Entry and Exit Times** The entry times "\( t_1 \)" and exit times "\( t_2 \)" are summarized as:

\[
\begin{align*}
t_1[c_{jk}] & \leq t_{jk} \cdot c_{jk} & \text{[38]} \\
t_2[c_{jk}] &= t_1[c_{jk}] + t_0[c_{jk}] & \text{[53]} \\
t_1[g_{jk}] & \geq \sum \frac{\epsilon}{f} t_2[R_{jk}] & \text{[40]} \\
t_2[g_{jk}] &= t_1[g_{jk}] + t_0[g_{jk}] & \text{[54]} \\
t_1[y_{jk}] &= t_0[g_{jk}] + t_1[g_{jk}] + t_0[c_{jk}] + t_1[c_{jk}] \\
&+ t_2[y_{jk}] + t_0[m_{jk}] + t_0[x_{jk}] & \text{[55]}
\end{align*}
\]
\[ t_2[\text{Ydm}_{jk}] = t_1[\text{Ydm}_{jk}], \text{ for all the } j, k \]

elements in the set \( d \)

\[ t_1[\text{Yd}^*_{mjk}] = t_2[\text{Yd}^*_{mjk}], \text{ for the elements} \]

in the set \( d^* \) obtained from the set \( d \)

\[ t_2[\text{Yd}^*_{mjk}] = t_1[\text{Yd}^*_{mjk}] + T[d^*] \cdot \text{Yd}^*_{mjk} \]

\[ t_1[\text{M}_{mk}] = t_2[\text{G}_{mk}] + t_2[\text{C}_{mk}] + t_2[\text{Yd}^*_{mjk}] \]

\[ t_2[\text{M}_{mk}] = t_1[\text{M}_{mk}] + t_0[\text{M}_{mk}] \]

\[ t_1[\text{X}_{jk}] = t_2[\text{M}_{mk}] \]

\[ t_2[\text{X}_{jk}] = t_1[\text{X}_{jk}] + t_0[\text{X}_{jk}] \]

\[ t_1[\text{P}_{jk}] = t_2[\text{X}_{jk}] \]

\[ t_2[\text{P}_{jk}] = t_1[\text{X}_{jk}] \text{ for the } j, k \text{ elements} \]

in the set \( h[RGr] \)

\[ t_1[\text{P}_{jk*}, 0] = t_2[\text{P}_{jk}] \text{ for the } j, k \text{ elements} \]

in the set \( h[RGr, j*0] \)

\[ t_2[\text{P}_{jk*}, 0] = t_1[\text{P}_{jk*}, 0] + T[\text{P}]. \text{P}_{jk*}, 0 \]

The Loading and Unloading Process The capacity loading variables and unloading variables are defined on the \( t_1 \) of the sequential functions. Thus, \( a_1[\beta_1] \) is a loading variable for function \( \beta_1 \) and \( a_1[\beta_2] \) an unloading variable for \( \beta_1 \).
\[ a_1[\beta_1] - a_1[\beta_2] \leq A[\beta_1] \] for each element in each capacitated function in each time period. \[ 67 \]

Using \( \beta \) to represent any of the capacity limited functions and \( t \) as the time index:

\[
a_t[\beta_{jk}] = (t+1) a_t[\beta_{jk}] - \sum_{\gamma=0}^{t-1} a_{\gamma}[\beta_{jk}] \quad [68]
\]

Where:

\[
\frac{[\beta_{jk}]}{a_t} = (\beta_{jk}) - \frac{t[\beta_{jk}]}{t+1} \quad [69]
\]

The capacity limits, \( A \), will serve to hold down the values of \( a_t \) against the effects of \( \gamma_t \). When the element \([\beta_{jk}]\) is not in the solution, equations \([69]\) are satisfied by zero values of \( a_t \). If \((\beta_{jk})\) is in the solution it will have a value of unity and \( a_t \) will assume the values \( \frac{1}{t+1} \), \( \frac{2}{t+1} \), \( \frac{3}{t+1} \) in periods \( t = t^*[\beta_{jk}] \), \( t^*+1 \), and \( t^*+2 \), respectively.

Then \( a_t \) will assume values as follows in the same periods:

\[
a_{t=t^*} = (t+1) \frac{1}{t+1} - 0 = 1
\]

\[
a_{t=t^*+1} = (t+1) \frac{2}{t+1} - 1 = 1
\]

\[
a_{t=t^*+2} = (t+1) \frac{3}{t+1} - 2 = 1
\]

Summing, in each function, over all the elements and subtracting out the loading of the next function in the sequence will give a period-by-period measure of total loading, which is constrained by the limit \( A \).

**Summary of the Completed Formulation** The inclusion of capacity constraints further complicates the formulation and again
introduces new possibilities for infeasibility.

The essential basis for the capacity limitation equations is the necessary assurance that the elements will be constrained to zero or unity values. These conditions are defined in equations [6], [15], [21], [33], and [35].

The resolution of infeasibilities due to the capacity bounds includes all of the previously suggested relief actions and, in addition, the possibility of direct increase in capacity.

7. The Nature of the Solution

The purpose of the formulation was to demonstrate that the data element approach was justified in the sense that it provided a model which could be expressed in a formulation which is known to possess a solution. Models, such as this one, in which the necessary constraints can be expressed in a set of linear equalities and inequalities and for which there is a linear objective function to be maximized or minimized are known to possess such mathematical properties as to permit determination of the existence of a feasible basic solution and, if feasible, determination of an optimum solution.

The additional restrictions on integer solution values cannot be explicitly accommodated in the linear programming problem and it is necessary to provide a means of revising the solution if any integer restrictions are not met. This requirement is also accommodated within proven theorems on linear programming.

Thus, the purpose of the demonstration is served. The formulation is linear and there exists, if the system design is feasible, an optimum solution.
This formulation of the problem is not amenable to practical application within the present state-of-the-art. Apart from the difficulties in obtaining the assumed cost and relationship data, the formulation contains more equations in more variables than can presently be accommodated by computation equipment.

Despite this complication, the formulation provides assurance that a data system, constructed in accordance with the model, which is found to be feasible, is in fact a basic solution. Any improvements, within the context of compatibility with the model, then provide an evolutionary approach to the optimum system.
CHAPTER V

IMPLICATIONS OF THE THEORY AND THE MODEL

1. **Summary**

The essence of the argument in support of the data element theory encompasses five points:

First, that any data, whatever their character or purpose can be precisely identified by specification of the subject involved and of the substance of the message about the subject.

Second, that the data of concern in the economic decisions of an industrial firm can readily be classified in such a manner as to facilitate identification, analysis and comprehension of their character and purpose.

Third, that the specification of the subject and substance of data elements and the relationship between elements, permits precise identification and consistent manipulation of all the data pertinent to a system problem.

Fourth, that there is an optimum solution to a data system design problem when data is considered on an elemental basis.
Fifth, that the implications of the data system theory and model are operationally meaningful in the sense that they are subject to empirical validation or refutation.

There are, in addition, several ancillary points of especial significance.

There is an explicit basis for distinguishing between a decision and a derivation of data, and a compact way to treat both situations.

There is a close analogy between the data system model and the models of the job -- machine sequencing problems.

It is appropriate to consider the consequences of certain modifications of the initial assumptions and to inquire into the possibilities for finding a good approximation to the optimum design. The first point is significant because of the apparent interrelation of the data system to its environment; the second, because the state-of-the-art in solution of linear programming problems does not encompass the techniques needed to handle the great number of equations and variables implicit in the data system formulation.

A. Environmental Impact

The essential distinction between the data system and the managers which it is to serve has been drawn in terms of a distinction between decision data and derivation data. Any modification of the concept or reality of the managers as pre-existing factors will have an impact on the data system design. There are at least three circumstances in which the preconceived decision duties of managers may be incorporated in the data system as data derivations:
a) Elimination of unnecessary decisions.

In terms of the linear programming model this condition arises when (1) a generated element \( G_{ij} \) is not required in any of the selected reports, or (2) when the same element is obtainable without the need for a decision. The assumed sovereignty of the managers assures that choice of generated elements is always dominant over observed or derived elements, but there is at least the possibility that relaxation of the sovereignty of the managers could lead to choice of non-decision elements.

The real counterpart to the mathematical situation is that of "standards" or other decisions that are not considered in subsequent actions or of simply unnecessary decisions. No basis has been established by which to found a judgement on the desirability of either eliminating the "standard" or of assuring its constructive use. Strictly redundant decisions can be resolved by redefinition of the managers' field of decision making with attendant impact on the system design.

b) Reduction of decisions.

By viewing the purposes which are to be served rather than existing concepts and practices (the methods of service) it may be possible to accomplish the ends by a different set of means -- incorporating some data derivation in lieu of decision-making.
The reduction need not be complete. The derivation rule may well be supplemented by managerial review and optional revision of the derivation.

c) Substitution of decisions.

The third case is that of a true decision requirement -- in the sense that an equivalent rule for derivation cannot be defined -- in which an approximating derivation rule is substituted because the expected increment of advantages by use of a decision is not justified by the additional effort required to obtain the decision. An example would be the determination of the optimum order sequence in a job-shop operation. It is certainly possible that a routing derivation rule -- that is known to be less advantageous than careful managerial sequencing -- may be preferable when the margin of advantage is compared to the real difficulties of making the decision.

On the other side of the problem there is also the possibility that a decision by rule-of-thumb, or otherwise, may be preferable to a derivation that involves a complex or lengthy process.

Each of the preceding circumstances suggests that the preparation of the data system design may lead to changes in the organi-
zation and accounting pattern and functions with consequent redefinition of data needs and system content.

A related area of concern is the correlation of the stated data needs with significant operational data. The alternative means of deriving the output data may provide sufficient flexibility to permit the use of that event data which is recognizable as most meaningful at the source, but, more likely, there will continue to be an impression that much "good" data is ignored by higher echelons.

Although the initial assumptions provide for the sovereignty of the managers in the determination of data needs, it is possible to attack the problem from both sides. If the data inputs resulting from the managers stated needs suggest illogical observation pattern, a "better" pattern and the resultant new report outlines may be proposed to the manager. This procedure will permit both views to be considered. The basic opportunity for disagreement, which arises from the imprecision of the decision process or from conscientious differences of opinion as to the degree to which data actually represents events, may thus provide the basis for a system that is preferable to the one obtained on the assumption of managerial sovereignty.

The degree of environmental impact depends upon the extent to which it is desired to pace organizational and accounting evolution by systems evolution. In an abstract problem of data system design it may be appropriate to assume that the organization and the personnel can accommodate any system complexity, thereby including the problems of organizational and accounting design with that of data system design. In an operating firm there is a conceptually valid
but indeterminate pace at which existing patterns can be modified and different -- perhaps more complex -- data systems accommodated.

Thus, it appears that the immediate attainment of an optimum system design is not only infeasible -- because the design model is too large to be solved by present techniques -- but may also be inappropriate to the speed of response of the institutions and personnel who must use the system. Lacking a once-and-for-all solution, there is greater need for a systematic approach since each problem will necessarily be reviewed and redesigned several times.

B. The Impact of Related Fields of Study

All of the fields of research related to the data problem may contribute to the further clarification of its objectives and the practical application of theories and concepts. Of particular concern are those which deal in problems analogous, or nearly analogous, to those encountered in data system analysis and design.

One of these is of special interest because it involves a similar design problem intimately associated with a data system.

It is the problem of sequencing a series of jobs on a set of machines.

As typically stated, the schedule-sequencing problem involves the following features.

Suppose there are \( m \)-machines of different types \( M_1, M_2, \ldots, M_m \) and \( n \)-Jobs \( J_1, J_2, \ldots, J_n \), each job requiring use of a subset of the machines in some ordered sequence. For example, Job 1 may require processing for varying periods of time on machines \( M_3 \), \( M_5 \), \( M_2 \). In a still more complicated sequence there could be a return to a machine previously used -- for example, \( M_2, M_5, M_2, M_3 \). It is also possible to have
several alternative sequences for the same job.
We shall consider the problem of how to schedule
the jobs on the machines so as to finish all the
jobs in the smallest lapsed time. Other optimality
criteria may, however, be used instead. ¹

The solutions to this problem have been extended to encompass
n-jobs on two machines ² and, in special cases, on three machines. ³

A general n job, n machine solution is not available.

Although the usual objective is stated in terms of minimizing
the time required to complete all the jobs other objectives, including
cost minimization, are equally applicable. ⁴

There is a readily apparent analogy to the linear programming
model of Chapter IV. The collectors comprise one machine group
(more may be added); the storage function is analogous to an in-
process inventory; in the reproducer I = I; the derivation function
is an assembly operation (which can be recycled); and the report
preparation functions accumulate different "orders" to be shipped
to a single "customer".

¹ George B. Dantzig, "A Machine-Job Scheduling Model", Management
Science, Vol. 6, No. 2, Jan 60, pp. 191-196.
² L. G. Mitten, "Sequencing a Jobs on Two Machines with Arbi-
³ S. M. Johnson, "Optimal Two and Three Stage Production Sched-
dules with Setup Times Included," Naval Research Logistic Quarterly,
⁴ Edward R. Bowin, "The Schedule - Sequencing Problem",
In the simplest case all except the collection functions are dropped. Assembly is not usually treated. A series of machines (collectors) can be provided and complete flexibility of the sequence obtained by considering an "explosion" of the machines into a sequence network. Thus a four-machine arrangement with fixed sequence would provide:

\[
\rightarrow A \rightarrow B \rightarrow C \rightarrow D \rightarrow
\]

Exploded, the network for a job which may begin on any machine would appear as in figure 12.

FIGURE 12

The Sequencing Network
The job can enter on any machine and progress in any sequence. In addition to the equations that specify permissible sequences it is necessary to assure that integer coefficients are obtained and that the number of times a job is on a machine is correctly constrained. (i.e. that $A_1 + A_2 + A_3 + A_4 = 1$, etc.)

It is, of course, true that the solution of the equation set for a sequencing problem is no more feasible than that of a data system design unless a severely restricted number of jobs and machines are being considered. The sequencing problem with a few machines and jobs is a more reasonable model than a data problem of the same scale; its rigorous solution may provide some insight that will permit better approximations to the optimum in larger problems.

2. Conclusions

From the demonstration that the data element theory can be meaningfully interpreted in the context of a data system model for the support of the economic decisions of an industrial firm and that the model possesses a theoretical optimum solution it is concluded:

First, that a classification and coding scheme can be devised that will permit any existing or proposed economic decision data system to be reduced to a standard characterization.

---

Second, that such a characterization would disclose patterns of similarity for the systems which are now recognized as having common characteristics, whether the systems so characterized involve manual or machine components.

Third, that a catalog of the essential characteristics of data problems of industrial firms can be developed, thereby permitting preparation of a basic data system design from a knowledge of the objectives and purposes of the firm.

These points suggest an approach to empirical validation of the theory and the model. Each will require a substantial development effort and considerable field work.

The potential returns from improved system design, which can realize the advantages of communication and computational technology suggest that the effort required to reduce a valid theory of data systems to practical application in industry is well justified.
BIBLIOGRAPHY


**Articles and Periodicals**


BIOGRAPHICAL NOTE

Paul B. Henderson, Jr. was born in McKees Rocks, Pennsylvania, November 20, 1928.

He attended schools in Pennsylvania, Arkansas and Missouri, receiving his diploma from the Ritenour High School in Overland, Missouri in 1945 and the degrees of Bachelor of Science in Mechanical Engineering and Bachelor of Science in Industrial Engineering from Washington University in 1948. Continuing with graduate studies at Washington University, he received the degree Master of Science in Business Administration in 1950. During the last year of graduate work in Business Administration and for one year after graduation he was instructor in Production Management and Office Management in the School of Business and Public Administration.

From September 1951 until July 1958, he was employed by the Department of the Navy, Bureau of Ordnance, first as an Industrial Engineer in the Management Control Branch of the Shore Establishment Division and later as Assistant for Management in the Research and Development Division.

He entered the doctoral program of the Massachusetts Institute of Technology, Department of Economics and Social Science in 1958 for studies in economic theory and industrial management.

Since resigning from the Bureau of Ordnance, he has been under contract to the Stanford Research Institute of Menlo Park, California as a consultant in program analysis and data processing.

Upon graduation he will join the Westinghouse Electric Corporation for further work in the design and installation of integrated data systems and in industrial management.


Mr. Henderson, in 1951, married the former Miss Betty D. Langewisch of St. Louis, Missouri. They have three children and currently reside at 5 Bow Street, Arlington, Massachusetts.