INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”. If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in “sectioning” the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from “photographs” if essential to the understanding of the dissertation. Silver prints of “photographs” may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Xerox University Microfilms
300 North Zeeb Road
Ann Arbor, Michigan 48106
MOFFATT, John Howard, Jr., 1943-
AN INTEGRATED DATA BASE FOR USE IN MANAGEMENT
OF HIGHER EDUCATION.

The University of Arizona, Ph.D., 1974
Education, administration

University Microfilms, A XEROX Company, Ann Arbor, Michigan

© 1974

JOHN HOWARD MOFFATT, JR.

ALL RIGHTS RESERVED

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED.
AN INTEGRATED DATA BASE FOR USE IN
MANAGEMENT OF HIGHER EDUCATION

by
John Howard Moffatt, Jr.

A Dissertation Submitted to the Faculty of the
DEPARTMENT OF EDUCATIONAL ADMINISTRATION
In Partial Fulfillment of the Requirements
For the Degree of
DOCTOR OF PHILOSOPHY
In the Graduate College
THE UNIVERSITY OF ARIZONA

1974
I hereby recommend that this dissertation prepared under my direction by John Howard Moffatt, Jr. entitled AN INTEGRATED DATA BASE FOR USE IN MANAGEMENT OF HIGHER EDUCATION be accepted as fulfilling the dissertation requirement of the degree of Doctor of Philosophy.

Dissertation Director Date

After inspection of the final copy of the dissertation, the following members of the Final Examination Committee concur in its approval and recommend its acceptance:*

*This approval and acceptance is contingent on the candidate's adequate performance and defense of this dissertation at the final oral examination. The inclusion of this sheet bound into the library copy of the dissertation is evidence of satisfactory performance at the final examination.
STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

SIGNED: John H. Maffett
ACKNOWLEDGMENTS

I would like to express my appreciation to my dissertation committee, Dr. Roy F. Blake, Dr. Howard T. Roberts, and Dr. Robert T. Grant for their perceptive questions and enthusiastic support. The model is much better as a result of their valued comments.

A special note of thanks is extended to Dr. Blake, my dissertation director, for providing the needed guidance, assistance, and cooperation to make this study a very rewarding effort. His unselfish devotion of time, and his genuine interest in the topic have made the model more meaningful to the field of education.

Particular thanks are also extended to Dr. Robert L. Baker and Dr. Fredrick J. Hill, who served as my minor committee members from the Department of Computer Science. Dr. Baker's interest and encouragement during my entire graduate program is greatly appreciated. Dr. Ralph Griswold, Department Head of Computer Science, is also thanked for his support and cooperation.

Others to be acknowledged are Dr. Richard L. Mann, The University of Illinois, and John W. Rust, The University of Cincinnati, for their prompt and thorough response to requests for information. Mrs. Aleen Klaas and David Butler
of the Registrar's Office, The University of Arizona, are thanked for their unending moral support.

My deepest appreciation is extended to my parents, J. H. "Curley" Moffatt and Helen Moffatt, for their inspiration and constant thoughtfulness. My mother served as typist for the study. Her skill in editing and her desire for perfection are responsible for the study's appearance. Her eternal patience and dedication to the task are truly appreciated. My father, who also acted in an editing capacity, is thanked for providing me with the practical knowledge and ambition so necessary to succeed in this undertaking.

I am truly fortunate to have had the able assistance of these many persons and others, who willingly provided help and encouragement. I sincerely thank them all.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF ILLUSTRATIONS</th>
<th>viii</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>ix</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>x</td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Purpose of the Study</td>
<td>3</td>
</tr>
<tr>
<td>Significance of the Problem</td>
<td>3</td>
</tr>
<tr>
<td>Assumptions</td>
<td>7</td>
</tr>
<tr>
<td>Limitations</td>
<td>8</td>
</tr>
<tr>
<td>Definition of Terms</td>
<td>9</td>
</tr>
<tr>
<td>Summary</td>
<td>11</td>
</tr>
<tr>
<td>2. REVIEW OF THE LITERATURE</td>
<td>12</td>
</tr>
<tr>
<td>Data Base Oriented Information Systems</td>
<td>12</td>
</tr>
<tr>
<td>Current Educational Information Systems</td>
<td>18</td>
</tr>
<tr>
<td>Use of Data in Managing Higher Education</td>
<td>24</td>
</tr>
<tr>
<td>Student Data</td>
<td>26</td>
</tr>
<tr>
<td>Financial Data</td>
<td>28</td>
</tr>
<tr>
<td>Personnel Data</td>
<td>30</td>
</tr>
<tr>
<td>Facilities Data</td>
<td>31</td>
</tr>
<tr>
<td>Curriculum Data</td>
<td>35</td>
</tr>
<tr>
<td>Summary</td>
<td>37</td>
</tr>
<tr>
<td>3. METHODS AND PROCEDURES</td>
<td>40</td>
</tr>
<tr>
<td>Summary</td>
<td>47</td>
</tr>
<tr>
<td>4. THE INTEGRATED DATA BASE MODEL</td>
<td>49</td>
</tr>
<tr>
<td>Description of Requirements.</td>
<td>50</td>
</tr>
<tr>
<td>First Dimension--Organizational Structure.</td>
<td>52</td>
</tr>
<tr>
<td>Second Dimension--Data Types</td>
<td>54</td>
</tr>
<tr>
<td>Third Dimension--Functional Users</td>
<td>71</td>
</tr>
<tr>
<td>Data Definition</td>
<td>79</td>
</tr>
<tr>
<td>Data Element Dictionary</td>
<td>80</td>
</tr>
<tr>
<td>Other Standard Terminology</td>
<td>84</td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS—Continued

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Structuring Techniques</td>
<td>96</td>
</tr>
<tr>
<td>Terminology</td>
<td>96</td>
</tr>
<tr>
<td>Security</td>
<td>101</td>
</tr>
<tr>
<td>Data Access and Structure</td>
<td>109</td>
</tr>
<tr>
<td>Retrieval</td>
<td>118</td>
</tr>
<tr>
<td>Data Base Content and Structure</td>
<td>121</td>
</tr>
<tr>
<td>Tables and Directories</td>
<td>121</td>
</tr>
<tr>
<td>Data Sets, Records, and Logical Relationships</td>
<td>130</td>
</tr>
<tr>
<td>Retrieval of Information</td>
<td>151</td>
</tr>
<tr>
<td>Summary</td>
<td>159</td>
</tr>
<tr>
<td>5. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS</td>
<td>161</td>
</tr>
<tr>
<td>Summary</td>
<td>161</td>
</tr>
<tr>
<td>Conclusions</td>
<td>168</td>
</tr>
<tr>
<td>Recommendations</td>
<td>171</td>
</tr>
<tr>
<td>APPENDIX A. DATA / FILE / REPORT MATRIX FORM</td>
<td>174</td>
</tr>
<tr>
<td>SELECTED BIBLIOGRAPHY</td>
<td>176</td>
</tr>
</tbody>
</table>
# LIST OF ILLUSTRATIONS

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Representation of Organization Levels</td>
<td>53</td>
</tr>
<tr>
<td>2.</td>
<td>Manpower Assignment Code</td>
<td>63</td>
</tr>
<tr>
<td>3.</td>
<td>Representation of Data Types</td>
<td>71</td>
</tr>
<tr>
<td>4.</td>
<td>Representation of The Integrated Model</td>
<td>79</td>
</tr>
<tr>
<td>5.</td>
<td>Program Classification Structure</td>
<td>87</td>
</tr>
<tr>
<td>6.</td>
<td>Hierarchical Representation of Data Base Terminology</td>
<td>97</td>
</tr>
<tr>
<td>7.</td>
<td>Tree Structure Diagram</td>
<td>112</td>
</tr>
<tr>
<td>8.</td>
<td>Network Structure Diagram</td>
<td>112</td>
</tr>
<tr>
<td>9.</td>
<td>Double Linked List</td>
<td>116</td>
</tr>
<tr>
<td>10.</td>
<td>Shared Sublist</td>
<td>116</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Activities at Organizational Levels</td>
<td>55</td>
</tr>
</tbody>
</table>
The purpose of this study is to construct a generic model of a data base that can be used as the nucleus of an information system for higher education.

The structure for the data base is represented by a pyramid. The pyramid is derived by adding a functional users' dimension to the classical triangle representation of information systems where three levels of management and several data types are characteristically shown. The resulting model contains three levels of management: Executive, Managerial, and Operational; five types of data: Student, Financial, Personnel, Facilities, and Curriculum; and four functional users: College/Department, Business Affairs, Admissions and Records, and Institutional Studies.

The review of literature and analysis of existing information systems at several major universities demonstrate the requirements that the data base and the resulting information system must meet. The improvement of both horizontal flow between functional units and vertical flow between management levels is provided through increased availability of data to all users. Retention of redundant data is eliminated as a result of analysis and creation of a Data Element Dictionary. Fragmentation of data by
specialized information systems is eliminated through integration of all data-items into one data base with the capability of a broad range of relationships.

The model has the capability of enforcing standardization while allowing a certain amount of local modification. The resulting flexibility allows an institution to designate the contents of tables in the model, and to structure the Program Classification Structure to fit programs and the accounting structure within the institution. Further standardization is achieved in the model through the description of all data, their aggregation, and their relationships in standard terminology defined for the data processing industry by the Conference on Data Systems Languages--Data Base Task Group.

Technical considerations of the model such as data accessing, file structuring, and retrieval are described for the non-technical reader of the study.

The three major components of the model consist of the definition of tables and directories; the description of the data and record relationships that exist among the five data types; and the description of four types of retrieval that satisfy the user's diverse needs. The tables and directories provide commonality of usage of terms, and the definition and control of each detail data-item in the data base. Four degrees of access authorization are included in the Data-Item Directory for every data-item.
The description of the data and their relationships provide the capability of relating data-items not readily related by current specialized systems design. The four retrieval processes utilize the many relationships that exist between data-items to enhance the flexibility of the model. The retrieval capability is very broad in range and satisfies the information needs at the three levels of the organization; Executive, Managerial, and Operational.

The integrated data base contained in the model provides the conceptual nucleus from which higher education management information systems can be constructed.
CHAPTER 1

INTRODUCTION

In the management of higher education, as in all management functions today, the computer and its associated capabilities are being given an increasing role in record keeping. The use of computers in decision making and control of educational institutions is ill defined and limited to specific functional areas. According to Fried (1973, p. 32) the ability to make effective decisions varies inversely with the amount and quality of data available. The computer is a significant contributor to the voluminous mass of data that descends on management and the attendant clerical functions. There is a limit to the amount of unclassified data that can be successfully comprehended. Fried (1973, p. 35) points out that once the volume of data exceeds the manager's limits of comprehension, he will revert back to his previous method of making decisions using the data that is convenient, but not always applicable.

The concept of providing decision makers with the relevant information they need rather than mountains of unstructured data is a major management function according to Gooding (1973, p. 85). To complement Gooding's opinion, Pendleton (1973, p. 495) states: "Most current
computer-based information systems are not meeting users' needs; particularly timely response, and the availability of coordinated outputs." The four general instructional and facilitative functions of higher education that are involved in this study are:

1. College/Departments.
2. Institutional Studies.

The three levels of management in higher education are defined as:

1. Executive.
2. Managerial.
3. Operational.

The mass of data in higher education can be categorized into five data types which are:

1. Student.
2. Financial.
3. Personnel.
4. Facilities.
5. Curriculum.

It is often necessary for all four functions to have access to portions of all five data types at each of the three levels of management. The term, coordinated output, referenced by Pendleton requires that each function must receive
timely information consistent with both the functional need and management level of the recipient. This study addresses the issue of providing a means of coordination of dissimilar data categories in such a way that information is available in a precise form that will be useful to the component parts of higher education.

**Purpose of the Study**

The purpose of this study is to construct a generic model of a data base that can be used as the nucleus of a university information system. This model can provide the necessary structure for categorizing the information necessary for efficient decision making at the Executive, Managerial, and Operational Levels of each of the instructional and facilitative functions.

**Significance of the Problem**

Alcorn (1973, p. 37) defines a management information system as:

... a set of procedures, people, equipment and materials (a complex of elements) which, when operating together (standing in interaction) collects administrative data and transforms it into management information for the purpose of allowing institutional decisions to be made with timeliness and accuracy.

Higher education has always had administrative data, but these data are not condensed into management information for use in decision making. Usually, no distinction is made between those data required for day-to-day operating control
and data required for long range planning. Kornfeld (1968, p. 18) and Manning (1973, p. 7) both cite the need to recognize the differing information needs at each level of management. Mosmann (1973, p. 128) defines the structure of information in the form of a two-dimensional triangle. The triangle includes finance, student, personnel, academic, facilities and auxiliary data types at the "Transactions and Operations" (Operational) level of the structure. The "Control and Coordination" (Managerial) level and "Policy and Planning" (Executive) levels are both described as part of Mosmann's structure. This two-dimensional representation used by Mosmann (1973, p. 128), Manning (1973, p. 7) and others does not provide the necessary delineation of the diverse requirements of various administrative functions. Therefore, the four functions that represent major contributors to, and users of, administrative data have been incorporated into this study and form the third dimension of the structure. With a three-dimensional model, it is possible to show the interrelationships of various data types with the three levels of management in each of the functions.

Does information flow between these three levels of management? Manning (1973, p. 8) indicates that it does not. She describes the lack of continuity between the management levels as the "Information Gap." A gap exists between the raw data needed at the Operational Level and the summaries needed at the Managerial Level. Another gap exists between
the summaries needed at the Managerial Level and the integrated information needed at the Executive Level.

An effective information system must provide a well-ordered set of connections between the levels. Mosmann (1973, pp. 128-129) cites the necessity of refined information being passed upward from the Operational Level to the Managerial Level. In addition, Mosmann describes the further refinement of Managerial Level data to be provided for Executive Level decisions. An understanding of this upward flow of selective information through each level requires a careful description of the interrelationships and needs of administrators operating at each level. The levels are described below in terms that relate them to the model.

The Executive Level is where the objectives of the institution are formulated, and all detail data from the functional areas are compiled for executive summaries. This level is also where a retrieval scheme with the capability of providing a broad range of interrogation is needed most since the types of decisions that arise often require information that is not available on existing reports. Therefore, it is necessary that the structure be designed to facilitate retrieval at the Executive Level.

The Managerial Level is where policies are interpreted and carried out in accordance with the objectives established at the Executive Level. Data requirements at
this level are related to the more diverse functional areas of colleges and departments, Institutional Studies, Business Affairs, and Admissions and Records. These functional units may each have their own information system which consists of data generated by and for their operation. Pendleton (1973, p. 495) defines these as "Independent Information Systems— No data passes directly from one to another in automated form." The data structure within the proposed model will allow for independent system concepts, but also will make specific data items available for access by members of other functional units at the Managerial Level.

The base of the structure is the Operational Level where the detail data is characteristically gathered in each functional unit by multiple application systems. It is at this level, according to Bontempo (1973, p. 31) where a strong tendency to fragment data resources occurs. Kornfeld (1968, p. 17) described this fragmentation as internal inconsistency, where a given item has different meanings and interpretations by different systems, but still carries the same mnemonic identifier. The basic data categories gathered at this level are student, financial, faculty/staff (personnel), facilities and curriculum. These categories of information must be maintained without redundancy and in a manner that facilitates both day-to-day operations and summarization for the other two administrative levels. Kornfeld (1968, p. 18) and Mosmann (1973, p. 128) both
indicate that information provided to a department head or principal research investigator should be summarized for the next higher level administrator.

Administrators in higher education must have certain information available to them to make correct and timely decisions. Often the data needed are in the data base, but not readily available for reporting. Thus, the proper structuring of the data base is extremely important as Kornfeld (1973, p. 17) indicates when he says that a properly coordinated and integrated master file can relate data in ways which are not available through any other means.

Assumptions

The assumption of the study is that the following conditions must exist for the successful implementation of the model.

1. That educational administrators are interested in development and utilization of tools that allow them to perform their duties more efficiently.

2. That the chief executive officer of an institution has developed long-term objectives, and is the motivating force behind the implementation of an information system.

3. That management will make the necessary commitment of time, effort and resources to the development and understanding of the data base concept.
4. That the concept of program budgeting and reporting is accepted as a planning and management technique.

5. That adequate computing and communication facilities exist to support the technical requirements of the model.

6. That sufficient data gathering processes exist to provide the required detail data.

Limitations

The model developed as a result of the study will provide a tool that will assist workers at the Operational Level, and at the same time provide information for administrators at the Managerial and Executive Levels. While there are many entities in the organizational structure, the areas of College/Departments, Institutional Studies, Business Affairs, and Admissions and Records were chosen as being representative of the majority of information system users. Detail implementation plans for subsystems to support any specific areas defined will be left for adaptation to the specific organizations in which they are to be used. No specific computing hardware or software is implied.

The major emphasis is placed on the identification of key data items in the five major data categories of Financial, Student, Personnel, Facilities, and Curriculum; defining a data base structure; and developing the
relationships needed to satisfy retrieval requirements for detail as well as summary information.

**Definition of Terms**

Data processing terms used in the description of the model are defined below.

1. **Alphanumeric**—Used to describe the characteristic of a data field; meaning the field may contain alphabetic, numeric, or special characters.

2. **Data Base Administrator**—One individual in the organization that is responsible for the definition, structure, control, integrity, and privacy of the data base.

3. **Free-Form Field**—A data field that is often quite large and has no requirement for placement of data in any specific format or sequence. Usually used for text.

4. **Generalized Retrieval**—A process of specifying requirements for data retrieval in parametric terms utilizing terminology familiar to the requestor.

5. **Hardware**—A term used in a generic sense meaning the electromechanical equipment that is part of, or associated with, a computer.

6. **Key**—A data field whose contents are defined as the identifier for a record.
7. Link—Normally the logical address or key of record that is associated with the record containing the link.

8. Logical Address—The address used to retrieve a record from a storage medium. Normally, the device number and location within that device.

9. Ordered—Another term for the word sorted, meaning placed in a predefined sequential pattern, usually ascending or descending.

10. Output—A generic term referring to the data derived from a computing system. Most commonly refers to printed reports, but not restricted to that category.

11. Packed Decimal—A method of storing two digits in one storage position, thereby greatly reducing storage space requirements for numeric data.

12. Picture Descriptor—The description of the way data is logically organized in a field. Most commonly used with COBOL (Common Business Oriented Language) programs.

13. Sequential—A description of the logical relationships among a set of records implying ascending or descending order.

14. Software—A term used in the generic sense meaning the programs that cause the computer to operate in a certain manner.
15. Terminal--A generic term referring to an electromechanical device that is connected to a computer via a communication line. The device may perform a limited number of input or output functions normally performed at the computer site, such as printing a report.

**Summary**

This chapter has established the need for investigation of the information requirements of administrators at the Executive, Managerial, and Operational Levels of the various functional units of an institution of higher education. Representative functional units of the instructional and facilitative facets of the institution are included. Terms commonly used in computer science were defined in addition to the three levels of management.

Chapter two will present a review of the literature divided into three areas of importance: data base oriented information systems; status of current educational management information systems; and the use of data in managing higher education. Chapter three is a summary of the procedure used to develop the model. Chapter four describes the model by discussing the detail data area functional unit requirements; data standardization techniques; the CODASYL method of describing a data structure; and finally, the data structure and retrieval possibilities. Chapter five makes recommendations for utilization of the model.
CHAPTER 2

REVIEW OF THE LITERATURE

The review of the literature found in this section concentrated on three areas of importance: data base oriented information systems; status of current educational management information systems; and the use of data in managing higher education. Review of literature regarding the use of data is divided into the five major data categories.

Data Base Oriented Information Systems

Information systems have existed throughout centuries in one form or another. The concept of an information system is basically the presentation of data in a meaningful manner so the user can make use of the result. There have been many examples of information systems throughout history. Hieroglyphics represent one of the earliest methods of formal communication or representation of information. The representation of information has many forms. The various written and spoken languages offer the greatest amount of diversity. Scientific communication methods have been developed to meet the requirement the scientist has for representing an entity that is not precisely defined in any other medium he knows. The Morse
Code developed for electronic communications systems, and the Hollerith Code used to process census data in the 1890 Census, are examples of the most elementary level of data representation. Just as it takes several dots and dashes to constitute a single letter in Morse Code, it takes two or more holes in a data processing card or paper tape to represent a single character. This concept of data reduction also relates to the electromechanical and magnetic representation of data used by computers. These elementary bits, holes, or electronic impulses are collected into what is called data. As Bontempo (1973, p. 37) says, "Data make up the building blocks of meaningful information."

According to Gwynn (1969, p. 10): "The data base is a reservoir of data or collection of facts related by a file structure, or at least catalogued." Gwynn further indicates that the value of the information system depends on the user being able to get what he wants out of it. Ultimately, the data base will become more highly particularized to the needs of the user. The computerized data base concept is a recent development and has not achieved this highly specialized level. Bruun (1973, p. 36) goes further in stating: "Some 600 or 700 companies may already be using data base information systems, but ironically, coherent data base information useful to executives at top decision making levels seems to be almost non-existent." The data base oriented information system is used at the Operational Level of
business and educational institutions. It offers the advantages of consistency among data items, improved control of data and increased flexibility at the operational level. However, the true benefit of drawing all of this data together lies in the ability of personnel at the Managerial and Executive levels to obtain a comprehensive look at data that was at one time widely scattered. Bontempo (1973, p. 36) indicates that data resource fragmentation implies a need for a higher degree of actual data integration achievable only by means of a central repository of data, such as a data base. According to Van Dyke (1968, p. 10), the management information system is said to offer "... significant advantages of a single accurate, timely and readily accessible file that can replace dozens or even hundreds of files scattered among various campus offices."

One of the prerequisites to the development of a data base oriented management information system is the ability to monitor and control the data resources. "The data element dictionary is recommended as a management tool for monitoring, controlling and planning the utilization of data as a system resource (Bontempo 1973, p. 31)." The use of the data element dictionary has been widely interpreted. Bontempo (1973, p. 34) defines the dictionary as: "... a single, authoritative source of information on data elements, their use, and their interrelationships in a standard organization and format. It is a way of monitoring
and controlling data resources without actually integrating and centralizing the data itself." Bontempo further indicates (1973, p. 35) the data element dictionary has eleven major categories of information about each item of data. These are:

1. Identification and Definition—Name, abbreviations, mnemonics, relationships, and definition.
2. Usage—Restrictions and security.
3. Origin—One organization responsible for its creation.
4. Destination—All final users and purpose.
5. Integrity—Responsibility for integrity, validity criteria, degree of accuracy and reliability.
6. Characteristics—Size, type (alphabetic or numeric), data value ranges.
7. Relevant Documents—Documents that describe the item's use, or procedures relating to it.
8. Machine Processing—Applications, programs relative location on input and in data base.
9. Input Processing Frequency—Receipt and processing schedules.
10. Output Processing Characteristics—Output identifiers, frequency of output, format, headings.
11. Data Base Characteristics—Location, special processing requirements, derivation process.
Wearing (1973, p. 29) sees the data element dictionary as a tool in upgrading documentation of a data processing organization. He cites the basic advantages to a data element dictionary as being:

1. Simple, standard information is available for files and data elements.
2. Control of data is simplified.
3. Error rate is decreased through standard validation procedures.
4. Data are standardized for programmers and analysts.
5. Data Base construction is greatly simplified.

There are additional advantages generally outlined by Bontempo (1973), Lawrence (1969) and Wearing (1973). The first is that data redundancy is more readily identified and analyzed. Second, a data element dictionary is useful to types of information systems other than automated systems. Finally, the most important advantage of the concept is that it causes exhaustive analysis and documentation of each element of data which in turn makes development of any kind of system (manual or automated) more complete.

The importance of the data element dictionary to higher education information systems is demonstrated by the dictionaries created by the National Center for Higher Education Management Systems (NCHEMS). The development of a dictionary for the student, staff, finance, facilities, and course data types is one of the Center's objectives. NCHEMS
director, Ben Lawrence (1969, p. 2) describes a data base as an array of data elements which are collected and filed by a management information system. For Lawrence, the compatibility of management information systems: "... depends not only on the collection of identical data elements to comprise the data bases of the systems, but also the standardization of the definitions and formulas used to derive comparable pieces of information from the identical elements." Lawrence et al. (1971, p. 6) further define data compatibility as data sets composed of data elements that are:

1. Defined in exactly the same way.
2. Measured in exactly the same way.
3. Aggregated and summarized by the same procedures.

These criteria, added to the categories established by Bontempo provide a more thorough guide for developing a data element dictionary for higher education. It is not possible to accept only the elements defined by the National Center for Higher Education Management Systems, since they are primarily required for use in comparative models and do not address the data requirements at the daily transaction level. The NCHEMS effort is contradicted by Bruun (1973, p. 36) who indicates that little work has been done at the Executive Level. A combination of the two concepts should provide a data base concept that would support
management information system operations at the Executive, Managerial, and Operational Levels of the organization.

**Current Educational Information Systems**

While there is a great deal of information available on different approaches to a data base oriented management system, not a great deal of that information describes educational information systems. There are several explanations for the lag in educational information systems. Collier (1972, p. 5) points out that historically, the primary need for financial information has been stewardship, i.e., "Did you do with the dollars what you promised to do when you accepted them?" This stewardship is not the profit and loss consideration found in industry. Collier further states that "Planning and management needs haven't exerted the pressure on data collection and maintenance as that exerted by fiduciary needs." The result has been fragmented and specialized systems that satisfy fiduciary needs of specific functional units. These systems do not interrelate the needs of all functional units with regard to planning and management requirements. Rourke and Brooks (1966, p. 21) indicate that in 1966 only 53% of all state institutions of higher education were using computers for administrative purposes as opposed to 56% for private institutions. Of the 53% using computers, 97% were using them for student affairs, 89% were using them for financial administration, 69% for policy planning, and 66% for physical
plant management. According to Gwynn (1969, p. ix): "Widespread management information system implementation in higher education is now on the horizon."

Manning (1973, p. 8) describes one of the major obstacles to successful implementation of management information systems as an "Information Gap." This gap she defines as the "... chasm between capability and reality in information flow to decisions." This information gap is categorized into two subsets. The definition provided for Information Gap--Type I is "... the chasm between raw data and data summarized in a meaningful way, or statistics." The Type I gap describes the break between data required at the Operational Level of the model and the Managerial Level. Manning (1973, p. 8) describes Information Gap--Type II as existing "... between statistics and information." The Type II gap can be equated to the difference between the needs at the Managerial Level and the Executive Level. According to Manning (1973, p. 8) Information Gap--Type II can be closed through "forecasting techniques such as regression, exponential smoothing, time series, and index numbers, coupled with modeling efforts such as simulation, Monte Carlo, and network analysis. However, Information Gap--Type I is still a real dilemma in higher education today." A solution to the Type I gap lies in the development of an integrated data base involving Operational and Managerial Level data.
Another opinion that essentially parallels Manning's is that of Kornfeld (1968, p. 18), who states that an almost universal defect of educational information systems is that the system is not integrated, and that parts of the system operate independently with a good deal of duplication and unnecessary work. The integration of functional unit needs for all types of data, at all levels of management, would end a great deal of duplication of effort. Gwynn (1969, p. 14) attempts to identify some of the factors involved in integration of operating systems:

The relationship between the operating data and information systems in institutions of higher education, such as registration systems, payroll, inventory, and the student information system, and the information system devised for the purpose of assisting in management decisions needs to be clearly established. While the integration of operating systems is technically possible in order to produce information for management information systems across the full spectrum of the college or university, control of operating systems presently remains in various departments or functional responsibilities within the institution. Difficulties with human relations, as opposed to technical difficulties, stand in the way of resolving the management information systems.

Associated Universities, Inc. (1969, p. IV-4) also describes some of the major difficulties of creating an integrated university administrative information system. Some of the difficulties are listed below:

1. The university is a very complicated organization; therefore, it is difficult to represent in a formal manageable language.
2. The design and installation costs per institution are very high.

3. There is apprehension by students, faculty and administrators that the computer will:
   a. Dehumanize the campus environment.
   b. Decrease freedom of action at the local level.
   c. Develop strong central administration control.
   d. Erode the privacy of certain information.

In addition to the difficulties above, Associated Universities, Inc. (1969, p. III-4) suggests a goal of developing a framework of information systems to support decision making and to help implement subsequent operations at all administrative levels. This goal compliments the objectives of the management information system at the Western Interstate Commission for Higher Education (WICHE) which are "... to encourage development of management information systems within higher education institutions which will, at the same time, be sufficiently individualized to serve the unique needs of the respective institutions and be sufficiently compatible to allow valid inter-institutional data comparisons (Lawrence, 1969, p. 1)." To follow through on the objective outlined above, the National Center for Higher Education Management Systems (NCHEMS) has been established within WICHE. The objectives as published on the inside of each of the NCHEMS publications (Martin, 1972a) are:
To design, develop, and encourage the implementation of management information systems and data bases including common data elements in institutions and agencies of higher education that will:

1. Provide improved information to higher education administration at all levels.

2. Facilitate exchange of comparable data among institutions.

3. Facilitate reporting of comparable information at the state and national levels.

Arnold (1973, p. 32) summarizes this as a goal "... to bring up-to-date management techniques within reach of every institution." NCHEMS has provided a great stride forward in the development of management information systems. The scope of NCHEMS has been on a national scale since December, 1969 (Arnold 1973, p. 31). The two areas of initial emphasis chosen are program planning and resource allocation, and a communication base. In order to perform the planning and allocation tasks, a common communication base must be established. Lawrence et al. (1971, p. 7) state: "One of the goals of the NCHEMS Data Element Dictionary and Program Classification Structure is to establish, through concensus, a set of data element definitions and aggregation procedures that will lead to 'NCHEMS compatible' data systems for major areas of institutional activity."

There are preliminary versions of all five data element dictionaries for course, facilities, finance, staff, and student type data (Martin 1972a, 1972b, 1972c, 1972d, 1972e). The Program Classification Structure (PCS) (Gulko, 1972)
has been developed to provide a consistent means of identifying and organizing the activities of higher education in a program oriented manner, and provide a standard framework for arranging the broad range of institutional data. The Program Classification Structure is not a new chart of accounts; it is only designed to supplement each institution's own unique data classification system. It contains the Higher Education General Information System (HEGIS) discipline sectors for compatibility in data for government reports. Gulko (1972, p. 2) comments that standard college and university accounting structures tend to identify line item expenditures with organizational units. These structures rarely provide the means for aggregating data in relation to institutional programs and objectives. Therefore, the basic foundation for developing a common communication base is well under way. The data element dictionaries and Program Classification Structure are still in the preliminary state, but moving forward in a well planned effort. This progress has also made possible the development of a Resource Requirement Prediction Model computer program (Martin 1972f).

Another effort is under way in the area of sharing management systems. The College and University Systems Exchange (CAUSE) was established in February, 1971. CAUSE performs services for member institutions such as providing a directory of administrative data processing directors at
most of the junior colleges, colleges, and universities in the United States and Canada. CAUSE also serves as a clearinghouse for systems to be exchanged and publishes a current list of systems available for exchange as well as an abstract of each system. According to Manning (1973, p. 7) "... the goals of CAUSE are the study, development, and exchange of information systems in institutions of higher education." The exchange of information systems provides an opportunity for the reduction of development costs of redundant systems. If everyone doesn't have a completely customized system, each institution can concentrate on specific application areas to strengthen the entire concept for everyone. This cooperative effort will address the Information Gap--Type I discussed earlier in this chapter. Alcorn (1973, p. 41) summarizes the status of the quest for good management information systems as "... it (MIS) is a useful concept and educational institutions are closer to realizing at least parts of a management information system than ever before."

Use of Data In Managing Higher Education

In order to develop a concept for providing information to the management of higher education, it is necessary to understand some of the demands that are placed on the information system. These demands will come from all corners of the institution. Bruun (1973, p. 32) states that: "Data base systems must contend with many specialized
groups throughout the organization." Associated Universities, Inc. (1969, p. IV-1) provide a more specific explanation of the needs that an information system should satisfy:

In its broadest context, an administrative information system is needed by the faculty to provide each teacher with current information on students to help realize a better learning situation for each student. The information system must also enable each faculty member to reduce to a minimum, the time he spends on paper work. It must provide current information in an easily useable form to those faculty members charged with administering departments, schools, and research projects in order to free them of routine and provide more time for the educational development aspects of their work.

It must provide the President and his staff access to a wide variety of factual data regarding people, the physical plant, equipment, courses, research projects and proposed undertakings in order to develop strong planning efforts and sound information as to alternative courses of action on which to make important educational and management decisions.

Gwynn (1969, p. 11) includes another dimension in the management area by saying that: "The data base is necessary regardless of the anticipated management activity. Those who expect to be doing analytical management functions such as modeling will need essentially the same information as those who are involved in day-to-day decision making."

Alcorn (1973, p. 41) further differentiates the needs of the analyst and decision makers. He defines Administrative Data as being data used on a day-to-day basis while Management Information is usually concise summaries, analyses, digests, charts, etc. of volumes of day-to-day data. Kornfeld
(1968, p. 17) indicates that, in general, few data have been generated on which major planning decisions can soundly be based. These conflicting definitions indicate little agreement upon how data are utilized or prepared for utilization at the Managerial and Executive levels. Mann (1973, p. 10) concludes that the most frequent assessment of the impact of a computer-based management information system was that it has provided the ability to make better management decisions through accurate and readily accessible information. Mann's study also indicates a trend toward increased centralization of administrative decision making and more objective management considerations, as compared to political considerations, when making decisions.

**Student Data**

Mann (1973, p. 5) indicates that the student information area appears to be the most advanced in incorporating new techniques of data processing. The large number of new student system concepts presented at the Annual College and University Machine Records Conference (Ksar and Settle 1972) tends to support Mann's suggestion that student information data processing is an area of great activity. Probably the most advanced overall application involving student information was developed at Stanford University under Project INFO, and titled OASIS. Although the words behind the acronym OASIS (**Online Administrative Information System**)
would indicate otherwise, the system was initially developed with student information (Ksar and Settle 1972, p. 276). Other administrative aspects are now included and implementation of these areas into the OASIS data base concept is one of the primary objectives of the project.

The integration of new administrative aspects into OASIS requires the capability of relating these aspects to the existing student files. A great deal of cross referencing is possible with many indices included in each OASIS file (OASIS System Description, 1970, p. I.2). The student data files in the OASIS Student Records System contain the student data essential to operate most colleges and universities. Ksar and Settle (1972, p. 276) describe the data files as including:

1. Admissions data (scores, institution codes, previous academic achievement).
2. Financial Data (payments, fees, grants).
3. Academic History (courses, grades, averages honors, degrees).
4. Current academic commitments (courses, grades, majors).
5. Housing information.
6. Address information (student, parents).

The files are used for registration, fee assessment, Course enrollment, grade processing, compilation of reports, directories, and rosters, as well as automated input to the alumni
records system. There are many manual and automated systems of lesser sophistication performing the same task at nearly every institution of higher education.

Financial Data

Mann's findings (1973, p. 5) indicate that financial applications are the second most prevalent area of development in computer-based management information systems for higher education. McIlroy (1972, p. 391) describes a very comprehensive online financial system at Texas A. & M. University that is built around a data base concept. However, only the financial data currently exists in a data base mode. Other systems are being designed to interface with the tables and account descriptors in the financial system. The overall data base does not include all types of data. Cermak and Sawin (1972, p. 322) describe an on-line payroll system that has four main modules:

1. Update the Monthly Budget.
2. Update the Payroll Address and Group Insurance.
3. Update the Faculty Status.
4. Update the Payroll Deductions.

In the detail description of each module, it becomes obvious that isolated files are used rather than an integrated data base containing all of the necessary data in one file structure. The needs for financial data are similar in most institutions; however, there are nearly as many solutions to
satisfy these needs as there are institutions. One of the most comprehensive approaches has been that of Washington University. Benson (1973, p. 348) describes the objectives of this system as those which relate the needs for financial data to the anticipated use of the data. The objectives of this approach are to:

1. Provide the university departments with timely financial information; in particular, provide the ability to process and report new fiscal year activity without first closing the old fiscal year.

2. Eliminate or reduce effort spent by departmental and university personnel in keeping handposted records.

3. Provide effective budget control on the broadest possible basis consistent with institutional and granting agency policies.

4. Provide flexible, powerful report preparation capabilities for management information, financial analysis, institutional planning, and external agency reporting: especially, eliminate dependency on meaningful account structures for report generation.

According to Benson (1973, p. 348) these objectives provide a positive step toward building a system that has the flexibility needed to be incorporated into a totally integrated, data base oriented, information system. Collier's
(1972, pp. 31-52) analysis of the type of financial data needed for managing higher education includes the following: identifying fund sources; allocation of resources; and establishing program analysis capability by using Gulko's Program Classification Structure (Gulko, 1972). Ziemer, Young and Topping (1971, p. 93) describe a technique using a summary of the accounting classifications established in College and University Business Administration (Van Dyke 1968), as well as Gulko's Program Classification Structure (Gulko 1972) in an attempt to relate an institution's current chart of accounts to the Program Classification Structure. This technique is called "Activity Crossover" and is developed in support of analysis of the purpose for which expenditures were incurred (Ziemer, Young, and Topping 1971, p. 116) The Activity Crossover presented in Cost Finding Principles (Ziemer, Young and Topping 1971) offers the basis for determining the many cross references required in creating the complex structure of an integrated data base.

Personnel Data

In Mann's study, he found that another field of emphasis in the use of computer-oriented information systems is the area of processing personnel records (1973, p. 5). Most of the personnel systems described are isolated from the systems that process other types of data. The most frequent relationship discovered is the tie-in to the
payroll and/or budget information in the financial system. Los (1972, p. 462), in describing the objectives of the system at The University of Massachusetts cites an elimination of duplication of records and clerical processing; the creation of an operating budget for the Board of Trustees on a timely basis; and the ability to meet new demands and growth without requiring a substantial increase in operating personnel as some of the requirements that must be met by an integrated personnel/payroll system. Although these are desirable objectives, the problem of integration of the personnel data with the student; curriculum; and facilities data, as well as the budget data, must still be solved before a totally integrated information system can exist.

Facilities Data

A great deal of attention is being given to the physical facilities by public institutions (Mann 1973, p. 6). A good indication of the interest generated by public institutions is the joint effort between the Planning and Management Systems Divisions of WICHE and The American Association of Collegiate Registrars and Admissions Officers (AACRAO) which culminated in a detailed series of manuals (Dahnke et al. 1971). The Higher Education Facilities Planning and Management Manuals are organized by type of space and organizational unit where appropriate. The space types are:
1. Classroom and Class Laboratory Facilities—The facility types that support instructional activities are broken down as follows: classroom, class laboratory, special class laboratory, and individual study laboratory.

2. Office and Research Facilities—The two facility types included are office and office-related facilities, and research and graduate training facilities.

3. Academic Support Facilities—The facility types are divided into study facilities (libraries); museum, gallery, and other exhibition facilities; audio/visual facilities; and computing facilities.

4. General Support Facilities—The facility types in this category include athletic/physical education, residential and dining, student health, other student services, physical plant, and miscellaneous general use and special use facilities.

In addition to manuals concentrating on each of the four areas above, an index, glossary, and exhaustive bibliography are provided for easy use and clear interpretation of the other manuals. Finally, a manual is provided that outlines the basis for facilities planning procedures at the institutional level. This allows the user to differentiate between the processes and data requirements for analysis at the institutional and state levels. Dahnke et al. (Manual 6, 1971,
p. 66) define the minimum data required for the analysis of facilities data as:

1. Organizational unit to which the room is assigned.
2. Room type.
3. Function.
4. Area.
5. Number of stations, where appropriate.

The interrelationship of facilities data with student, course, and staff data is clearly identified as an integral part of the analytical process (Dahnke et al. Manual 6, 1971, p. 62).

Bareither and Schillinger (1968, p. 19) state that: "The key to properly analyzing the existing physical facilities of any institution of higher learning is the establishment of an inventory system containing the proper data, with provisions for rearranging, reporting, and updating." Their description of the necessary facilities data requirements includes characteristics of the data as well as the kinds of data to be retained. The key contents of an inventory are:

1. Description of all facilities and their use.
2. List of all facilities assigned to each department or administrative unit.
3. Breakdown of the space according to the various functions of the university (teaching, research, and extension).
4. Breakdown of data in elements which will permit rearrangement and flexibility to produce information in the desired format.

5. Provision for updating existing data.

6. Compatibility with the methods used in projecting space requirements.

The requirements for course, student and staff data as well as facilities data for internal analysis are also described by Bareither and Schillinger (1968, p. 42). Some sample reports are developed to show how many of these data may be combined for practical analysis and other uses (1968, Appendix B).

Bareither and Schillinger's concept of standardization of coding structures (1968, p. 19) coincides with the codes and categories used by Martin's Resource Requirement Prediction Model (RRPM-1) (1972f, p. II-8). These codes and categories are contained in the Higher Education Facilities Inventory and Classification Manual (Romney 1972) which Bareither and Schillinger assisted in developing. The integration of many types of standardized data takes place in RRPM-1 resulting in an excellent tool reflecting types of instruction, space, staff, student, and course data all defined by either Gulko's Program Classification Structure (1972) or some aggregation of detail data element categories defined in Martin's series of data element dictionaries (1972a, 1972b, 1972c, 1972d, 1972e).
Curriculum Data

In general, there are three major categories of Curriculum data which are:

1. Courses available in the past.
2. Current courses available.
3. Planned courses.

Analysis of past, and current courses, their demand, and enrollment projection provides the foundation for prediction of future curriculum requirements. The RRPM-1 module used for facilities analysis may be used to predict curriculum needs. This is done by means of an Induced Course Load Matrix (ICLM) which is an array that describes the distribution of the average load placed on the various academic departments (disciplines) by students of various student levels and majors (Dahnke et al. Manual 7, 1971, p. 16). The ICLM must be analyzed with a number of years of data to identify any instability and pressure of identifiable trends (Dahnke et al. Manual 6, 1971, p. 44). While the large volume of data required may be difficult to gather initially, the use of an integrated data base organized around the WICHE Data Element Dictionaries (Martin 1972a, 1972b, 1972c, 1972d, 1972e) will allow the creation of the ICLM with automated procedures. Dahnke et al. (Manual 6, 1971, p. 46) suggest two ways of reducing the magnitude of gathering data to support curriculum research.
1. Reduce the number of student levels used.

2. Reduce the amount of course data by aggregating the data on individual courses by levels of courses, (i.e., lower division, upper division, and graduate).

Bareither and Schillinger (1968, p. 43) define a similar, more simplified concept of computing Full Time Equivalent (FTE) students by class level (freshman, sophomore, etc.), through a summarization of all Semester Credit Hours (SCH) by class level. The resulting FTE amount is an index of instructional effort by department and by class level. The level indication is useful in evaluating the requirements of students as they progress from the freshman through the senior and graduate levels.

A developing use of curriculum data is in the area of automated degree check systems. The retention of the student's academic history; the catalog or curricular data, past, current and future; and multiple degree requirement profiles (Moffatt 1971, p. 19) will allow a great deal of analysis to be performed by the student, the advisor, the department head, and Registrar's Office. The Pennsylvania State University degree audit system includes 236 majors and provides the data base for transcript production and updating; grade discrepancy checking; registration and current enrollment reporting; and drop-add and late registration processing (Hussey, Berg and Bennett 1971, p. 111).
Although the curriculum area is still relatively new in the data base concept of managing higher education, Mann (1973, p. 5) found that a number of respondents to his survey were including curriculum in their computer-based management information system. The RRPM-l program uses curriculum planning as an integral part of its operation, and most schools have their current catalog automated. However, there remains a large gap between actual curriculum implementation and possibilities using the data base concept.

Summary
This chapter has provided an overview of the literature relevant to the development of education management information systems. First, the concept of data base management information systems was discussed. Second, the status of current educational management information systems was reviewed with regard to deficiencies in their current concept. And finally, the needs for each of the five types of data was reviewed with regard to their prospective use in managing higher education.

The advantages and disadvantages found in literature regarding information systems are summarized below. Associated Universities, Inc. (1969, p. IV-3) has created a list of the advantages of a widely used administrative information system for higher education. Use of such a system will:
1. Provide more and better information.
2. Provide information in time to be useful in current decision making.
3. Lower the cost per unit of information handled.
4. Reduce the amount of file duplication.
5. Reduce the number of errors.
6. Free faculty and administrator's time for more important purposes - better use of scarce resources.
7. Help the institution discover:
   a. Where it is.
   b. Where it is headed.
   c. Where it should be going.
8. Cause the institution to think about itself as a system rather than as a collection of arbitrarily related units.
9. Consider alternatives and monitor progress toward goals.
10. Provide a means of comparing one institution with another.

Pendleton (1973, p. 503) cites significant disadvantages of those same administrative information systems. The disadvantages are that:

1. The system must be designed as a whole unit.
2. Correction of design errors is very difficult.
3. Consolidated data allows propagation of undetected errors.

4. Higher level of computing hardware and software involvement is necessary.

5. Security, backup, and recovery problems increase greatly.

As Kornfeld says: "It is not enough to use a high speed, expensive computer to provide the same data which we have been obtaining in the past (1968, p. 17)." The well designed, data base oriented management system goes several steps beyond providing the same data that had been available in the past. The system can provide a new dimension in operational management if designed, implemented, and used properly. The interrelationship of the various data items required to manage higher education are extremely complex. Alcorn summarized the situation well when he said, "Even though the concept of a MIS is not dependent upon computers per se, an operational MIS is, in most cases (1973, p. 38)."
CHAPTER 3

METHODS AND PROCEDURES

An overview of the procedures used to analyze the components of the data base model is provided in this chapter. An inquiry into the various approaches used to gather, retain, and report data used in the management of higher education was conducted. The two-dimensional, triangular structure set forth by Mosmann (1973, p. 128) served as the basic model for gathering data. A third dimension of four basic functions generic to institutions of higher learning was added to the present model.

Data types as well as detail data-items were obtained from the publications of the National Center for Higher Education Management Systems (NCHEMS) at the Western Interstate Commission for Higher Education (WICHE) and from institutions of higher education throughout the United States. Institutions were chosen on the basis of their success in automating one or more of the functional units under study. After the detail data were gathered, an analysis of the various types of data structures (i.e., list structures and tree structures) that were technically feasible was performed. Concurrently, the Conference on Data Systems Languages (CODASYL) Data Base Task Group Report was reviewed to insure that
terminology utilized in describing the data base model would be consistent with data processing industry standards. When all of the alternative methods of data organization had been considered, the three-dimensional, descriptive model was developed which included data types, functional units, and levels of operation.

The first step in the model construction was the development of the pyramid structure that encompassed the four functions to be studied on one horizontal axis; five data types on the other horizontal axis; and the three levels of operation on the vertical axis. Mosmann's triangle model (1973, p. 128) was a two-dimensional model involving data types and the three levels of operation, which although useful, did not provide the necessary delineation of the diverse requirements of various administrative functions. Therefore, the four functions felt to represent the major contributors to, and users of, administrative data were incorporated into the model as a third dimension.

The inclusion of the functional units made it possible to create a model that not only embraced detail data, but also provided a method of showing the interrelationships of various data types with the functional areas that use them. The four functional areas have been included because of their broad range of data-item requirements. Each functional unit used many of the same items used by the other functional
units; however, these units also contained many items that were mutually exclusive, and thus required a broader construct for their categorization.

The top of the model described the Executive Level. Functional and data categories were not relevant at this level due to their high degree of summarization; also, the need for the Executive Level to observe all functional areas of the institution required a different category system.

Once the structure of the model was determined, the next operation was to determine the categories for placing detail data-items needed to support operation of the institution at all three levels. The data types chosen had been selected by the National Center for Higher Education Management Systems (NCHEMS) for their work in developing management systems for higher education.

The primary purpose of the NCHEMS effort was to help develop program planning and resource allocation techniques, and a common base of communication through the definition of specific data items in data element dictionaries. The defined purpose of specific data-items was to serve as input to simulation models and for the comparison of similar kinds of data between institutions. The data were well defined and provided an appropriate format for establishing a data element dictionary required as part of the data base concept.
These data-items did not provide the detail necessary to operate the institution on a daily basis at the lowest or Operational Level.

A rationale for classification, grouping, and summarization using HEGIS Discipline Sectors and a Program Classification Structure developed by Gulko (1972) provided a useful scheme for categorizing all the items of data gathered within each data type. Methods of utilizing the summaries and models suggested by the National Center for Higher Education Management Systems were available from Chaney (1969) and Collier (1972).

More detail level data was obtained by reviewing the detail data elements used in functional subsystems by Stanford University, The University of Cincinnati, The University of Illinois, and The University of Arizona. Much of these data were analyzed by use of a Data/File/Report Matrix form which readily showed redundancy of data items, (A sample of this form is in Appendix A). The dual purpose of redundancy analysis was: first, to identify possible common data-items that would provide linkage between two or more types of data; and secondly, to eliminate the retention of the same data-item a multiple number of times in the same file structure. The criterion for elimination of multiple entries was established as part of the integration process described on the next page.
The actual relationships of the data within a data base were considered extremely critical; therefore, careful analysis of the advantages and disadvantages of each type of file structure was necessary to arrive at the optimum combinations of data relationships. Flores' (1970) exhaustive review of file structuring techniques were important to the model developed in this study. Suboptimizing a file structure of data-items as they applied to a single functional area, though seemingly appropriate, did not always satisfy the requirements of other functional units. The necessity of the model serving all functional units with all necessary data dictated multiple accessing paths. Organizational conflicts, and attitudes toward continuing with suboptimization have resulted from attempts at providing multiple access capability. The structure of the data base must be supportive of the immediate needs of the institution and allow for smooth progress in the implementation of the system development plan (Gwynn 1969, p. 12).

A significant step forward in the computer industry occurred when the Data Base Task Group was formed as a part of the Conference on Data Systems Languages (CODASYL). The group issued a preliminary report in 1969; and then provided a more definitive report in 1971 (CODASYL Data Base Task Group Report 1971). The 1971 report had two purposes: first, to define those elements that would be necessary for an independent data description language; and secondly, to
define those elements that would enable the COBOL programming language to interface properly with data described in that independent data descriptive language. If the model was to be readily adapted to the CODASYL standards, then it had to be defined in the proper terminology. Since many of the terms were new, their definition and use in proper context was of great importance. The terminology describing the data base and the structure of the model often were directly related; thus, the CODASYL terminology was used to describe the logical relationships in the model.

The final step in the design of the model was to integrate the data-items so that they were representative of the organizational structure of institutions. In addition, the model had to meet criteria designed to provide a maximum amount of information with a minimum amount of associated costs with respect to computer time, storage media, and personnel time. The combination of machinery, technology, people, and data had to be appropriately integrated to provide the optimum information for the decision maker, regardless of his level (Executive, Managerial, or Operational). According to Alcorn (1973, p. 38), information for sound decisions must come from all aspects of the institutional program. The criteria used to judge the usefulness of institutional information from the administrator's position were indicated by Alcorn (1973, p. 38) as:
Accuracy—Does it represent reality?
Timeliness—Is it received in time to act effectively?
Comparability—Can it be compared to standards to assist in evaluation?
Comprehensiveness—Does it include all pertinent data?

A building block approach was used to develop a pyramid structure for the model. Detail level data were used in the foundation of the pyramid with subsequent higher level data reflecting summaries of the detail data. In order to define the various detail data-items consistently, a Data Element Dictionary was used. The Data Element Dictionary provided a clear, concise description of each detail data-item, as well as summary items that were retained and summary items that were derived. Once the task of creating a Data Element Dictionary for each functional area was reviewed, it was then possible to analyze the various dictionaries for discrepancies and redundancies. Criteria presented by Dyba (1973) and other authors for evaluating the Data Element Dictionaries were:

1. Data-item definition must be consistent regarding representation, editing, measurement, and summarization of common terms.
2. Data-items at the lowest level must be in discrete, irreducible terms.
3. Data-items that are derived must be the result of combining items that are consistently defined.
4. Data-items must not occur in the data base more than once unless redundancy is valuable from a processing standpoint.

5. Data-items must be relevant from the standpoint of:
   a. Needs of users.
   b. Automation requirements.
   c. Security requirements.

If the data-items met the criteria above, they were then incorporated in the model at the location and level that they most effectively represented. The criteria developed from Krauss (1970) and Kanter (1972) for placement were:

1. Data collection.
2. Data classification.
3. Data demand.
4. Data volume.
5. Data privacy.
6. Data integrity.
7. Data organization.

Summary

This chapter has outlined the process used to develop the model. The process began with analysis of the model structure required, and was followed by identification of selection criteria and data-items to be included in the structure. Data structuring techniques currently available to support the model were reviewed in order to determine
optimum organization criteria for specific data-items. Upon completion of the data structure review, predefined placement criteria were used for placement of data-items in the structure. In addition, a retrieval scheme was developed. The result was a model of a data base structure that can respond to the information needs at all levels of the institution.
A description of the three-dimensional model is presented in this chapter. First, the general needs for such a model are summarized, followed by a summary of the requirements of each component of the three dimensions. The requirements summary is followed by a discussion of the importance of standard data definition and a description of the methods employed in the model to insure the implementation of those standards.

Due to the technical nature of the data base model, the inherent parts of data structuring techniques are described to familiarize the reader with terminology and concepts that are used in describing the internal structure and relationships of the model. The model is described in three major sections. The first section describes the tables and directories used to insure standard interpretation of codes and generally support the operation of the model. The second section contains the description of each of the five data types in terms of relationships with other data types as well as relationships among records within any one type of data. The third section describes the retrieval process which is included in the model to demonstrate four basic
methods to be used by functional users, at all levels, to obtain desired data from the data base.

**Description of Requirements**

The requirements for information in higher education have been discussed briefly in the introduction to this study. This section of the study discusses: current conditions that necessitate use of a data base; general requirements the model satisfies; and the requirements of each component of the three dimensions of the model.

The conditions that currently exist in most institutions of higher education are a result of specialized systems being developed to solve a problem in one functional unit of the institution. As the use of data processing equipment has grown in the administrative area, more specialized systems have evolved. The result of the emphasis on specific solutions to problems in specific functional areas is that the data resources of the institution are fragmented. A great deal of redundancy exists among data files used by different functional areas. Fragmentation of data and data redundancy demonstrate a need for coordination and standardization of data-items among information systems serving the institution. The data base model addresses that need.

In order to provide data and controls for operation of the institution, four major categories of data must be maintained by the model. These data categories are classified as:
1. Historic--Data are recorded once for the purpose of description of past events, and normally does not change.

2. Operational--Data are generated in day-to-day operation of the institution, which includes modification.

3. Special--Data are generated which are pertinent, but these data do not necessarily fit standard classifications, such as a catalog entry in text format.

4. Control--Data are generated that are used to validate reliability of specific data items, transaction processes, or the entire data base.

The historic, operational, and control data are normally retained in both detail and summary form to satisfy various demands placed on the model by functional users at each of the three levels.

The final general requirement that the model satisfies is that of flexible reporting. The specialized systems mentioned above are capable of satisfying the reporting requirements of administrators that use the data in that system. The model not only satisfies those same requirements met by specialized systems, but also provides a reporting capability that is flexible enough to meet the needs of all functional users at each of the three levels of the model.

In addition to the needs of functional users, the number of reports to funding agencies external to the
institution is growing rapidly. The model provides the flexibility to respond to funding agency requests through two principles. First, the data are organized in such a manner as to relate all components of the institution that are involved in a specific program. Secondly, the model includes a retrieval mechanism that is capable of satisfying both routine reporting and special interrogation reporting requirements.

The acceptability of the model is dependent upon satisfaction of the general requirements stated above. In addition, the requirements of each component of the three dimensions must also be described and subsequently satisfied by the model.

First Dimension—Organizational Structure

To provide information that is useful to management, it is important that the organizational structure of the data base approximates the organizational structure of the institution. Manning (1973, p. 8), Mosmann (1973, p. 128), Kanter (1972, p. 6) and many other authors in the field of management depict the typical organizational structure in the form of a triangle. These same authors identify three distinct levels within the triangle. The triangle represents the hierarchical form of organization structure which is prevalent in higher education with a president or chancellor presiding over the institution, and vice-presidents responsible for various administrative and academic functions. The
organization structure broadens as the various heads of colleges, departments, and administrative offices are included in the organizational structure. Each of these administrators, in turn, has several employees who together carry out the detail objectives of that specific unit. The three levels in the figure below are commonly referred to by many authors as Top Management (Executive Level), Middle Management (Managerial Level), and Operating Management (Operational Level). The three levels in the model are represented in Figure 1 below.

Figure 1. Representation of Organization Levels
The requirements for information differ between the three levels in the organization structure, thus the model provides the necessary information to support the activities of personnel at each of the three levels. Information requirements vary, depending on the level, from detail accounts payable transaction data at the Operational Level to five-year admission trends at the Executive Level. Table 1 compares the range of activities of personnel from each of the three levels for several characteristics. This table shows the basic differences in activity and consequently the differences in data requirements of the three levels.

In addition to satisfying differences in information needs discussed for the three levels, it is important that the model provide the necessary upward flow of data through categorization and summarization techniques to insure that summaries at the Executive and Managerial Levels correctly reflect detail data at the Operational Level. The categorization of programs and activities at the lower levels reflects the intent of those programs as defined at the Executive Level. Therefore, the model facilitates flexible categorization of programs and subsequent summarization of the categories for administrative review through use of the Program Classification Structure.

Second Dimension—Data Types

The three vertical levels of the model have now been described. The second dimension of most models of an
Table 1. Activities at Organizational Levels

<table>
<thead>
<tr>
<th>CHARACTERISTIC</th>
<th>EXECUTIVE LEVEL</th>
<th>MANAGERIAL LEVEL</th>
<th>OPERATIONAL LEVEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning Activity</td>
<td>Heavy</td>
<td>Moderate</td>
<td>Minimal</td>
</tr>
<tr>
<td>Control Activities</td>
<td>Moderate</td>
<td>Heavy</td>
<td>Heavy</td>
</tr>
<tr>
<td>Time Reference</td>
<td>One to Five Years</td>
<td>Up to 1.5 Years</td>
<td>Daily</td>
</tr>
<tr>
<td>Scope of Activity</td>
<td>Multifunctional</td>
<td>Functional Unit</td>
<td>Task</td>
</tr>
<tr>
<td>Structure of Activity</td>
<td>Unstructured</td>
<td>Moderate Structure</td>
<td>High Structure</td>
</tr>
<tr>
<td>Product of Activity</td>
<td>Policies &amp; Plans</td>
<td>Schedules and Measurements</td>
<td>Instruction and Detail Paperwork</td>
</tr>
<tr>
<td>Level of Complexity</td>
<td>Very complex</td>
<td>Moderately Complex</td>
<td>Very Simple</td>
</tr>
<tr>
<td>Information Utilized</td>
<td>Internal Summaries and External Data</td>
<td>Internal Controls and Summaries</td>
<td>Detail Data</td>
</tr>
<tr>
<td>Mental Attributes</td>
<td>Creative and Good Judgment</td>
<td>Responsible and Administrative</td>
<td>Efficient and Detail Oriented</td>
</tr>
</tbody>
</table>
information system describes the types of data included. The
types of data used in higher education are described by
Mosmann (1973, p. 128) at the Operational Level. Martin
(1972a, 1972b, 1972c, 1972d, 1972e) lists five categories of
data for higher education. The categories that both Mosmann
and Martin describe which have been incorporated in the model
are:

1. Student.
2. Financial.
3. Personnel (Faculty/Staff).

A brief summary of the kind of information that can be ex­
pected to be found in each data type will follow.

Student Data. Student data included in the model is
defined by Martin (1972e, p. 5) as: "An individual who has
formally applied for admission at the institution. Or: An
individual who is making a demand on the instructional re­
sources of the institution." The student data category con­sists of five groups of related items:

1. Demographic and Biographical Data--Those data-items
that describe the student's current demographic and
biographical characteristics.
2. Previous Educational Experience Data—Designed to define information about each previous educational experience of the student, including the student's terminal secondary educational experience.

3. Admittance Data—Designed to define information regarding the student's admission to a specific institutional program or degree objective. There may be multiple occurrences of some items.

4. Term Data—Designed to define information about a student for each term that a student is enrolled in a course.

5. Course Data—Designed to define information about each course attempted by the student. Usually the basis for generating transcripts.

Certain subsets of each of the five groups or blocks must be maintained historically for proper analysis of trends as well as transcript reproduction and other services for the student. Martin (1972e, p. 7) states that: "Institutions should structure their data bases so as to link these five data blocks appropriately."

**Financial Data.** The accepted standards for financial data used in the model are those established by the American Council on Education and summarized by Van Dyke (1968, p. 141) as follows: "Certain principles of classification and presentation of accounting as well as a standard terminology for institutions of higher education have come to be
accepted, and colleges and universities should maintain their accounts and present their financial reports accordingly."

WICHE has adopted the American Council on Education categories for the basic data items for financial information which further the comparison capability between institutions. The financial resources of educational institutions come from five basic sources (Van Dyke 1968, p. 142) These sources are:

1. Revenues for general institutional operations.
2. Unrestricted gifts and bequests.
3. Gifts, grants, and governmental appropriations that have restricted use.
4. Gifts, grants, and bequests that must be invested intact. The resulting income will provide the revenue source.
5. Proceeds from public borrowing programs such as bonding.

The classification and resulting summarization capability must reflect the appropriate fund groups. These groups are used by Collier (1972, p. 9) in describing the basic use of the fund accounting principle. The origin of the major fund categories is the American Council on Education. The fund groups are:

1. Current Funds.
2. Loan Funds.
3. Endowment and Similar Funds.
4. Annuity and Life Income Funds.
5. Plant Funds.
6. Agency Funds.

It is also necessary to reflect the daily activities of the institution appropriately. There are twenty-two principles recognized as basic requirements for financial accounting and reporting. Van Dyke (1968, pp. 143-154) explains these principles with broad guidelines for including various types of activities. These principles provide a basic understanding of the types of summarization and reporting that will be required.

WICHE (Martin 1972c, p. v) describes data-items that are required only for current or anticipated WICHE program products as well as a few items needed by institutions for operational and reporting purposes. The items described include the Program Classification Structure (PCS) code (Gulko, 1972) which is rapidly becoming the accepted standard for structuring the accounting system to reflect functional classifications and emphasize program management. However, the PCS code is not a replacement for the chart of accounts. The detail data-items to implement budget, payroll, student-aid, accounts payable, accounts receivable, and other financial systems are not specifically included in the WICHE Data Element Dictionary: Finance (Martin, 1972c). Therefore while the model includes all necessary items to meet WICHE requirements, it also includes Operational Level
data-items for daily operation. All data records will include the Program Classification Structure item to insure proper aggregation of values contained in those records for reporting at the Executive and Managerials Levels.

**Personnel Data.** The personnel data for an educational institution often are divided into the teaching (faculty) category and supporting (staff) category. Martin's (1972d, p. 5) definition of personnel is: "All persons considered to be employees of the institution." He excepts two categories from the term *all*. These categories are:

1. Those persons hired for brief periods of consultation.

2. Those persons employed by organizations under contract to provide a service to the institution.

Depending upon the nature and length of time a service is being rendered, contracted workers often are given benefits associated with the institution; e.g., reduced tuition. If such benefits are to be extended, and accounted for within the personnel system, a minimal record must be kept for personnel receiving them. The grouping of data items in logical blocks makes it feasible to collect and retain only the blocks of information necessary to properly represent each employee's status. Martin (1972d, pp. 5-6) defines each block of related data-items as:
1. Demographic and Biographical Data—Those data-items that describe the employees demographic characteristics and cumulative professional background.

2. Previous Educational Experience Data—The data-items that describe each educational experience of the employee.

3. Appointment Data—Those data-items that describe each appointment of the employee.

4. Assignment Data—Those data-items that define the employee's assignments within the institution. Coding of the information is discussed as part of the Manpower Accounting Structure (Minter 1972).

5. Faculty Activity Data—Those data items that describe each faculty member's activities.

It is possible to combine these blocks to allow only the required data types to be stored for each employee. There are many personnel data items that relate to other major data categories such as facilities (office and classroom space), financial (budget position), student (advisees) and curriculum (courses taught). It is important that the cross reference capability of the personnel file be maintained through proper file structures and data-item content.

One method of identifying various relationships arising out of the personnel data is the use of a Manpower Accounting Structure (Minter 1972, p. 6) which has two
dimensions. The first dimension includes seven broad occupational activity categories. The categories are:

1. Executive--Administrative--Managerial.
2. Instructional.
3. Professional.
4. Technical.
5. Office.
6. Crafts and Trades.
7. Service.

These occupational categories can be broken into subcategories. The second dimension of the Manpower Accounting Structure ties to the Program Classification Structure (PCS) by using the program attributes of the PCS, and is identified as a "Program Function Category (Minter 1972, p. 27)." The seven program categories are:

<table>
<thead>
<tr>
<th>TITLE</th>
<th>PCS CODE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instruction</td>
<td>1.0</td>
</tr>
<tr>
<td>Organized Research</td>
<td>2.0</td>
</tr>
<tr>
<td>Public Service</td>
<td>3.0</td>
</tr>
<tr>
<td>Academic Support</td>
<td>4.0</td>
</tr>
<tr>
<td>Student Service</td>
<td>5.0</td>
</tr>
<tr>
<td>Institutional Support</td>
<td>6.0</td>
</tr>
<tr>
<td>Independent Operations</td>
<td>7.0</td>
</tr>
</tbody>
</table>

Minter (1972, p. 29) states: "The purpose of classifying manpower assignments by program function is to identify more
precisely the kinds and levels of manpower committed to specific institutional programs." Program and Subprogram categories from the PCS structure are included in the breakdown of a Manpower Assignment Code. These two codes as well as the inclusion of an Assignment FTE field allow codification of each assignment every member of the faculty or staff is given. According to Minter (1972, p. 9), the system provides the flexibility required by administrators in making assignments on the basis of individual ability and the ability to account for manpower in relation to authorized positions and institutional goals. Minter's Manpower Assignment Code (1972, p. 11) is shown below.

![Figure 2. Manpower Assignment Code](image-url)
In addition to the Manpower Assignment Code, a separate Compensation Survey Title Code consisting of a one-digit faculty rank indicator and a two-digit administrative title code is included in the codification of personnel data. The major purpose of including this code in the model is to provide the capability to respond to surveys from the National Education Association, the American Association of University Professors, the College and University Personnel Association, and the U. S. Office of Education which adopted the codes for the Higher Education General Information Surveys (HEGIS).

The benefits of including a classification structure in the personnel data of the model are outlined by Minter (1972, p. 3), when he states that the system, "... must be capable of accounting for manpower resources assigned to program activities that have been established to accomplish institutional objectives." Minter (1972, p. 3) also indicates that the system will provide a basis for consistent use of the standardized categories throughout the institution. The incorporation of the standardized manpower data structure into the data base provides a large number of potential interfaces with other facets of a management information system.

Facilities Data. In the WICHE Data Element Dictionary (Martin 1972b, p. 5) a facility is defined as: "Any physical structure or space required by the institution for the performance of its programs and related activities."
Facilities data in the model includes two categories. The first category conforms to Martin's definition for a facility. The second category of data includes all equipment used by the institution for the performance of its programs and related activities.

The physical structures and space portion of Facilities data has three basic classifications that are organized hierarchically as follows:

Level 1. Facility Data.

Level 2. Building Data.

Level 3. Room Data.

As required, multiple level two, and level three types of data will be included in the data base; however, facilities that are not buildings will not have these unnecessary data items retained.

Dahnke et al. (Manual 6 1971, p. 66) summarize the data categories required for proper facilities utilization as:

1. Organizational Unit to which the room is assigned.
2. Room type.
3. Function.
4. Area--Assignable square feet.
5. Number of stations--where appropriate.
Bareither and Schillinger (1968, p. 20) divide a sixth category of data, the Designation Element, into the four following parts:

1. Campus Identification Code—Used with multi-campus or multiple institutions such as in a statewide system.

2. Building Number—A unique number assigned each building in a given campus or installation.

3. Building Name—A brief, but meaningful, abbreviation is recommended by Bareither and Schillinger to be carried in the detail record. The model will contain tables that will address the building abbreviation and full name as a function of the Campus Identification Code and the Building Number.

4. Room Number—A unique number assigned to each room including corridors, restrooms, janitors' closets, and mechanical rooms. The numbering scheme must be flexible to allow for subdivision of rooms.

The organizational unit to which the room is assigned is identified by the standard division/department identifier for the institution. This identifier must coincide with the division/department coding used in other segments of the data base to insure proper linkage between segments. The first six digits of the Program Classification Structure (Program, Subprogram, and Program Category) are also suggested as a data-item by Martin (1972b, p. 47). The purpose of the
inclusion of the Program Identification is to provide information as to which program activities are supported by a specific room, building, and facility.

Bareither and Schillinger described the Designation Element as a primary requirement for facilities management. In addition, this combination of data-items also provides the link to the facilities file from the personnel (office), curriculum (classroom number and time), and student (course-room schedule) segments of the data base.

In addition to the physical structures described in the preceding paragraph; the inventory of all equipment used in connection with the institution also is maintained as part of the facilities data. Equipment is categorized into five major categories.

1. Moveable Equipment--Desks, chairs.
2. Livestock--Cattle, chickens, horses.
4. Fixed Equipment--Radio or Television broadcast antenna.
5. Collections--Paintings, books.

The equipment must be further delineated according to ownership category for inventory and reporting purposes. The general categories of equipment in the model are as follows:

1. Owned--The institution holds title to the equipment or property.
2. Leased—Another organization holds title. The institution has agreed to pay for the use of the equipment for a specific period of time.

3. Rented—Another organization holds title. The institution has agreed to pay for the use of the equipment for an unspecified period of time.

4. Contracted—The equipment is provided as part of an ongoing contract with the institution and the funding agency holds title.

5. Government Furnished Equipment—The equipment is federal or state government surplus and is effectively on loan to the institution. If title is transferred to the institution, then the equipment is transferred to the owned company.

In addition to the equipment type and ownership categories, the location (building and room), using departments and person responsible, item property number, and source of funds must be included in the equipment facilities data for cross referencing purposes. The Program Classification Structure also is used to relate specific equipment to a program objective where applicable.

Curriculum Data. The data contained in the curriculum segment of the model is termed a Course and is defined by Martin (1972a, p. 5) as: "An organized set of activities pertaining to instruction in a particular subject matter, which is conducted during a given period of time (usually a
quarter or semester) and for which credit toward graduation or certification is usually given." A subdivision of a course is defined as a **Section**. Martin (1972a, p. 5) defines a section as: "A group of students assembled for instruction in a regularly scheduled meeting of a course." The hierarchical representation of the data is described by Martin (1972a, p. 6) as:

1. **Course Attribute Data**—Designed to describe information about each course offered by the institution. Example: Course identifier, Number of credit hours.

2. **Section Attribute Data**—Designed to describe information on all sections associated with each course. Multiple section attributes may occur for each course. Example: Contact hours, instruction type.

3. **Section Resource Data**—Designed to describe information pertaining to the resources associated with each section. Example: Instructor, meeting time and place.

One of the items included in the course attribute data is the Program Identification which is made up of the first two positions (program and subprogram) of the Program Classification Structure. The Program Identification allows inclusion of the course under the proper category for program analysis. The HEGIS Subject field (Martin 1972a, p. 17) is the third portion of the Program Classification Structure, which is defined as the Program Category. This is a
breakdown of conventional academic subdivisions of knowledge and training and is commonly called the HEGIS Taxonomy (Martin 1972a, p. 53). The HEGIS Subject field further identifies the activity category the course fulfills in a designated program.

One component that is a necessity for a useful curricular data base is the term (semester, quarter, mini-semester, etc.) and actual dates the course is offered. In order to retain multiple terms in the same file structure, the differentiation between historical, current, and future course offerings is critical. The distinction is necessary in the model's structure to provide for differing requirements for these three chronological groupings.

Finally, the course identifier, and often the section attributes, are the primary links between the student data segment and the faculty data segment when it is necessary to retrieve a student's instructor or to produce a final grading sheet. In addition, the course relates the student and facilities segments in regard to rooms and times. The course segment relates faculty and their classrooms (facilities), and provides the capability to analyze the use of facilities for various educational programs.

The curriculum area is ignored in many educational management information systems; however, the model incorporates curricular data as a key link between several other data types in the data base.
The five data types in the model are represented in Figure 3 below. Since Executive Level data requirements are for broad summaries and aggregates, the five distinct data types are not represented at that level of the model.

![Figure 3. Representation of Data Types]

Third Dimension—Functional Users

The two dimensions normally included in classical information systems' models have been described. The third dimension in this model is important because it identifies the users of the system, each having differing information requirements. Proper structuring of the data base in the model provides specialized information to meet the differing functional requirements, while also providing institution-wide summaries for use at the Executive Level.
The complexity of representing every functional unit in a large university necessitates the inclusion of four representative units. These four units provide academic (College/Department), financial (Business Affairs), administrative (Admissions and Records) and Executive Level (Institutional Studies) representatives of the organization. The first three units have operations at the Managerial and Operational Levels of the model, and Institutional Studies represents the needs of many other functional units due to the diverse demands for information placed on the office for data from all three levels in the model. A summary of the requirements of each unit follows.

**College/Department.** The four primary problem areas facing an academic department head are:

1. **Matching faculty and courses to student demand.** This matching process requires information regarding the demand for regular, extension, and special short courses that are currently offered or to be offered in future academic terms.

2. **Maintaining student records.** Student records must be maintained on a current basis for advising, scholarship, and academic probation purposes.

3. **Monitoring financial activities.** Financial activities include the process of developing and monitoring the budget, classification of programs and projects
for distribution and monitoring of funds, and administra­tion of funds received through research grants.

4. Managing personnel. Consideration is given to each faculty member's teaching load and area of speciali­zation, faculty member's promotion and tenure, and faculty member's remuneration.

College requirements parallel those of the depart­ment in areas of student records, budget, and personnel. A broad summary of course availability and demand is required to review courses provided by all the departments within the college. The capability of summarizing detail data by discipline, program and subprogram provides the college dean with meaningful data regarding specific programs established by the college.

There is much greater dependence on historic data at the college level than at the department level. Since the college dean deals with aggregates of departmental figures, it is important that he have historical data from departments and the college as a whole for analysis.

Another major function of administrators at the college level is the projection of future resource require­ments. Information on current requirements, admission trends, financial trends and many other variables is required by forecasting models, such as the Resource Requirement Pre­diction Model (Martin 1972f). The data within the structure
of the data base must provide the necessary detail to allow automated input to the forecasting models for rapid analysis of many combinations of variables.

**Business Affairs.** The Business Affairs Office has the following primary tasks:

1. Accounting for all revenue sources and expenditures.
2. Acting as a clearinghouse for all financial aid and many student loans.

The Business Affairs Office also provides program and project summaries for all levels of management and outside agencies.

In addition to the tasks mentioned above, developing and monitoring the budget requires that budget information be available at the three levels of management for all functional units. Historical budgets and actual expenditures broken down by program and subprogram, department, expenditure type, and fund category are required for prediction of possible changes in fund requirements. Analysis of this data is facilitated by the use of the Program Classification Structure and the Chart of Accounts to identify various combinations of programs and accounts within fund categories (Gulko 1972, p. 27). This analysis serves as a planning tool and can be supplemented by simulation models used for
predicting future resource requirements such as the Resource Requirement Prediction Model (RRPM-1) described by Martin (1972f).

Another analytical tool is the Faculty Activity Analysis where faculty salaries are distributed to cost centers and projects within cost centers; the distribution is based upon the results of the development of a activity/program matrix (Ziemer, Young and Topping 1971, p. 153) from information developed for the personnel segment of the file. The activity/program matrix consists of program categories, from the Program Classification Structure on one axis and faculty activity categories on the other axis. The activity categories are mutually exclusive and exhaustive of all faculty functions that apply to institutional objectives (Ziemer et al. 1971, p. 150. Total program cost by cost center may be computed by adding staff activity categories to the matrix.

Through the cost center concept and other time series analyses of expenditures against budget, the Business Affairs Office has improved tools to perform the fiduciary function and monitor all financial operating aspects of the institution. The data base concept of the model provides the proper data relationships to facilitate use of the techniques described above.
Admissions/Records. The Admissions and Records function is representative of the administrative functions because of the involvement of this office in the areas of:

1. Admission processing. Processing of applications for admission results in the gathering of initial student data-items for the data base.

2. Transfer credit evaluation. The coordination of the evaluation of transfer credit involves all academic departments and colleges. The results of the evaluation are entered into the student's data in the data base.

3. Student record maintenance. While the student is active, his records are constantly updated. The current version of a student's record is used by several other functional areas such as the Financial Aids Office, Student Health Center, Campus Security Office, and the student's college, department, and professor.

4. Course scheduling. The demand for courses is coordinated among students, departments, audio-visual service, the curriculum coordinator's office, and other functions for current and future courses.

5. Statistical analysis. Information from student and course records is supplied in support of studies performed within the institution.
The Admission and Records function deals with the data base at all levels. The clerical functions of processing admissions and investigating errors in course registration are examples of Operational Level activity. The planning of the registration process and reporting of admission statistics and trends are examples of Managerial Level tasks while establishing admissions criteria and goals are Executive Level activities. The Admissions and Records function serves a broad range of users from clerk to president and a wide range of functional activities from academic departments to campus security. The supportive data to serve these users and functions must be well integrated to provide the service in an economical and accurate manner. The model provides a structure to accomplish this task.

Institutional Studies. The Institutional Studies function is included in the model for two major reasons:

1. Retrieval Benefits. The majority of information required at the Executive Level and for response to outside agencies is provided by the Institutional Studies function.

2. Integration Benefit. This functional unit achieves the greatest advantage from an integrated data base allowing cross referencing of data-items since the majority of the studies performed require several types of data that are related in some manner.
The data requirements of administrators at the Executive Level vary. The retrieval process in the model provides flexibility in requesting special reports. The structure of the data base provides multiple entry points for retrieval of the same basic data-item. In addition, summarization criteria tied to the Program Classification Structure provides summary files for use by generalized retrieval programs. The capabilities described above assist Institutional Studies in selecting, organizing, and reporting information requested or required by outside agencies, or members of the Executive Level of management within the institution. Information is available for all data types and selectively summarized for both the Executive and Managerial Levels. Automated input capability to planning models such as RRPM-1 (Martin 1972f) is provided to minimize the manual data preparation required for analysis of a specific discipline or area of concern. The structure of the data base within the model will support these activities adequately.

The four functional users are added to the organizational levels and data types in the model and all are represented in Figure 4. Since Executive Level data requirements are not closely associated with functional users, the four functional users, like the five data types, are not represented at that level of the model.
Figure 4. Representation of The Integrated Model

Data Definition

The definition of the detail data items to be included in an integrated data base involves gathering of detail data-items from various subsystems and establishing standards to be used. The data utilized in numerous subsystems is not normally coordinated from an overall system standpoint. This lack of coordination is the result of most information systems being developed with a specific use and customer in mind. Isolation of systems is due primarily to the parochial attitude of personnel defining the system requirements. The primary consideration in the design of a sub-system is the user's department, rather than the institution as a whole.
The lack of standardization has led to the same data-items being organized or classified differently in different departments.

A solution to the problems of system isolation, data fragmentation, and data redundancy is based upon the concept of a centralized, integrated data base. Under the integrated data base concept, departments are no longer the owners of their own data. An analysis of data sources is made with the result that data-items are collected at the most efficient point. The institution as an entity is the primary point of consideration rather than the department. The central data base becomes a resource that departments contribute to and select from as the need arises. The data base is an institutional resource which replaces multiple departmental resources.

**Data Element Dictionary**

The primary requirement for successful development of a centralized data base is the common definition of all data-items. This common definition makes it possible to integrate the various data types for cross referencing. The medium most frequently utilized for presentation of the common definitions is called a Data Element Dictionary. This dictionary is normally maintained by one individual or office in the organization that has responsibility for coordination of all data-items.
There are many methods of organizing a data element dictionary. The WICHE Data Element Dictionary (Martin 1972a, 1972b, 1972c, 1972d, 1972e) is broken down by data type consistent with the data types used in the model. The University of Cincinatti and The University of Illinois break their data element dictionaries down by data file; the University of Illinois uses a "KWIC Index" to gain access to all the data elements. This index offsets a major deficiency of the usual file oriented dictionaries where the user must know what file contains a data item before the description can be located. In addition, an integrated data base is based on data types, not specific data files. Krauss (1970, pp. 162-163) developed a data standards glossary, which describes the data-item from the users' standpoint. Krauss' glossary also includes the standard representation of the data-item as it is stored, edited and reported, as well as the programming mnemonic used for standard programming routines, and the security level of the data. This technique of including technical standards with the general user definition provides a single source for both the technician and the user to interrogate, thus providing a common point of understanding. The information recommended for the data element dictionary for the model is grouped by data type; i.e., student, financial, personnel, facilities, and curriculum. In areas of overlap where the same data-item is an integral part of two or more data types, a cross-reference entry referring to a
single data-item definition by data type and identification number insures the singularity of definition possibilities, yet associates the data-item with each appropriate data type.

The contents of the data element dictionary required as an integral part of the model are described as follows:

1. Identification Number—Prefixed by an alphabetic identifier for each data type. Unique number for each data-item.

2. Item Name—A descriptive name that reflects the terminology normally used by the users of the item.

3. Mnemonic Name—The coded name used in accessing the data-item with a computer program.

4. Definition—A definition in the users' terms of what the item represents.

5. Access Level—The level of access prevention to be assigned to this item. The level may vary between users.

6. Use—Application systems and program identifiers for every program having access to the data item. A brief discussion of the various uses the item will receive including derivation of other items.

7. Origin—One organization responsible for creating the contents of the data-item and also responsible for its accuracy.

8. Codes or Recording Instructions—Detail codes and a brief description of the meaning of each code; also
the method of converting raw data to the specified codes.

9. Representation--Description of the recording mode for the data-item. Usually described as Alphabetic (A), Numeric (N), Alphanumeric (X), Binary (B), Text (T) or Packed Decimal (P), etc.

10. Validation Criteria--The range of values the data-item may assume. Includes automated correction criteria if the data-item can be corrected automatically or a default value inserted, if found to be in error. Origin field (Number 7 above) indicates the organization to be contacted should automatic correction not be possible.

11. Field Size--Description of the number of positions required to retain the item on an automated storage medium.

12. Picture--The representation of the item in a program format. Decimal places are stated if not implicit in the picture descriptor.

13. Output Format--The format in which the data-item is always represented in output media. If different for different media, each format and its respective medium should be represented.

14. Where Maintained--Every SCHEMA, SUB-SCHEMA, AREA, RECORD and SET; where the data-item is included. File name for non-data base use.
15. Comments--General comments regarding planned use, editing problems, systems external to the data base that are impacted by this item, etc.

16. Effective Date--Date the item was first implemented in the data base or placed in the dictionary.

17. Date of Last Revision--Date the item was last revised for any reason.

18. Revision Number--A cumulative revision number that relates to a log of every revision made to the entire data element dictionary.

There are several references to files and non-data base uses in the eighteen items described. The purpose of a data element dictionary is to provide a common definition for all data-items whether or not they are included in the data base. The common definition is the key to better understanding, and eventual integration of most data-items into the data base, over a period of time.

Other Standard Terminology

The use of a data element dictionary to define all data-items provides a common understanding of internally used terminology. The numerous reports that higher education must provide to outside agencies require that common or broadly defined terminology be included whenever reasonable. The item contents must include terminology commonly
understood by educators. A very simple example is the inclusion of the class level title "Freshman" for first-year students.

WICHE has made common definitions one of the primary steps in developing an information system concept which is capable of providing comparable information from different institutions. Most of these common definitions are gathered from governmental agencies external to WICHE. Many of the definitions or coding structures apply to several types of data, and are discussed first. Then, specific definitions are discussed where they apply to only one or two data types.

**HEGIS Taxonomy.** The basic unit for breaking academic activity into instructional programs is "The HEGIS Taxonomy" (Martin 1972a, p. 53). The separate instructional programs are grouped into general categories with a brief explanation. Specific disciplines are identified and coded within the numeric code range of each general category. An example follows:

- **0700** Computer and Information Sciences, General.
  Includes those subject field designations associated with the design, development, and application of computer capabilities to data storage and manipulation and related computational procedures.

- **0701** Computer and Information Sciences, General.

- **0702** Information Sciences and Systems.
0703 Data Processing.
0704 Computer Programming.
0705 Systems Analysis.
0799 Other, Specify.

The categories and disciplines defined are used in student data to identify the student's major field of study. The categories are part of the Program Classification Structure (PCS) used in financial data. The PCS is defined next. The disciplines of the HEGIS Taxonomy also define the major field of study for each degree in a faculty or staff member's personnel record. The final use of these disciplines in the model is in the area of curriculum. The codes are used to identify the primary subject field of each course contained in the curriculum segment of the model.

Program Classification Structure. The Program Classification Structure (Gulko 1972, p. 13) provides a mechanism which will facilitate the organization of data for reporting purposes, (as mentioned earlier, the PCS is a supplement to the accounting structure, not a replacement). The Program Classification Structure is made up of program elements representing the smallest, unique collection of resources which are output-producing activities. Through the coding structure, it is possible to combine similar program elements which serve to accomplish a defined institutional objective. The structure is a hierarchical representation of
program categories within the institution. Figure 5 presents the various levels that are available for use in the Program Classification Structure (Gulko 1972, p. 21).

The differentiation between Primary Programs and Support Programs is based upon the first digit (Program) of the coding structure. Program codes 1--Instruction, 2--Organized Research, and 3--Public Service represent primary programs. Program codes 4--Academic Support, 5--Student Service, 6--Institutional Support, and 7--Independent Operations comprise support programs. The composition of the 16 digit code is defined by Gulko (1972, p. 26) as:
<table>
<thead>
<tr>
<th>Field</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program</td>
<td>1</td>
</tr>
<tr>
<td>Subprogram</td>
<td>2</td>
</tr>
<tr>
<td>Program Category (HEGIS Taxonomy)</td>
<td>3–6</td>
</tr>
<tr>
<td>Program Sector</td>
<td>7–8</td>
</tr>
<tr>
<td>Program Element</td>
<td>9–14</td>
</tr>
<tr>
<td>Unassigned</td>
<td>15–16</td>
</tr>
</tbody>
</table>

The Program Category contains the categories and disciplines described above for the HEGIS Taxonomy. In addition to the HEGIS categories, PCS program categories for non-academic activities are defined to cover various institutional programs which, while not necessarily academic, contribute to the overall objectives of the institution (Gulko 1972, pp. 105-106).

The Program Sector and Program Element are defined by the institution and are used to provide a linkage to the organization coding structure and Chart of Accounts. The final two unassigned positions are also available for institutional use in developing alternative summarization and reporting categories.

The Program Classification Structure is used to gather information from the financial, personnel, and curriculum data areas. The relationship between the Chart of Accounts and the Program Classification Structure is made possible through the Program Element Field. The Program, Subprogram, Program Category, and Program Subcategory fields
are used to relate each assigned personnel activity to a specific program and objective within the institution. This relationship is further described by the addition of the Occupational Category, Occupational Subcategory and Assignment FTE (Full Time Equivalent) fields to create the Manpower Assignment Code which was defined as part of the Manpower Accounting Structure portion of the Personnel Data section. The first two positions of the Program Classification Structure, Program and Subprogram, are used with the curriculum data to identify the institutional program that the course supports. The HEGIS Taxonomy can also be joined with the first two positions to identify a course through the Program Category level of the Program Classification Structure thus creating an analytical tool for studying the contribution a course makes to a specific program discipline.

Organizational Unit. Another common data-item that affects the financial, personnel, facilities, and curriculum data is the organizational unit category. An internal, standard coding scheme for identifying the division and department associated with each record or record segment, is a primary requirement for integration of data types. Often development of isolated systems for various functions precludes the integration of data due to incompatible coding schemes.

Federal Standards. Several federal agencies have established standards for use in reports that are provided
for their use. The U. S. Department of Commerce, National Bureau of Standards, publishes the following standards used in the model:

1. Two-letter abbreviation for state names.
2. Two-letter abbreviation for each country, dependency, and area of special sovereignty.
3. Five-digit ZIP code for specific geographic locations.
4. Six-digit calendar date in the form of year, month, day; based on the Gregorian Calendar.

The U. S. Department of Justice, Immigration and Naturalization Service publishes standard codes and descriptions for visa-type, which are retained for interpretation and reporting of all alien students and personnel. The U. S. Department of Health, Education, and Welfare reports include five basic ethnic groups. The model's coding scheme for ethnic groups allows for more specific categories that can be summarized into the five groups for federal reports. Experience of several institutions indicates the need for more specific categories for analysis purposes.

All of the data-items described above are used extensively throughout the model to insure consistent information on all reports, not just reports for federal agencies.

**FICE Code.** A unique six-digit numeric code exists for precise identification of every institution of higher education. The code is commonly known as the FICE Code, is assigned by the Federal Interagency Committee on Education,
and is found in an annual publication titled *Education Directory 1972-73 Higher Education*. The number is unstructured and is assigned to an institution permanently. Specific assignment rules are designated to handle branch campus expansion and consolidation or the incorporation of an institution into a new system. This code is used in student data to identify transfer work and institutions awarding prior degrees. Personnel data also contain the FICE Code for each institution awarding the faculty or staff member a degree. Finally, the FICE Code is used as the institution identifier for facilities data in institutions with branch or extension campuses.

**Social Security Number.** The Social Security Number is required for various governmental reports on financial aid for students and earnings reports for personnel. The American College Testing and Scholastic Aptitude Test scores are forwarded to selected institutions with the student applicant's Social Security Number acting as the identifier. An increasing number of institutions are making use of the Social Security Number. One advantage is the ability to use the Social Security Number for easy linkage between financial records and the student's or employee's personal data. The number is theoretically unique for each individual; therefore, it can act as a non-redundant key for file organization purposes. The key to the personnel file is then a part of the facilities data where the office occupant is identified.
This condition follows for the curriculum data where the instructor's Social Security Number is retained as part of the detail section data. Finally, the Social Security Number is used to link students to advisors by retaining the advisor's number in each student's record; and by retaining each advised student's number in a list in the advisor's personnel record. The Social Security Number is a requirement for effective file integration. Other identification numbers will provide much of the same capability with the exception of the financial-aid, earnings, and entrance examination data, all of which are dependent upon the Social Security Number for proper correlation and reporting.

The data-items discussed above affect several data types. The standards set forth for specific data types are as critical as general standards due to the requirement of meeting standard reporting formats for each type of data. The standards that follow apply to one or possibly two data types.

**Financial Data.** The implications of the Program Classification Structure have already been discussed. The Chart of Accounts for the institution must be taken into consideration when developing the financial data items. The items listed below are categorized in general in *College and University Business Administration* (Van Dyke, 1968). Often there are subcategories (division and department) developed for specific institutional needs, and the size of the data
field must allow for this possibility. The general categories are:

1. Fund Group.
2. Balance Sheet Category.
3. Fund Source.
4. Functional Classification.

Contents of these and other data items for accounting and reporting purposes are guided by twenty-two basic principles outlined by Van Dyke (1968, p. 141-154). In addition to the data described above, budgetary categories for classification of expenditures must be included for development of an effective budgeting and cost reporting mechanism.

**Personnel Data.** Although many of the general data items have been described previously, it should be noted that there are three areas that require coordination with other segments of the data base or with specific reporting and analytical processes. These areas are:

1. Use of standard accounting and budgetary categories used in the financial system. The use of these categories facilitates implementation of the Manpower Accounting Structure (Minter, 1972).
2. Use of standard building and room identifiers. The identifiers reference the work station occupied by the staff member and office for faculty members. Classrooms are associated with their respective faculty members via the curriculum file.
3. Use of standard course identifiers. The course identifiers for every course taught reside as part of the faculty member's data, creating a two-way path between the curriculum and personnel data structures.

**Facilities Data.** The model contains coding categories and techniques which represent physical plant facilities in a format usable for internal reporting as well as for comparison between institutions. The coding categories consist of:

1. Program Classification Structure including HEGIS Disciplines.

Through use of the categories above and basic definitions of data and codes provided by Romney (1972, pp. 43-45), the model provides consistent representation of physical plant facilities.

Equipment facility data includes two standard codes established by the Defense Supply Agency (1967), which are:

1. Federal Supply Item Code—A unique number assigned each type of item that might be purchased. The model contains a six-position code to allow subdivision of the four position federal code.
2. Manufacturer Code--A unique number is assigned to each manufacturer of equipment purchased by the institution. The federal numbering scheme is expanded to accommodate local suppliers not on the federal list.

The names and addresses of suppliers, manufacturers, and contracting agencies are all retrieved from standard directories based upon a code assigned to each organization, and carried in the record for each piece of equipment. The location indicator for equipment is the standard institution, building, and room number. The link to the financial segment of the data base is through the responsible division, department, and appropriate account number. Another possible link is through the source of funds data-item, which is the indicator of the account the equipment was charged to, when it was acquired. These data-items provide links for integration of physical plant and equipment facilities into the data base.

Curriculum Data. There are two additional standards required for curriculum data. These are:

1. Instruction Type--This code should be included for each section segment and the coding structure should meet the standards established by the WICHE Faculty Activity Analysis project (Romney 1971) and (Martin 1972a, p. 39).
2. Building and Room Identifier--The standard building and room identifiers are used to indicate the meeting place of each course and section and also, to relate each specific section to the facilities data.

**Data Structuring Techniques**

A brief survey of the security procedures and data structuring techniques to be used as well as the terminology used in the description of the data relationships demonstrates the technical implementation considerations of the model.

**Terminology**

The terminology outlined in the CODASYL Data Base Task Group Report will be presented for an understanding of the standard terms to be used by the technicians in describing the data base for use by a data base management system on a computer. Figure 6 demonstrates how the CODASYL standard terminology describes the basic components and relationships in a limited data base. A similar organization, on a much larger scale, would represent the data base of the integrated model to the data base management system.
Figure 6. Hierarchical Representation of Data Base Terminology
The definition of terms shown in Figure 6 proceeds from lowest to highest level in the naming hierarchy:

1. A DATA-ITEM is the smallest unit of named data.

2. A DATA-AGGREGATE is a named collection of data-items within a record. There are two types of data-aggregate.
   a. VECTOR--A one-dimensional, ordered collection of data-items, all having identical characteristics. Example--A list of Social Security Numbers for all students in one course section.
   b. REPEATING GROUP--A collection of data-items that may occur an arbitrary number of times with a record occurrence. Example--A list of all courses and sections for a student.

3. A RECORD is a named collection of data-items or data aggregates. A specific record type may occur an arbitrary number of times.

4. A SET is a named collection of record types. Each set type must have one record type declared as its OWNER and one or more record types declared as its MEMBER records. Each occurrence of a set must have only one owner record.

5. An AREA is a named sub-division of the addressable storage space in a data base. This term refers primarily to a physical placement of the data on a
specific device (run-unit) within a computing system. However, in the general data base concept, it could also refer to a filing cabinet.

6. A SUB-SCHEMA describes only the areas, sets, records, data-aggregates and data-items that apply to a specific application. The declaration of a sub-schema sets no limitation on the declaration of, or number of other sub-schemas, and one sub-schema may overlap another as shown in Figure 6.

7. A SCHEMA is a complete description of a single data base. It includes the names and descriptions of all areas, set occurrences, record occurrences, and associated data-aggregates and data-items.

8. A DATA BASE consists of all record, set, and area occurrences controlled by a specific schema (see Number 7). The contents of two or more data bases are considered to be mutually exclusive.

The terminology above is essential to the understanding of the structure of the model.

The operational requirements placed on the data base by the information system are directly related to the structure that is developed. The dynamic nature of higher education institutions requires a great deal of flexibility in the structure of the data base. Retention of various data types requires the ability to deal with the following types of data-items or data-aggregates:
1. Fixed Length Field—Social Security Number is always nine digits in length.

2. Variable Length Fields—Title of Dissertation is not the same length for every faculty member.

3. Lists—Titles of Publications do not contain the same number of publications for each faculty member.

4. Null Entries—Laboratory Sections are not included for every course in the curriculum area.

The volume of transactions against a specific data-item or record often determines the location of the item or record in the structure. The volume considered is broken into categories of preformatted inquiry, maintenance, and unstructured retrieval.

The preformatted inquiry process consists of a well defined procedure whereby the user requests a specific report, the appropriate data-items are retrieved, and the report is produced. Items that must be retrieved rapidly or frequently are located in the OWNER record for that type of data.

Maintenance of a data-item is more complex as the number of relationships between records increases. In this respect, careful reconstruction of all links when a record or data-item is deleted or modified is required. Also links must be established when adding records or data-items.

Unstructured retrieval requires highly sophisticated file structuring and maintenance techniques to provide all of
the indices, cross references, and links necessary to support a comprehensive retrieval system. Complex file structure organization is critical to efficient operation of the unstructured retrieval technique.

The density of the information to be retained is a significant factor in deciding how an area of the data base should be structured. If each new incoming record is given a new, unique, sequentially assigned identifier, the contents of the area are very dense, and searching for a specific record is facilitated. The use of a Social Security Number normally creates less dense area contents since the records are sorted by a key derived from the Social Security Number; for nearly every application system, except in the Social Security Administration, the numbers will not run consecutively. Lack of density requires an indexing process to locate the record rapidly. The density of list structures determines whether they are handled sequentially or through an index.

Security

The controls placed upon the data base must provide two types of security:

1. Data Integrity.
2. Privacy (Unauthorized Access Prevention).

Data Integrity. Data integrity refers to the ability to recover the data base with minimal interruption and cost
once it has been destroyed. There are infinite possibilities of sources for data base destruction. Three common sources are:

1. Mechanical failure.
2. Programmer's error.
3. Improper input editing.

The model cannot prevent the first two; however, the improper editing of input is prevented by validation of the contents of every data-item input to the system according to criteria established for the data-item in the Data Element Dictionary. The validation process guards against accidental or deliberate entry of erroneous data. The model also provides complete transaction logging by recording every inquiry, update, addition or deletion, and associated error message along with the type of operation, time received, user identification, and terminal number. There are three primary uses for this transaction log.

1. Statistical analysis of transactions by user, terminal, operator, operation type, error occurrence, and type data.
2. Analysis of the transactions for an indication of the cause of the failure in the data base.
3. Restoration of the data base to a condition that existed just prior to the time the failure occurred.
While the model provides for statistical analysis and transaction analysis data to be retained, their specific use is a function of the implementing institution. Restoration of the data base is an integral part of the model. Restoration can be extremely costly from the user's viewpoint. If a failure occurs during the normal working day, the users in the functional areas must have access to correct data in a minimum amount of time.

The most effective restoration process is called the Edited Input Method developed by Bayless and Sherman (1973, p. 15). In support of this method, the transaction log contains input that has been edited and accepted, and ordered by time received. The log also contains the key for finding the records to be updated and the new field contents and field number for all fields updated by a single transaction. The backup file of the data base is read, and the latest version of each updated field is inserted as the record is restored to the primary storage medium. This process has the advantage of being able to restore selective data areas. The transaction log is small compared to several other techniques, e.g., retaining a copy of each updated record before and after the update action. Restoration time is minimal since only specific fields in specific records are changed as they are restored; and only the affected portion of the data base must be restored as opposed to the entire database requirement under a total rerun concept (Bayless and
The disadvantage of this method is that several transactions affecting the same field must all be handled chronologically in the restoration process in order to create the proper audit trail.

In addition to the transaction log, another key factor in file restoration is the creation of the backup file or files. A backup file of the entire data base is normally created for a given cyclical period. The period used in the model is one week. In addition to the general backup file, a selective backup file is created for each area, set, or specific record type on a more frequent cycle. The duration of the interim cycle is inversely related to the volume of changes that occur to a given area, set, or record. For example, the set of records that contain budget data is backed-up weekly during the seven months of the year when change volume is minimal, but during the budget preparation and review process, the set is backed-up daily. Extremely high volume on-line data entry areas can be backed-up every few hours. As described earlier, the backup file for the lost or destroyed area, set, or record type is read and applicable transactions off of the transaction log are applied to the file for restoration of a useful, up-to-date file.

Privacy. The second security consideration of any type of data base is prevention of unauthorized access. In higher education, there is a real and a philosophical concern
regarding inappropriate release of information. In Hearings on Federal Data Banks, Computers, and The Bill of Rights, Dr. Malcolm Moos, President of The University of Minnesota, testified (Subcommittee on Constitutional Rights):

There is solid evidence, Senator, that military intelligence agents not only secured information from records systems within the university, either from the (campus) police department or from the office of Admissions and Records, but also engaged in what I would describe as photosurveillance of demonstrations and events on the campus . . . (1971, pp. 538-539).

At the same hearing, Eugene Eidenberg, the Assistant Vice President for Administration at The University of Minnesota presented a series of exhibits (1971, pp. 559-569) that outlined the written policy of that institution for release of information. The written policies were useless in the case where office workers became familiar with agents who performed routine security checks for former students who had provided the military with a signed authorization form. Once a pattern was established, the agents were allowed "... total access to confidential student records without demonstrating that the student had granted permission for the investigation (Subcommittee on Constitutional Rights 1971, p. 533)." This example is presented to emphasize that attempts are constantly being made to obtain access to records maintained throughout institutions of higher education.

The model uses the multiple sub-schema capability defined for data base management systems by having an access level coding scheme to match user, terminal and operator to a
a specific sub-schema. A second level of access must be in-
voked after a sub-schema has been accessed. This second
access level consists of the determination of the operator's
authorization for specific types of action such as modifying
or deleting data-items and records. The determination is
made by comparing the access level code assigned to the user,
terminal, and operator, to the access level code established
for the desired data-item.

The data element dictionary items that relate to
automated processing are included in a table in the data base
to allow general access to the dictionary and also to provide
a method of checking the rights of a specific application or
user to access a data-item. This table is called the Data-
Item Directory and contains the following items:

1. Data Element Dictionary Number.
2. Mnemonic name.
3. Item name.
4. Size and representation (numeric, alphanumeric).
5. Output report heading.
6. Output data format.
7. Data value range or pointer to table of acceptable
   values.
8. Exact location in the data base structure (or
   organization for manual files).
9. Access Level—Broken down by user, (College of Engineering), terminal identifier, operator identification (usually the employee's Social Security Number), and sub-schema number.

With the four levels of access protection, it is possible to code extremely sensitive data-items so that only authorized employees may access them. The four access levels discussed above are in excess of the normal password used to obtain access to a specific portion of a terminal oriented computer system.

The Data-Item Directory is also readily adaptable to manual files. The location, name, and access level for every data-item can be retrieved in support of a search for a specific type of material. The data base administrator is the only user allowed access to modify the Data-Item Directory. The location of the Access Level in the Data-Item Directory allows the data base administrator complete control over changing protection levels. The access passwords are changed periodically for additional protection.

There are basically three types of privacy violations which are:

1. Accidental disclosure—This violation normally occurs through program error or the user requesting improper information.

2. Passive infiltration—This violation consists of
observation of normal reports or terminal displays to learn whether some information might be of interest.

3. Active infiltration—This violation consists of a deliberate attempt to secure information through normal procedures or devious manipulation of normal procedures.

Although the four levels of access protection for each data-item introduce an extra cost in processing overhead to an automated system, such protection reduces accidental disclosures and makes deliberate violations much more difficult. Passive infiltration can be a function of terminal location and office procedures of securing all sensitive material. Although technical data security is of great concern in military and commercial organizations, the security of personal data in education is also a matter that database oriented systems must respect and protect vigorously. The access levels designated in the Data-Item Directory offer a comprehensive method of providing the necessary privacy. Thus, the model provides a method for technological protection, but it does not have the means to protect against the employee who misuses authorized information.
Data Access and Structure

An understanding of the methods used to access and structure the data in the model is essential to the comprehension of the relationships that exist between the five data types.

**Data Access.** The terms defined below describe four basic data accessing techniques which are used in the model. The accessing techniques that support data base oriented management information systems are the following:

1. **Sequential**—All records are read no matter how many are of interest. The records are physically ordered in a sequence determined by an identifier which is part of the record.

2. **Indexed Sequential**—Records are physically arranged in some sequence just as the sequential file exists. As the sequential file is created (and modified), an index file of each record's address on the storage medium is created with the identifier for that record. Normally the index is ordered, and when a record is to be retrieved, the identifier is matched to the identifier in the index file. When a match is found, the address of the record is used to retrieve the record directly.

3. **Random**—Records are not arranged in any particular sequence. The identifier is normally converted by a mathematical algorithm to a number which is a logical
address on a direct access device such as a rotating
disk. This method allows for fast retrieval, but
wastes a great deal of the storage space available.

4. Direct—The placement of records under the direct
scheme is the opposite of the random scheme. In di­
ext file accessing, the identifier assigned to a
record, and in turn, the student, is the logical
address on a direct access device. This method
allows the storage space to be utilized sequentially
and access time to be minimized because the logical
address does not have to be derived from some other
number. One disadvantage of the method is that the
resulting identifier often does not make logical
sense to the human users of the system.

**Data Structures.** Each of the access methods avail­
able is useful in implementing techniques for dealing with
complex data under differing needs. The CODASYL Data Base
Task Group Report (1971, p. 2-30) outlines the three basic
data structures to be supported by a data base management
system which are:

1. Sequential Data Structure—This structure is one in
which each element has a relationship to the element
preceding it and the element following it, with the
only exception to the rule being the first and last
elements. A single set occurrence is a representa­
tion of a sequential data structure, or list, whose
elements are complete records. An example of a set occurrence is an encumbrance data file that has an OWNER record for each student and then a list of sequentially ordered encumbrance records that are MEMBERS of that specific student's set of encumbrances.

2. Tree Structure—This structure (See Figure 7) is a direct hierarchical representation where each element may be related to \( N \) other elements at any strata below the stratum of the OWNER element. No element may be related to more than, or less than, one element at a higher stratum (OWNER) except the highest stratum element which is called the "root" of the tree and only has dependents, i.e., is not related to any higher stratum element.

The example of a tree structure in Figure 7 shows the breakdown of contents of buildings on a campus. The campus (institution) identifier would serve as the root with buildings being the elements under the institution identifier. Each building has rooms and each room would have only one building as its OWNER. Most rooms have equipment in them and that equipment normally relates to only one room.

3. Network Structure—The network structure, although similar to the tree structure has one very basic difference; i.e., any element in a network may have
Figure 7. Tree Structure Diagram

Figure 8. Network Structure Diagram

--- Facilities Relationship --- Personnel Relationship
any number of OWNERS and any number of MEMBERS. It is because of this characteristic that the network structure is used to integrate the various data types in the model. Figure 8 demonstrates the difference between a tree structure and a network structure.

In Figure 8 personnel assignments have been added to the rooms in the tree structure shown in Figure 7. Each room stratum element now has two OWNERS, the building, and the instructors that hold their classes in the room. In the case of offices, the resident or residents of the office are the other OWNERS.

Combinations of the three structures can be used to develop the relationships which exist in a highly integrated data base that must serve the entire institution.

A common concept used in many complex file structures is the List, which is normally a type of sequential file appearing at one of the lower levels of a tree or a network structure. Although a list can also be a random file, most applications do not readily adapt to random list processing. There are four types of list structures used in the model which are:

1. Ordered List--An ordered list is essentially a sequential file with no vacancies in it, thus adding or deleting records requires the mass movement of all
records following in the list. Because of the sequential nature of an ordered list, a tree structure cannot be represented by such a list.

2. Linked List—The linked list is preferred when file alteration predominates. Instead of relocating all following records, as required for ordered lists, the pointer(s) in the associated record(s) would be changed. The linked list offers a great deal of flexibility. The records do not necessarily reside in contiguous locations, but contain a pointer to the next record according to the order relation of their identifiers or some other rationale. Often this pointer is called a link. A link may be the logical address on a direct access device, or it may be an identifier which may be used in another area or set of the data base to retrieve a different type of data. Both addressing techniques have an important role in the model.

3. Double Linked List--The model will require the use of a double linked list where a specific record in a list will contain a pointer to the next logical record and the preceding logical record in the list. The double linked list concept allows forward and reverse processing of the list regardless of the entry point. A requirement for complete circular lists is that the first and last records in the list must
carry a pointer to each other. One disadvantage of a double linked list is that two pointers must always be maintained. The ability to satisfy a request for the names of all students enrolled in a class with a specific student whose last name begins with the letter M, is an example of the use of a double linked list. Figure 9 shows a class enrollment list presented in alphabetic order, so the entry is made in the middle of the list. It should be noted that it is possible to begin following the pointers in either direction, and when the last pointer processed points to the original student, the entire list has been captured.

4. Shared Sublist—The shared sublist presents many technical implementation problems, but has the capability to reduce redundancy drastically if it is properly implemented. The sublist normally consists of components that are common to several members of one or more lists. An example (See Figure 10) is a lecture hall ordinarily used by several faculty members and hundreds of students each day. The lecture hall is related to its users, and in addition, is also incorporated into the facilities data. The lecture hall thus can appear in a sublist with a single direction pointer from each student's schedule record for the class(es) he attends in the subject room, and another
Figure 9. Double Linked List

Figure 10. Shared Sublist
single direction pointer from the faculty member's teaching assignment record. The lecture hall is an integral portion of the list of rooms in the building in which it is located. This example (Figure 10) shows the flexibility offered by list structures.

The highly integrated data base management system uses directories to locate areas, sets, records, and sometimes even data-items within a record. The use of a directory minimizes the detail data shuffling required to handle many types of data formats and transactions. The major identifiers and logical addresses of areas, sets, or records are included in the directory. The directory is sequenced by the identifier while the list it refers to may be in random sequence. The concept is essentially the same as that described earlier for index sequential processing except, in this case, every record can have an entry in the directory, and the detail records need not be in sequence or of a specific length.

The combination of rapid retrieval through use of the directory, and flexibility of a linked file structure provides the linked directory concept with advantages of both concepts and only a minimum amount of additional processing. According to Flores (1970, p. 132), the structuring of the directory concept into levels of subdirectories for lists can be optimized by using the formula:
\[ X = \sqrt[N]{M}, \text{ Where} \]

\[ X = \text{Number of directory entries.} \]
\[ M = \text{Number of items in the list.} \]
\[ N = \text{Number of levels of subdirectories.} \]

Flores called this concept a Table of Contents (1970, p. 129). The application of the Table of Contents structure to the model varies by data type.

**Retrieval**

Discussion of retrieval techniques so far has assumed that the primary identifier (Social Security Number) is known by the inquirer. The major reason for building an information system around a data base is to enhance retrieval of several data types in support of one request. Such a request might ask for names of all graduate students that are pursuing a particular major (student data); serving as teaching assistants (personnel data); and receiving another form of financial aid (financial data). Under the fragmentation of files concept discussed in Chapter Two, each file has to be searched sequentially to find matches on each criteria, gradually eliminating all students except the desired group. However, there is an easier method of achieving the same outcome.

**Inverted File.** Through the use of an inverted file, only very short searches are required to satisfy the request
above. "An inverted file consists of one ordered subfile for each value a key (identifier) or subkey may have in the field (Flores 1970, p. 318)." An inverted file is essentially created for every data-item for which a general search is anticipated. Locating the correct subfile for a given data value is accomplished either with a directory or by searching the entire sequential file. The directory method, although quicker than searching the entire file, requires more processing overhead to maintain the directory.

Linked, inverted files are also used with each subfile carrying a pointer to the next subfile, or next block. Linking the inverted files allows easy addition of entries to the inverted file and its associated subfiles. Use of linked subfiles provides a quick response to the request discussed above.

Thesaurus File. Much of the terminology used to discuss structures, areas, sets, files, subfiles, and data-items will have little meaning to the average college dean, registrar, or accountant. If data is to be retrieved easily by the users described in the model, the users must be able to request information in terms that mean something to administrators and educators. Thus, a Thesaurus file is included in the model to find the term supplied by the user and convert it to a standard term defined in the Data-Item Directory, which can then be used by the retrieval programs in the information system.
Homographs. The opposite of the Thesaurus file is the process of distinguishing the meaning implied by the user when a term is input that may have several meanings; e.g., a homograph. The model provides the capability to analyze homographs and determine proper categorization by the following methods:

1. Self-elimination--The context of the request is analyzed and the determination is based on other fields requested.

2. Qualification by User--The user is instructed to give a qualifier for the homograph when it is being supplied to the system. The qualifier then is used to match the Thesaurus or go directly to the directory to retrieve the items.

3. Inquiry--If the user does not supply a qualifier for items requiring one, the model notes this occurrence and provides the user with the possible qualifiers to the original homograph. When one qualifier is chosen, the retrieval process continues.

The retrieval of data from a data base requires a great deal of sophisticated logic unless the data base has been created for the sole purpose of retrieval. The data base in the model supports Operational Level activities requiring on-line file maintenance as well as the retrieval requirements at the Managerial and Executive Level. The three
techniques just discussed enable the model to support activities at all three levels of the organization in an efficient manner.

Data Base Content and Structure

The primary purpose of the model is to organize the data base with an emphasis on flexibility and a high degree of integration among the component parts. This section of the chapter describes the connections among the components of the data base in the standard terminology defined earlier, (Key CODASYL words will be capitalized for emphasis). The common tables and directories are described first; then data SETS, records and their logical relationships are described; followed by a description of the retrieval mechanism using Thesaurus file and inverted file techniques.

Tables and Directories

The use of tables and directories provides the model with a great deal of flexibility as well as standardization. Changes in table and directory entries allow modification of operating criteria without the expensive programming required to maintain less flexible automated systems. All of the common data-items discussed in the Data Definition section of this chapter will be incorporated into either a table or a directory.
Tables. Table entries are used to encode and decode information input or output respectively. The table includes the following data-items.

1. Table Identifier.

2. Numeric and/or Alphabetic Code.

3. Code Descriptor or Title.

An example used earlier for the HEGIS Taxonomy is included in a table entry as follows:

<table>
<thead>
<tr>
<th>Table Identifier</th>
<th>Code</th>
<th>Descriptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>HT</td>
<td>0705</td>
<td>Systems Analysis</td>
</tr>
</tbody>
</table>

Since the majority of the tables are used to provide the descriptor for reporting; and they all have a very low level of transaction activity, and require rapid retrieval; the Direct file organization scheme is most desirable. However, the codes for these tables are already standardized and cannot be the actual address of the table entry, eliminating the use of the Direct method. Thus, the indexed sequential access method is used with all table entries unless noted in the table description. Some hierarchically oriented tables such as the Program Classification Structure and the Manpower Accounting Structure tables utilize a network structure with ordered lists at the Program Element and Occupational Sub-category levels of the two structures respectively.
Table entries are only changed with the highest level of access authorization, as they affect all users of the database and must be carefully coordinated. Due to the static nature of the tables, backup files are created only after a modification, not daily or weekly as is the case for the remaining portions of the database.

The model includes the following tables to insure system flexibility and standardization of input and output:

1. Chart of Accounts—Descriptions of each entry in the systematic classification of accounts.

2. HEGIS Taxonomy—Numeric codes describing conventional subdivisions of academic and supportive activities. Table entries are linked to the Program Category portion of the Program Classification Structure table.

3. Program Classification Structure (PCS)—"A hierarchical representation of activities of higher education in a program oriented manner (Gulko 1972, p. 1)."

The table entries are broken down through the use of directories (OWNERS) established for each major subdivision (MEMBERS) in the hierarchical level, as shown in Figure 6.

The directory for Program Category is a pointer to a location in the HEGIS Taxonomy Table since those entries make up the Program Categories. This relationship between tables reduces redundancy, and
insures consistency of terminology although complicating the life of the programmer implementing the use of the two tables.

The contents of the Program Element portion of the table are ordered lists containing pointers to the appropriate account entries in the Chart of Accounts Table. Redundancy is eliminated; consistency of terminology is insured; and reports based on the Program Classification Structure have the flexibility of incorporating classifications consistent with the Chart of Accounts, rather than with some arbitrary classification technique.

4. Manpower Assignment Code Table--The first six positions of the Manpower Assignment Code coinciding with the Program, Subprogram, and Program Category and Subcategory, in the Program Classification Structure. The references to Program and Subprogram are related to these descriptors in the Program Classification Structure (PCS) Table. The Program Category and Subcategory are HEGIS Taxonomy codes and are related directly to that table for purposes of more efficient retrieval.

The Occupational Category entries in the table are defined as OWNERS of the Occupational Subcategory (MEMBER) entries in a hierarchical relationship. The
Subcategory entries consist of a list structure of codes with their descriptors.

5. Organizational Unit Table—A hierarchical structuring of the institution into divisions and departments within those divisions. A limited tree structure consisting of directory (OWNER) entries for each division and list structure (MEMBER) for each department within the division provides rapid retrieval with proper organizational relationships.

6. State/Country Table—A table containing the standard two-letter abbreviation for states, countries, dependencies, and areas of special sovereignty. These abbreviations are expanded into the complete title for reporting purposes. This table is required when reporting these data-items in a format other than an address for mailing. The opposite is true in arriving at the proper two-letter code from a complete country name when the entire name is input to the system.

7. Visa-Type Table—A small table containing the title and legal description of each visa-type a student or staff member can possess. The categorical descriptions are necessary for reporting and also, codifying of inputs.

8. Racial Category Table—A hierarchical table subdividing the five basic Civil Rights categories. The
five basic Civil Rights Categories often do not provide the necessary breakdown for detailed analysis by racial category. The model incorporates a hierarchical approach with each of the five basic categories acting as OWNERS of subcategories (MEMBERS) within the range of the basic category. An example of this is a list structure containing a code and descriptor for each tribe of American Indians, each of which would be MEMBERS of the American Indian SET.

9. Month Table--A table of month names. The calendar date standard used in the model provides for numeric representation of year, month, and day. It is often necessary to print the name of the month on reports. The twelve table entries represent the month number and name, and are organized sequentially with the month number acting as the subscript for Direct retrieval.

10. FICE Table--A table of codes and institution names designated by the Federal Interagency Committee on Education. This table contains a unique six-digit numeric code for every institution of higher education. The table contains several thousand entries and is updated annually. There is no structure to the coding scheme. The institution name is in the descriptor of each table entry.
11. Building Table—A table of building numbers and names. Depending upon the institution, this table could be a hierarchical structure with the OWNER record being the Installation and the MEMBER record consisting of a list of records with building numbers as codes and building names as descriptors; otherwise, the table is a simple list of building numbers and names.

12. Instruction Type—A table of codes, titles and descriptions of each type of instruction designated by the institution. In order to maintain the flexibility required to meet WICHE reporting standards, this small table describes instruction type. The codes are normally carried in the section segment of the Curriculum SET. The descriptor is the title and description for each type of instruction. The description is useful in coding the type of instruction for sections as they are created.

13. Manufacturer Table—A list, supplied by the Defense Supply Agency, of manufacturers serving the federal government. Local manufacturers are added to the coding scheme if their equipment is acquired by the institution. No specific structure exists in the numbering scheme.

14. Supply Item Code Table—A list of equipment items by category, supplied by the Defense Supply Agency.
Most users have found the original four-digit code to be too restrictive; therefore, an additional two positions are included in the model for flexibility. The codes are established by category and a hierarchical structure is used to retain the codes and the associated item descriptions. A two-way link with potential suppliers of the item is also maintained.

15. Supplier Table--A table containing the supplier's name, address, representative name, and two-way link to the Supply Item Code Table for determination of the items the supplier has available. The table is accessed through the links with the Supply Item Code Table or by Supplier Code.

Directories. A great deal of the operational flexibility of the model results from the use of directories to determine acceptable data ranges, access rights, field sizes and headings, and the general definition of every data-item in terminology familiar to the user. The directory portion of the data base contains the following entries.

1. Data Element Dictionary--The Data Element Dictionary is included in the directory portion because the directory portion contains: standard definitions of codes and data; linkage between the Data Element Dictionary and the Data-Item Directory; and the
identification of all of the SCHEMA, SUB-SCHEMA, AREA, RECORD and SETS containing the data-item.

The contents of the Data Element Dictionary and Data-Item Directories do not conform to the same format discussed for table entries. The structure of the Data Element Dictionary consists of one record for every element entry. Fields required for use in the Data-Item Directory are not part of the Data Element Dictionary, but are represented by a link to the corresponding data-item entry in the Data-Item Directory. The ability to modify Data Element Dictionary items is restricted to the data base administrator. All users are authorized to read all entries except Access Level, which is safeguarded from general use for security purposes.

2. Data-Item Directory--The Data-Item Directory is the key to the model's flexibility as it controls:
   a. Privacy of each data-item based upon the Access level for each user, terminal, and operator.
   b. Generalized retrieval of data-items based upon the field name, size, format, report heading, and location in the data base.
   c. Input of data-items based upon the value range list used to validate input data.

Due to the frequent use and critical response time required of the Data-Item Directory, the
directory is structured in the following manner. The primary (OWNER) record of the directory contains the data-item mnemonic name; a DATA-AGGREGATE in the form of REPEATING GROUPS, each group containing access levels for a user, terminal, and operator; and a VECTOR of links to secondary (MEMBER) entries. The secondary entries consist of:
   a. A record containing the items required to support the generalized retrieval function.
   b. A list of validation criteria for the data-item.
   c. A list of all SCHEMA, SUB-SCHEMA, AREA, SET and RECORD occurrences where the data-item is included.
   d. A record containing the item-name, Data Element Dictionary number, and a pointer to the associated entry in the dictionary.

The complex structure of the Data-Item Directory results from the necessity to serve the primary requirement of access control and still support other activities with the secondary entries.

Data Sets, Records, and Logical Relationships

The many relationships possible in an integrated data base provide a challenge for logical representation. The record types within each SET of records are defined. Relationships between related record types in the various SETS
are included as each record is described. The relationships described are the substance of the integration effort for the five basic data types. These steps provide the paths for users at all three levels of the model to acquire data useful in the performance of necessary tasks.

The sets of record types correspond to the five types of data defined earlier: Student, Financial, Personnel, Facilities, and Curriculum. Data-items that fit more than one category are retained in the category where they are maintained most readily, and where access time is minimized. Other references to the data-item are made via link from another SET of records, or possibly, from another record within the set.

Student SET. The first set of records in the model contains student data. The records are established in a hierarchical relationship within the set. The record types correspond to the five categories defined for student data types earlier in this chapter. They are:

1. Demographic and Biographical.
2. Previous Educational Experience.
3. Admittance.
4. Term.
5. Course.

The hierarchical relationships among the five record types are shown on the following page.
The Student Demographic and Biographical (SDB) record is defined as the OWNER of the SET named Student. The SDB contains the student identifier, demographic and biographical information and a direct link to the Current Term course records. The Previous Educational Experience, Admittance, and other Term records are accessed indirectly through indices contained in the SDB record.

The Previous Educational Experience record is the OWNER of individual MEMBER records that exist for each previous educational experience. The records are ordered
chronologically and contain the FICE (Federal Interagency Council on Education) Code representing each institution.

Admittance records are MEMBERS of the Demographic and Biographical record's SET. One record exists for each application for admission submitted by the student. The record is organized in DATA-AGGREGATE form and contains admission requests and action data.

Term records are broken into three categories as shown below.

1. Historical.
2. Future.

Historical Term records are organized chronologically, carry a directory for location of the desired term, and also are OWNERS of a SET of MEMBER Term records in list form. These Term records carry the course information, and instructor's Social Security Number for courses completed in the OWNER record's term; however, they do not carry room and time information. The Term records also carry a link to the Course record in the Course History set of the Curriculum SET which provides minimum storage medium requirement and consistent course titling. Historical Term records are quite static and are often maintained under control of a master student directory on another storage medium when not required for on-line access.
The second category of Term records is for Future terms. A Term record is possible for every term in the next calendar year. Future term records are of the REPEATING GROUP format and contain a link to each desired course in the Future Courses subset of the Curriculum SET. This link provides the course identifier, room, and times in a single place which minimizes maintenance problems when future courses change status.

The third type of Term record is for the Current term and is related directly to the Student Demographic and Biographical OWNER record. Course entries are organized in a two-way linked list structure. The link to the Current Courses subset of the Curriculum SET provides the capability to use standard course information that is retained once. It is via this link that the instructor and room number are retrieved. High access volumes dictate that the course records, in the SET owned by the Current Term record, have a direct pointer associated with the Student Demographic and Biographical OWNER record.

Financial Set. The second set of records in the model contains financial data. There are several distinct types of financial data represented in the model. To insure proper separation with rapid retrieval, the financial data are divided into five SETS which are:

1. General Ledger.
2. Budget.
3. Accounts Receivable.
4. Research and Other Sponsored Programs.
5. Accounts Payable.

The rationale for inclusion of a data-item in its primary set relationship is based on the organization structure of the financial area, and the operational demand placed on the information system. The account number and Program Classification Structure number are the common data-items used to summarize the various sets for reporting purposes. The five subsets of financial data are discussed in more detail as follows:

1. General Ledger--The first SET of financial records represents the general ledger accounts. The hierarchical relationships among the records in this SET are shown below.

   SET NAME--Financial.
   General Ledger.
   Fund Group.
   Account 1--Summary.
      Detail Pointer 1, Detail Pointer 2.
   Account 2--Summary.
      Detail Pointer 1, Detail Pointer 2.

The Fund Group record is the OWNER of Account records (MEMBERS) which apply to the specific fund group. There are two types of account records. Account 1 shown in The General Ledger hierarchical
representation is the OWNER of a record in VECTOR form that contains bidirectional pointers to detail records in other subsets of the Financial SET. Account 1 contains a summary of the detail amounts in the records referenced by the pointers in the VECTOR. The summarization process, known as the Current Summarization Technique, occurs as each detail record is updated, leaving the current amount in the Account 1 record. The Current Summarization Technique keeps data available for rapid Managerial and Executive Level use.

Account 2 shown in the General Ledger hierarchical representation is a standard account record that contains the account number, Program Classification number, and appropriate amount data items. Account 2 is a free standing account and differs from Account 1 in that it does not have an associated VECTOR record pointing to detail entries.

2. Budget--The second SET of financial records incorporates Budget records. The Budget record SET is ordered hierarchically in network form with both the Budget Master Record and Budget Category Record acting as OWNERS of the Departmental Budget Record. The MEMBERS represented in the second Departmental Budget Record SET are repeated to emphasize their network relationship.
The Budget Master Record (BMR) is defined for each appropriate division and summary account number; contains one total and twelve monthly increments in REPEATING GROUP form for allocation, encumbrance, and actual expenditure amounts. The Budget Master Record is used for reporting at the Executive Level.

There are six Annual records that are MEMBERS of each BMR. The Annual records contain the amount budgeted and expended by month for the BMR division and account for five prior years, and the budgeted amount for one year in the future.
Departmental Budget Records (DBR) are also MEMBERS of each BMR. A DBR is established for each department and account within the division and contains the allocation, encumbrance and expenditure amounts in annual and monthly increments for the current fiscal year. The DBR is the OWNER of two MEMBER record types. The first MEMBER is the departmental equivalent of the six Annual records defined for the BMR.

The second MEMBER of the DBR SET is present if the Account in the DBR is a salary account. The records represent Personnel Line Item (PLI) entries in the department's budget. The PLI records are arranged in a linked list structure with additional pointers to the Personnel SET and when appropriate, to the Research and Sponsored Program SET.

In addition to the SET relationship with the Budget Master Record, the Departmental Budget Record and all of its MEMBERS are also MEMBERS of another SET titled the Budget Category Record SET (BCR). The Budget Category Record relationship with Departmental Budget Records is based upon the organization and presentation of the budget. Budget categories and line items make up the coding structure with summarization being performed for each budget category in the hierarchy and is accomplished by having a MEMBER
list structure consisting of all related departments and accounts for each BCR.

3. Accounts Receivable--The third SET of financial records involves Accounts Receivable data. Records reflecting receivables, their fund category and the account to be credited are carried in an indexed sequential file. The records contain a link with appropriate entries in the Student, Personnel, Financial Aids, and Research and Other Sponsored Programs data records depending upon the source and receiver of the funds receivable. There are no OWNER-MEMBER relationships among the Accounts Receivable SET; therefore, no outline of record relationships is presented.

4. Research and Other Sponsored Programs--The fourth SET of financial records includes Research and Other Sponsored Programs (RSP) data. The SET is structured hierarchically in the following manner.

SET NAME--Financial.

RSP Master.

Investigators--SSN, SSN, . . . SSN.
Revision.
Abstract
Cost Breakdown

Allocation--Department (DBR Link).
Allocation--Department (DBR Link).
The RSP Master record contains the identifier and the Program Classification Structure number and is the OWNER of three MEMBER records.

The Investigators record is the first MEMBER of the RSP SET. The Investigators record is a VECTOR of Social Security Numbers for personnel associated with the project and links the RSP to the Personnel SET.

The second MEMBER of the RSP SET is the Revision record. This record is the OWNER of a SET that contains the topical abstract in one MEMBER record and the cost breakdown in the other MEMBER record. The identifier of each Revision record is an extension of the identifying number assigned to the RSP Master with the original abstract and cost breakdown placed in the "00" extension record. Each subsequent Revision record is stored in a double linked list which allows access to the first, last, or any other revision with a minimum amount of wasted access time.

The third MEMBER of the RSP SET is the allocation record which is generated once the program funding is negotiated and accepted. Each allocation record in the list contains the identifier of the revision number to which it applies, the Program Classification Structure number (including account number), the department receiving the funds, the amount of funds time phased, and a bidirectional pointer to the
Departmental Budget Record. If the program is re­negotiated, another entry is placed in the revision list and the new Allocation records are inserted in the Allocation list.

5. Accounts Payable--The fifth SET of financial records includes Accounts Payable data which is organized in the following hierarchical relationship.

SET NAME--Financial.

Accounts Payable.

Purchase Order.
   Item--Pointer to Facilities SET.
   Encumbrance--Pointer to DBR.

The Accounts Payable record contains the Program Classification Structure number and General Ledger Account number as well as a summary of the detail amounts in the MEMBER records, and is used in Managerial and Executive Level summaries.

There are ordinarily multiple Purchase Order Records (POR) that are MEMBERS of each Accounts Payable record. The POR contains the information found on each purchase order and has the capability of retrieval by requisition number, purchase order number, vendor name or code, and Supply Item code. The POR is the OWNER of the Item and Encumbrance records.
Item records are organized in list form and contain the Supply Item Code, Manufacturer Code, quantity, description, price per item, expected delivery date, date received, and a pointer to the Equipment Record in the Facilities SET. An Item record is generated for each separate item on a purchase order.

Encumbrance records, also organized in list form, contain the account numbers, distribution of charges to each account, and a pointer to the Departmental Budget Record.

Upon receipt of the ordered items, the appropriate expenditure accounts are charged the actual cost and the encumbrance is removed via the Departmental Budget Record pointer; and the item automatically enters the inventory via the pointer to the Facilities SET.

Personnel SET. The third major set of records in the model represents personnel data. There are six major categories of data that correspond to the six record types described below. They are:

1. Demographic and Biographical.
2. Previous Educational Experience.
3. Appointment.
4. Assignment.
5. Activity.
6. Payroll.
The hierarchical relationships among the six record types are shown below:

SET NAME—Personnel.

Personnel Demographic and Biographical.

Previous Educational Experience.
  Institution 1.
  Institution 2.

Appointment.

Assignment.
  Activity 1--Link to Student SET.
  Activity 2--Link to Curriculum SET.

Assignment.
  Activity 3--Link to Financial SET.

Payroll.

The Personnel Demographic and Biographical (PDB) record is defined as the OWNER of the SET named Personnel. The PDB contains the employee identifier (Social Security Number), demographic and biographical information and a direct link to the Payroll record. The Previous Educational Experience and Appointment records are accessed indirectly through indices carries in the PDB.

The Previous Educational Experience record is the OWNER of a MEMBER record for each educational experience of the faculty or staff member. The records are ordered chronologically and contain the FICE Code (Federal Interagency Committee on Education) representing each institution name,
as well as the employee's degree and major, which is coded in accordance with the HEGIS Taxonomy.

Since Appointment record can occur several times for one employee, these records are retained in a sequentially ordered list, keyed on the Department Budget Record. This procedure allows the Appointment record to be associated with the Budget Master Record for summarization, and the Budget Category Record for detail budget reporting. The Appointment record can also be associated with future budget records when the request budget is prepared.

In addition to the budgetary considerations, the Appointment record is the OWNER of the Assignment records. An Assignment record is established for each assignment an employee receives. The Manpower Assignment Code (Minter 1972) is used to categorize each assignment in terms of the type of assignment (Academic Support), and the occupational activity (Technical), required to perform the assignment.

The Activity records are organized in a list structure and are coded by Activity Categories established for use in the Faculty Activity Analysis Model (Romney 1971, p. 82), as well as other analytical models. The Activity record also contains the link between the personnel record and the following records:

1. Student--Linked to the Student Demographic and Biographical record for advising activities.
2. Curriculum—Linked to the Section record for current courses to indicate teaching activities.

3. Financial—Linked to Research and Other Sponsored Programs for research and other programs.

The Payroll record is a MEMBER of the SET whose OWNER is the Personnel Demographic and Biographical record. The contents of the record are controlled by personnel, thus, it is placed in the personnel section of the model. Through linkage to the Budget SET, the Payroll record can be accessed by authorized financial users. For periodic payroll operation, the Payroll records are directly associated with the PDB record to minimize access time.

Through linkage, a sublist is shared with the Appointment record to provide each account with the proper proration of each payroll disbursement. These prorations are then related to the expenditures portion in the Departmental Budget Record via the link with the Appointment Record.

Facilities SET. The fourth SET of records, the Facilities SET, consists of five basic record types:

1. Institutional Data.
2. Facility Data.
4. Room Data.
5. Equipment Data.
These five records are related hierarchically as shown below:

**SET NAME—Facilities.**

**Institution.**

**Facility.**

- **Building.**
  - Room--Links to other SETS.
  - Room--Links to other SETS.

- **Equipment--Links to other SETS.**
  - Room--Links to other SETS.

The Institution record is the OWNER record for all Facility and Equipment records for a given institution. The FICE Code (Federal Interagency Committee on Education) is used to identify the institution. Summary amounts of book value by category for the institution are maintained at this level for ready reference and retrieval.

A Facility record is referenced via a directory based upon a facility number that in most cases corresponds to Building number. For non-building facilities, the name and description, cost of acquisition, cost of improvements, and a linked list of pointers to equipment items located at the site are retained. For buildings, the facility name, gross area, book value, replacement cost, year of construction, condition, ownership and location are retained in the Facility record. The Facility record is the OWNER of the Building record.

The Building record contains summary values for total area and area by use category, such as, mechanical and
custodial. The summaries are used in building use analysis from an architectural standpoint. The Building record is accessed via an index and acts as the OWNER of Room records for each room in a building.

The Room records describe characteristics of each room and contain a sublist that links:

1. The classroom to individual course sections taught in the classroom.
2. The office to the faculty or staff member(s) that occupy the office.
3. The dormitory room to the student(s) that occupy the room.
4. The space to the division and department responsible for the space, and
5. The equipment assigned to the room.

As shown in the hierarchical relationships of Facilities records, the Room record is a key element in the integration of the data base. Except for the division and department responsible for the space, each of the categories above is a double linked list structure that is a MEMBER of the SET defined by the Room record.

The Equipment record is the second MEMBER of the Institutional record's SET. The property tag number is the identifier used to access the record using the indexed sequential mode. Categorized descriptive information regarding the piece of equipment and cross reference capability...
based on the Federal Supply Item Code are included for searching the data base for retrieval of a specific category of equipment. The appropriate Program Classification Structure code is included for equipment associated with a specific program. The Equipment record emphasizes the value of the network concept. A VECTOR, appended to the record, contains pointers to the Room record, General Ledger Account record(s), Research and Sponsored Program record, Purchase Order record, and other appropriate Accounts Payable records that deal with the specific equipment item.

Curriculum SET. The fifth set of records is named the Curriculum SET. This SET consists of six basic record types which are:

1. Catalog.
2. Current Course.
3. Section.
4. Course History.
5. Future Course.
6. Catalog Text.

The records named above are related hierarchically on the following page.
The Catalog record is defined as the OWNER of the SET named Curriculum. A Catalog record exists for every course in the Current, Future and Course History records and acts as a directory record to the appropriate record depending upon the term and course identifier.

A Current Course record exists for every course offered during the current term and contains: the division and department responsible for the course; cross-listing departments; the applicable Program Classification Structure number; course level; pass/fail and audit availability; credit hours; and total current enrollment. The Course record is the OWNER of the Section record.

The Section record provides detail information regarding each section offered such as: time and days offered; type section (laboratory, lecture, discussion); and beginning and ending date of the section. An appended VECTOR carries
pointers to the Room record and instructor's Activity record. Multiple pointers exist for sections taught by teams of instructors. The Section record is the OWNER of the Student List for that specific section.

The Student List is a double linked DATA-AGGREGATE list that contains the following information for each student: a bidirectional pointer to the student's Current Term record; and the student's registration status in the class, i.e., withdrawn, audit, or pass/fail. The double linked list expedites creation of class rolls.

A Course History record is created for every course offered at the institution. Since course numbers are often reused, maintenance of Course History records in chronological order within course number allows the retrieval program to match on the term date and course number to assure correct retrieval. The file is organized in an indexed sequential manner. Primary access is via links from students' Term records for purposes of generating transcripts and degree progress checks.

The Future Course record is generated for each course to be offered in a future term and contains the same data-items that are required for Current Course records plus approval status and beginning and ending dates. The Future Course record is the OWNER of a linked list of pointers to students requesting preregistration for the course. This
list is used to measure demand and expedite preregistration processing.

The Catalog Text record carries the detail course title, description, and information regarding placement of Future Course entries in the printed catalog. The text record is accessed via the directory in the Catalog record. Although the Catalog Text records are defined as an integral part of the data base, they are not normally maintained on direct access media.

The logical relationships described for the five data types form the integrated data base structure required for the model to support an information system that is useful in managing higher education.

Retrieval of Information

The structure of the data base is developed with the retrieval requirements being one of the primary placement criteria. The records and relationships provide the necessary data to satisfy four types of retrieval which are:

1. Data-item retrieval.
2. Preformatted display retrieval.
3. Structured interrogation.
4. Unstructured interrogation.

The methods of performing each type of retrieval are discussed on the following pages.
Data-Item Retrieval. Acquisition of a single data-item such as a student's cumulative grade point average requires that the requestor must qualify for access to that specific data-item, and that the identifier for the student be presented. This elementary type of retrieval is possible without an integrated data base, and is normally performed as part of a specific task at the Operational Level.

Preformatted Display Retrieval. The acquisition of all data-items required for a specific report poses a problem of greater complexity. The steps in generating a preformatted display are:

1. The requestor's authorization to request the specified display is validated.

2. The data base management system is instructed to retrieve the necessary data-items that are associated with the identifier. The retrieval is performed by invoking a SUB-SCHEMA that includes all the necessary record types and their logical relationships, such as OWNER-MEMBER, and links.

3. The data-items retrieved are validated individually against the requestor's authorization.

4. Only the data items to which the requestor has authorized access are included in the display. In order to distinguish between missing data and unauthorized data on the display, asterisks are inserted in place of unauthorized data-items.
There is no limit to the number of SUB-Schema that can be defined for the data base. Therefore, when a pre-formatted display is programmed, a SUB-Schema including all of the necessary relationships is generated at that time. The OWNER-MEMBER relationships used for the display are described in the Schema for the data base; they cannot be established as each display is programmed. This requirement is satisfied in the model since all relationships among records with common data-items already exist.

An example of this type of request is a report that displays appointment and past, current, and future information regarding a specific employee. The request can be made two ways. First, the employee's identifier regarding Division, Department and line number associated with the Budget Category Record is used with appointment data obtained via the link from the employee's line item Departmental Budget Record to the employee's Appointment Record in the Personnel SET. Second, the employee's Social Security Number is used to request the display, in which case the retrieval is via the reverse link from the Appointment Record to the Departmental Budget Record line item entry.

Structured Interrogation. The demands for information at the Managerial and Executive Levels require the capability to select all records meeting a specified set of criteria. The criteria are obtained through a dialogue with the user where categories of data, value ranges, and display
formats are chosen from a menu of available interrogation combinations. Once the user's request is obtained, his access authority to that information is verified. If access is authorized, the solution to the request is implemented. The model has the capacity to satisfy structured interrogation problems in two ways; through use of an Inverted File, or through a categorical search of all pertinent records.

An Inverted File is established for each data-item to be interrogated when a great deal of structured interrogation is anticipated. The Inverted File records consist of the data-item or a derivative of the data-item, and a pointer to the record in which the data-item resides. The file is ordered on the data-item field and is sequential or indexed sequential, based upon the number of entries to be searched.

An example demonstrating this technique is the periodic generation of an Inverted File containing pointers to Master Budget Records where expenditures plus encumbrances exceed the allocation. The data-item in the Inverted File can be either one or the other of the following types: percentage overexpended, or amount overexpended. In order to reduce search time and allow compound searches, multiple Inverted Files are established in a directory or hierarchical relationship where all overexpended records for a specific department would be related to the department record and all departments having overexpenditures would be related to the
division record in the directory. This relationship allows at least the following interrogation possibilities:

1. All budgets in a division that are overexpended.
2. All budgets in a department that are overexpended.
3. All departments and their budget line items that are overexpended.
4. All departments in a specific division and their budget line items that are overexpended.
5. All budget line items that are overexpended by an amount greater than a given percentage of allocation.

The use of Inverted Files requires anticipation of the type of retrieval request. Generation of the files is normally performed in a batch mode, and the files reloaded each day. Another approach to maintain current Inverted Files is to keep the Inverted File in linked-list form and update the file as the source records are updated. Using the Current Summarization Technique described for Budget Master Records, the amounts in a master record are tested after each update transaction. If an overexpenditure condition exists, the application program inserts the Budget Master Record's identifier into the linked list. The identifier is similarly removed when the condition is corrected.

The second way to handle structured interrogation is to search each record in the desired category of data for a match to the criteria presented. Due to the complex nature of the data base structure, the search procedure is not as
cost effective as the Inverted File method for locating detail records in hierarchical file structures. Inefficiency in the search procedure is avoided by utilizing the Current Summarization Technique, which updates selective data-items in the detail MEMBER records and also updates the summary data-item in the OWNER record near the top of the hierarchical structure, such as described for the Budget Master Record and General Ledger Account Record. Simultaneous updating and summarization adds little to total processing time since the OWNER record is accessed by the update program in route to the MEMBER record, and the two records can be modified simultaneously.

By taking advantage of OWNER records in the hierarchical structure that contain summary amounts, the structured interrogation can be efficiently performed for a wide range of requests. The advantage over the Inverted File method is the reduction of storage space required to hold the directories, links, and inverted files. The disadvantage is that every OWNER record of the specified data type(s) is interrogated, which is normally a much slower process. Structured interrogation requests are normally stored in a queue to be processed at night, making the results available the next morning. The categorical search technique obviously offers less rapid response but provides a much wider range of retrieval capabilities. Files required for simulation models
such as RRPM-I are also generated by the structured interrogation technique.

**Unstructured Interrogation.** Regardless of the planning that is performed in developing Structured Interrogation capability, the decision making and planning requirements at the Executive Level often necessitate analysis of diverse categories of data. To support such decision making and planning, an Unstructured Retrieval capability is included in the model, which is built around several key components including a:

1. Retrieval Language Processor.
2. Thesaurus.
3. Homograph processor.
4. Data-item directory.

The Retrieval Language Processor allows a great deal of latitude on the part of the user when requesting information. The Processor interacts with the user until the user has provided the necessary criteria for the Retrieval Language Processor to structure an interrogation program and resulting display. Steps needed to complete the interrogation and to provide the requested output include:

1. Scanning the request for recognizable words.
2. Picking out key operator words; for example, SORT, EQUAL, FROM, TO, and ALL.
3. Comparing requested data names to the Thesaurus file to determine the exact data-item names that are
recognizable by the Data-Item Directory. This step involves homographs (words with multiple meanings) which are analyzed in accordance with the process described under Retrieval in the Data Access and Structure section of this chapter (page 120).

4. Validating access authorization for each data-item requested and notifying the user of invalid requests.

5. Compiling a list of SUB-SCHEMAS containing the data-items requested. If the data-items do not have a common SUB-SCHEMA, the minimum number of SUB-SCHEMAS necessary to retrieve each requested data-item is invoked.

6. Analyzing of key words and their logical relationships by the Retrieval Language Processor invokes the necessary generalized routines to search, select and display the data-items requested.

7. Estimating the amount of time required for the search and the volume of possible responses is performed by an algorithm that analyzes the search criteria and samples the initial output.

8. Displaying the estimated time requirement, output volume, and processing recommendation for the user. A complex search or high volume output estimates result in a recommendation for a batch oriented search during idle hours. A simple search or small output volume estimate results in a message telling
the user to expect results in the amount of time estimated, and the search process continuing automatically.

9. Choosing the processing option to be used. The user chooses to either proceed with the search, place the request in a queue for processing during non-peak hours, or discontinue the search. As users gain experience with the Unstructured Interrogation process, they recognize when they have made too broad a request by the amount of predicted output.

10. Producing the requested information in the format and on the medium (cards, paper, video terminal) desired by the requestor.

The Unstructured Interrogation Process has potential access to every data-item in the data base with all of the logical relationships that are defined in the SCHEMA. The user is given latitude in the terminology used in the request for information, and he has feedback regarding the results of the request in a matter of seconds.

**Summary**

The requirements placed on the model by the Executive, Managerial, and Operational Levels of the organization are described first in this chapter. The five basic data types are described, as are four representative functional users. The concepts of standard data definition and
terminology, and their place in the model are discussed. A building block approach describing terminology, security, data access, structuring, and retrieval techniques is presented for readers not familiar with data base technology. The terms and techniques presented are then used to describe the content, organization, and relationships that make up the basic components of the data base model. Finally, the retrieval capability of the model is described to portray the value of the model in terms of availability of meaningful information to perform tasks in all functional units and at all levels in the institution.
CHAPTER 5

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Three sections are included in this chapter. The first section summarizes the purpose of the study, the literature reviewed, methods used, and the descriptive model that was constructed. The second section draws some conclusions regarding the value of the model in managing higher education; and the third section includes recommendations for possible implementation of the concepts contained in the model, and topics that merit further investigation.

Summary

A model of an integrated data base which can serve as the nucleus of a university information system is developed in this study. The basic structure, after which the model is patterned, is a pyramid. The pyramid structure is a derivative of the classical triangle representation of an organization structure and data types presented by Mosmann (1973, p. 128) and others. The organization structure is divided into three levels of management:

1. Executive.
2. Managerial.
3. Operational.
The Managerial and Operational Levels of the organization are divided into five data types:

1. Student.
2. Financial.
3. Personnel.
4. Facilities.
5. Curriculum.

To represent the users of an information system, the third dimension is added to Mosmann's triangle resulting in the pyramid. The number of users was limited to four representative instructional and facilitative functions to make the model manageable. The users are:

1. College/Departments.
2. Institutional Studies.
4. Admissions and Records.

The literature reviewed is concentrated in three major areas; data base oriented information systems; status of current educational management information systems; and use of data in managing higher education. The survey of literature regarding data base oriented information systems reveals that while 600 or 700 companies are using data base information systems, useful information for Executive Level decision making seems to be almost non-existent (Bruun 1973, p. 36). Bruun's opinion is contradicted by literature
regarding current efforts on the part of the National Center for Higher Education Management Systems to develop comparative models for use at the Executive Level with little support for the Managerial and Operational Levels in higher education. This contradiction demonstrates a need to join the two concepts to support all levels of the institution.

Another key component of data base oriented systems is the Data Element Dictionary. Bontempo (1973, p. 35), Wearing (1973, p. 29), and Lawrence (1969, p. 2) all stress the importance of the Data Element Dictionary in identifying and describing standard, non-redundant data-items. The basic concept of a data base is the non-redundant storage of data-items which are normally defined and controlled through the use of the Data Element Dictionary.

Literature regarding the current status of educational management information systems reveals that historically, stewardship prevailed in the need for financial information, as opposed to the profit and loss emphasis in industry. The lack of emphasis on controls and planning in education is, in turn, reflected by a lack of emphasis in developing integrated information systems.

Another obstacle to implementation of information systems is what Manning (1973, p. 8) terms an "... Information Gap," which is described as existing between the raw data found at the Operational Level and statistics at the Managerial Level; and also existing between statistics and
information needed at the Executive Level. Manning (1973, p. 8) states that the gap between statistics and information can be solved through forecasting and simulation techniques; however, the gap between raw data and statistics: "... is still a real dilemma in higher education today." Kornfeld (1968, p. 18) indicates that a problem also exists due to the lack of integration among parts of the information system.

The National Center for Higher Education Management Systems (NCHEMS) is beginning to address the problems stated above. The NCHEMS emphasis is on development of a Data Element Dictionary and Program Classification Structure (Gulko 1972) to provide standard definitions and aggregation procedures for educational data. Some of the forecasting models that Manning describes as closing the gap between statistics and information are NCHEMS products.

Another effort to standardize and exchange ideas in management systems is being coordinated through the College and University Systems Exchange (CAUSE). The joint efforts of data processing specialists at many institutions are providing higher education administrators with the management information systems required to deal with complex organizational problems.

The third area of emphasis in the review of literature concentrates on the use of data in managing higher education. The review is further divided into use of each of the five data types that are included in the model. Mann's
study (1973, p. 5) indicates that the greatest current emphasis on computer based management information systems is in the use of student data, followed by financial data, personnel data, and facilities data. Curriculum data, while previously receiving little consideration, is an area of increased attention. Most of the information systems reviewed serve a single functional area, and provide the basis for establishing the requirements included in the model. Equipment inventory is an area with minimal information as compared to the large volume of information on physical plant facilities management.

The advantages of information systems found in the literature are summarized to demonstrate the benefits derived from an information system. Conversely, some disadvantages cited in the literature are included to caution the reader of the magnitude, complexity, and impact of an integrated information system. Several authors cite the parochial attitudes on the part of designers and users as being the greatest deterrent to the integration process. Data elements from functional subsystems used at Stanford University, The University of Cincinnati, The University of Illinois, and The University of Arizona are included in the analysis of the data-items necessary to integrate the various data types. The Program Classification Structure (Gulko, 1972) is used for classification, grouping, and summarization purposes because of the flexibility built into the structure for institutions to
define their own subcategories as well as use the program oriented categories established in the structure. Many terms and standard definitions are also included to enhance understanding and comparability of data retrieval from the model.

A conceptual review of the automated file structuring techniques used in the model is included in the study. In addition, terminology for representing the data base structure is contained in the model. The terminology used is defined by the Conference on Data Systems Languages (CODASYL) for implementation throughout the computing industry.

The component parts of the model are then defined in detail. First, the requirements to be satisfied by the model are described for the three levels in the organization structure, the five types of data, and the four representative functional users.

Second, the concept of common data definition is described through a detail requirements description of the Data Element Dictionary used in the model. In addition to the Data Element Dictionary, other standard terminology to be carried in tables within the model are described. The structural makeup of the HEGIS Taxonomy (Martin 1972a, p. 53) and Program Classification Structure (Gulko 1972, p. 21) are defined in detail due to their importance in developing the summarization and classification records within the data base. Explanation of other standard coding schemes and their
interpretation is carried out first for those codes affecting only one of the data types. These coding schemes often constitute the linkage between different types of records, and are defined as an integral part of the model's internal structure.

Data structuring techniques and associated issues of terminology, security, data access, and retrieval are dealt with in the third component of the model description. The data structuring component is included for two purposes. First, to provide a cursory understanding of the data structuring, accessing, and retrieval techniques employed in the model; and second, to describe the security process which is essential for proper protection of data under all conditions.

After the component parts of the model are either defined or described, the content and structure of the data base are described using a building block approach. The tables and directories common to all parts of the model are described first. Each data type and its logical relationships are then described in CODASYL Standard terminology. Categories of data within each data type are defined as records; then, the data-items used to access the record or link it to another record are identified. Where appropriate, the Current Summarization Technique of carrying summary amounts in SET OWNER records is included.

When the data types, logical relationships, and pre-defined summaries are all described, the four types of
retrieval available as part of the model, are discussed. The retrieval of data from the model is related to the demands of the three levels. Functional units are not discussed specifically because the model is designed with cross referencing capability to provide access to any data-item for which a user is authorized; therefore, if an administrator or clerk in any functional unit has a legitimate need for a data-item, and receives authorization, retrieval of that data-item can occur in any one of the four ways described. Flexibility of access is emphasized as the essence of the data base concept employed in the model.

Conclusions

Conclusions regarding the value of the model in managing higher education institutions are presented in this section. Benefits that are derived from the components of the model are listed below:

1. Satisfaction of data requirements of all aspects of the organization. The capability to satisfy the data requirements of clerical personnel and administrators at all levels of the organization is included in the model. Each user is provided with the data required by means of four types of retrieval defined in the model. The Current Summarization Technique, and the large number of links between records provide rapid response to complex interrogations by users throughout the organization.
2. Addition of a third dimension to the classical management information system model. The addition of functional users allows the model to better represent the complete management information system.

3. Standardization of terminology and data definition through use of the Data Element Dictionary, Data-Item Directory, and numerous tables. As redundancy is decreased through use of the techniques above, the reliability of the data in general is increased.

4. Identification of an integrated data base structure. A great deal of flexibility is allowed as a result of the integrated design of the data base. The horizontal flow among functional users is identified. The capability of implementing new record relationships and adding modules to the data base is enhanced by the linkage concept in the model.

5. Specification of security procedures. Four access authorization checks are available for every data-item and preformatted request. The access authorization process in the model provides extensive privacy protection safeguards. The integrity considerations of the model are provided by the selective backup capability and the Edited Input Method of file restoration.
The benefits listed on the preceding page are partially offset by three disadvantages associated with the model.

1. Understanding the complex structure of the model. The many relationships possible in the model are difficult to comprehend, thus, utilization of the model's full capabilities can be impaired.

2. Monitoring the data base. Periodic modifications to the relationships in the data base are required as a result of the dynamic nature of the organization. Restructuring the data base requires intimate knowledge of each relationship. The need to restructure the data base is determined by analysis of average response times to requests. Transactions are also analyzed for the purpose of identifying potential security violations.

3. Updating the data base. Although minimized by organizing data types into SETS of records resembling input categories, the updating process is complicated by the large number of links among the records. Proper restoration of links during file maintenance activity is required in order to insure the integrity of the data base.
Recommendations

This section of the chapter identifies potential application areas for the model and suggests areas that may warrant further study.

The model is generic in nature, and the concept contained in it can be employed in a number of ways. The prerequisites to implementation of the model are defining the management and clerical functions which fall within each level; describing the data types to be included; and arriving at the data input and retrieval requirements for each of the functional users to be included in the model.

The model is applicable to institutions ranging in size from small community colleges to major universities. Implementation of the model also is possible for a statewide network of institutions that all report to the same governing body.

The scope of the model can be reduced to satisfy the requirements of a subdivision of an institution such as an engineering college or medical school. It should be noted that uncoordinated implementation of multiple versions of the model leads to renewed conditions of fragmentation and lack of standardization within the institution. Coordinated implementation of the model in modular form allows for orderly analysis of each module and the integration of each subdivision.
The data base presented in the model serves as the foundation for further work in the field of information systems and their use in managing higher education. The topics listed below range from broad concepts to specific applications:

1. Develop and implement the retrieval language component of the model.
2. Develop new simulation techniques that take advantage of the relationships in the model for historical, current, and future data.
3. Develop the model for the secondary and elementary education application.
4. Develop a process of gathering, categorizing, and retrieving all of the diverse data-items required for the annual report produced at all levels of the institution.
5. Develop specific applications in all functional areas of education.
6. Develop an algorithm to calculate the complete cost of instruction or operation for any identifiable discipline or program, using data in the data base.
7. Develop the procedures required to use the Catalog Text record for maintaining and printing the institution's periodic course catalog, and possibly the schedule of classes for each term.
8. Develop an algorithm to analyze the data base and determine the distribution of effort for faculty and multiple appointment staff personnel.

9. Develop an extension of the Student Course record to simplify record keeping and grade reporting procedures for the faculty.

Gwynn (1969, p. ix) states that: "Widespread management information system implementation in higher education is now on the horizon." The standard terminology defined in the tables and directories; the five data types and all of their logical relationships; and the retrieval capability of the model, provide the conceptual foundation upon which higher education management information systems can be constructed and utilized, rather than appearing as a mirage on the horizon.
APPENDIX A

DATA / FILE / REPORT MATRIX FORM
### DATA / FILE / REPORT MATRIX

<table>
<thead>
<tr>
<th>Data Element</th>
<th>FILE OR REPORT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SELECTED BIBLIOGRAPHY


176


178


