C. 1

DATA ADMINISTRATION AND CONTROL:
A FRAMEWORK FOR DESIGN

by

BRIAN L. FINLEY

B.Sc., University of British Columbia, 1965

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF BUSINESS ADMINISTRATION

in the Faculty

of

Commerce and Business Administration

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

September 1973

Department of Commerce and Business Administration

The University of British Columbia,
Vancouver 8, Canada.

Date: September 15th, 1973.

ABSTRACT

Data is an important resource of an organization and is one of the fundamental building blocks of an effective information system. The failure of top-level management to define a framework for information systems and to recognise the potential of the data resource has a serious impact on information systems costs and development.

This thesis attempts to identify some of the problem areas associated with unmanaged data and proposes a framework for the design of a Data Administration and Control System (DACS). Existing data analysis techniques have been reviewed and were found to be inadequate to meet the general requirements for data definition and documentation. DACS, when implemented, will assist in the identification and definition of the data resource, how it is used and where it is stored throughout the organization. It provides a tool to monitor and control the data and to assist in the design of information systems.

DACS has applicability in the growing field of computer-aided information systems analysis and design. DACS itself is an automated approach to the definition of data and its uses. Extensions to the basic design are discussed which would further contribute to the development of computer-aided design tools.

TABLE OF CONTENTS

CHAPTER		PAGE
I	INTRODUCTION	1
II	INFORMATION, DATA AND MANAGEMENT	7
	2.1 Information Systems Development	7
	2.2 The Management of Data	12
	2.3 The Problems of Data Mismanagement	15
	2.4 Integrated Data Banks	22
	2.5 Automated Systems Design	26
III	ELEMENTS OF A DATA ADMINISTRATION AND CONTROL SYSTEM	29
	3.1 Objectives of DACS	32
	3.2 Questions to be Answered	34
	3.3 Requirements of DACS	36
IV	DATA ADMINISTRATION AND CONTROL SYSTEM	39
	4.1 Definitions	40
	4.2 Data Specification Methodology	42
	4.3 Record Specification	54
	4.4 Document Specification	59
	4.5 Data Element Specification	65
	4.6 File Considerations	71
	4.7 System Output Reports	76

CHAPTER		PAGE
V	THE IMPLEMENTATION AND USE OF DACS	84
	5.1 The Data Administrator	84
	5.2 Standards Concerning Data	89
	5.3 Collection of the Data Information	91
	5.4 The Users of DACS	94
VI	SUMMARY	99
	BIBLIOGRAPHY	102
	APPENDIX I	104

LIST OF FIGURES

FIGURE		PAGE
2.1	Information Systems Requirements	10
3.1	Data Element Identification	29
4.1	Data Specification Sheet	43
4.2	Supplementary Specification Sheet	44
4.3	Title Statement for Record and Segment Specifications	56
4.4	Title Statement for Document Specification	61
4.5	Title Statement for Element/Group/Array Specification	66
4.6	Element Function Definitions	68
4.7	Logical DACS Record Structure	72
4.8	Data Dictionary/Directory	78
4.9	Keyword Index	80
4.10	Data Requirements Analysis Report	81
5.1	Information Systems Organization	85

CHAPTER I

INTRODUCTION

"Data constitute the sine qua non of information processing".[1]

Throughout organizations people depend on information to assist them in performing their functions. They use information of various forms and in various ways in strategic planning, management control and operations control.[2] Information from internal and external sources, historical, current and projective, is employed in goal formulation by the strategic planners. It is used in the management control function to assure that resources are used effectively and efficiently in the accomplishment of the organization's objectives and it is used by operations control in the day-to-day operation of the organization.

It can be, and has been, argued, that information systems of an organization are analogous to the nervous system in a human. It consists of stimulii (inputs), processing and responses (output and action). Formal and informal information and communication channels are vital to the success and health of an organization. The information and control systems are a cohesive network that bind the organization together for a common purpose.

---

[1] C.J. Bontempo, "Data Resource Management", Data Management, February 1973, p.31.

[2] R.N. Anthony, Planning and Control Systems, A Framework for Analysis, Division of Research, Graduate School of Business Administration, Harvard University, 1965.

An effective information system is one of the requirements for a successful organization and for this reason, the field of information systems has been receiving considerable attention at the present time and during the last decade.

However, if information is analyzed it is apparent that one of the building blocks consists of data elements. Close examination will reveal that most information and control systems are based on a relatively small collection of these data elements and without these elements there would be no way to meet the information requirements of strategic management, management control or operations control.

Considering the importance of information to an organization, it could be assumed that the data elements which constitute information would be a highly valued and managed resource; one that is considered with the same degree of attention that is given to the other resources such as people and capital.

It is evident, however, from personal observations and perusals of the literature, that too few companies consider data as a valued resource of the organization. Often, they have no established philosophy towards the data that exists within their organization and they have few structures, standards or control mechanisms to ensure that the data resource is used effectively and efficiently throughout the company. In fact, it is often apparent that the data resource suffers from management neglect.

This apparent neglect can cause severe problems for an organization. The information systems may suffer from a lack of integrity and consistency, considerable cost can be incurred in duplicated efforts, opportunities may be lost when data is thought to be unavailable and increased burdens are placed on information systems architects and analysts who are trying to design and implement systems for use across functional areas of an organization.

The subject of "management information systems" has been receiving significant attention by management and systems professionals. As a fundamental requirement in the design and implementation of MIS, there are three major tasks which must be accomplished:[3]

- the information requirements of management have to be identified.
- the data elements which are potentially available must be defined as to technical description, meaning, storage location and retrieval.
- the data relationships among the various data elements must be identified.

The first of these requirements, that management must be able to identify its information needs, is a basic component of an effective information system. An information system can only be meaningful if it reflects and supports the organization's objectives, strategies, policies, and procedures. It is the responsibility of management to determine and define the framework for the development of information systems.

---

[3]R.V. Head, "Management Information Systems; A Critical Appraisal", Datamation, May 1967, p. 23.

In the context of this thesis, it is assumed that upper-level management can and is willing to define this requirement framework and to recognize the potential of information and its underlying data resource.

It is the undertaking and completion of the latter two tasks that will significantly contribute to better data management and control of the existing data resource. If the definition and relationships are accomplished in a disciplined, standardized manner, a strong foundation will have been established on which to build an effective management information system.

It is the purpose of this thesis to examine the role of data as a resource of an organization; to trace the development of information systems with the effect this has had on data and to discuss the problems encountered as a result of data mismanagement. In order to better manage and control the data resource, it is contended that a formal system is required which will identify data elements, where and how they are used and stored in the organization and with whom the responsibility of data integrity resides. For this purpose the requirements for a Data Administration and Control System (DACS) are examined and a proposal for the design of such a system is forwarded.

The scope of this work includes the input and output requirements and conceptual file organization for the Data Administration and Control System but does not include the writing of the support programs to make it operational. It is felt that an in-depth analysis of what is required of a system for the management of the data resource will contribute to the state-of-the-art in the ever growing field of information system analysis and design.

It is emphasized that the Data Administration and Control System is not a file organization technique for the retrieving, analyzing and displaying of data in a data bank. This belongs to the area of data management systems whose integral function is the storage and retrieval of data in support of an information system. The Data Administration and Control System is concerned with identifying the unique data elements themselves, where they are located in the various files, and where and how they are used.

Chapter II examines the development of information systems and the effect on data elements, how data is managed (or mismanaged) in many organizations, the problems encountered because of management neglect of the data resource and the need for better management and control. The philosophy of integrated data banks is also discussed in perspective with data control.

Chapter III analyzes the objectives for a Data Control System, what information it should provide to its users and what requirements are needed for its implementation.

Chapter IV discusses the proposed Data Administration and Control System. Through a Data Specification Methodology the required information regarding data elements is captured. The various types of informational output which must be provided by the system is described as well as a conceptual analysis of a permanent file organization to store and retrieve the data information.

In Chapter V, the role of the Data Administrator, the standards which must be developed and the approach to data collection are discussed. The potential users of the system and how they can employ and interface with the system are also described.

CHAPTER II

INFORMATION, DATA AND MANAGEMENT

2.1 INFORMATION SYSTEMS DEVELOPMENT

The requirement for data processing and the need for relevent information, both internal and external, has always been a characteristic of an enterprise. Before the growth of the presently large and complex organizations a business was composed of relatively few people and operated in a rather static and structured environment. Each businessman conducted most of his own affairs and through personal involvement and observation intimately knew the required information about his business and its environment. His information system was composed of his first-hand knowledge and personal discussion with his employees and people in his market. Each of his decisions was based on a fairly complete knowledge of the situation from the environment, market, economic condition and state of production and he could quite accurately predict the interrelating effects his decisions would have on these components.

As organizations became larger and more complex, communication lines longer, slower and more difficult to maintain, the enterprise began to fragment along functional lines. At the same time, the quantity of data and information requirements increased at a seemingly exponential rate. Data collection, data processing and information dissemination became a very real and costly concern. Many manual systems and procedures were developed to meet these increasing problems. The first systems were based on accounting transactions and the data base consisted of journals, ledgers and accounts organized in specialized recording and filing systems to serve selected purposes. Additional systems began to develop in the other functional areas such as marketing, production, and personnel. A characteristic of these manual systems was that they generally served one functional area and one level in the organization. It was very difficult to obtain information which crossed functional or vertical lines from these subsystems.

The advent of computers in the early 1950's was considered by many to be the panacea for the data processing and information problems which had arisen.

The first systems to be automated when computers were introduced into business were those of high volume and high clerical activity. Such activities as payroll, invoicing, order processing and general ledger were systems that could easily be identified as to their requirements and could be justified on a cost-benefit basis. These systems were designed to replace the manual systems and in scope did not offer much more than the manual systems except faster processing and lower cost. Each system was still independent unto itself and served a single function and single purpose. These systems were implemented at the operational level of the firm and did not have included in their design any integration with other operational systems horizontal in the organization, nor did they provide any significant amount of vertical information to higher levels of the organization. Each of these systems had their own input, files and output and the data these systems captured and processed were used exclusively for that system.

With the increasing capability of hardware and software, systems designers recognized that the proliferation of single function systems might be profitably co-ordinated into more comprehensive systems. For example, it was desirable to integrate the payroll function, the employee benefits function, the distribution of costs to the accounting systems and the personnel records keeping function into a cohesive manpower system. The co-ordinated systems often, however, fell short of their objectives. This resulted mainly due to the fact that the individual systems maintained their single function approach and uniqueness and data was passed through each system and processed where it was required. Each system retained its own files with the data organized as required. In addition, the systems were usually agreed to and developed by functional area management with little involvement by top management and their impact was primarily at the operational level.

Today, the great majority of computerized systems maintain the characteristics of the earlier manual systems. These systems were developed independently over an extensive period of time, have little regard for future developments and concentrate on limited functional areas of the company controlled by the same management that had controlled the non-mechanized system. The context and structure of a specific system and its data is well understood by its users and system support personnel. However, these data generally exist isolated from the rest of the organization and knowledge about them is very limited. Other users are ignorant of their existence, location or the fact that the data may be exploited by them. The bottom portion of figure 2.1, illustrates some of these closed functional systems which have been developed.

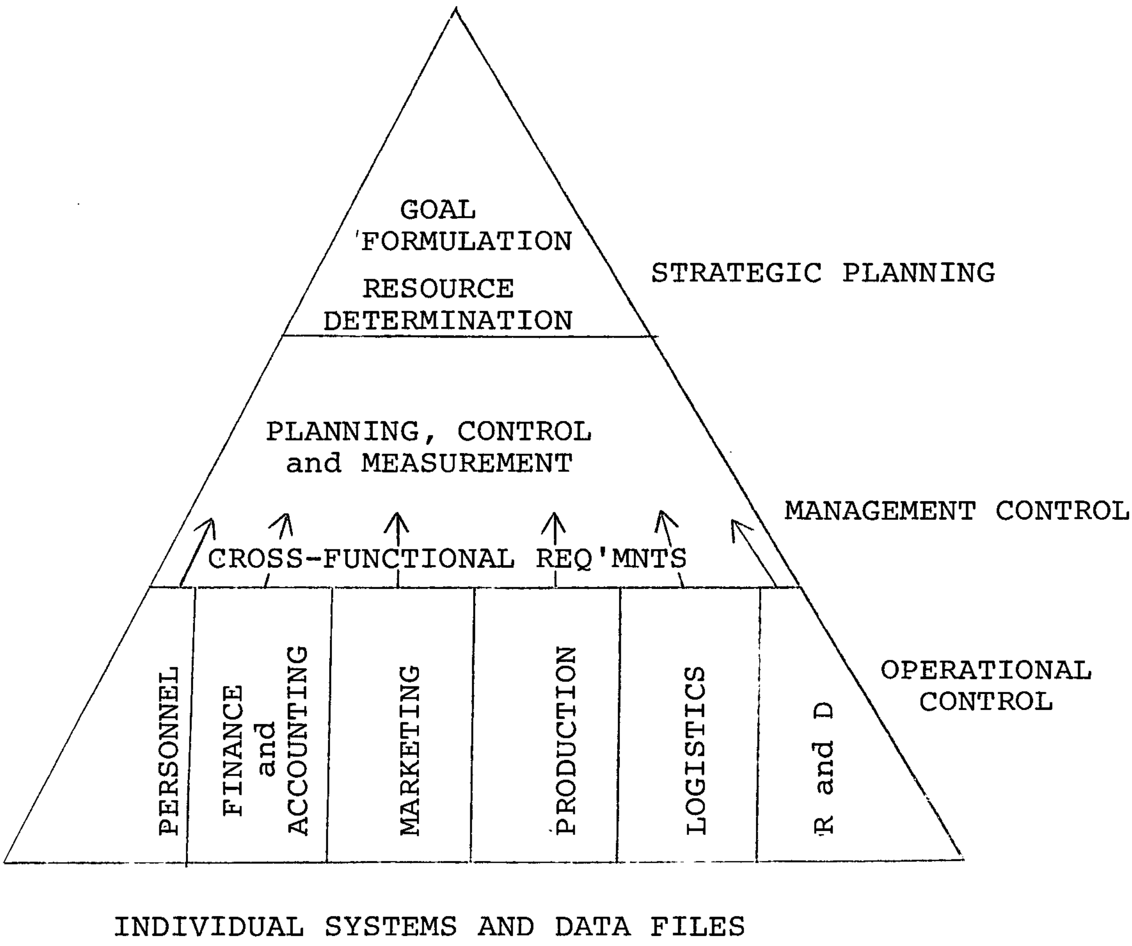

Figure 2.1
Information Systems Requirements

At the present time there is great appeal to the "integrated systems approach"; to develop information systems which would serve not only the functional areas but would have the capability to cross the vertical and horizontal lines of figure 2.1, and serve the entire organization.

There is a realization that the data elements existing in various subsystems can be profitably exploited to meet the information requirements of management control and at a higher level, to aid in the policy setting and decision making functions of strategic management. Often the higher level information requirements are of an "undefined" and demand nature which require as one component, a clear, definitive description of the available data resources.

Therefore, a common means for describing data that would lend itself to a common understanding across divisional, functional and vertical lines is required. The problem of uniformly defining the meaning, structure and use of data located within individual data banks or files of various information systems and of providing a vehicle to facilitate common understanding becomes apparent. The desire to develop integrated systems must be based on sound understanding of what is required, what is presently available, what the real problems are and the economies involved.

One of the first steps that is required in the quest for better information systems is the adoption by top management of a concern and philosophy towards the data in their organization, their recognition that data is a very valuable resource and that a means is required for its administration and control.

2.2 THE MANAGEMENT OF DATA

When data is viewed as a primary and very valuable resource of an organization, it is readily apparent that it suffers from a lack of attention and proper management. This is highlighted when the attitude towards the other valuable resources - people and capital, is considered. With these resources, their value to the organization is illustrated by the concern and constructive management afforded to them. Well established tools, techniques and procedures are implemented in an effort to obtain, allocate, utilize and monitor these resources in such a way as to maximize their utility.

It is easy to assume that once the crucial role that data serves is recognized, similar management practices and attention would be directed towards data as it is given to people and capital. However, data resources are left virtually unmanaged. It is a rare manager who has ever given serious thought or consideration to the data resources. He is concerned with informational content in the form of reports and displays and neglects the fundamental requirement of administering the underlying data elements. This task is left to the analysts and programmers who generally are concerned with individual applications in separate functional areas. Frequently, there is very little communication between analysts when identifying data elements in order to avoid unnecessary redundancy and duplication and to ensure the data resource is utilized for the optimum benefit of the entire organization.

No conscientious manager would allow his personnel or capital resources to go unmanaged, yet the dangers and problems involved in allowing data to suffer neglect have not been properly recognized. The complexities of modern data processing and information systems, multi-functional systems, on-line processing and dynamic user requirements, all compound the far reaching problems associated with the unmanaged use of the data resource.

The quantitative value of information and its contribution to decision-making is very difficult to determine. It is even more difficult to allocate a value to the data element components of information since allocations must also be made to other components such as programs and processes. However, the identification of the costs associated with data may be less difficult to quantify. Jarvinen[1] has analyzed these cost factors and summarizes them as:

- design of files
- programming of file manipulation programs
- generation of files
- maintenance of files
- data processing
- required or reserved space
- outer properties
  i.e.
  - frequency of input and output
  - security and protection
  - query capability
  - equipment

---

[1] P. Jarvinen, "Design of Information Systems", Computer-Aided Information Systems Analysis and Design, Bubenko, Langefors and Solvberg (eds.), p.83.

No attempt is made in this thesis to quantify either the cost or the value associated with data. Generalities can not be made as each situation must be evaluated in its respective environment. However, certain problems associated with data can be highlighted. These problems have frequently been observed in the author's personal experience and they are discussed on a general level to avoid presenting examples which may be interpreted as a fabrication.

2.3 THE PROBLEMS OF DATA MISMANAGEMENT

The mechanization of previous manual systems, the continuance of the functional and single application approach to new systems design and the failure of management to provide direction and guidelines has had serious impact on the data resource. It has frequently led to an uncontrolled data environment which has associated problems and costs.

Data Fragmentation

A severe problem arising from the development of independent information systems to meet the limited functional areas of an organization is the fragmentation of data resources. Data is introduced on an "ad hoc" basis to meet the requirement of each of the application areas. The primary mission of each application is to collect, process and disburse the data resources necessary to satisfy the information needs in its own area. Thus, each of these application areas or systems tend to introduce, process and maintain data in their own sets of files without regard for the needs of the other systems. These other systems, of course, must meet the pressures exercised by their users and will follow suit in their data utilization practices. Applications will also sometimes introduce data without determining first whether they can capitalize on the fact that other systems may be using the same data and have it already in a form which may be useful.

Once this cycle of fragmentation, duplication and proliferation of data throughout the various systems has begun, there is a lack of control over the data resource which leads to several problems.

Data Duplication

With each application area or system processing and maintaining its own data in isolation from the other systems, identical data elements become distributed over many files. This not only leads to additional costs of storage for the data elements but more important, it results in the duplication of effort in the capture, processing and maintenance of the data elements. The degree of duplication is often not known and management is unaware of any data or processing redundancy. However, there are circumstances in which it is desirable to maintain and process duplicate data. This is the case when the effectiveness of the information system would be impaired or the processing cost would be increased in an effort to eliminate the duplication. What is required in order to determine when duplicate data is, in fact, redundant, is a clear description of the data resources, the uses to which these data are put and the cost of capturing, storing and processing these data. Only then can an intelligent decision be made on the costs and benefits of duplicate data.

In addition to the costs involved, the maintenance of duplicate data also gives rise to the problem of data inconsistency.

Data Inconsistency

Since the same data elements may exist in several systems, these data elements are subject to the processing requirements of each respective system. A particular data element in one system may be updated on a daily basis while the same element in another system may be updated on a monthly cycle. The data elements thus become out of phase with each other and although each purports to be the same thing they represent the situation at different periods in time. This may present no problem when the information generated is used exactly for the purpose it was intended and designed. However, there is a tendency for information to migrate outside these intended boundaries. The problem often rears its head when personnel from two or more different areas discuss a problem or decision area and discover that their respective information, although supposedly the same, in fact, differs in time.

When the same data is carried in several systems the problem of maintaining the data so it is consistent becomes very difficult. There is always the chance that a transaction which updates the data element in one system does not find its way through all the subsystems to update them accordingly. If this happens, the entire information system suffers an integrity problem from which it is often difficult to recover.

In these cases, management develops an uneasy feeling about the integrity of the information they receive and about the information systems themselves. There has become a lack of control over the consistency of data that is used by various areas and inevitably information is generated which suffers an integrity problem whose source is the same data inconsistency.

Communications

Every organization develops its own vocabulary of terms which it uses in day-to-day communications. Similarly, the various functions and aspects within the organization develop vocabularies and technical languages. If effective communication is to take place, information systems must employ a common language and produce a common understanding. There is a danger with segmented and independent systems that the same terms are used to represent different entities or different entities are referenced by the same names. This leads to misunderstanding, misinterpretation and poor communications between people, especially those from different areas of the organization.

A serious concern arises when information coming from different sources, although identified similarly, has a different meaning and context.

For an information system to be effective there must be a common vocabulary of data and information definition which is understood and provides a non-ambiguous frame of reference for data and information.

Data Isolationism

The fact that the data resides in separate files, serving unique purposes with no central control, makes it difficult for knowledge of the data resource to be communicated throughout the organization. Situations often occur when new information requirements in one area may demand data that is currently unfamiliar to this area, even though the data may be, in fact, resident and available for processing in the files of other existing application areas. Since the availability of this data is unknown there may be misguided responses to the request for the new information. The prospective user may be told that his request cannot be satisfied since the data is not available or erroneous estimates may be made regarding the cost of obtaining the data and satisfying the new requirements. If the information is worth the cost, a new system may be built duplicating the data collection and storage and further contributing to data isolationism.

If the content and location of the data resource were centralized, the first step in responding to new requirements would be to determine whether or not the data already exists in a form that will satisfy the requirements.

The ability of an information system to respond to random requests and changing requirements is seriously hampered by isolated data. New techniques such as RPG and special retrieval languages have done much to lower the interface barriers between the data and the end-users. However, these languages still require knowledge and understanding of the available data and much effort is involved in these tasks. If the data resource is clearly defined these techniques would be much more valuable. The time required to respond to unanticipated requests would be appreciably shortened since the effort to define and locate the required data would be reduced.

Design and Implementation of New Systems

The fragmentation and isolationism of data elements causes analysts and programmers to expend much effort and time familiarizing themselves with the data resources that are required to meet their own program specification. It may be easier for an analyst to design a new system with data collection, files and output, than it is to try to determine if the data already exists. Even if he does attempt to accomplish this through informal communication and analysis, data element definition is so non-standardized, that the may not succeed.

The information systems architect faces an extremely difficult task when he analyzes the organization to determine what information is required where, and who uses what data. Since he is viewing the information system requirements from an overall viewpoint, the fragmentation of data throughout the existing systems may seem like a giant jigsaw puzzle. He would welcome a technique to lend some order to the maze of the data resource, to provide some standardization in the identification of the elements and to shed insight into the relation between data, information and users.

Increasing Data Processing Costs

Total data processing costs have continued to increase in spite of gains in computer hardware efficiencies with the ratio of people costs to equipment costs continually growing larger. Much of this is due to increased personnel costs. However, a significant amount of data processing dollars are spent inefficiently identifying, maintaining and controlling the data resource. The fragmentation of data elements, the resulting duplication of efforts, the data redundancy and the inflexibility of the systems all contribute to the costs of information systems. Analysts and programmers spend too much of their time identifying and defining data requirements to satisfy informational needs. They could be more productive and effective if supported by a system such as DACS.

2.4 INTEGRATED DATA BANKS

The recognition that one or more of the above problems exists with regard to data has often led organizations to consider, or to adopt, the concept of an "integrated corporate data bank" as a panacea for the problems associated with data management and control. All too often data resource management is viewed simply as a technical matter requiring a technical solution. This approach is understandable enough, since there is a strong tendency for people in the data processing field to solve data processing problems with more and faster hardware or more complex software.

Recognition of data problems often causes an over-reaction by the personnel involved and they may hastily agree that their problems can be solved by eliminating data redundancy and creating program independent data banks. Frequently,the approach considered is to incorporate all data resources within the framework of a single "data base management system" (DBMS) which eliminates data redundancy and locates all the data in a central repository.

Bontempo[2] highlights the danger of viewing redundant data as the main problem and warns that duplicate data does not in itself constitute conclusive evidence of redundancy. He suggests that the view that it does so is based on one of two non-sequiturs in the analysis of data problems: "that any two occurrences of the same data constitute a gratuitous duplication or redundancy

---

[2]C.J. Bontempo, "Data Resource Management", Data Management, Vol 11, No.2, February 1973, pp.33-34.

of data".

He continues by saying: "There are circumstances in which it is desirable to maintain and process duplicate data. What is required in order to determine when duplicate data is, in fact, redundant, is an unequivocal, clear description of data resources and of the uses to which these data are put. Without this evidence, any remedies invoked to eliminate redundancy..... can serve merely to compound the original error which is the failure to monitor and control data utilization in a systematic and deliberate way - in the same way we monitor and control the use of other resources".

The view that a single data base management system is the only solution or the best solution to data control, is based on his second non-sequitur: "data resource fragmentation implies a need for actual data integration achievable only by means of a centralized repository of data, i.e. a data base, managed by a complex and elaborate set of programs".

Once this decision is made, management has committed, perhaps unconsciously, to integration and centralization as their data processing goals and the cost involved is high. In addition to the costs of the software, they must re-educate analysts, programmers and users, add new systems support, restructure and convert data and reorganize the entire systems flow.

Unfortunately, the approach to a DBMS is often taken very irrationally and without the required concern for planning, analysis, selection and implementation. There is no doubt that the data base management system approach and the investment it requires can well be worth its cost and effort when implemented in the proper manner, in the right environment and with the necessary management involvement. Very often, however, data banks are constructed with the goal of data processing efficiency and the models stress input data format, flows and files with little attention to the specific output requirements of the users. There is a tendency, as well, to approach the DBMS with the same philosophy as with previous designs; that is to concentrate in functional areas with limited scope without consideration at the outset for the requirements and general needs of the entire organization. The concentration is often on the techniques structuring the data in hierarchies, networks, chains, lists, rings etc, and too little attention is paid to the information content of the data, where it is used, and the need for control.

It is often in such cases that the data residing in data banks is no better understood, is not better managed or controlled and as Dearden[3] observed, does not provide for the expected increase in the quality and value of information.

The decision to implement corporate data banks, centralized or distributed[4] is not one to be considered lightly or naively. A considerable amount of effort, planning and control must be

---

[3]J. Dearden, "MIS is a Mirage", Harvard Business Review, January-February, 1972, p.96.

[4]C.H. Kriebel, "The Future MIS", Business Automation, June 1972, p.42.

undertaken by all levels of management to ensure that the implementation meets the requirements, produces economical returns and is justified through better and more timely information with reduced costs of new systems.

Management, when considering solutions to its data problems, should regard its primary objective as the need to monitor and control its data resource rather than the integration of the data via a DBMS.

What is needed is a tool by which management can analyze the data, determine if, in fact, problems do exist and thus develop evidence on the basis of which rational decisions can be made with regard to the remedies required and the approach most suited to its particular environment.

In the succeeding chapters of this thesis a proposal is forwarded for a Data Administration and Control System (DACS) which serves as such a tool. The implementation of this system can provide management with a vehicle which serves not only as an aid in identifying data, its uses and associated problems, but which, by itself, may be adequate to meet the objectives of sound data management and control.

If, on the basis of rigorous analysis and planning, a DBMS is considered as an economical approach to data management, DACS will provide valuable assistance in defining existing data as to location, characteristics, relationships and use. It may be used to advantage in improving the order, uniformity and discipline of the data before and during the implementation of the Data Base Management System.

2.5 AUTOMATED SYSTEMS DESIGN

The analysis and design of computer-based information systems is generally characterized as a manual process. From the time of an initial request for a new or modified system, to the completion of a set of detailed specifications for programs, hardware, output and processing requirements, there are few tools or automated processes to assist the analysts and designers. Recently, however, there is increasing research under study to develop automated procedures to assist in the various design phases.[5,6]

One of the most ambitious projects in this field is the development of the Information System Design and Optimatization System (ISDOS) by Teichroew. This system is envisaged as encompassing the entire design process from the initial specification of the requirements via a Problem Statement Language (PSL), through to the structuring of the data and the production of the object programs. One of the modules of ISDOS is the "Data Re-organizer" which accepts specifications from various other modules to structure the data in the form required. In order to accomplish this phase, a "meta-databank" is needed which describes and defines data as they exist in the various data banks and files of the information systems. The Data Re-organizer then interrogates this meta-databank or directory to map the logical data requirements on the

---

[5]R.V. Head, "Automated System Analysis", Datamation, August 15, 1971, p.22.

[6]D. Teichroew and H. Sayani, "Automation of System Building", Datamation, August 15, 1971, p.25.

existing data base and to indicate any missing or incomplete requirements.

CODASYL[7] and Guide-Share[8] have published proposals describing the requirements for such a directory as a means to define the physical and logical characteristics of a data base and to act as an interface between the actual data, and the users and programs which employ these data. This interface, although designed to facilitate the definition and retrieval of data, could also serve as the data directory required in techniques for automated systems design.

The Data Administration and Control System (DACS) as presented in this thesis could be extended to provide a facility to meet the requirements of Codasyl's Data Description Language and Guide-Share's Data Base Descriptive Language (DBDL).

As DACS goes beyond the identification and definition of the data as they exist in the data structures in the various systems and also includes the identification of data elements as they are captured and employed throughout the organization, it may be extended to provide automation to the process of file and data base design. The automated analysis of the data elements, how they are used, their time and retrieval requirements, could conceptually result in the definition and structuring of the data base itself.

---

[7]CODASYL Systems Committee, "Feature Analysis of Generalized Data Base Management Systems", May 1971.

[8]Joint Guide-Share Data Base Requirements Group, "Data Base Management System Requirements", November 11, 1970.

A system such as DACS could serve as the nucleus of one of the components of Automated Management Information Systems[9]. It would accept the data specifications from the Problem Statement Language, scan the data directory for the required data and construct logical and physical data structures.

If the design of the data base can be automated and there is no reason why it can or will not be done, a significant step towards automated MIS will have been taken.

[9] H.J. Will, "A System of MIS Concepts", p.12, June 1973.

CHAPTER III

ELEMENTS OF A DATA ADMINISTRATION AND CONTROL SYSTEM

A data processing or information system conceptually consists of input data being collected and maintained in files and this data in turn being manipulated by procedures to produce the required output and information. The building block or raw material of these systems are data elements. They exist in the input, files and output phases with the data collection and processing programs massaging and organizing them to meet various requirements. (Figure 3.1). It is with these data elements that the Data Administration and Control System is concerned.

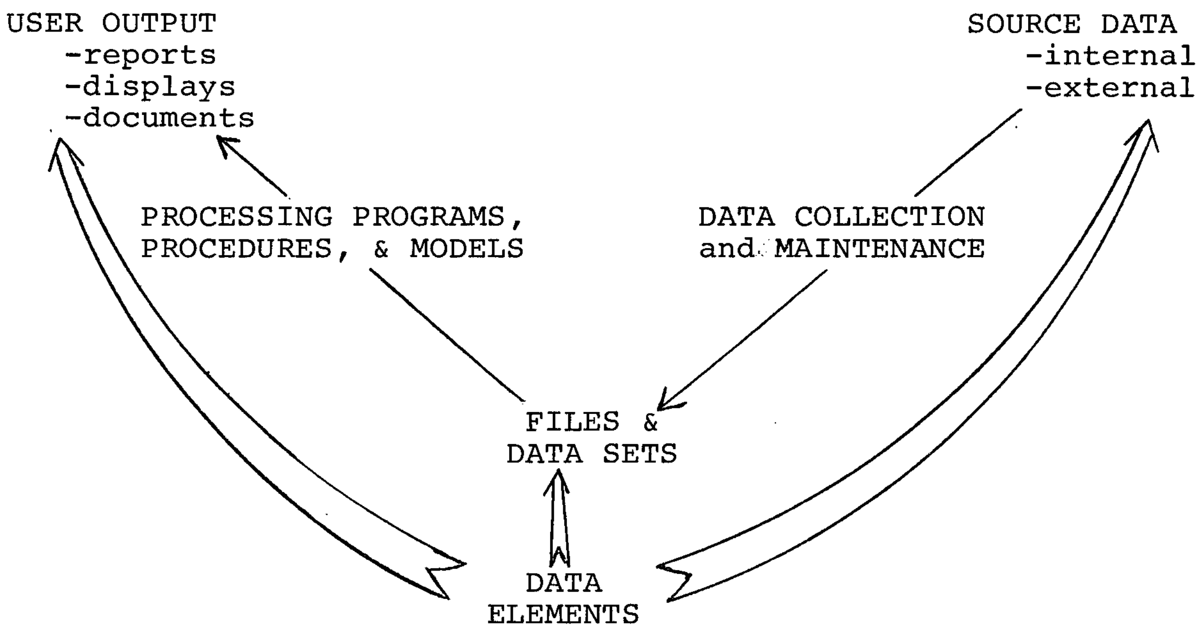

Figure 3.1

Data Element Identification

In considering the design for DACS, other data analysis techniques were reviewed. These included AUTOSATE and TAG (see Appendix I). AUTOSATE was one of earlier techniques developed and its primary emphasis was on the documents existing in the organization and the flow of these documents between the various stations. Secondary importance was given to the actual data elements which comprised these documents and the files in which they were stored. AUTOSATE's emphasis makes it difficult to analyze the data elements in order to identify redundancies, to gain insight into the uses of these data and to structure them into more effective data organizations. The main emphasis of DACS is on the data elements; employing document analysis to determine where and how the data are used.

The TAG technique uses the input and output documents of an application in order to develop data requirements and file structures for each of the applications. TAG, however, does not have a facility to describe data as they currently exist and, moreover, has limited ability to analyze data as they are used across application boundaries.

For these reasons, it was determined that a system could be developed which would concentrate on the data elements as they currently exist, could have the generality of analyzing these data in all their locations and use and could also have the flexibility of being able to determine the impact of new requirements on existing data structures.

It is recognised that the processing programs, procedures, models, and the data collection and maintenance processes are an integral part of the information system (figure 3.1). They must be included in a comprehensive tool for the design and documentation of information systems. The emphasis of this thesis and the design of DACS are on the data elements and for these reasons the definitions of the processes are beyond the present scope. However, it would be feasible to extend the facilities and incorporate the ability to include process and program definitions as well. If this were done, the definition of the total information system could be rigorously documented and automated.

3.1 OBJECTIVES OF DACS

The basic objective of DACS is to provide a tool to assist management in the planning, monitoring and controlling of the organization's data resource. It accomplishes this by identifying the data elements as they exist in the present systems, where they originate, where they are stored and how and by whom they are used. It provides a knowledge about the data resource as it is employed throughout the organization. Where data is shared among various systems and across organizational or functional lines, DACS serves as a central repository of information about business and operating data for all persons in the organization.

DACS uniformly defines the meaning, structure and use of data, provides a vehicle for common control, communication and understanding and assists in improving the order, uniformity and discipline of the data resource.

The centralized control of data element definitions, locations and uses does not imply the integration and centralization of the data elements themselves. Instead, information on data is integrated and centralized and is available through data resource reports to all interested people in the organization. It may be desirable for effectiveness and efficiency to retain decentralized files and to employ the distributed data processing (DDP) approach as suggested by Kriebel[1].

---

[1] C.H. Kriebel, "The Future MIS", Business Automation, June 1972, p.42.

DACS will provide a tool by which management can define the data resource, identify existing problems and provide information which can be used to make more effective and efficient use of this resource.

3.2 QUESTIONS TO BE ANSWERED

To meet the objectives as outlined, DACS must provide answers to key questions about the data for management, operational and technical personnel.

Identification and Definition

1. Is the data element currently available in our system?
2. What is the element's formal name and description?
3. Is it derived from other data elements?
4. What are its technical characteristics?
5. Where is it located for retrieval and use?

Usage

1. What reports use which data?
2. Who are the users and where are they located?
3. What is the frequency of use?
4. How is the data used?

Source

1. Where does this data originate?
2. How often is it received?
3. How often is it processed?

Responsibility and Control

1. Who is responsible for the specification of the data?
2. Who is responsible for its integrity and consistency?
3. Who has authority to update the data?

Processing

1. How is the data structured and accessed for processing?

2. What programs or procedures use what data?

Redundancy

1. Is there unnecessary data element redundancy?

2. Is there unnecessary document or report redundancy?

3. Is there unnecessary file or record redundancy?

3.3 REQUIREMENTS OF DACS

In defining DACS, the following requirements have been considered.

Identification and Definition

Each data element, report, document and record must have a label, a name, a textual and a technical description.

The label is a short and unique identifier used to refer to the data item and to cross-reference it to other data items in the system.

The name is a string of key words which meaningfully identifies the data items and can be used to prepare indexes to facilitate retrieval.

The description is a free-form English text which allows the data to be described in as much detail as required so as to be non-ambiguous and convey a clear meaning and understanding.

The technical characteristics of the data provide information such as length, format and precision .

Location

The permanent location of the data element and how it can be accessed must be specified. If data resides in more than one location, all locations must be specified. This allows the analyst to draw upon data that already exists in processable form and not to propose recapturing the data which may result in possible duplication and redundancy.

Responsibility for Specification and Integrity

It must be determined exactly who is responsible for specifying the data possessed in the various areas of the organization (i.e who "owns" the data) and who is responsible for, or has the authority to, update, change or delete the data.

If one area has the responsibility for maintaining the integrity of a data element and its relationships, the problem of inconsistency is reduced.

Determination of Data Users

The system must include the capability of registering all the users of each data item. This provides insight into the data flows in the organization, how the data is used and the time requirements for it. It also prevents the change or deletion of data items until the impact on all users has been considered.

Determination of Data Source

In order to efficiently capture source data and to avoid unnecessary duplication it is required to know in what manner and from what source the data items originate.

Data Duplication and Redundancy

Identical or very similar items of data occurring in different parts of the organization indicate potential areas for concern. There may be duplication in files, source documents or output information. Detection of these cases is necessary so each may be studied in detail. Those responsible for the data items involved must have a tool to enable them to determine if inconsistency or duplication actually exists. If such cases are found, there must be an ability to determine whether such constitutes redundancy or is desirable.

Retrieval of Data Information

Users must be able to selectively retrieve the items of data that are of interest to them and review or analyze only the specifications or relationships of those individual items. This requires the support of an appropriate data processing facility so that the characteristics of each item may be simply and selectively analyzed.

If the above requirements are satisfied through the Data Administration and Control System, the framework for controlling the data resource has been built. Efforts must be made to review the system output, define cases of inconsistency and redundancy and determine where effectiveness and efficiency can be improved.

CHAPTER IV

DATA ADMINISTRATION AND CONTROL SYSTEM

The Data Administration and Control System (DACS) is designed to satisfy the requirements which have been identified in the preceding chapter. It is conceptualized as a computer supported system which has the following components:

- A Data Specification Methodology (DSM) which facilitates the definition and specification of the identity, technical and relational attributes of documents, records and data elements.
- A file structure to facilitate the storage and retrieval of the data specifications and their related information.
- A series of output displays or reports which enable the user to analyze and control the data resource as to its identification, terminology, usage and relationships.

In developing DACS, the underlying concept is that the basic units of information are data elements and it is on these elements and how they are combined and used that the system is built.

The definitions used throughout DACS are common IBM definitions and closely correspond to Guide-Share requirements. As there is no universally accepted terminology at the present time for data definition, the author has attempted to provide keywords which will best represent data concepts. CODASYL has attempted to outline a data definition language and several of their terms correspond to DACS definitions. Explicitly, DACS definitions for element, group, segment, record and file correspond to CODASYL's data item, group, entry, record and file. CODASYL definitions do not include document specifications and, therefore, there is no correspondence in this area.

4.1 DEFINITIONS

The basic levels of data definition are the element, group and array.

The element is the smallest unit of data specification and cannot be separated into smaller components, i.e. month.

A group consists of two or more data elements which can be considered as one logical entity or unit, i.e. date is a group made up of the data elements day, month and year.

An array is a series of recurrences of data elements or groups which are identical in their meaning or attributes, i.e. a series of amounts showing sales of products by month is an array.

Elements, groups and arrays can be combined into segments and records for storage in an automated file or into documents for operational or informational purposes.

- A segment is a portion of a data record containing one or more logically related data elements, groups or arrays. A segment is typically the smallest portion of data which can be physically retrieved from a file.

- A record consists of one or more logically related segments, elements, groups or arrays, i.e. an employee record contains all data pertinent to that employee.
- A file or data_set is a collection of logically related records, i.e. an employee file would contain all records for all employees.
- A document is a combination of data elements, groups and arrays put into such a form as to be humanly understood and usable for informational or operational purposes, i.e. invoices, purchase orders and reports are documents.

4.2 DATA SPECIFICATION METHODOLOGY

In order to capture information about data, its attributes and its use, DACS uses Specification Sheets and Specification Statements to describe the data. Information about the data being analyzed is entered onto the preprinted Specification Sheets by completing the appropriate Specification Statements. There are two Specification Sheets that are used to gather information about data; the Data Specification Sheet and the Supplementary Specification Sheet. The latter is used when there is not sufficient room on the first sheet to complete the data specification. Sample Specification Sheets are shown in figures 4.1 and 4.2.

Specification Statements

The specification or description of data to be included in DACS is accomplished by means of Specification Statements. Each short, concise statement communicates specific characteristics of the specified data item. The Data Specification Methodology provides three sets of Specification Statements; one set is designed for elements, groups and arrays (E/G/A), one for documents and the other for records and segments (R/S).

The Specification Statements for each of the data entries fall into five general classes:

1. Label: The label, a short and unique tag, associated with each item of data in DACS, is intended primarily for computer manipulation and retrieval.

ACTIV		DATA LABEL							SEQ.		
1	2	3	4	5	6	7	8	9	10	11	12

DATA ADMINISTRATION AND CONTROL SYSTEM

DATA SPECIFICATION SHEET

PAGE ____ of ____

Columns 1-12 to be punched into each card.

CARD (13–14)	FILE LABEL (15–21)	VER (22–23)	STAT. (24)	PREPARED BY (25–39)	PREP. DATE (40–45)	CERTIFIED BY (46–60)	CERT. DATE (61–66)
01							

CD (13–14)	ENTER SPECIFICATION STATEMENTS AND TEXT BELOW (15–79)
02	
03	
04	
05	
06	
07	
08	
09	
10	
11	
12	
13	
14	
15	
16	

Figure 4.1 Data Specification Sheet

CARD		DATA LABEL							SEQ.		
1	2	3	4	5	6	7	8	9	10	11	12

DATA ADMINISTRATION AND CONTROL SYSTEM

SUPPLEMENTARY SPECIFICATION SHEET

PAGE ___ of ___

Columns 1-12 to be punched into each card.

CD		ENTER SPECIFICATION STATEMENTS AND TEXT BELOW																																																																
13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79

Figure 4.2 Supplementary Specification Sheet

2. Name: The name consists of a number of keywords assigned to provide descriptive information about the data. It is the human identifier for a piece of data and must be unique. The name appears in the Keyword Index report and provides a means for identification and retrieval.
3. Description: The description includes a free-form text which unambiguously describes each item of data and provides additional information deemed necessary for clarity.
4. Technical Data: Technical data, mainly of interest to the programmers and analysts, includes such information as the length, precision and mode of a data element. It also includes access method and language processor in the case of a mechanized record.
5. Relational Data: Relational data refers to such information as the ordering of a data element within a record, the use of a unit of data such as an element by another entity such as a record or a document, the source and destination of the data and the use of the data in the organization

The DSM uses the name and label in conjunction to achieve selective retrieval capability. If an analyst knows the label of a particular item of data, an element, record or document, he may reference the label for information on that item. If he does not know the label associated with the item of data, he may consult the Keyword Index which is ordered by the keywords of the name, to find the data items and labels he is interested in.

He may then use the label to refer to the data specification he wishes to study.

The Title Statement, the first statement on the Specification Sheets, is required on every specification and is the only one which has a fixed positional format. It must always be completed for every data specification.

The remaining specification statements are essentially free-form and are designed to permit as much flexibility as possible to the person describing the data. These statements have three components: the operator, the delimiter and the operand.

The operator is a keyword which identifies a single attribute of the data such as the mode, the frequency or the access method. Full words are used as the operators in the Specification Statements, however, the first four letters may be used for brevity.

The delimiter follows the operator and is always an equal sign (=).

The operand is the information supplied by the analyst and is the actual data which will be entered into DACS. The operand may consist of one or more values and the interpretation of the values is determined by their position within the list. Commas (,) are used to separate the values in a multi-value operand list.

A Specification Statement is always terminated by a semi-colon (;).

Other than the Title Statement, the Specification Statements may be completed in any desired sequence on the Specification Sheets.

Title Statement

The Title Statement, the first line of the Specification Sheet, identifies to DACS the label of the data item to be submitted to the system and the type of processing to be performed. Data items can be added to DACS, have existing specifications changed or be deleted. It also provides for the inclusion of administrative information such as the name of the analyst specifying the data and the date of specification. The Title statement is recorded on the Specification Sheet in a fixed format and must always be completed for the data specification to be accepted.

1. Activity Type (positions 1-2)

This field indicates whether the specification is for a document, record or element/group/array and how the specification is to be processed.

Position 1 can contain one of:

E - to indicate the specification is for an E/G/A.

R - to indicate the specification is for a record.

D - to indicate the specification is for a document.

Position 2 contains one of:

B - to indicate a new specification is being entered.

C - to indicate an existing specification is being changed

D - to indicate a specification is being deleted.

In building a specification, the Title Statement is followed by the remaining Specification Statements required to specify the data item.

To change a specification, the Title Statement need only be followed by those statements required to effect the desired changes.

In deleting a specification, only the Title Statement is required to remove the entire specification of the data item from the system.

2. Data Label (positions 3-9)

The label is created by the analyst at the time the specification is prepared. It may contain up to seven alphameric characters and serves as the unique identifier for the particular E/G/A, record or document. It is used for machine manipulation, retrieval and appears in several reports.

3. Sequence Number (positions 10-13)

The sequence number consists of three numeric characters and is used when completing E/G/A specifications and indicates the relative position of the E/G/A within the record or document.

4. Card Number (positions 13-14)

The card number is printed on the Specification Sheets. Its purpose is to sequentially number the statements for a given data specification.

5. Record or File Label (positions 15-21)

This field serves one of two purposes depending if an E/G/A, a record or a document is being specified. When specifying an E/G/A this field contains the label of the record or document on which the E/G/A appears. When specifying a record, the file or data set to which the record belongs is placed in this field. For a document this field is left blank, or if a manual file is used to store the documents, its name can be placed here.

6. Version Number (positions 22-23)

The version number consists of two numeric characters and is required when specifications differ for the same item. It is used when multiple usages create a need to show different characteristics of data.

7. Status (position 24)

This field shows the status of the item at various phases.

P - indicates a Proposed status and means that the addition, change or deletion of a data specification is being proposed to meet a requirement. The proposed specification is reviewed by the data administrator for accuracy and is entered into the system. Any affected users must be notified of the new specification and their agreement obtained. Any conflicts with other items in the system must also be identified and corrected.

A - indicates an Approved status where the proposed specification has been agreed to by all concerned. The specification is now formally entered into the system and is available for testing but not for formal use.

E - indicates Effective status and the item is now available for use. All changes affecting various systems throughout the organization have been made and the new specification is permanently resident in the system.

8. Prepared By (positions 25-39)

This field contains the initials and last name of the analyst preparing the specification.

9. Date Prepared (positions 40-45)

This field indicates the date the specification was prepared.

10. Certified By (positions 45-60)

This field contains the name of the manager of the department who is primarily responsible for the data. New or changed specifications must be approved by him before entry into DACS.

11. Effective Date (positions 61-66)

This date indicates the date at which the status of the item becomes effective.

Name Statement

NAME = keyword string;

The Name Statement provides a humanly sensible identifier for the data item, E/G/A, record or document. Each item must then have two identifiers; the Label and the Name. The Name gives the data item an identifier which indicates the use of the item. It provides a retrieval mechanism for the user who is able to describe the definition being sought but does not know its Label.

The structure of the Name consists of a series of keywords, chosen by the analyst, which will describe the item at the time of specification. Each of the keywords used must be highly definitive and easily understood to insure effective usage.

When building the Name, the analyst should consider which is the prime word, i.e. the most general term that can be used to describe the item. He then should consider the modifiers or next general terms etc, until he has completed the least general or most specific term.

The Name is the data specification which is input to the Keyword Index. In order to make this report as complete and useful as possible, the analyst should be very conscientious when he constructs the Name for the data.

The Name Statement should be the last statement completed by the analyst. The creation of an effective Name is dependent on how well he knows the data and can identify the most descriptive keywords. During the process of completing the other Specification Statements he will gain insight into the data and its characteristics which may be helpful in identifying keywords for the Name.

Description Statement

The Description Statement supplies DACS with the textual description which unambiguously describes the E/G/A, record or document. The analyst should compose a complete, clear and concise Description Statement. He should avoid abbreviations, mnemonics or obscure phraseology which may confuse a user who is not familiar with the data. He may include in the Description Statement, any information which he judges to be relevent and is not included in any of the Specification Statements. The structure of the Description Statements takes the form:

DESC = text;

The statement is begun in column fifteen and, if required, can be continued on succeeding lines. The first character of any continuation line must start in column sixteen or after.

Reference Definition

Before proceeding to identify a data element, it should always be determined if the element already exists in the system. This is done by checking the Keyword Index to see if there are any similar elements using the associated Labels and determining if the element has been defined. If it has, a Reference Definition can be used to simplify the repeated specification of the data. If the element to be specified is identical to an existing one, only the Title Statement need be completed giving the same Label and Version number. If the element differs in some of its attributes, a new Version number is assigned and only the attributes which differ from the original entry need to be supplied.

Origin, User and Responsibility Codes

In completing the Source, User and Responsibility Statements, it is required to identify a person, function or department in the organization. The Source Statement identifies the originator of a document, the User Statement identifies who uses the document and the Responsibility Statement determines who is responsible for the content and integrity of the data.

The code may be structured to suit the requirements of the specific organization. The structure of the code, therefore, must allow for the clear identification of the location and user of the document or the person reponsible for the data integrity.

A possible structure may consist of division, function, location, department and position numbers. An organization chart may be helpful in determining the code and, in fact, a classification structure may already be in existence. Schruben[1] suggests that this code can be similar to an accountant's chart of accounts or budgetary system.

Although the formalization of these codes may entail considerable effort in specific organizations, they are a necessary requirement for a successful implementation of DACS in each case.

---

[1]L. Schruben, "The Information System Model", Datamation, Vol.15, No.7, July 1969, p.93.

Data Specification Sheets

The methodology followed to complete the data specifications consists of first completing record or document specifications and then completing a Specification Sheet for each of the data elements which appear on the document or is contained in the record. In this way linkage is set up between records, documents and E/G/A's which is later used in the output reports.

4.3 RECORD SPECIFICATION

The Data Specification Sheet (fig. 4.1) is used to define the machine processable records that exist in the information system. These records are stored in files or data sets on cards, magnetic tape or magnetic disc and are processable by a computer or business machine. A distinction between documents and records as defined here, can be made by determining if they can be used as they are (i.e. are humanly intelligible) or if they require machine processing in order to reveal the data or information captured on them. If they are humanly intelligible, they are documents, if not, they are records.

The specification of a record includes identifying it by a Label, Name and Description and providing information as to where the record is located and how to access it.

A distinction must be made between records and segments. A segment is a portion of a data record containing one or more logically related elements, groups or arrays. A record then, is defined as a logical collection of one or more segments, elements, groups or arrays. Segments are included in the definition of records since they are the basis on which many Data Management Systems are built and may be the smallest portions of a data set or file which can be separately accessed and stored.

Title Statement

This statement takes the general form as the Title Statement for documents or elements; is described in section 4.2 and illustrated in figure 4.3.

Name Statement

The Name Statement is described in section 4.2.

Description Statement

The Description Statement, providing textual information about the record, is described in section 4.2.

Type Statement

TYPE = Record/Segment;

The Type Statement specifies if the specification is for a record or a segment.

Class Statement

CLASS = Perm/Temp;

The Class Statement identifies the record/segment as being a member of a permanent file or a member of a temporary file

FIELD	POSITION	DESCRIPTION
Activity	1 - 2	RB: to create a new record RC: to change an existing record specification RD: to delete an existing record specification
Record/Segment Label	3 - 9	Up to seven alphameric characters giving a unique label to the record/segment
	10-12	Blank
Card Code	13-14	Card Code
File Label	15-21	Up to seven alphameric characters identifying the file or data set in which the record/segment resides
Version	22-23	Two numeric characters which indicate the version number of the record/segment
Status	24	The status of the specification in the system P: proposed A: approved E: effective
Prepared by	25-39	The name of the person preparing the specification
Preparation date	40-45	The date the specification was prepared
Certified by	46-60	The name of the manager of the prime user department
Effective date	61-66	The date the specification becomes effective.

Figure 4.3

Title Statement for Record and Segment Specifications which exists for only a limited duration.

Access Statement

ACCESS = ISAM/BDAM/SAM/ etc:

This statement provides data organization and access method for a record/segment. Any access method pertinent to the installation is acceptable.

Language Statement

LANGUAGE = COBOL/ALGOL/FORTRAN/ etc;

The Language Statement indicates the language used when the record/segment was initially created. The operating system used (i.e. DOS or OS) should also be indicated.

Media Statement

MEDIA = TAPE/DISC/CARD;

The Media Statement describes the media on which the record/ segment is stored.

Format Statement

FORMAT = F/V/U/FB/VB/FS/VS;

The Format Statement identifies that the record is either fixed length or variable, blocked or unblocked, undefined or consists of fixed or variable length segments.

Computer Statement

COMPUTER = IBM360/UNIV9400/ etc;

The Computer Statement identifies the system used to create the record/segment.

Count Statement

COUNT = number;

The Count Statement, in reference to a record, indicates the number of records in the file or data set. If a segment is being specified, the count refers to the number of occurrences of the segment in the record.

Retention Statement

RETENTION = number of periods;

The Retention Statement provides the length of time that the data in a record/segment is maintained, i.e. two months would be indicated by 2M.

End Statement

This indicates the end of the Record/Segment Specification. After completing the Record/Segment Specification each of the individual data elements are defined by means of the Data Element Specification.

4.4 DOCUMENT SPECIFICATION

The Data Specification Sheet shown in figure 4.1, is also used to identify data and information that appears on various forms, documents, reports and displays throughout the organization. Its main purpose is to record each document by label, name, description, volume, frequency of preparation or use, source and identity of users.

A document can be defined as a medium carrying information in the form of a collection of one or more data elements. The medium in this context not only includes paper but also cards, machine processable forms, micro film, telex messages, video displays, etc. Any document that exists in the information system, which serves a purpose and provides data of informational or operational content to a user or process, must be identified by means of the Document Specification Statements.

ARDI[2] classifies documents according to their characteristics:

A source document is a document which introduces new data to the system. Generally, data collection and conversion will be necessary when a source document is used in conjunction with computers.

A work or intermediate document is a document such as a worksheet or a summary card, used mainly to summarize a large quantity of source data or to facilitate the editing of transactions or reports.

---

[2]W. Hartman, H, Matthes and A. Proeme, Management Information Systems Handbook, McGraw-Hill Book Co., 1968, sect. 3.2, p.31.

A record is a document carrying recordings of a set of related data elements. It is usually part of a manual file which is regularly updated to permit the supply of current information for use in the preparation of reports.

A report is a document carrying managerial or operational information. Such a document may call for a managerial decision or may serve to initiate a necessary operation.

Title Statement

The Title Statement has the same format as described earlier and is shown in figure 4.4.

Name Statement

The Name Statement, as described earlier, consists of a series of key words which unambiguously identifies the document.

Description Statement

The Description Statement provides textual information and a description of the document.

Type Statement

TYPE = source/work/record/report;

The Type Statement identifies the document according to the description outlined in section 4.4.

Class Statement

CLASS = status/activity/transaction;

Status indicates a report recording status at a given date,

FIELD	POSITION	DESCRIPTION
Activity	1 - 2	DB: to enter a new document specification DC: to change an existing document specification DD: to delete an existing document specification
Document Label	3 - 9	Up to seven alphameric characters giving a unique label to the document
	10-12	Blank
Card Code	13-14	Card Code
File Label	15-21	If applicable, up to seven alphameric characters defining the file where the document is stored for reference; otherwise blank
Version	22-23	Two numeric characters which indicates the version number of the document
Status	24	The status of the document in the system P: proposed A: approved E: effective
Prepared by	25-39	The name of the person preparing the specification
Preparation Date	40-45	The date the specification was prepared
Certified by	46-60	The name of the manager of the prime user department
Effective date	61-66	The date the specification becomes effective.

Figure 4.4.

Title Statement for Document Specification i.e. a balance sheet. Activity indicates a report recording activity over a given period, i.e. a Profit and Loss Statement. Transaction indicates a document recording some action or event. i.e. a purchase order.

Form Statement

FORM = paper/card/video/micro;

The Form Statement further defines the document by describing the type of media on which the document appears.

Count Statement

COUNT = number;

The Count Statement represents the quantity of the documents received, processed or used in the time interval defined by the Frequency and Period Statements.

Preparation Statement

PREPARATION = manual/computer;

The Preparation Statement indicates whether the document is manually or computer produced.

Frequency Statement

FREQUENCY = number;

Frequency represents the number of times this document is received, processed or used in the period defined by the Period Statement.

Period Statement

PERIOD = SS/MM/HH/DD/WW/BW/SM/MO/QQ/SY/YY/RR;

The period is the shortest time period in which this document is received, processed or used.

SS = seconds	SM = semi-monthly
MM = minutes	MO = monthly
HH = hours	QQ = quarterly
DD = days	SY = semi-yearly
WW = weeks	YY = yearly
BW = bi-weekly	RR = as required

Origin Statement

ORIGIN = source-code;

This statement describes from whom or where the document originates. The source-code specifies the department or person who initially submits the document to the system. It can be either internally generated by the organization, or come from an external source such as a customer. The format of the source-code conforms to the structure established for the source and user specification described in section 4.2.

Users Statement

USERS = user1, user2, user3, ..., user n;

This statement defines all the users of a particular document. As there may be more than one user, it is important to ascertain the identity of all the users even though they may only use certain pieces of data on the document. The format for the user code is described in section 4.2.

Retention Statement

RETENTION = number of periods;

The Retention Statement indicates the number of time periods the document is stored, or retained, between its initial processing and eventual destruction. Periods are indicated by:

D - days	M - months
W - weeks	Y - years

For example, a document that is kept for six months would be described as:

RETENTION = 6M;

End Statement

The End Statement indicates the end of the document specification. When the document specification has been completed, the data elements appearing on the document must each be separately specified.

4.5 DATA ELEMENT SPECIFICATION

A Data Specification Sheet is completed for every data element on the document or record after the document or record itself has been specified. The data elements are linked to the corresponding documents or records through the Document/ Record label in the Title Statement. The use of reference definition and the version number can be used to simplify the specification of a data element if it already exists in the system. As described in section 4.2, only the Title Statement and those statements needed to specify attributes which differ from the initial entry, need to be supplied.

Title Statement

The Title Statement is similar to the description in section 4.2, and is shown in figure 4.5.

Name Statement

The Name Statement is described earlier in section 4.2.

Description Statement

The Description Statement is described earlier in section 4.2.

Type Statement

TYPE = element/group/array;

This statement specifies the data item as an element, group or array.

FIELD	POSITION	DESCRIPTION
Activity	1 - 2	EB: to enter a new element specification EC: to change an existing specification ED: to delete an existing specification
Element Label	3 - 9	A seven character alphameric label giving a unique identifier to the element
Sequence Number	10-12	Three numeric characters that indicate the relative position of the element within the document or record
Card Number	13-14	Card number
Document/Record Label	15-21	An eight character alphameric label identifying the document/record on which the element appears
Version Number	22-23	Two numeric characters that indicate the version of the element
Status	24	P: proposed A: appeared E: effective
Prepared by	25-39	Name of person preparing the specification
Preparation Date	40-45	Date the specification was prepared
Certified by	46-60	Name of the manager responsible for specification
Effective Date	61-66	Date the specification becomes effective.

Figure 4.5

Title Statement for Element/Group/Array Specification

Class Statement

CLASS = name/code/amount/date/text/quantity/flag/control;

The Class Statement identifies the general use of the data. The classes have been chosen to portray information to a user or analyst. They are an extension to the object (entity), property and time concepts as discussed by Langefors[3] and required for automated design.

Name	- data which identifies specific entities
Code	- data which identifies a classification of an entity
Quantity	- the number or quantity of anything (excluding monetary units)
Amount	- the quantity of monetary amounts
Date	- calendar date
Text	- data having undefined content
Flag	- an indicator showing a yes-no condition

Function Statement

FUNCTION = FI/FF/VF/VR;

This statement further defines the function of the data element.

The Function Statement is applicable to Solvberg's[4] process relations which identify the relationships between data elements.

The definitions are determined by figure 4.6.[5]

---

[3] J. Bubenko, B. Langefors and A. Solvberg, Computer-Aid Information Systems Analysis and Design, Studentlitteratur, Lund, 1971, p.22.

[4] Ibid, p.98.

[5] J.F. Kelly, Computerized Management Information Systems, The MacMillan Company, New York, N.Y., 1970, p.372.

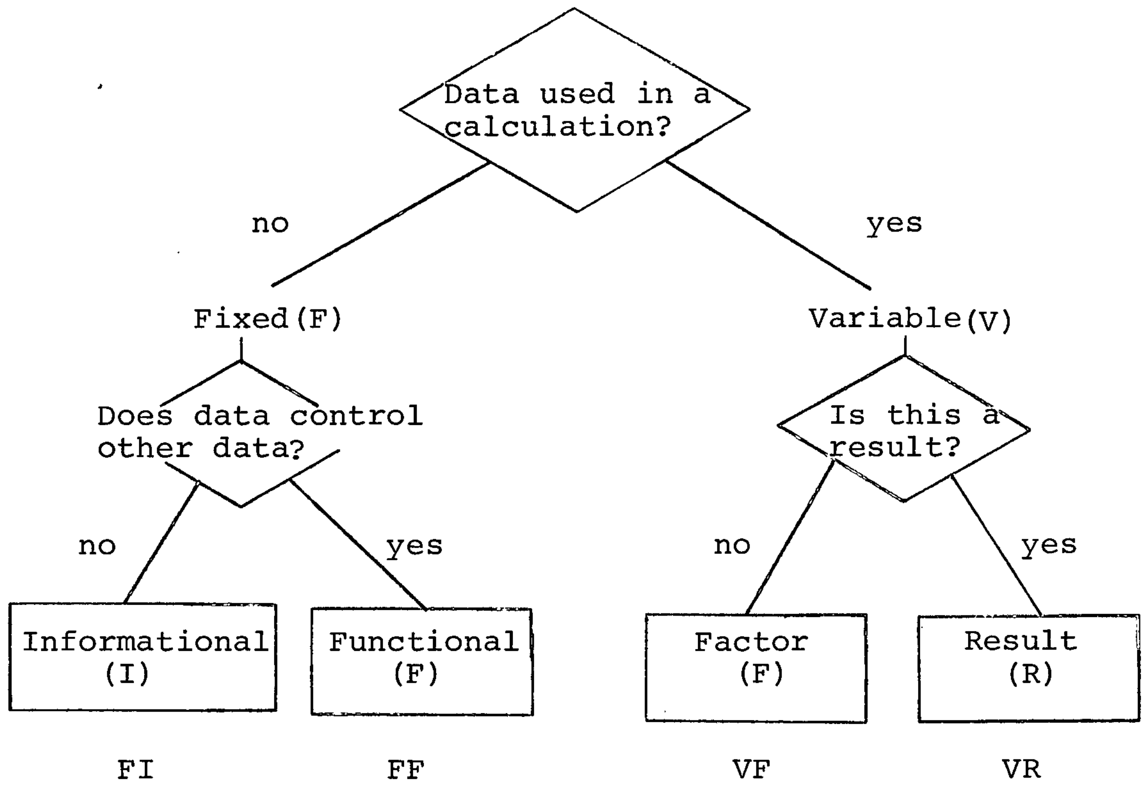

Figure 4.6

Element Function Definitions

Count Statement

COUNT = number, number, number, ...;

The Count Statement, used when defining groups or arrays, provides a numerical count of the number of data elements within the group or array. If one number appears in the count statement, it indicates the number of elements in the array or group. If more than one number appears, they express the number of planes in a multidimensional array.

i.e COUNT = 3, 2, 4; is a three dimensional array having 3 planes, 2 rows and 4 columns.

Responsibility Statement

RESPONSIBILITY = code;

The Responsibility Statement indicates the person responsible for the content and integrity of a given element, group and array and who is authorized to update or delete the element. This statement, in fact, determines who "owns" the data and is defined in section 4.2.

Mode Statement

MODE = format, length, scale;

The Mode Statement describes the format of the data as it appears on a document or is contained in a record. The format operand identifies the element as:

CH - a string of alphabetic or numeric characters
BI - a string of binary bits
FD - a zoned decimal number
PD - a packed decimal number
BD - a binary fixed point number
DF - a decimal floating point number
BF - a binary floating point number

The length operand specifies the number of characters, or in the case of binary numbers, the number of bits, in the element. The scale operand, specifies the number of fractional digits for fixed point numbers.

Edit Statement

EDIT = picture;

The Edit Statement furnishes an edit mask for use when displaying the element on a report or video display. COBOL or PL/I edit words may be used for the operand.

Key Statement

KEY = digit;

The Key Statement is used to identify a data element as an access key. In addition, it provides a means of indicating the relative importance of the element as a key in relation to other keys. The digit in the operand indicating this relative importance may be from 1 (major) to 9 (minor).

Code Statement

CODE = symbol, meaning, symbol, meaning ...;

The Code Statement is used to specify coded data. It should appear in the specification whenever CODE appears in the Class Statement. Example:

CODE = 01, Eastern Region, 02, Western Region;

End Statement

This indicates the end of the data element specification.

4.6 FILE CONSIDERATIONS

Permanent computer files must be designed, organized and structured in order to store the data information collected and allow a means for manipulating and retrieving this information. The physical format of these files will depend on the implementation of DACS and on the computer hardware and software the user has at his disposal. If the user has a Data Management System available, the structure of the DACS files will undoubtedly be physically different from a user who has only the more conventional organizations and access methods, such as Indexed Sequential, at his disposal. For these reasons, the discussion on file design here will be at a rather general and conceptual level and will concentrate on the logical organization and content of the files rather than on the actual physical structure.

This conceptual design, however, does make the assumption that a direct access storage device is used to store the information and that some type of indexed or direct file organization is available to allow for the selective retrieval of information by data labels.

The conceptual record structure for the DACS file is shown in figure 4.7. The Label of the record, document or E/G/A serves as the primary access key to the individual records. The records can be considered to be made of a number of segments, each containing some of the information provided by the Specification Sheets. These segments may be fixed or variable length

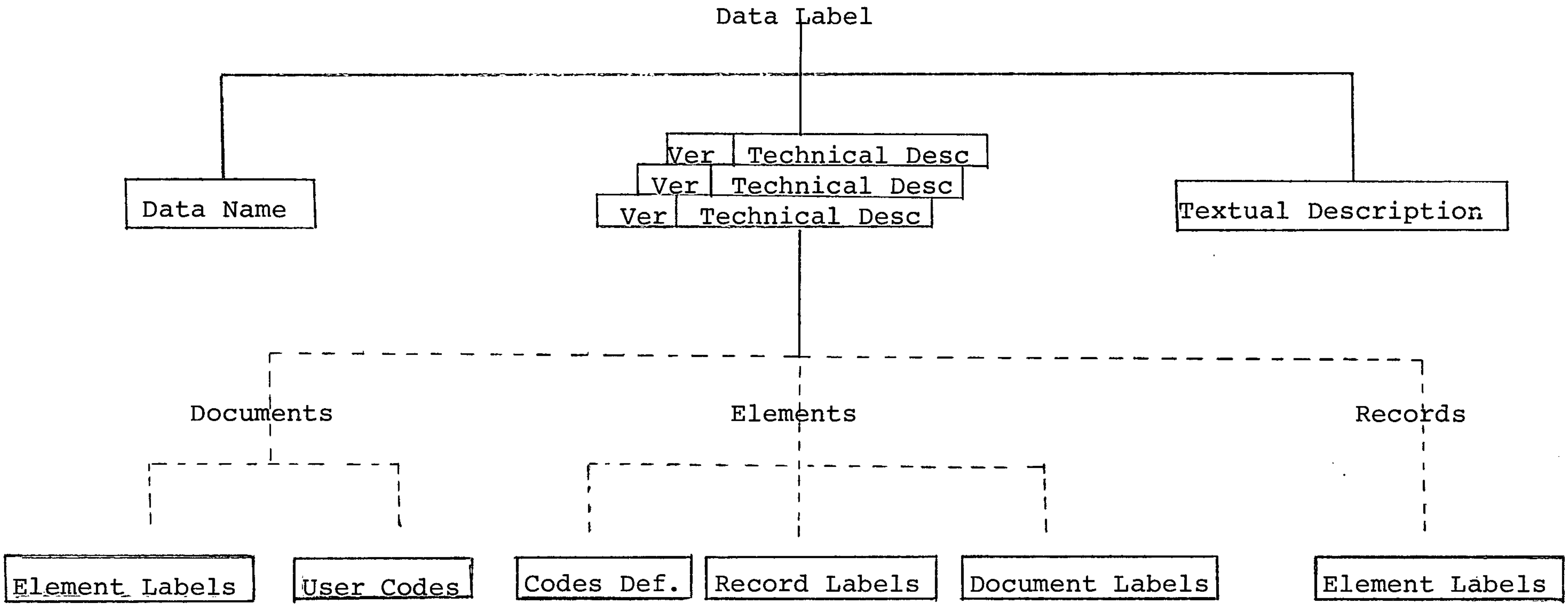

Figure 4.7

Logical DACS Record Structure

depending on the capabilities of the organization and access method employed.

For all data items there is the Name, Description and technical data segments.

Data Name Segment

This segment contains the Keyword name of the data item and is the input to the Keyword Index.

Description Segment

This segment contains the textual description of the data. It may be necessary to have one or more of these segments in order to contain the complete description.

Technical Segment

This segment contains the Version number and status for each data item as well as the technical information. Since the specifications and characteristics vary for different versions of the data, there may be several of these segments.

For each version this segment will include:

- Status
- Effective Date
- Type
- Class

In addition it will include for Records:

- Access
- Language
- Media
- Format
- Computer
- Count
- Retention

For Documents:

- Form
- Count
- Frequency
- Period
- Source
- Retention

For Elements:

- Function
- Count
- Responsibility
- Mode
- Edit

For each version of the data item there will also be associated other segments. These segments will differ depending upon whether the item is a record, document or E/G/A.

For documents the following segments are defined:

Element Segments

These segments contain the Labels, Version numbers and sequence numbers of the elements which appear on the document.

User Segments

These segments contain the user codes of the departments, or persons, who use the document either as generators, processors or recipients.

For records, the element segments are defined to indicate the elements which make up the record. These segments will contain the element Label, Version and sequence number and the key indicating the relative importance of the element in the record.

For elements there are three more segment types.

Record Segment

Each of these segments contains the Label of the records where this element can be found.

Codes Segment

If the element is classed as a code, segments should exist which identify the various codes and their definitions.

Document Segment

The Labels of the various documents in which the element appears is contained in these segments. An indicator showing that the document is source, work or report could also be included.

These various segments could be combined into physical records in varying ways. They may all be grouped together to form one variable length record with the Label as the key; or each segment may be treated as a separate record using the Label and a suffix code as the key.

If a hierarchical-oriented Data Management System such as IMS[4] is available, this conceptual segmented approach could be utilized directly.

Two other files may be considered to simplify retrieval and processing. One would be a User File organized by user codes and for each user, would contain a list of documents which that particular user employed. This could be a separate file or could be an inverted index file maintained by the file organization method.

The second file would be the Name File which would be in sequence by each Keyword of the data names and would be utilized to produce the Keyword Index Report.

---

[4]IBM CORPORATION, Information Management System V2, General Information Manual, GH20-0765, 1972.

4.7 SYSTEM OUTPUT REPORTS

The output reports or displays produced by DACS fall into two general categories:

- Verification and Edit Reports
- Usage Reports.

Verification and Edit Reports

These reports are produced when the Specifications for the data are processed by the system as additions, changes or deletions. They provide the following information:

- Errors: Any errors found in processing the Specification Statements or unsuccessful update transactions are identified. This could include format errors, duplicate labels or references to labels which do not exist in the system.
- Ommissions: Any missing specifications or characteristics that have not been supplied should be identified.
- Statistics: Show the current content of the system by number of entries and indicate the additions, deletions and changes that have occurred during the last processing cycle.

The Verification and Edit reports are used mainly by the Data Administrator to maintain the completeness and integrity of the contents of DACS. It is his responsibility to check these reports and ensure that any errors or ommissions are rectified.

Usage Reports

These are the significant reports produced by DACS and are used to answer the questions that were raised in section 3.2.

- Data Dictionary and Directory: This lists all the specifications for a given record, document or element; provides information as to the content of a record or document and the location and use of the data elements.
- Keyword Index: This is an alphabetic listing by the keywords of the Name Statement and displays the complete data name and its associated label.
- Data Requirements Analysis: This report shows, by user, the source, work and report documents employed by them.

Data Dictionary/Directory

The Data Dictionary/Directory (figure 4.8) provides a reference for all data elements, documents and records that have been specified to the system. It is divided into three sections; one section each for documents, records and E/G/A's with the data labels in alphabetical order within each section.

The dictionary section of the report shows for each data item, the label, name, description and technical characteristics.

The directory portion of the report shows:

For Documents - the elements which appear on the document and the users of the document.

For Records - the elements which are contained in the record.

For Elements - the documents on which the elements appear and the records that contain the elements. If the element is classed as a code element then the code information is also included.

```
                          DATA ADMINISTRATION AND CONTROL SYSTEM
                                DATA DICTIONARY/DIRECTORY

 DATA ELEMENTS

CUSTNO   CUSTOMER NUMBER                                                      )Label and Name

         A NUMERIC IDENTIFIER ASSIGNED TO ANY COMPANY OR PERSON PURCHASING    )
         PRODUCTS OR SERVICES.                                                )Description

         VER  STATUS  TYPE  CLASS  FUNC  COUNT  MODE  EDIT    RESPONS  EFFECT    )Technical
         01     E     ELEM  NAME    FI   1600   CH,6          02160903 01/01/69  )Characteristics

      LOCATION
         LABEL    VER   SEQ   KEY   NAME

         CUSTMST  01    01    1     CUSTOMER NAME AND ADDRESS RECORD          )Records containing
         ACCREC   01    01    1     ACCOUNTS RECEIVABLE STATUS RECORD         )this element

      USAGE
         LABEL    VER   SEQ         NAME

         CUSTORD  01    01          CUSTOMER PURCHASE ORDER                   )Documents containing
         INVOICE  01    03          CUSTOMER INVOICE                          )this element

DATA RECORDS

ACCREC   CUSTOMER ACCOUNTS RECEIVABLE STATUS

         RECORD CONTAINING CUSTOMER INFORMATION SHOWING OUTSTANDING  ACCOUNTS
         RECEIVABLE STATUS BY CURRENT, 30-60, 60-90, AND OVER-90 CATEGORIES

         VER  STAT  TYPE  CLASS  COUNT  FILE   ACCESS  LANG   MEDIA  FORMAT  COMP    RETENT  EFFEC
         01    E    REC   PERM   1600   ARFIL   SAM    COBOL  TAPE     FB    IBM360          01/01/68

      CONTENTS

         LABEL    VER   SEQ   KEY   NAME

         CUSTNO   01    01    1     CUSTOMER NUMBER
         CURAMT   01    02          CURRENT AMOUNT OUTSTANDING
         ETC.
```

Figure 4.8

DATA DICTIONARY/DIRECTORY

The Dictionary/Directory serves as the main source of reference for the data resource. It is used for identification, definition and clarification purposes for anyone requiring information about a particular data item. It also serves to highlight duplication and redundancy in documents and records and can be used by an analyst to improve the efficiency of data use and storage.

When new information requirements arise, the Dictionary/ Directory can be referenced to determine if the data elements already exist in the system and if there are source documents and reports which can be utilized to satisfy the new need. If not, the report will be helpful in determining how the new information can most effectively be produced.

Keyword Index

The Keyword Index (figure 4.9) is a listing of data labels and their corresponding names. The name of each document, report or data element appears in the Index by alphabetical order of each keyword in the name. This enables the user to locate any DACS entries, listed in alphabetical order, through a knowledge of any word in the data name. Once having found the correct name, he can, by the associated Label, locate the data item and any information he requires in the other reports of the system.

Data Requirements Analysis

This report (figure 4.10) shows, by user or department code,

DATA ADMINISTRATION AND CONTROL SYSTEM

PAGE 1
01/07/73

KEYWORD INDEX

LABEL	E/R/D	NAME
CUST002	E	ADDRESS *CUSTOMER BILLING
CUST020	R	ADDRESS MASTER *CUSTOMER NAME AND
CUST030	E	ADDRESS *CUSTOMER SHIPPING
SA001	D	ANALYSIS BY PRODUCT BY CUSTOMER *SALES
AR001	D	ACCOUNTS RECEIVABLE STATUS
CUST002	E	BILLING ADDRESS *CUSTOMER
CUST002	E	CUSTOMER BILLING ADDRESS
CUST005	D	CUSTOMER INVOICE
CUST020	R	CUSTOMER NAME AND ADDRESS MASTER
CUST001	E	CUSTOMER NUMBER
CUST004	E	CUSTOMER PURCHASE ORDER NUMBER
SA001	D	CUSTOMER *SALES ANALYSIS BY PRODUCT BY
EMP001	E	EMPLOYEE NAME
EMP002	E	EMPLOYEE NUMBER
EMP003	E	EMPLOYEE SALARY
SA002	D	HISTORY BY PRODUCT BY YEAR *SALES
INV001	E	INVENTORY NUMBER
INV003	E	INVENTORY QUANTITY ON ORDER
INV002	E	INVENTORY QUANTITY IN STOCK
INV010	D	INVENTORY STATUS
CUST005	D	INVOICE CUSTOMER
CUST020	R	MASTER CUSTOMER NAME AND ADDRESS

LABEL	E/R/D	NAME
CUST020	R	NAME AND ADDRESS MASTER *CUSTOMER
EMP001	E	NAME *EMPLOYEE
CUST001	E	NUMBER *CUSTOMER
CUST004	E	NUMBER *CUSTOMER PURCHASE ORDER
EMP002	E	NUMBER *EMPLOYEE
INV001	E	NUMBER *INVENTORY
INV003	E	ORDER *INVENTORY QUANTITY ON
CUST004	E	ORDER NUMBER *CUSTOMER PURCHASE
SA001	D	PRODUCT BY CUSTOMER *SALES ANALYSIS BY
SA002	D	PRODUCT BY YEAR *SALES HISTORY BY
INV003	E	QUANTITY ON ORDER *INVENTORY
INV002	E	QUANTITY IN STOCK *INVENTORY
AR001	D	RECEIVABLE STATUS *ACCOUNTS
EMP003	E	SALARY *EMPLOYEE
CUST003	E	SHIPPING ADDRESS *CUSTOMER
SA001	D	SALES ANALYSIS BY PRODUCT BY CUSTOMER
SA002	D	SALES HISTORY BY PRODUCT BY YEAR
AR001	D	STATUS *ACCOUNTS RECEIVABLE
INV010	D	STATUS *INVENTORY STOCK
INV002	E	STOCK *INVENTORY QUANTITY IN
INV010	E	STOCK STATUS *INVENTORY
SA002	D	YEAR *SALES ANALYSIS BY PRODUCT BY

Figure 4.9

Keyword Index

DATA ADMINISTRATION AND CONTROL SYSTEM

PAGE 1
01/07/73

DATA REQUIREMENTS ANALYSIS

USER		LABEL	VER	STAT	NAME
02061301	Origin	INV001	01	E	INVENTORY RECEIPT NOTICE
		INV002	01	E	INVENTORY SHIPPING TRANSMITTAL
	Report	INV010	01	E	INVENTORY STOCK STATUS
		INV020	01	E	INVENTORY WAREHOUSE ORDERS
01071312	Origin	PAY001	01	E	EMPLOYEE PAYROLL CHEQUE
		PAY004	01	E	EMPLOYEE PAYROLL REGISTER
		PAY010	01	E	EMPLOYEE F-4 FORMS
	Work	PAY020	01	E	PAYROLL FILE UPDATE AND EDIT LIST
	Report	PAY050	01	E	EMPLOYEE TIME TICKETS
		PAY055	01	E	EMPLOYEE PAYROLL CHANGE NOTIFICATION

Figure 4.10

Data Requirements Analysis Report

the documents and data that they generate, process or receive. It can be used, when analyzing a particular area, to determine the data requirements for that area and if any redundancy, duplication or inefficiency exists. This report would also be of value to the Information Systems Architect when determining what data and information is being employed by various functions or departments within the organization. Concentration on a particular subsystem of the total information system may be achieved by analyzing the data elements and documents which serve this particular subsystem. For example, the marketing information system may be analyzed by obtaining a report retrieved by the marketing function code, which shows the data required by this function.

If, as Langefors[5] suggests, lists could be constructed showing the functions of firms of different types and their direct information needs, the Data Requirements Analysis Report could be compared to a standard for data requirements. In this way it could be determined as to whether certain functions are above or below this standard with respect to the information they receive.

---

[5]B. Langefors, Theoretical Analysis of Information Systems, Studentlitteratur, Lund, Sweden, 1966, Vol.2, p.225.

Redundancy and Similarity Reports

AUTOSATE[6] and Schruben's[7] model contain reports which identify identical or similar documents and files by the commonality of data elements.

AUTOSATE determines duplication by comparing a calculated value for each file with the values for other files and reporting the files which have identical or similar values.

Schruben, when analyzing each document in the system, also identifies similar documents by listing the documents which satisfy a "prespecified criterion of similarity". The criterion he used was to list the ten reports which were missing the least number of data elements on the report under study. He suggests that other criteria could include percent of identical data elements or frequency of generation.

Such reports showing similarity would be very useful to the analysts but may be difficult to program and produce. A more realistic approach may be to define a set of data elements and then to scan the documents and files to determine which ones contain a certain percentage of the specified set.

These examples cited do indicate that duplication and similarity can be identified through automated analysis and do provide a valuable tool in analyzing the data resource.

---

[6] O.T. Gatto, "AUTOSATE", Communications of the ACM, Vol.7, No.7, July 1964, p.430.

[7] L. Schruben, "The Information System Model", Datamation, Vol.15, No.7, July 1969, p.99.

CHAPTER V

THE IMPLEMENTATION AND USE OF DACS

5.1 THE DATA ADMINISTRATOR

The Data Administration and Control System cannot achieve its potential in effectively and efficiently monitoring the data resource unless certain other requirements are met.

- Management must be informed and educated to regard data as a corporate resource and to understand the value and potential of DACS as a tool in data management.
- DACS must be assiduously maintained and updated as to content and integrity. It will quickly lose its value if allowed to become outdated or neglected.
- DACS must be made available to all people concerned and must be used regularly in identifying problem areas and defining new data requirements.

The requirements and responsibilities of the above functions make it necessary to establish a specific organizational position identified as the Data Administrator.

The Data Administrator should be considered as a staff function reporting directly to the person responsible for Information Systems and Data Processing as shown in figure 5.1.

One of the attractive features of a system such as DACS, is that once it has been set up, it is relatively easily administered by a single person. Such may not be the case with more complex

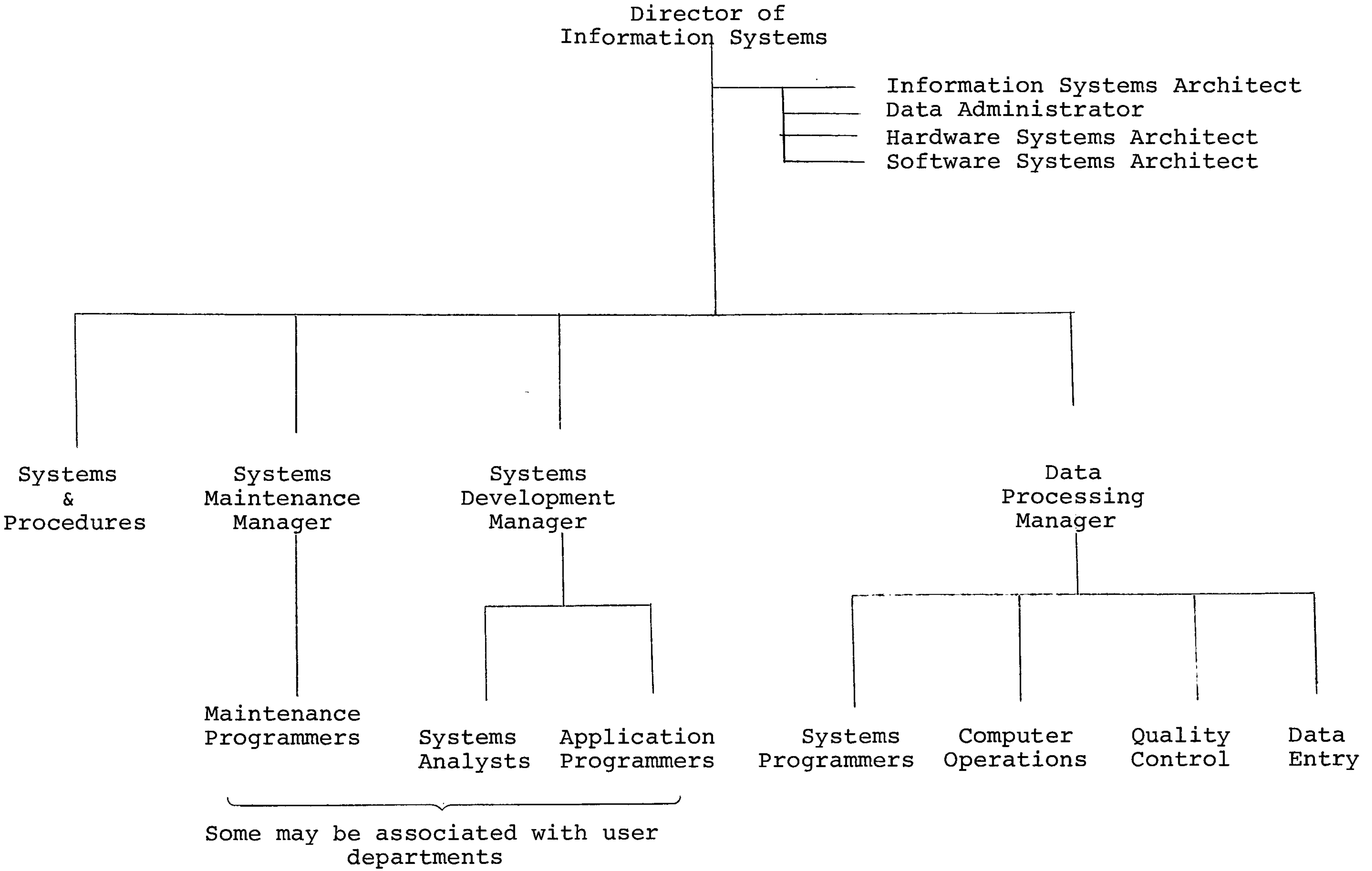

Figure 5.1

Information Systems Organization

Data Base Management Systems and their requirements for definition, organization and structure, access and retrieval, security and recovery procedures.

The Data Administrator is concerned with the data that presently exist within the company and analyzing problems that may be inherent in their present organization and use. He is not the owner of the data but only the custodian. He is responsible for ensuring that the data resource is utilized according to certain standards and objectives which have been developed. He may, in fact, be deeply involved in defining these standards and objectives.

Supported by a system such as DACS, the Data Administrator is the focal point for the administration and control of the data resource and will have unique knowledge on the overall extent, content and location of this resource.

His functions and responsibilities will include:

- To inform and educate management on the extent and value of the data resource.
- To assist in the establishment of data resource objectives and standards.
- To analyze and define existing data elements as to name, source, use, location and responsibility.
- To implement a system (i.e. DACS) by which the above analysis and definition may be formalized.

- To analyze the data for unjustified inconsistency and redundancy and to make recommendations for its rectification.
- To ensure that he is informed of all new and changed data requirements.
- To monitor and concur with new data element definitions and the source, file and output requirements for new applications.
- To have continuous liason with the Information Systems Architect and his objectives, plans and requirements.
- To keep aware of data users' needs and requirements.
- To make recommendations regarding file organization and retrieval techniques including Generalized Data Base Management Systems.
- To evaluate the data resource from an overall organizational viewpoint, considering effective and efficient utilization.

The Data Administrator will hold a very responsible and demanding position in the realm of Information Systems and Data Management and may perhaps be considered analogous to the comptroller[1] in the financial area. Certainly he has the same mission of developing and maintaining a control system for definition and recording, measuring of effectiveness and efficiency, auditing for accuracy and consistency and disseminating of statistics and information concerning the resource.

---

[1]M.J. Gordon and G. Shillinglaw, Accounting and Management Approach, Third Edition, Richard D. Irwin Inc., Homewood, Illinois, 1964, pp.527-529.

The data administrator has the responsibility of ensuring that the data resource is maintained and organized to support the information framework that is defined by the higher level management and that it meets their requirements, objectives and standards. It is management who must develop the standards and controls for the data resource, and it is on these standards and the ability of the data to fulfil the information framework, that the data administrator will be evaluated.

5.2 STANDARDS CONCERNING DATA

Many of the required standards associated with data are inherently implied in the design and structure of the Data Administration and Control System and are discussed in Section 3.3, under Requirements for a Data Control System.

It is through the implementation and use of DACS, that the standards for identification, definition, source, use and responsibility of data are satisfied. The data specifications must be complete and accurate and any conflicts resolved before entry is made into the formalized system. The Data Administrator may be required to act as an arbitrator in disputes over the ownership or responsibility of certain data that may be shared across functional boundaries.

Any data duplication or apparent redundancy must be closely examined to determine if it is justified on the basis of cost, timeliness or effectiveness and if not, the redundancy must be corrected through the consolidation of files and the redesign or elimination of source or report documents. Although some resistance may be encountered from user groups or data processing personnel, an examination of the costs and problems resulting from unmanaged data and its effect on the operational and informational content, should justify the need for redesign and a more disciplined approach to data.

It must be stressed to users, analysts and programmers, that DACS serves as a central repository for data and that its output should be examined first when new or changing data requirements arise.

DACS provides a control mechanism to preclude the introduction of redundant or inconsistent data elements. It should be coupled with procedures which require that new data elements be checked against DACS to ensure that they do not duplicate existing elements unnecessarily and that they are not inconsistent with these existing elements

It is only through the formalized use of this tool that its value can be realized and its objectives be attained.

5.3 COLLECTION OF THE DATA INFORMATION

A most important and demanding task in the implementation of DACS, is the collection of the information about the data itself. Most of this information probably exists in some fashion in various areas of the company but it will have to be searched for and collected in an organized manner.

Useful approaches to this task are suggested by Kelly[1] and in ARDI[2] in the discussions on starting points for information system analysis and design. One approach is to start with the operating system or production flow. It is based on an in-depth analysis of the manufacturing and production processes involved in the company's operations and the data required to perform and control this function.

The other approach is to start with the information flows and systems and analyze the order processing, sales, marketing and financial functions, with emphasis on the documents and data that are used in the operation, planning and decision-making roles.

The choice between the two approaches will depend on the type of organization - whether it has a manufacturing or service orientation. However, as Kelly makes clear, the study should begin at the operational level rather than at the management control, or strategic planning levels. This is the level at which a majority of the data elements can be identified most

---

[1] J.F. Kelly, Computerized Management and Information Systems, The MacMillan Company, New York, 1970, pp.73-75.

[2] W. Hartman, H. Matthes and A. Proeme, Management Information Systems Handbook, McGraw Hill Book Co., 1968, Sects. 3-2.2. and 3-2.3.

clearly as to definition, source, responsibility and use. If the elements at this level are rigorously defined, it will simplify the task of identifying them as they appear in informational content, at higher levels in the organization.

The determination of the point at which to begin the study should be the most logical system input and the one that a large part of the company would be dependent upon. For example, the service-order would be a logical starting point for a utility such as a Telephone Company.

Once a starting point has been determined, the various departments are analyzed as to their data use and requirements. An organization chart may be useful to define the data as to user, source and responsibility and to determine the interfaces and relationships between other departments.

The data collection cycle will proceed through interviews with department managers, analysis of the documents and information used and review of any documentation which exists in the data processing department, including input forms, record specifications and output reports.

Care must be taken to identify both manual and computer prepared reports. Discussions with people in the departments and with analysts will help identify the informational requirement and significance of the data elements.

As data items are identified and submitted to DACS, increasing use is made of the output reports to determine if the data elements have already been defined and if so, reference definition can be employed.

Once the operational levels have been reviewed, the analysis should proceed up the organizational hierarchy, noting new source data that enters the system and the varying employment of the data elements. Higher levels of management will be characterized as information consumers and there may be special purpose systems developed to suit their needs. These systems are important as they are often a significant cause of data fragmentation.

The data collection process may be a complex, tedious and time-consuming endeavour. However, the standardization and discipline of the Data Specification Methodology and the automated ability of DACS, should make this process simpler and more complete.

5.4 THE USERS OF DACS

The Data Administrator has previously been discussed as to his use and interface with DACS and the responsibilities he has concerning the Data resource. However, Management, Information Systems Architects, Analysts and Programmers can all make beneficial use of DACS.

Management

One of the earliest benefits of DACS to the management of a company comes during the data collection phase. As management personnel should be involved in the identification of the data in their department or area, they will become aware of the data resources that they provide, process or consume. As the data specification requires responsibility to be assigned to the data elements, the individual managers assume this responsibility for the consistency and integrity of the data within their control. With this increased awareness and concern for the data resource, they will pay more attention to it and be conscious of problems associated with data neglect. Once the content and extent of the data is realized, management will be able to provide a significant contribution in the analysis of data problems and their solutions. They will be more aware of data redundancies, inefficiencies and misuse and should draw attention to possible areas of concern.

After DACS has been implemented, several of its uses will furnish tangible benefits. The interface to DACS can be direct consultation with the output reports but probably will be through the Data Administrator.

When new data or information requirements arise, DACS can be consulted to determine the availability of the data elements and this information can be factored into development lead times and cost and manpower requirements to meet the new need. This practice will eliminate the problem of refusal of information demands when the data may already exist in a usable form.

The ability of a manager to define his requirements explicitly in the form of standardized data elements, will narrow the communications gap that sometimes exists between users and data processing personnel. As a result of better communications and data identification, his requirements should be met more quickly and with a better quality product. As inconsistency and duplication are reduced, management will have more confidence in the information they receive and in the data and systems which underlie it.

Information Systems Architect

A fundamental problem confronting the designers of information systems, is the identification of the information requirements of the horizontal and vertical structures in the organization. Their concern is whether different functional portions of the company can share a common data base or whether the data base must be separated into vertical and horizontal segments. A major challenge lies in the integration of the data resource so that it can be utilized by all levels and components.

The Information Systems architect is also faced with determining management's "information threshold"; the level of detail which a given person may require or expect.

Although the Information Systems architect may be more concerned with the eventual design of the information system rather than its present structure, he can extract a large amount of useful information from DACS.

DACS will provide him with the identity of the data elements which are used by the various segments of the organization and the context and relationships among them. It can also provide clarification of the informational interfaces between these segments. During analysis of the reports, weaknesses in the existing systems may become apparent and may provide insights into more effective design approaches.

Whether he takes a "top-down" or "bottom-up" approach to information systems design, at some phase he must document, analyze and structure the data elements and define models or procedures with which to manipulate the elements into information. With a tool such as DACS, this task may be greatly simplified. The existing systems have grown and developed over many years and analysis of the present status can help in identifying and determining future requirements.

DACS relieves from the architect, the necessity of determining "here is what exists" and allows him to concentrate on "what is really needed of an information system".

<u>Analysts</u>

Analysts generally concentrate in a more limited and detailed environment than Information Systems architects. The latter is concerned with the overall view and philosophy of the information system while the former, given the general requirements and design, directs his efforts to more specific subsystems or functional applications. It is the analyst who details input formats, file requirements, output specifications and user interfaces.

For analysts working on new requirements or extensions to existing applications, DACS provides a central source of data information. Once the data output and requirements have been defined, the analyst should consult DACS to determine whether or not existing documents can satisfy his needs and whether the data elements already exist somewhere in the system. If the data requirements do exist in some form, the analyst should exploit them in a manner consistent with the established data objectives. He must preclude the introduction of redundant or inconsistent data elements. If new elements are to be introduced into the system, he should first consult with the Data Administrator and determine if any users will be affected.

The analyst, by drawing on existing data resources while meeting the objectives of the Information Systems architects, plays an important role in maintaining the integrity, consistency and control of the data.

Programmers

The application programmer's function consists of translating the analyst's design specifications into a computer intelligent language. He can utilize DACS to gain more knowledge of the data he is to employ and to familiarize himself with the way this data is stored and accessed in the computer files. By utilizing DACS labels and names in the source programs, standardization and communication can be enhanced.

An interface to DACS could feasibly be incorporated to place record definitions from the DACS file into a source library which could then be "called" or "included" by COBOL or PL/I source programs. This ensures standardization of data usage and minimizes programmer time required in making up these statements whenever a data file is used in an application. A further extension of this interface would enhance DACS by incorporating the Data Description Language of CODASYL and satisfying the Guide-Share requirements.

CHAPTER VI

SUMMARY

In the quest for better information systems, numerous problems are encountered by management and systems personnel. Some of the earliest questions to arise are expressed as follows:

- Where is the logical starting point if relatively little is known about the data that are presently in use in the various systems throughout the company?
- How can the data, which should constitute the nucleus of the information system, be determined?
- There is a tremendous investment in the present systems. Can these systems be improved without a complete redesign and reprogramming effort?

In the preceding chapters of this thesis, it has been emphasized that a fundamental requirement in developing effective information systems, is to introduce discipline and control to the data resource. Once this is accomplished, the task of melding the information requirements and the data elements into a management information system will become a less formidable endeavour.

A review of the development of information systems and the management of data has indicated that the data resource suffers from management neglect, has become severely fragmented and is often not recognized as a valuable corporate resource. In order to alleviate these problems, a Data Administration and Control System has been proposed to introduce standardization and discipline to the existing data elements.

DACS, by itself, does not promise to provide more effective or improved information. It is a tool by which management and systems personnel can better manage, control and utilize the data resource. Through this tool the data elements are defined and described and data relationships are identified. This definition and identification is a primary requirement in information systems development. In conjunction with information requirements determination, data base design, file creation and maintenance, model development and information retrieval techniques, it constitutes one component of the complete development effort.

DACS can be considered to serve two purposes:

- It will provide a method by which the data resource can be better managed and controlled on a continuing basis.
- It can provide a starting point in the development of a comprehensive management information system.

The implementation of DACS will highlight the extent of the data resource and make management more aware of the importance of consistent, reliable data to corporate well-being. It is only through the realization of the value of the data resource and its effective management that this resource can be exploited to achieve its potential of meeting the informational requirements of an organization.

In the realm of Management Information System design, DACS provides a method by which the data elements can be identified in a disciplined manner. Through the Data Dictionary/Directory, redundancy and duplication are identified and rectified. Standardized data definitions are developed to be employed throughout the organization with DACS serving as a central repository for data information. With the Data Dictionary/Directory and the Data Requirements Analysis reports, designers can determine how each data element is employed and what the users' data requirements are. Through this facility, he will have more knowledge and insight of the data when new data base structures are designed.

As computer-aided system design becomes more developed, DACS could be extended to construct model data banks to meet the MIS requirements. This model data bank could then serve as input to a simulation model constructed to evaluate the performance of various generalized data management systems. Another necessary task, if automated design is to become a reality, is the development of techniques to define the models, programs and process relationships between various data elements and resulting information. Incorporating this technique into DACS will further contribute to information system documentation and design.

Computer-aided system design is certainly a development of the near future and DACS would require many extensions and modifications to contribute to this field. However, as an interim step, DACS, as described in this thesis, could be programmed with less than six man-months of effort and would provide a much needed and desired tool.

BIBLIOGRAPHY

Ackoff, Russel L., "Management Misinformation Systems", Management Science, Vol.14, No.4, p.B147.

Anthony, Robert N., Planning and Control Systems, A Framework for Analysis, Division of Research, Harvard School of Business Administration, Harvard University, 1965.

Blumenthal, Sherman C., Management Information Systems; A Framework for Planning and Development, Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1969.

Bontempo, Charles J., "Data Resource Management", Data Management, Vol.11, No.2, February 1973.

Bubenko, J., Langefors, B., Solveberg, A. (eds.), Computer-Aided Information Systems Analysis and Design, Studentlitteratur, Lund, Sweden, 1971.

CODASYL Systems Committee, "Feature Analysis of Generalized Data Base Management Systems", May 1971, Available from Association for Computing Machinery, New York, N.Y.

Dearden, John, "MIS is a Mirage", Harvard Business Review, January-February, 1972, p.90.

Dodd, George G., "Elements of Data Management Systems", Computing Surveys, Vol.1, No.2, June 1969, p.118.

Gatto, O.T., "AUTOSATE", Communications of the ACM, Vol.7 No.7, July 1964, p.425.

Gordon, M.J. and Shillinglaw, G., Accounting: A Management Approach, Richard D. Irwin, Inc., Homewood, Illinois, Third Edition, 1964.

Hanold, Terrance, "A President's View of MIS", Datamation, Vol.14, No.11, November 1968.

Hanold, Terrance, "An Executive View of MIS", Datamation, Vol.18, No.11, November 1972.

Hartman, W., Matthes, H. and Proeme, A., Management Information Systems Handbook, McGraw-Hill Book Co., 1968.

Head, Robert V., "Automated System Analysis", Datamation, August 15, 1971.

Head, Robert V., "Management Information Systems: A Critical Appraisal", Datamation, May 1967, p.22.

IBM Corporation, "Information Management System V2", General Information Manual, GH20-0765, 1972.

Joint Guide-Share Data Base Requirements Group, "Data Base Management System Requirements", November 11, 1970, Available from SHARE Inc., New York, N.Y.

Kelly, Joseph F., Computerized Management Information Systems, The MacMillan Company, New York, 1970.

Kriebel, Charles H., "The Future MIS", Business Automation, June 1972, p.18.

Langefors, Borge, Theoretical Analysis of Information Systems, Studentlitteratur, Lund, Sweden, 1966.

Morton, Michael, S.Scott, Management Decision Systems, Division of Research, Graduate School of Business Administration, Harvard University, 1971.

Nolan, Richard L., "Systems Analysis of Computer Based Information Systems Design", Data Base, Vo.3, No.4, Winter 1971, p.1.

Paul, Norman L., "MIS .... Are You Ready?", Data Management, October 1972, p.26.

Price, Gerald F., "The Ten Commandments of Data Base", Data Management, May 1972, p.14.

Sage, David M., "Information Systems: A Brief Look into History", Datamation, Vol.14, No.11, November 1968.

Schruben, Lee, "The Information System Model", Datamation, Vol.15, No.7, July 1969, p.93.

Teichroew, Daniel, and Sayani, Hasan, "Automation of System Building", Datamation, August 15, 1971, p.25.

Will, Hartmut J., "A System of MIS Concepts", condensed from H.J. Will, "Towards a System of MIS Concepts", Working Paper No.194, Faculty of Commerce and Business Administration, University of British Columbia, June 1973.

Will, Hartmut J., "Management Information Systems as a Scientific Endeavour; The State of the Art", Working Paper No.84A, University of British Columbia, Vancouver, B.C.

Will, Hartmut J., "MIS: Mirage or Mirror Image?", Working Paper No.146, University of British Columbia, Vancouver, B.C.

Zani, William M., "Blueprint for MIS", Harvard Business Review, Vol.48, No.6, November-December, 1970, p.95.

APPENDIX I

EXISTING DATA ANALYSIS TECHNIQUES

There have been several approaches, cited in the literature, which have attempted to analyze the data and information resource, establish relationships and bring more control and discipline to the area of information system analysis and design. Most of these techniques have been concerned with automating parts of the design process in the development of data processing systems which replace existing manual ones. A common characteristic of each, is that at some phase they analyze the present system as it currently exists. Some of the concepts employed in these techniques have been adopted in the formulation of the Data Administration and Control System and a brief comment about them is warranted here.

AUTOSATE

One of the first attempts at automated systems design was AUTOSATE developed by O.T. GATTO, of The Rand Corporation, for the United States Air Force. The technique "is geared to determining workload, relationships and storage characteristics of documents in the information network automatically"[1]. AUTOSATE examines "stations" as nodes in the information network and identifies data activity in and out, and information stored in these stations. Stations are defined as a person, a group

---

[1] O.T. Gatto, "AUTOSATE", Communications of The ACM, Vol 7, No. 7, July 1964.

of people, a work centre or a specific thing. Through interviews with station personnel, information used, processed or stored at each station is identified on document specification sheets. The specifications are translated into machine readable form and a series of analysis reports are produced. These indicate by station, document activity as to input, processing, output and storage. Other reports trace the flow of documents through the various stations and the activities which are performed with them. A series of "redundancy" reports is also produced to show identical and similar documents.

AUTOSATE's main emphasis is on the document flow through a manual information system and provides the analyst with more insight and knowledge on which to base his design of a new system. Although its major contribution was intended to assist in the design of new computer systems, several of the concepts are applicable and have been used in the identification and use of data elements in DACS. The use of data specification sheets, the removal of artificial boundaries of the data by function or application and the insights provided by the redundancy reports, have all contributed to the design of DACS.

Schruben's "Information System Model"[2] is similar to AUTOSATE in that it uses existing documents to trace the information flow through a system. "Document analysis", "where used", and "data-

---

[2]L. Schruben, "The Information System Model", Datamation, Vol 15, No 7, July 1969.

flow" reports, assist in the analysis and design of new information systems.

ARDI

ARDI[3] is a technique developed to provide a detailed guide to the Analysis, Requirements determination, Design and development, and Implementation and evaluation phases of the system design effort. In one phase, identifying the information flow, description sheets are used to document forms, files and fields existing in the system. While automated analysis is not used, the techniques used to identify the data elements and illustrate relationships have been useful in determining DACS requirements.

TAG

The Time Automated Grid[4] technique developed by IBM is intended to aid the systems analyst in the design of information systems. Data to be analyzed is recorded on "input/output analysis forms" that describe the characteristics of the inputs, outputs or files". The use of TAG begins with a description of required outputs and then works backwards to determine what inputs are necessary at what periods of time. When inputs and outputs have been defined, the next iteration of the program produces file and system flow descriptions.

---

[3] W. Hartman, W.H. Matthes, A. Proeme, Management Information Systems Handbook, McGraw-Hill Book Co., New York, 1969.

[4] J. Kelly, Computerized Management Information Systems, The MacMillan Co., New York, 1970, Chapter 8.

Although time is obviously one of the factors in file design, a study of the example provided in the reference indicates that a proliferation of files result from TAG analysis. These files are constructed to meet very structured requirements and do not appear to consider redundancy or commonality of use of data items. The technique seems to create very functional and application oriented files, contributes to the problems of common data base definition and defeats the desire to make the data resource flexible in order to meet various, and perhaps, unforeseen requirements.

Although the techniques described above do not directly focus on the problems of data resource management, they all have the common characteristic of attempting to identify the data resource in a disciplined standardized manner so that it may more readily be analyzed. However, none of these systems answer all the questions or satisfy the requirements that have been identified to achieve the desired level of control but the contribution of each to the development of DACS is acknowledged.