R E P O R T   R E S U M E S

ED 014 902   24   EM 006 002

A COMMUNICATIONS SYSTEM FOR HIGHER EDUCATION. FINAL REPORT.

BY- WILD, J. DONALD   DOUGHTY, DONALD D.
CALIFORNIA STATE COLL., DOMINGUEZ HILLS

REPORT NUMBER CRP-E-034   PUB DATE   AUG 67
REPORT NUMBER BR-5-0791
CONTRACT OEC-5-10-300
EDRS PRICE   MF-$0.50   HC-$4.76   117P.

DESCRIPTORS- *MANAGEMENT, *ADMINISTRATIVE ORGANIZATION, *DATA PROCESSING, CODIFICATION, *HIGHER EDUCATION, *INFORMATION STORAGE, *INFORMATION RETRIEVAL CALIF. STATE COLL., DOMINGUEZ HILLS, CALIF.

THIS PROJECT SOUGHT TO DEVELOP AN OPERATIONAL "TOTAL" DATA SYSTEM FOR AN INSTITUTE OF HIGHER EDUCATION. THE REPORT SYNTHESIZES MANY PREVIOUS EXPERIENCES IN DATA PROCESSING, IN ORDER TO PRESCRIBE A MANAGEMENT INFORMATION SYSTEM SUITED TO COLLEGES AND UNIVERSITIES. IT WAS FELT THAT A BROAD BASE, GENERAL PURPOSE INFORMATION SYSTEM WOULD CENTRALIZE AND STRENGTHEN THE MANAGEMENT OF THE INSTITUTION. THE FUNDAMENTAL DESIGN PRINCIPLES OF THE SYSTEM WERE (1) THE DATA BASE ITSELF MAY BE REGARDED AS A SET OF FILES, (2) MAINTENANCE OF THE DATA BASE SHOULD BE EFFECTED ACCORDING TO SPECIFIC SCHEDULES, (3) INFORMATION RETRIEVAL SHOULD BE SCHEDULED TO CONSIDER THE TOTAL REQUIREMENTS OF THE INSTITUTION, (4) INTRODUCTION OF NEW DATA SHOULD BE ORDERLY, WITH MINIMAL CHANGE TO EXISTING PROGRAMS, (5) EDUCATIONAL ADMINISTRATORS AND RESEARCHERS SHOULD BE TRAINED IN THE USE OF THE DATA BASE. (MS)

FINAL REPORT
Project No. E-034
Contract No. OE5-10-300

A COMMUNICATIONS SYSTEM FOR HIGHER EDUCATION

August 1967

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

A COMMUNICATIONS SYSTEM FOR HIGHER EDUCATION

Project No. E-034
Contract No. OE5-10-300

J. Donald Mild
and
Donald D. Doughty

August 1967

The research reported herein was performed pursuant to a contract with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

California State College, Dominguez Hills

Dominguez Hills, California

CONTENTS

	Page
LIST OF ILLUSTRATIONS	v
PREFACE AND ACKNOWLEDGEMENTS	vi
INTRODUCTION	1
The Problem	2
Organization and Management	4
Governance and Policy Formulation	5
Personnel Selection	6
Academic Freedom	6
Total System	8
Flexible System	8
Responsive System	9
Diagnostic System	9
Growing System	10
Projective System	10
Economic Considerations	11
Objective	15
Students	16
Curriculum	16
Faculty	17
Staff	17
Facilities	17
Fiscal	17
METHOD	20
Preliminary Study	20
Survey of the Literature	21
Report Inventory	24
Field Work in the California State Colleges	26
Other Investigations	29

	Page
System Design	30
Data Element Inventory	31
File Design	36
Data Acquisition	39
Storage and Retrieval	42
Applications Programing	44
RESULTS	48
Hardware Requirements	48
Organization of Data Base	50
Key Files	54
Working Files	55
File Linkage	56
SR:SF	59
SR:FS	59
SR:BF	59
SR:CM	59
SR:LB	60
SF:FS	60
SF:BF	60
SF:CM	60
SF:LB	61
FS:BF	61
FS:CM	61
FS:LB	61
BF:CM	61
BF:LB	61
CM:LB	61
User-Operator Tools	61
Data Catalog	69
Index	71
File Specification	71
Classified Information	72
Storage	72
Storing Historical Information	72

	Page
Storing New Records	73
Introducing New Data	73
Information Retrieval	76
Implementation Phase I	77
Implementation Phase II	78
Ultimate Implementation	79
Retrieval Program	79
Match/Merge Program	81
File Maintenance	82
Update 1, Location Known	85
Update 2, Location Unknown	87
Applications Programing	91
Program and Procedure Standards	92
Program Documentation	92
DISCUSSION	94
CONCLUSIONS AND IMPLICATIONS	97
SUMMARY	100
REFERENCES	102

LIST OF ILLUSTRATIONS

Figure		Page
1.	Management Information System	12
2.	Data Inventory Form	32
3.	Hierarchy of the Data Base	51
4.	Files A and B as Parts of the Institution's Universe of Data Base	52
5.	File A Segmented Into Four Subsets	53
6.	Files A and B Linked By Common Data	54
7.	Key File Linkage	58
8.	Data Catalog	63
9.	Data Index	64
10.	File Specification	65
11.	File Specification Title Sheet	66
12.	Reference Codes	68
13.	Representation of Historical Record Containing "Case Counter" (n)	74
14.	File Updating Through Key Files	84
15.	Data-Element/Report-Cycle Inventory Form	86
16.	Input for Update 1, Location Known	87
17.	Part of Student Record Showing Grades Earned	88
18.	Input for Update 2, Location Unknown	90

PREFACE AND ACKNOWLEDGEMENTS

This report is the result of nearly thirty months of activity directed toward the design and implementation of a data system for a new institution of higher education. A few words about the nature of this report seem appropriate.

This report contains certain ideas on the maintenance and exploitation of data inventories. It does not contain a massive inventory of data germain to educational administration. Such inventories are available and many are of excellent quality and detail.

The advantages and disadvantages of the use of codes in place of natural data are discussed, and certain conclusions regarding their use are contained in this report.

Although programing considerations have been held in mind throughout, and the advantages and disadvantages of general purpose programs are discussed, this report does not contain a collection of computer programs.

This report argues for the use of programing languages which have grown in use and power over the years and which offer the greatest prospects for survival. This report does not present a new "educationally oriented" programing language, primarily because the authors feel that such special purpose languages constitute a poor investment.

A general concept for storage and retrieval, which we believe has practical application for computer installations of a relatively modest scale, is presented in this report.

The purpose of this report is manifold. It is hoped that the discussion of the nature and complexity of a "total" system will emphasize the need for careful planning--to educational administrators and data processing technicians alike; that the discussion of the complexity of the system and attendant high costs of implementation may cause the exercise of restraint in the implementation of over-ambitious and under-supported programs; and that sufficient idea content is present to provide food for thought for experienced systems personnel.

There are no new, wholly revolutionary, ideas in this report. Rather, they are the synthesis of many attitudes, ideas, and experiences in institutional data processing obtained through discussions with educational administrators and data processing technicians, as well as the literature.

This report is the outgrowth of many studies and insights relating to information systems in higher education. Of the many significant contributions made to this report, those which are most deeply appreciated address the higher education environment (William H. Cowley, Professor of Higher Education, Stanford University, in his Random Writings, 3 volumes) and the scope of the information system problem (Rev. R. J. Henle, Vice-President, St. Louis University, and others, in the report To Devise and Test Simplified Adequate Systems of Measuring and Reporting Financial, Manpower, Facilities, Research and Other Activities in Colleges and Universities). Important insights on what to do (and what not to do) were obtained from systems implementors, among whom James Blakesley, Purdue University; Charles Clark, California State College: Los Angeles; James Farmer, San Fernando Valley State College; George Hackney, California State College: Long Beach; William Mason, San Francisco State College; Theodore Syrgley, Florida Atlantic University; and Calvin Wright, Office of the Chancellor, The California State Colleges; were most imaginative and helpful. Other valuable assistance on current state-of-the-art machine processing methods was obtained from technical personnel with whom the project staff came in contact during the course of this study: primarily, systems analysts with General Electric Corporation, International Business Machines Corporation, Matrix Corporation, Radio Corporation of America, and many of the California State Colleges not mentioned above.

The most significant direct contribution to this report was that of Irene Okuda, Senior Stenographer, of the project staff, California State College, Dominguez Hills. Without her diligence and attention to detail, the many ideas and examples contained here would not have been effectively assembled.

JDM
DDD

INTRODUCTION

The California State College, Dominguez Hills, was established by statute in 1960, as South Bay State College. In 1962, Dr. Leo F. Cain was appointed President of the College and initial members of the planning staff were selected. In the Fall of 1965, the College began operation in temporary quarters on the Palos Verdes Peninsula, and in September of 1966 it was relocated to Dominguez Hills, to another temporary facility located next to its permanent site.

During the three years from the inception of the planning staff to the commencement of operations, a thirty-five year Master Plan for curricular and physical development was drawn. A planning figure of 20,000 FTE students, in the year 2002, is the basis for the planning thus far accomplished.

During the early days of the planning activity, the staff turned its attention to the definition of the management information which would be vital to the coordination of curricular planning and growth, campus development, and the effective allocation of the institution's resources. The staff recognized many significant advances in the use of computers in the administration of institutions of higher education, but it noted that the more significant systems appeared to be segmented so as to accommodate only limited aspects of planning and administration.

In the Spring of 1964, a proposal entitled "A Communications System for Higher Education," initiated by M. Milo Milfs, Dean of Students, was submitted to the U. S. Office of Education and subsequently approved as Project E-034. The proposal and the resultant research, planning, and design activity was undertaken with the full support and participation of the planning staff, which continues to address itself to the long-range objective of the development of the information system described herein.

The Problem

In recent years the population explosion has made a tremendous impact on institutions of higher education. A greater responsibility for providing college and university educations to the youth of the nation has been accepted by federal, state, and local governments, as well as private institutions and organizations. In the main, this growth has resulted in the rapid expansion of construction programs and the recruitment of increasing numbers of faculty and support personnel. At the same time, more institutions have sought to find ways to improve the utilization of plant and personnel, so that a high level of operating effectiveness would be obtained through the wise allocation of institutional resources.

A concurrent dimension to the problem of growth in higher education has been an increase in the number of technologies and curricula which are properly a part of higher education. New disciplines such as bionics, computing, and telecommunications have begun to appear, and entire new categories of facilities and faculty have been required to satisfy the needs of these disciplines.

In many instances institutional management, or the lack of it, has met the impact of growth by improvisation and the result has been relatively unplanned construction-modernization-modification programs which have left college campuses and catalogs in states of disarray. Recently, many institutions have embarked upon major catalog revisions in order to alleviate the problems which resulted from the rapid proliferation of curricula. It goes without saying that it is somewhat more difficult to clean up a campus cluttered with the results of hastily planned construction.

It is not our intention to devote this report to a discussion of decision making practices or philosophies, but in order to properly discuss an information system it has become necessary to make certain assumptions with respect to management practice as follows:

1. That management (or administration) is decision making in and of itself. (49, 57)

2. That regardless of the methods used in decision making processes (arbitrary action, committee deliberation, or staff recommendation with review procedures), the decision making process consists of the selection of that alternative which offers the greatest probability of the attainment of a desirable outcome. (53, 46, 48)

3. That decisions which are based upon information which is routinely collected and reported are frequently programed to the extent that the decision making process begins to resemble a clerical function. (51)

4. That decisions which are of a "one of a kind" variety are frequently made on the basis of a non-economic (wholly subjective) utility scale. (53)

5. That the availability of meaningful information (when brought to bear on the unique decision) tends to cause a more economically utilitarian decision to be made. (51, 27)

Both Simon and Luce treat the decision making process with respect to economic utility, Simon from the standpoint of Heuristic problem solving using models and man-machine systems, and Luce from the standpoint of subjective expected utility theories wherein the decision making must subjectively arrive at a utility value based upon a subjective or intuitive feeling of the probabilities of outcome. Obviously, the greater the availability of pertinent information, the greater the tendency on the part of the decision maker to adopt a systematic (objective) decision making procedure. (51, 53)

One of the peculiarities of the current growth situation in higher education is that historical data provide insufficient insight into future developments. The order of things is such that traditional information systems which define the scope and nature of operations no longer serve as an adequate guide in making unprogramed decisions. (14) The growth of emphasis on graduate study, for example, was not in evidence even twenty years ago. As a matter of fact, the State of California, with the largest and youngest population in the nation, is even now involved in a policy struggle concerning the offering of doctoral programs in the state college system.*

Individuals involved in the study and development of information systems argue that the data necessary for the detection of changes in trends are always available someplace within an organization. (21) They also contend that these data are maintained in a form which lends itself to analysis; i.e., these data are of the type which lend themselves to quantitative analysis. The problem, they will point out, is that a significant portion of these data reside in files in organizational components where they are neither readily identifiable nor accessible to the individuals who need them. Furthermore, these data are maintained through systems designed for the convenience of the immediate user, usually one responsible for programed decisions, rather than the entire organization. (29, 32, 45)

Organization and Management

It seems fair to say that the purpose of an institution of higher education is manifest in its implementors, the faculty, and that the decision making effectiveness of the academic administration ranges from salutary to strong, depending upon the administration's ability to negotiate effectively with the internal and external forces which are present. The relative vigor of the administration will have a lesser affect on its role than will the quality and quantity of the information which it may bring to bear on the problem under consideration, as the element of uncertainty of outcome diminishes with an increase in resource information. (53, 36)

---

*The Master Plan for Higher Education in California, adopted in 1960, authorized the California State Colleges to offer the doctoral degree jointly with the University of California. As recently as June, 1967, only one doctoral program had been approved and implemented. (8)

The primary activities of an institution of higher education may be defined in the broad categories of teaching, research, and public service.* Administration, including the planning, budgeting, and housekeeping functions, may then be viewed as supporting services which permit the primary activities to be accomplished.** However, unless performance in administrative or support services materially affects the primary activities of the institution, it is reasonably safe to say that the same "governing" rules apply to these services as to the primary activities. (23) Some of the traditions in higher education which tend to retard the development and implementation of the integrated information systems which would afford institutional planners effective bases for major decisions are related to governance and policy formulation, personnel selection, and academic freedom phenomena.

Governance and Policy Formulation.--At the risk of overgeneralization, governance and policy formulation are most commonly accomplished through the consultative process wherein leading members of the affected groups (usually through committees) develop, defend, discuss, and debate the issue at hand at increasingly higher levels, until the appropriate level for resolution is attained. (15)

---

*The report To Devise and Test Simplified Adequate Systems of Measuring and Reporting Financial, Manpower, Facilities, Research, and Other Activities in Colleges and Universities, Rev. R. J. Henle, S. J., et al., discusses ten categories of activities of academic/academic-professional personnel as constituting the full professional life within the university. While this study suggests that the total effort or time of the academic/academic-professional person is wholly contained within these ten categories of activities, it seems fair to say that the bulk of the time expended by the majority of the professionals in the college or university is expended in the three activities named above: teaching, research, and public service. (22, 18)

**The admissions process, because of its close relationship to the primary activities, may be viewed as an integral part of those activities.

Once policy has been established, the responsibility for the means of implementation becomes the province of individuals traditionally afforded high measures of autonomy. (15) Administrators, exercising the same measure of autonomy over the methods used by their staffs as are faculty members concerning their teaching, examining, and research methods, will only rarely be found to develop their data bases in such a way as to interface with those of other units of the institution, or control agencies, in order to provide the broader base required for higher level decision making.

Personnel Selection.--The consultative process, when applied to the selection of personnel, frequently results in "comfortable" appointments of personnel lacking in the decisiveness which may be required. (This, notwithstanding the fact that individuals so selected offered the greatest appeal to the broadest base of individuals representing organizational purpose.) (18) A common result of weak administrative assignments is the predisposition toward avoidance of conflict or disagreement with subordinate personnel as well as other components of the organization. (4, 17) Such a situation frequently leads to the making of decisions, which significantly affect the total scope of the institution, by individuals at the lower levels of the hierarchy where the perception of organizational purpose is limited.*

Academic Freedom.--The freedom of the faculty member to inquire, challenge, defend, and otherwise express himself, within the realm of his professional capability, is the very essence of the institution of higher education. That this freedom permits and possibly even encourages irrational behavior is of

---

*In the collected papers of Mary Parker Follett (Dynamic Administration) three methods of dealing with conflict are presented: domination, compromise, and integration. While we may assume that the typical administrator strives for integration, it is probably true that every administrator is constantly involved in all three methods. Because of the possibility of unhappy results, domination is something less than desirable. Because the process of compromise requires relatively little social importance, for the individual faculty member must accept the consequences of his action. (18) However, when irrational tendencies are allowed to invade the administration of an institution, the institution itself becomes responsible for these actions. (11) While it is desirable to introduce academically oriented individuals into administrative assignments in a college or university (in order to obtain communication, acceptance, and other working harmonies) this course may inappropriately bring with it a sense of personal autonomy which denies total organizational purpose. (3, 14)

Another manifestation of academic freedom is the deference accorded the individual on matters relating to his field of specialization. As decisions concerning the equipping of physics laboratories are made by the physicists responsible for the program, so are decisions regarding the form and content of student records left to registrars (and clerical personnel subordinate to the registrars), and decisions regarding the information collected on applicants for admission are made by the individuals responsible for processing those applications. Similarly, the structure of the institution's fiscal records is usually based upon the expedients required by the institutional business management function. (7) It follows that the programs (relating to instruction, research, and public service) are expressed in differing terms in the records of the administrative units responsible for housekeeping and support services. (41) The operational data defining the many activities within the academic departments do not, therefore, readily interface with the information bases developed to serve the institution's support functions. (22)

---

the giving up of one's own objectives, this resolution is not especially desirable. Because a significant additional effort is involved in obtaining both one's own objectives and those of a conflicting interest group, integration is frequently ignored in resolving conflict. The less dynamic and less courageous administrator, therefore, tends to avoid conflict through maintaining isolation, maintaining a status quo position, or ignoring the conflict should it develop. (17)

In summary, let us suggest that the academic freedom vital to an institution of higher education may well lead to autonomy within administrative units, and that this autonomy may materially affect an institution's capability in the assembly of information which is requisite to a decision making base.

Total System

McDonough and Garrett tell us to "beware of the total system," as a hazardous quest for the ideal. (2, 29) In the foregoing, we have touched upon some of the organizational complexities which seem to preclude the integration of the many information systems which operate within an institution of higher education. If we are to attempt to bring all of the information systems into a state where they provide management with the resources which are necessary for effective decision making, then we are looking at a problem of integration--the integration of the information resources of the organization. This we will call a "total system," with the understanding that the feature which distinguishes this total system from the one which now exists is the degree of integration of the data base and the completeness of information exchange which is afforded. In other words, a rapidly changing environment now forces us to adopt systems of information maintenance which will permit the fullest possible information base to be available to the decision making processes. (6, 35, 50)

What, then, should the characteristics of our total information system be? Taking the position of the systems analyst, who assumes that once the problem of the user is defined he may proceed, on a technical level, to develop all of the procedures and analyses necessary to arrive at a resolution to that problem, we venture that our system should have the following properties: flexibility, responsiveness, diagnostic capability, growth capability, and projective capability.

Flexible System.--By flexibility we mean here that the same system which provides for the traditional or routine operational reports for the organization should have the potential for exploitation in non-programed decision making activities. (50) For example, let us suggest that a new program of financial assistance for students whose family incomes are less than $4,000 per annum is under discussion. A flexible system, which routinely provides for the reporting of family financial data in respect to student scholarship, grant-in-aid, loan, tuition deferrment, and other student fiscal problems, must readily provide information on the significance of the proposed financial aid program to the students attending the institution.

Responsive System.--By responsive, we imply that the information system should operate without undue time lag; i.e., a capability for reaction to our "one of a kind" retrieval requirements should not require a massive effort where data collection, validation, and analysis is necessary. (55) (This might mean that the specific requirements of the decision maker may be so stringent that the existing information base cannot fully provide, but he may still obtain a response containing a significant amount of meaningful information.) For example, let us suppose that an administrator is called upon to report the number of out-of-state athletes being given financial aid, according to their race. While our data system may not accommodate a breakdown according to race, it should certainly be able to provide a roster of individuals receiving financial aid who have also lettered in one or more sports.

Diagnostic System.--The most common type of diagnostic system is the exception reporting system. (56, 30) An example of this type of diagnostic system is the inventory system which creates a report for all items where stock levels have fallen below a prescribed amount. In this instance, we feel that the diagnostic system should be extended in concept to include rates of performance which vary significantly from prior years, or from the overall institutional base. For example, let us suppose that the growth rate of the institution is 4% (in terms of FTE students), and that the growth rate of the mathematics department is 7%. Let us also assume that we have some indication of the relative growth of computer use as far as the total college enrollment and the enrollment within the department of mathematics is concerned. Let us say this growth figure is 5% based upon the last few years of experience. A diagnostic system should tell us not only that in the last year our computer utilization figure has jumped to, say, 11%, but where the additional use has been fostered. In this way, management may ascertain background information which would permit it to caution users about the problems of overconsumption of institutional resources, or to recognize a new and significant user of computing equipment, or to otherwise adjust to a new situation. (54)

Growing System.--By growing system, we mean one which is open ended. As in the aforementioned case, where a report which indicated the racial constituency of the athletic and financial aid programs, a flexible system should provide for the addition of data specifying the races of students, in the event that such information will be routinely required in the future. No major systems revision should be necessary for such an addition, as the addition of this element of information to the student record should be accommodated through an open-end device which does not involve record reformating and/or machine reprograming. (25)

Projective System.--A projective system possesses essentially similar characteristics to the diagnostic system mentioned earlier, requiring a well founded base in historical data. But it views this historical base slightly differently: It assumes that what is happening and what has happened will continue to happen at the same rate of change as has been encountered previously. A projective system combined with a diagnostic system would provide an administrator with two kinds of summarizations which would make his life much more livable: confronted with a projective system's reports which indicated the next one or two years of expectations (according to the last three, five, or seven years of experience) and a diagnostic system's reports showing serious or significant deviations in activities which occurred during the last one or two years, our administrator should possess the very essence of planning information.

Economic Considerations

Every information system has, as its foundation, the records of the transactions effected or supervised by task oriented organization units. The sum of all transaction records may be defined as the total data base for the organization. (21) Automatic data processing has greatly facilitated high volume record keeping operations. At the same time, the use of computers has permitted the production of various management and control reports as by-products of the record keeping function. (26)

The level of sophistication of a management information system is directly related to the comprehensiveness of the information made available to management, depending upon the decision to be made. For example, when considering a new program the following questions arise:

1. What is the need for this program (internal as well as external)?

2. What resources are required for the implementation of this program?

3. What existing resources are available for this program?

4. If existing resources are diverted to the new program, what affect might we experience on our existing operations?

5. What additional resources (over and above those already available) must be obtained in order to implement the program?

6. What will these additional resources cost?

7. How will the additional costs be met?

8. In view of the above, how should the implementation of the new program be scheduled?

Obviously, the breadth of the data base which is required for answers to the above questions varies significantly depending upon the nature of the program under consideration. If the new program is to be a course, the information required will probably be available within the department through which the course will be offered. If the new program is to be an entirely new curriculum, the resources of the entire institution should be considered. If the program under consideration is to be a new campus, the information required for answers to these questions would far exceed the information resources of the institution itself.

The structure of a management information system somewhat resembles a classic ladder of abstraction. The higher the level of decision making we consider, the more inclusive and generalized our information requirements must be. Conversely, the lower the level of decision making, the more exclusive and detailed are our information requirements. Figure 1 illustrates this phenomenon in terms of three levels of inputs: transactions, information requirements, and information requests. Corresponding outputs are records, reports, and studies. (21)

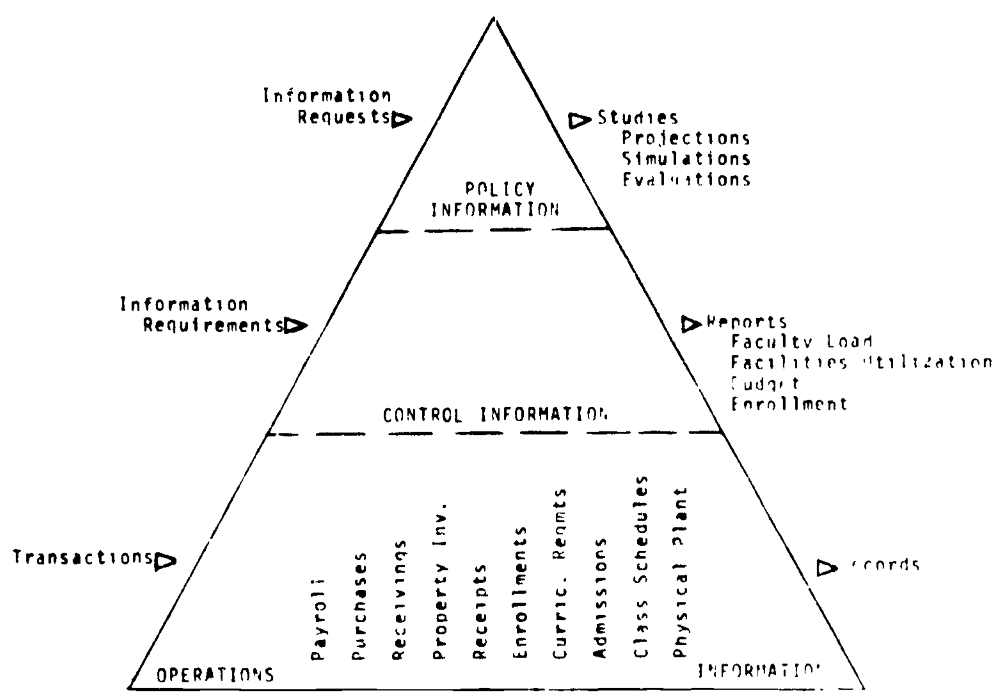

Figure 1. Management Information System

The information systems analyst must develop linkages so that the respective data base segments containing transaction records may be readily related. Some of the relationships are relatively easy to perceive (as would be the case in relating payroll to student enrollments, using class schedule as the link, in order to obtain information on the direct cost of instruction), but if experience is any guide serious problems may be anticipated (as would be the case when instructor name in the class schedule was not identical with the name used in the payroll operation). (22) In addition to the problems encountered in horizontally relating the organization's data base segments, the analyst must also know the nature of the information required by higher levels of decision making. Given a complete set of specifications on the form and content of the reports and studies required, the problem is essentially one of programing information retrieval and summarization. However, all too often, administrators unaccustomed to relatively new management techniques are unable to contribute significantly to the design of the system, or to the definition of its output. (10, 20) The organization may thus find itself in a situation where top management moves to implement an information system which is beyond the effective utilization of its staff. This problem is often furthered by a communications gap between the professional systems analysts and the administrators whom they desire to serve. (21, 26)

A current trend in higher education, and one most probably selected as a means of overcoming some of the difficulties of data base integration, is the creation of an office of "analytical" or "institutional" studies. (37) Its counterpart in private industry would be labeled management analysis or operations research. Such organizations commonly obtain parts of their data base in the form of copies of transaction records. (24) The transaction records are processed to add identification codes which permit linkage to be established with other elements of the data base. The remainder of the data base is usually obtained through special studies, inventories, polls, etc. (5) Such an approach often leads to duplication and, unfortunately, another incompatible data base within the organization.

Third generation computer systems provide great capacity for routine processing, and afford improvements in information flow at the record maintenance and control report levels. The current trend in the utilization of third generation computers is the pooling of most or all of the information processing of the organization in order to obtain sufficient workload to feed the ultra-fast machines. (20) At the same time that increased speed and computing capacity is made available to the existing record keeping and report preparation aspects of the operation, additional benefits in the form of improved and more permanent programing languages, improved compiler diagnostics, bulk storage, and direct access communications become available. (12) But third generation equipment is costly if one is to acquire the equipment which is necessary to obtain the secondary advantages mentioned above. Also, the systems and programing effort required to take full advantage of third generation capabilities is costly, for it means a redoing of that which has already been done and at a somewhat more advanced level of sophistication. The addition of another data system such as analytical studies merely adds to the cost, unless it is a by-product of the existing data system.

Progressive managements are currently approaching the problem of upgrading their management information systems by either: a) substantially refashioning their existing data systems or, b) by developing and implementing entirely new systems--while in the process of implementing third generation computing equipment. The decision to revamp or completely rebuild an information system in order to fully integrate the data base is most probably made on the basis of the availability of funds, the objective being acquisition of the highest level of management information at the optimum cost. (21) This report presents a plan for a data system which is appropriate in either of the above cases.

Objective

The objective of this project is the development and implementation of a total information communication system for an institution of higher education. The system concept requires the maximum practicable integration of the data base in order that:

1. Instructional methods may be researched and evaluated by the faculty and educational programs and policies improved;

2. The funds allocated to public higher education may be effectively described in terms of the educational objectives for which they are appropriated;

3. The routine burden of information processing in higher education which is now shared by students, faculty, and administrative staffs will be reduced and this energy may then be made available to more significant educational achievements; and,

4. The administrative and service responsibilities of higher education can be facilitated through effective interagency communication.

Recognizing the constant presence of change in report requirements and in programs and activities in higher education, and in deference to McDonough and Garrett who suggest that caution should be exercised in attempting to be all inclusive, the initial phase of this project includes the definition of: (29)

1. That core of the institutional data base which must be routinely maintained; which is commonly required for various control reports; and which is vital to major planning decisions within the institution;

2. Operational procedures and programs which provide for the maintenance of the data base; which produce routine control reports; and which provide meaningful management planning information;

3. A method for the extension and modification of the data base which will facilitate growth, flexibility, and responsiveness;

4. Methods for the establishment of linkages between elements of the data base, as well as between the college information system and the reporting requirements of other organizations and agencies.

Another dimension of this project is the selection of data base elements to be included in an initial phase of computerized operation. While the total information requirements of the institution must be held in mind throughout the planning and designing activity, the limitations of resources and the changing nature of the institution's information requirements preclude the implementation of a complete data system at a given point in time.

The criteria used in the selection of data elements to be included in the system were output oriented. In the early stages of project planning, it was agreed that the ideal institutional information system could be generally defined in terms of the information required by the operations, control, and planning activities of the organization. A brief listing of these information requirements categorized according to subject were:

Students.--What kinds of student information are essential for: the admissions process; the development of a program of study; the faculty in academic advisement; the student in self-assessment of progress; the administration in predicting course enrollments; academic rosters and lists; the identification of exceptional students; the administration of fees, loans, and scholarship funds; the preparation for transmittal of the student record of academic achievement, activities and work experience; and the evaluation of curriculum?

Curriculum.--What kinds of curriculum information are essential for: the catalog; facilities requirements and the scheduling of classes; the universal description of course content and credit (without standardizing course labels or course numbers); the machine determination of inter-institutional course equivalencies; and the preparation of reports for comparative purposes?

<u>Faculty</u>.--What kinds of faculty information are essential for: the recruitment program; the preparation of payroll; the evaluation for promotion; the administration of vacation and leaves of absence; the catalog; the scheduling of classes; the identification of research talent; and preparation of reports for comparative purposes?

<u>Staff</u>.--What kinds of non-teaching staff information are essential for: the recruitment program of staff; the preparation of payroll; the administration of vacation and leaves of absence; evaluation for promotion; the advancement schedule; the identification of special abilities; and the preparation of reports for comparative purposes?

<u>Facilities</u>.--What kinds of plant and facilities information are essential for: optimizing facilities utilization; planning campus construction; facilities maintenance; and preparation of reports for comparative purposes?

<u>Fiscal</u>.--What kinds of fiscal information are essential for: the preparation of monthly balance statements for all institutional accounts; the cost analysis of instructional, support, and administrative programs; billing for fees and charges; the preparation of the budget; the control of property and equipment; the inventory of materials and supplies; and the preparation of reports for comparative purposes?

This cursory inventory of information requirements brought to mind some less readily defined problems dealing with the interrelationships of the information itself. Questions arose concerning the manifestations of student curriculum requirements in faculty, facilities, library, and budgetary planning:

1. Did current techniques in the accounting and evaluation of enrollment data permit the best planning insight to be obtained;

2. Were existing methods of classifying faculty

qualifications and facilities capacities adequate for near and long-term planning; and,

3. Were current budgetary formulae adequate for the effective evaluation of workload, performance, and level of support?*

---

*The California State College system with 172,000 students and an annual support budget of nearly $200,000,000 has for many years employed the use of formulae in determining the requirements for support appropriations. (10,9) The basic statistic in the evaluation of workload (and resultant support requirements) was the FTE student which represented fifteen credit hours of student enrollment. A breakdown of FTE by type of instruction (lecture-lab) and by discipline provides the current basis for space planning. A similar breakdown with the additional dimension of undergraduate/graduate study introduced provides the criteria upon which the number of instructional faculty are based. Public service allocations are based upon total FTE students. Student services staffing is based on a combination of FTE and individual student enrollment. Library budgets are based upon a target figure of forty volumes per student, with staffing allocations based upon the number of acquisitions planned for a given year. Essentially, the approach taken to budgeting (via the various formulae) consists of taking each major program or activity of the college, analyzing relatively crude data which defines the scope of operations for the prior year, and applying a growth factor to prior year performance in order to establish current year funding requirements. While this system does permit adjustments to be made to accommodate increases in workload, it does little to provide for subtle differences in instructional methods (indeed, instructional objectives within the various state colleges). Exploratory or experimental activities must be supported through the diversion of resources which have been justified on the basis of existing budgetary policy. New programs of too great a magnitude to be thusly supported are submitted through channels to the Executive branch and the Legislature of the State of California where program justifications are put on trial in the court of current

Out of these questions came still another problem requiring resolution: To what extent should the development of this information system involve the questioning or challenging of existing operating, control and planning systems? It was decided that the focus of attention should be placed upon the development of a data system which would meet the needs of existing requirements, and that (if effectively designed) ample opportunity would be available for exploitation and experimentation.

---

political thinking. (State Senator George Miller, Chairman, Senate Committee on Finance, recently estimated that approximately seven years of deliberations were required before a new program of any significance was adopted and funded by the state government.)

METHOD

The implementation of this project may be divided into two categories of activity: The definition of the institution's information requirements; and the design and implementation of the operating data system. Although these two activities have been and will continue to be carried on concurrently, a gradual change of emphasis from preliminary study to system design has occurred during this project.

As the preliminary steps (data definition) were undertaken, it at once became obvious that a significant problem would exist in attempting to differentiate between data and system: When discussing coding methods and data acquisition procedures, we were frequently confronted with statements suggesting that the reason for certain data being coded in a specific fashion was due to the limitations of existing equipment. While such problems merely reinforced our position that an exhaustive inventory of information requirements was called for, it led us to the point where we began to question the viability of the data descriptions we were collecting. A disregard for the definition of the data being collected germinated and grew to the point where we began to discuss data only in terms of output.

Although the methodology for this project, which is discussed in the following, is not apropos of the scholarly exercise, it is fairly representative of an action program in the data processing area. We hope that the significance and benefit from first person communication (from which we have gained greatly) will be more fully exploited by those following this course.

Preliminary Study

Initial steps in the implementation of the study were addressed to the definition of the information requirements of the institution, including those reports which were prepared at the request or direction of other organizations and agencies. Preliminary study included a survey of the literature, an inventory of existing report requirements, and relatively informal contacts with both educational administrators and data processing specialists in educational organizations.

Survey of the Literature.--Research to determine the availability of appropriate literature was conducted along several topical lines, as follows:

1. Education

   a. Administration
   b. Data Processing

2. Management

   a. Control Systems
   b. Information systems
   c. Models
   d. Systems

3. Data Processing

   a. Data Management
   b. Information Retrieval
   c. Integrated Data Processing

The indexes, library catalogs, and bibliographies researched for writings along these topical lines are listed below.

Applied Science and Technology Index. New York: The H. W. Wilson Company, July, 1963 through June, 1966, Volume 54, Numbers 3, 6, and 8.

Business Periodicals Index. New York: The H. W. Wilson Company, July, 1963 through June, 1966.

Education Index. New York: The H. W. Wilson Company, July, 1963 through June, 1966, Volume 38, Number 7.

International Index to Periodical Literature. New York: The H. W. Wilson Company, June, 1964 through February, 1966.

Readers' Guide to Periodical Literature.
New York: The H. W. Wilson Company,
May, 1964 through August, 1965.

Card Catalog of the College Library, California
State College, Dominguez Hills.

Card Catalog of the College Library, California
State College at Los Angeles.

Card Catalog of the College Library, University
of Southern California, Los Angeles.

Catalog of the Los Angeles County Public Library,
Los Angeles.

Card Catalog of the Municipal Reference Library,
City Hall, Los Angeles.

Operations Research Group. Comprehensive
Bibliography on Operations Research.
New York: John Wiley and Sons, Incorporated,
1958.

Literature indexed under educational administration and educational data processing provided an abundance of information describing computerized methods for the processing of educational data. The bulk of the material was found in periodicals and the terms "total system," "integrated system," and "data bank," were found to be fairly common. However, almost without exception, total systems were limited to a specific area of operation, such as a comprehensive student scheduling and registration package; integrated systems were almost always discussed in terms of vertical integration where machine readable data was captured as a by-product of an existing operation such as cashiering or purchase order preparation, with succeeding layers of report and analysis preparation made possible without additional data preparation being required; and data banks were usually limited in scope to a single file, such as the academic record of the student.

By far the most significant literature obtained was a draft of the final report on the project, To Devise and Test Simplified Adequate Systems of Measuring and Reporting Financial, Manpower, Facilities, Research, and Other Activities in Colleges and Universities, and its supplements. This report prepared under the sponsorship of the National Science Foundation and the National Institutes of Health was based on a proposal submitted by Rev. R. J. Henle, S. J., Academic Vice-President of St. Louis University, and it involved the participation of the Universities of Arizona, Florida, Pennsylvania, and Washington, Michigan State University, Texas A & M University, and Rensselaer Polytechnic Institute. The report discusses the need for the collection and maintenance of data defining the resources of the institution (manpower, facilities, and funds) and its clientele (students). Inventories of data elements and discussions of their utility are included, together with suggestions on the definition of terms, coding techniques, and data collection orocedures. The latter chapters of the report, together with a supplement prepared at Texas A & M University, discuss the problem of file integration and offer general suggestions for the structuring and maintenance of data files. (22, 7, 18)

Another influential resource was a collection of writings by W. H. Cowley, which discussed the nature of the college and university environment. Of particular importance was Cowley's discussion of the affect of history upon the policies and decision making processes of the institution. (14) His argument that history itself is a causative factor in the decision making process was so persuasive as to convince these authors that a maximum of detailed historical information should be retained for an indefinite time period for purposes to be defined in the future.

Literature on management, control, and information systems and related topics was similarly abundant. Again, the bulk of the writings was contained in periodical literature, and it described successful attempts at automating information processing procedures. Again, "total," "integrated," and "model" systems were under discussion, and, as in the above case, a close examination usually revealed a relatively limited degree of horizontal integration. "Total" systems encompassing sales forecasting, production control, and materials warehousing and distribution, were common, but rarely integrated with each other or with other elements of the operational and control levels of the organization's information system. (10, 21, 55)

The literature on information retrieval, integrated data processing, and data management, although not as abundant as in the above cases, was found to be available in sufficient quantity to provide some direction in system design which was not otherwise obtained. However, information retrieval usually led to a discussion of the use of random access memory devices; integrated data processing was usually focused on real-time remote inputs, or the use of a card or tape producing slave device attached to a document creating machine such as a typewriter, cash register, or bookkeeping machine; (12) and articles concerning data management usually dealt with the cost/benefit analysis which should be undertaken before wholesale data collection projects were undertaken. (20, 21)

In summary, the survey of the literature made the authors aware of a high degree of interest and activity in the formulation of concepts and the implementation of segments of total systems, and it created the feeling that time would see their realization.

Report Inventory.--The inventory of the institution's report requirements was greatly facilitated by the release of drafts of the reporting system to be implemented by the U. S. Office of Education (known as HEGIS, Higher Education General Information Survey). Various other federal questionnaires, primarily dealing with the financial aspects of the institution, were collected and inventoried so as to obtain a guide to the content and structure of the data base according to the requirements of the federal government.

The definition of state and state college system reporting requirements posed much greater difficulty. Approximately one hundred fiscal reports required by the State Administrative Manual and the State College Administrative Manual were inventoried. (41) Several major changes in the form and content of many of these reports have been implemented during the last few months of this study* In addition, current policy in

---

*The California State College System, through the

the State of California is geared to the implementation of a program budget during the next few years. (33) Although such a budget construct must be based on the integration of fiscal data with that defining the activities and programs of the organization, no specifications defining either the form or content of future budgetary reports is yet available.

While the uncertainty of state reporting requirements was initially believed to be a serious handicap in the design of the data system, the realization that such a situation would probably exist at any given point in time, reemphasizing the need for system growth

---

Office of the Chancellor, is undertaking a major revision of the statistical reporting required of the eighteen colleges. The first phase of what portends to be a massive reporting system was relased to the colleges on March 15, 1965. Although the first reports specified in the initial phase of an "Information Service Guide" are limited to the term-by-term reporting of enrollment data, sufficient information is present to allow a number of audits and analyses to be performed: (24)

1. Biographical data on new students facilitates the study of the sources of students in each of the colleges, thereby providing a basis for the interpretation of public school enrollment data developed by the Population Research Section of the State Department of Finance;

2. Summary data describing the academic loads carried by students (according to sex and enrollment status) provides a basis for the auditing of the individual and FTE student enrollment which serve as budgetary criteria; and,

3. Summary enrollment and scheduling data for each class section provide a base for the study of facilities utilization, enrollment trends by level and discipline, and, at the same time, it provides a validation base for the reports mentioned above.

and flexibility, resulted. Techniques which would provide the necessary properties of orderly growth and flexibility became the chief topics for discussions in the field.

Field Work in the California State Colleges.--Initial field investigation was limited to twelve of the California State Colleges, the Office of the Chancellor of the State College system, and the (California) Coordinating Council for Higher Education. Administrators having involvement in, or responsibility for, data processing were contacted. In addition to the collecting of data acquisition forms and documents and samples of machine output, discussions were held concerning future machine processing plans, existing or proposed methods of integrating various data files, methods and techniques used in changing or expanding the data files being maintained, and the nature of the organizational framework within which data processing services were accomplished.

Almost without exception, data processing activities in the colleges visited were limited to the admissions and records area, and were characterized by low-speed tabulating equipment of limited capability.

For the most part, plans for the improvement of data processing services were limited to the acquisition of faster or additional tabulating equipment, so that existing applications could be accomplished in less time. Two colleges were exceptions to this rule: The California State College at Los Angeles and San Fernando Valley State College, both having recently installed small scale (tape-oriented) computer systems (IBM 1401 and GE 225, respectively,) were working to convert existing admissions and records tabulating operations to their small computers and, at the same time, beginning to develop plans for the implementation of applications for the business management functions of the colleges. Even these colleges, with relatively sophisticated equipment, were found to be planning and implementing data processing applications without a general plan. File integration, as such, was not identified as an objective, and computerized operations were based upon the requirements of operating and control personnel with little regard to the needs of the management and planning functions. Programers (usually students) were assigned to work with the lower and middle management personnel of various organizational units, where objectives were almost exclusively focused on speeding up existing operations. (It must be pointed out here that this approach was probably the only alternative open to these colleges due to the fact that staffing formulae did not permit the employment of systems analysts; machine processing was traditionally identified with operating levels where volumes of data were present; and the time requirements for major system studies--if implemented--would have denied immediate relief to overloaded operations.)

As long as three years after the most recent Coordinating Council Cost and Statistical Study (Fall, 1963), few attempts at the development of an integrated data base which would permit the assembly of a cost/performance profile on each class section had been made by the colleges: (5) Building designations, as contained in class schedules, only rarely correspond to those maintained in the college's facilities inventory files; faculty identification contained in class schedule records only rarely agreed with that established for the payroll; and, of course, department and course number designations presented an alphabetical matching problem with the curricular master file (if such a file existed).

Discussions which focused on the methods employed when a change in report requirements was to be effected brought to light the fact that the preparation of a new data base was often required. Simply stated, the reason for the existence of this problem was that the limited capabilities of the tabulating equipment then in use required a heavy reliance on coded representations of data rather than "natural" data. For example, the requirement that each laboratory section of lecture-lab courses would be separately reported beginning in the Fall of 1963, resulted in the need for an entirely new data base defining weekly student lecture and laboratory hours, related student credit hours, and laboratory section enrollments. A similar problem would be encountered in modifying student class level where a system used codes 1, 2, and 3, to denote lower division, upper division, and graduate levels, respectively, and a change which required a breakdown according to freshmen, sophomore, junior, senior, graduate, and post-Master's was to be implemented. (Natural data is usually avoided because the card columns required would far exceed those used for a code.)

Discussions concerning the organization and administration of data processing services brought out that, in most of the colleges, the tabulating installations were responsible for serious dislocations in other sectors of the college organization. Almost without exception, they failed to provide services for any operating unit of the college outside the admissions and records area. A review of the history of these tabulating installations makes the reason for this disposition clear: Tabulating equipment was approved for installation in the admissions and records area by the State Department of Finance in 1958, with the proviso that the personnel required for its operation would be taken from the existing staffs of the college registrars. Not only did this impose a severe limitation on the perspectives of the individuals charged with operating the equipment but, in the light of organizational autonomy discussed earlier, the tabulating installations began to function in unchallenged isolation. In 1964, at the behest of the State College Dean of Institutional Research, the Trustees of the California State Colleges adopted a policy consolidating the data processing activities (instructional and administrative) under a single (unnamed) administrative officer on each of the campuses. Although physical decentralization was to be permitted, the Trustees' resolution made clear that data processing was considered to be a collegewide service to be provided to students, faculty, and administrators alike.

At about that same time, some ten or twelve additional small computers were installed, bringing the total number of colleges possessing computing equipment to fourteen. Unfortunately, this equipment was justified on the basis of the needs of the instructional programs of the colleges, and it was installed without provision for professional staff for its administration. The lack of professional staff has given rise to poor machine utilization practices, piecemeal systems and programing work, and a lack of concern for the quality of the data entering the systems. In addition, the incidence of duplication and overlapping of systems and programing effort between the colleges has risen steadily due, at least in part, to the fact that such work is performed by student assistants working without the benefit of knowledgeable supervision which, if available, would certainly keep informed on the activities, successes, and failures of its counterparts in the other colleges.

Other Investigations.--In addition to this preliminary reconnaissance within the state college organization, contacts were made and information obtained from a number of colleges and universities outside this system.

A number of articles appeared in periodical literature describing recent successes in the development and implementation of scheduling programs. Because a scheduling system would require integration of the data bases characterizing student enrollment demand, faculty teaching specialty, and facilities types and capacities, several of the institutions reporting such systems were contacted by the project staff. With one exception, the scheduling packages were found to be systems where student registration data was submitted to a data bank which maintained a current inventory of space in classes scheduled for the following term. A computerized search was made and the student enrolled in the courses which he required, enrollment documents were produced, and the space inventory maintained by the data bank was updated. (1, 52) The notable exception to this was the CUSS (Comprehensive University Scheduling System) system developed by Purdue University for implementation at Seattle University, Washington. This latter system definitely represented a step forward in that it interpreted student enrollment demand, faculty teaching competency, and facilities capacities, in order to minimize time and place conflicts in the preparation of the class schedule itself. The result was a relatively complete integration of three files: Student Enrollment, Faculty, and Facilities. (2)

Another example of the integration of data bases is found at Florida Atlantic University where a library system is being installed in accordance with a design developed by the University of Illinois and General

Electric Company. This system interfaces the files associated with acquisitions (including fiscal data as well as on-order information), cataloging, and circulation control (including both book and borrower identification). One particularly interesting aspect of the FAU library system was that, nearly two years after initial implementation, a restudy was being implemented in order to increase the efficiency of operation, and to tighten the interface between files in order to permit a greater exchange of information from one to another--notwithstanding the fact that initial implementation was based upon an exhaustive and detailed systems study.

Other achievements in the integration of the institutional data base have recently evolved: the Southern Methodist University (resource allocation) system; (7) the IBM sponsored University-College Information System (primarily fiscal); (26) the AACRAO sponsored guidelines for the analysis of space utilization; (42) and the cost study of the California Coordinating Council for Higher Education. (5) With the exception of the work done at Southern Methodist University, none were found to expressly seek file integration as an objective.

System Design

The design of a machine based information system may be as simple or as complex, as basic or as sophisticated, as available equipment and personnel will allow. Prior to the commencement of this project, a consultant retained by the college planning staff had estimated that the system requirement for machine readable data concerning students totaled approximately 11,000 characters per student, a very discomfiting figure for an individual accustomed to working with punched-cards. In discussing this file size with data processing personnel during the preliminary phases of our study, reactions ran from awe and ridicule at one end to indications of sincere interest on the other. Questions ranged from "how will you ever build that data base when it is so difficult to acquire a keypunch?" to "Because you can't possibly maintain all of that data for each student within a single record, how will you keep track of the whereabouts of the data on your students?" (Both of these happen to be very good questions.)

Initially, system design was categorized into five basic activities: data definition; file design; data acquisition procedures; storage and retrieval procedures; and applications programing. Although the interdependence of each of these activities with the others is immediately apparent, it was felt that some division was necessary and that a precise definition of the data base was prerequisite to data acquisition, file design, storage and retrieval, and applications programing efforts. A form was developed for the inventorying of information items which came under discussion during personal interviews, as well as through readings in the literature and inventories of report requirements (Figure 2).

The following pages discuss the five categories of activity, relating to system design, in general terms. Emphasis is on operating concepts which were developed after many discussions (more detailed than those mentioned above) covering a broad spectrum of considerations, ranging from administrative decision-making to technical data processing problems. Discussions concerned with administrative decision making problems were primarily conducted within the college planning staff. Technical, or operating, problem discussions drew from a much wider base of experience, including individuals representing most of the California State Colleges, other colleges and universities, equipment manufacturers, consultants, and software houses.

<u>Data Element Inventory</u>.--The making of an item by item inventory of the information requirements of an institution of higher education is not a simple program for implementation. One is immediately confronted with problems in the definition of terms as well as very subtle nuances of meaning in the words used to define the terms under discussion: To a facilities planner, a room is something bounded by four walls, a ceiling, and a floor, while a professor of drama may see it as an area of 20 feet by 40 feet, equipped with certain types of stage lighting devices, props, recording and broadcasting equipment, and an

Inventory Item #______________

Data Descriptor ____

____

General Description ____________________________________

____

____

____

____

____

<u>Source (initial)</u>

Organization Unit ______________________________________

Process ____

Form Name & Number _____________________________________

Field Heading and/or Number ____________________________

<u>Source (updating)</u>

Organization Unit ______________________________________

Process ____

Form Name & Number _____________________________________

Field Heading and/or Number ____________________________

Figure 2. Data Inventory Form

audience capacity of 300; entrance examination scores obviously have different meanings to the director of admissions and the chairman of the Department of English; and high school of graduation obviously means one thing to the college's director of institutional relations and another to the State Department of Finance, Population Research Division--but all of these values are mere numbers to the keypunch operator, and data to be used in sorting, comparing and computing to the programer.

In attempting to inventory and define the data to be maintained and produced by the system, a high incidence of redundancy was detected. In investigating the student record keeping operations of fourteen distinct programs in the Student Affairs area, we found that only a small percentage (less than 20%) of the information being maintained (in twelve or more offices concerned with student affairs) to be unique to the institution's data base. Eighty percent, or more, of the data being maintained was found to be already in existence (in many cases in machine readable form) within the Admissions and Records offices. When questioning the need for this duplication, we were usually informed that the activity was necessary, due to the inability of Admissions and Records to provide the information which was required within the time requirements of the user. On deeper investigation, we found that much of this redundancy was not redundancy at all, but, rather, different data being maintained under the same label (e.g., "enrollment status" in one file might actually reflect grade level; in another it might reflect the difference between a continuing student, a first-time freshman, or a transfer; and in a third, it might indicate whether the student is currently enrolled in the institution or whether he is no longer active). The problem of attempting to build a data system when different kinds of data are inventoried under the same labels, or when identical kinds of data are inventoried under different labels, should be immediately apparent. This one problem, more than anything else, led to the abandonment of the data inventory form and the growth of an argument in favor of centralized record keeping for student data: If a unit of the organization possesses such unique information requirements as to warrant an independent data system, then the central system need not concern itself with peculiar or private data definitions; and, if a unit of the organization possesses an information requirement compatible with that of the major or dominant system, then it should accept the data names and descriptions employed by the dominant system. While in some cases this may require the invention of labels quite different from those now in use (in order to eliminate confusion during and after conversion), in most cases the problem may be solved by employing a set of data names and definitions designed for a specific purpose and having a high degree of internal integrity (such as those provided by AACRAO and the U. S. Office of Education). (13, 31, 38, 39, 40, 43)

The representation of data by the use of codes, which categorize different classes or properties, has long been used in machine processing. Primary advantages resulting from the use of codes are: conservation of space in machine records; expedient segregation and summarization; and significant reductions in the time required for the sequencing of the files. However, significant disadvantages lie in the use of coded representations of data: The need for more detailed classifications often arises, requiring the reconstitution of the entire file, or various special purpose uses of a particular category of data often result in multiple coding systems employed to define essentially identical data bases. For example, a student with seventy semester hours of credit may be considered a junior on the basis of the work completed, while at the same time he may be considered a sophomore on the basis of the work required to be completed prior to the award of his Bachelor's Degree; or a laboratory designed for physical chemistry courses may be defined by assignment to the physics department as a special purpose room containing sixteen stations, and it may be reported to the State College Trustees, the California Coordinating Council for Higher Education, the U. S. Office of Education, and the National Science Foundation, respectively, as a teaching laboratory coded 7728 (Chemistry), 2620 (Chemistry), 6807 (Chemistry except Pharmaceutical), and (no code) Physical Chemistry.

Noting that the above code numbers represent the broad subject matter classification of chemistry, and anticipating from past experience that a more specific breakdown will be required (i.e., physical chemistry) led us to the conclusion that coding systems would only be completely effective when they represented all of the uniqueness of the natural data. Obviously, this negates the philosophy of coding and argues for the maintenance of a data base in natural language form.

Another problem area encountered in the inventorying of data was concerned with the amount of data relating to one file which should be carried in a transaction record that was properly a part of another file. For example, how much information describing facilities, faculty, schedule, and student characteristics should be contained within an individual course enrollment record? The more data which is present, the greater the utility of that particular record: relatively complete facilities and schedule information would permit a comprehensive facilities utilization analysis to be made with the course enrollment record as the data base; important student bio-data could be provided instructors on class rosters prepared from the course enrollment records.

However, it is not physically possible to retain all desirable facilities, faculty, schedule, and student data in individual course enrollment records. Nor is it desirable to eliminate all elements of any type of data, for those elements which establish effective linkages with the files constituting the institution's data base must be retained.

In order to obtain maximum benefit from a large scale data system, it is imperative that the users (both line administrators and individuals engaged in research and institutional studies) be made fully aware of the data base which is available to them. One of the chief disadvantages in the traditional compartmentalized or simi-autonomous information handling systems in institutions of higher education is that individuals needing information lack the means to determine its existence and whereabouts. In addition, in instances where the availability of information is known, users frequently lack the tools by which to access this information.

In summary, the definition and inventory of the data base led to the formulation of three broad operating concepts:

1. Semantic problems and redundancy (both real and false) may only be overcome by the maintenance of a precise catalog of data being maintained in the system.

2. Ever changing report requirements (many of which are outside the institution's control) tend to make most coding systems ineffective over an extended period of time. Coding techniques should be employed only when they are designed to characterize the maximum number of classifications of the data being represented. In the last analysis, natural data should be preferred, notwithstanding the additional cost and difficulty of validation and interpretation.

3. Insofar as it is practicable, unit records (punched-cards) representing individual transactions, or items of activity, should contain information which facilitates the machine accessing of appropriate related files. The inclusion of interfile linkages within transaction records will afford the greatest degree of system flexibility, though somewhat reducing the direct utility of the transaction records themselves.

4. The multiple and diverse uses to which a comprehensive data base may be put may only be realized when its form, content, and location (or method of access) is made known to all prospective users.

File Design.--During discussions on the most effective means of designating file content through a convenient label, terms such as master file, working file, general purpose file, multi-purpose file, and special purpose file were explored. Master files were usually characterized by the term "stable" or unchanging. Working files were usually suggested to be those with only a temporary usefulness. No satisfactory definition for a general purpose file was forthcoming. Multi-purpose files, on the other hand, were frequently described as possessing information which could be put to a wide range of uses, such as a file on course enrollments which would permit facilities utilization and faculty load studies to be produced, along with class rosters and study lists. Special purpose files were generally regarded as those containing data extracted from master and/or working files but limited in content to the requirements of a particular processing operation.

The classification of files according to the data contained therein does not seem to pose a serious problem: data defining the physical plant is contained in the facilities file; data relating to students belong to the student file; faculty and staff data belong to the personnel file; and data defining revenues and expenditures belong to the fiscal file. In practice, however, these distinctions are not so easily drawn: data defining the course requirements for a particular degree, and which are considered to be a part of the curriculum file, may also be proper elements in the files of students pursuing that degree. Similarly, the characteristics of special purpose facilities which are included in the facilities file may also be integral parts of the curriculum file serving as a basis for the preparation of the class schedule.

Early discussions of file structure failed to produce insightful techniques for the elimination of redundancy of data between files, or for the determination of the form and content (and even the number) of working files which would be required. As a matter of fact, the development of the data catalog wherein each item of data deeme to be vital to the system was inventoried once and only once, led to the concept that the entire data base for the institution could be considered as a single file, with consideration of the form and content of subsets (or file segments) to remain for a later consideration.

It must be pointed out that at the point in time where file design was initially considered, random access peripheral equipment was available, but extremely costly. Insufficient economic benefit existed in justification of the use of random access equipment for the storage of the totality of the data base, especially considering the fact that much of the data would be of limited initial utility. File organization was, therefore, considered as a complex of sequential file segments (tapes) with segmentation based upon machine imposed constraints (for record and block lengths) as well as the requirements of application programs (for content). The consideration encountered in discussing the most effective means for the segmentation of files are broadly encompassed by two particular questions: What is the optimum content of individual file segments in terms of routine processing requirements; and, in what sequences must these file segments be maintained in order to facilitate interfile matching and minimize pre-process sorting?

Another aspect of file design which was discussed at length was the extensive use of natural data rather than coded representations of that data. Realizing that serious problems would exist in sequencing and matching files on the basis of natural language content, not to mention the quality control problems inherent in free form natural language data, it was concluded that those elements contained in the files which would be required for sequencing and matching operations should be represented in codes as well as natural form. For example, while the natural language description of a particular property item (a 32" x 60" walnut desk) might be described in various ways (desk, wal., 32 x 60; walnut desk, 32 x 60; 32 x 60 desk, walnut; etc.), a code (1820) indicating that the property item was a desk would be a great help. And the inclusion of the dimensions in the property code (1820032060--least dimension first) would greatly facilitate the grouping of like elements in the property file--but, of course, denying selected retrieval of any particular desk.

Another aspect of file design which was considered was the problem of keeping track of the presence of various data elements within the many file segments which would be in use. It was recognized that certain file segments prepared for specific purposes would have a much greater utility if their contents were readily identifiable. For example, a file on newly admitted students would be a much more convenient base for the assignment of faculty advisors than would a file containing the records of all of the students attending the institution, most of whom would have established advisor assignments.

In summary, several significant working concepts evolved out of the discussions devoted to the topic of file design:

1. File identification should be structured according to the institution's basic data categories. Working files (or subfiles) should be similarly identified, but as subsets of parent files.

2. Redundancy must be minimized in order to economize file maintenance operations. Insofar as it is practicable, a given data element should appear in the data base only once.

3. Because the size of some of the files, particularly the student record, will exceed economical processing requirements, the segmentation of primary files will be necessary.

4. Data which is to be used for matching or sequencing operations should be represented in codes as well as natural form. In addition, when other organizations and agencies require the reporting of data according to their established codes, these codes should be considered a part of the data base.

5. The use of a data catalog as an index of all data being maintained in the system should be implemented. The specific locations of each data element should be indicated for all files and subfiles which are regularly maintained.

Data Acquisition.--Technical discussions held on the subject of data acquisition procedures covered three primary topics: logical point of entry; verification/validation, and system feedback; and hardware considerations.

General agreement was found to exist on the most effective point of entry for data into the system. The consensus was that the organization unit which first encountered and used the information should, in most cases, provide it to the system. Several sound reasons for this thinking seemed to exist:

1. Individuals accustomed to the data and whose operations are based upon it have a greater appreciation for quality control.

2. By entering data into the system at the earliest possible point, some manual handling may be eliminated and machine processing substituted.

3. By entering data into the system as early as possible, a wider distribution of information may be effected at an earlier time.

A similar consensus exists regarding verification and validation. The thinking is that individuals who are initially responsible for handling and working with data and who provide it to the system should be best qualified for the verification and validation procedures. Dissenting opinions or, more accurately, notes of caution did arise: the primary purpose for data being initially handled by a particular organization unit was due to its urgent need for that data, and the additional burden of providing that data to the system (and verifying and validating it) could lead to a substantial increase in workload which could result in a detrimental effect on an important operation. It was noted that substantial benefits could be obtained through the application of computer based verification/validation programs which may very quickly isolate such common data preparation errors as alphabetic information in numeric fields, numeric data in alphabetic fields, improper justification, missing data, and coded data which falls outside acceptable ranges. Significant limitations do exist.

Many organizations are now using automatic data preparation devices in order to prepare data for entry into the system without the need for keypunching. Great improvements in the performance of optical scanning equipment and reductions in its cost suggest the continued expansion of such methods. The two most common problems cited in discussions on the use of optical scanning equipment were: a volume (for each application) sufficiently high to justify the cost of form preparation and programing was necessary; and rigid control over the configuration of the input document (preferably through machine preparation of source documents) was required. It was also noted that the verification and validation of data prepared by optical scanning devices must be continued with the same emphasis as with data which is prepared by other means.

Another data acquisition technique which offers significant potential and warrants full consideration is the use of equipment which provides machine readable records as by-products of an existing activity. Paper tape perforators, and slave keypunches, attached to typewriters or other document creating machines offer distinct advantages: the net cost (resulting from additional effort) of data preparation is reduced; and the individual operating the device has a greater opportunity to scan the information being produced, thereby combining elements of the verification/validation processes with that of initial data preparation. Again, a reasonably high volume of activity is required in order to justify the equipment and the programing costs incurred. Also, in deciding which type of equipment to use, the difficulty of correcting errata contained in lengths of paper tape must be weighed against the limited capacity of the punched-card.

One of the most common methods of reducing the cost of data acquisition, verification, and validation is the use of prepunched-cards. Because of the popularity of such systems, little time was devoted to a full exploration of advantages and disadvantages--the prevailing philosophy being that when documents prepared by the system are designed to precipitate action which should be communicated back to the system, the use of punched-card documents should be fully explored.

One last point with respect to data acquisition, particularly directed at magnetic tape based computer installations but also worthy of consideration in computer installations utilizing random access equipment: Problems and inaccuracies in input are always with us and these may be relieved somewhat by the batching of input data until it has completely passed the verification/validation processes. In other words, the updating of files with new data should be withheld until the accuracy of the new data has been certified. Failure to observe this general principle may lead to the need for correction of data after it has been distributed to various working files in the system--perhaps after important reports have been printed and provided to users.

In summary, points to be remembered in the development of data acquisition procedures are:

1. The point of entry for data and verification/validation responsibility should be logically selected on the basis of the organizational need for and familiarity with the data.

2. Allowances may be required in order to provide for increased workload for verification and validation responsibilities assigned to various organization units.

3. Automatic or by-product data acquisition should be used whenever economic justification exists. The same care must be maintained in verifying and validating automatically prepared data.

4. Machine readable data (output) should be reused whenever possible.

5. Most imputs should be batched in order to optimize verification and validation prior to the updating of files.

<u>Storage and Retrieval</u>.--Discussions of this project frequently led to the use of the term "data bank," an analogy which, however accurate, was usually improperly drawn. While we tended to discuss the general objectives of this project in terms of a central reservoir of information which could be made available to administrators and educational researchers, such a description merely likens the project to the creation of a depository (such as Fort Knox). But banks are not mere depositories: they are systems which provide for the acquisition and distribution of (monetary) symbols of property and resources.

In order for a large data system to be properly likened to a bank, devices must exist which permit: The complete merging and mixing of data; the ready acceptance of new resources; the orderly maintenance and safeguarding of the resources which are entrusted to it; and the timely distribution of the resources when warranted.

The problem of merging and mixing has been discussed earlier, in terms of interfile linkages or file integration. The data system envisioned here, if it is to be effective over an extended period of time, must contain provisions for the creation of interfile linkages as they become necessary. For example, if initial planning made no provision for the inclusion of information on a student's academic major in the data base used for decisions on housing assignment and it later became evident that such information should be considered, it should be provided without requiring major changes in operating procedures. Or, if for some reason it became necessary to add elements of data to a particular segment of the data base (or even entire subsystems), such a necessity should be accommodated without the need for major changes in status-quo operations or for the creation of independent and non-integrated subsystems.

Most data processing specialists considering the problem of information storage and retrieval immediately become immersed in the technical considerations of available hardware and software. Storage capability is regarded in terms of millions of words or characters of data storage and milliseconds of access time. The act of storing and retrieving data is considered in terms of machine generated storage addresses which eliminate the need for massive housekeeping activities on the part of the programer. However, these relatively technical considerations deal with two basic assumptions: that the amount of information to be utilized by the system may be determined prior to the selection of storage devices; and that the information which the programer has in hand points the way rather directly to the storage locations of needed additional information stored elsewhere. In the instant case, the first assumption is an invalid one in that, at the point in time where implementation must begin, no all-encompassing specification of the data base exists, and, further, the economics of data processing preclude the acquisition of equipment with massive capacity on the basis of the possibility of rapid growth and implementation. The second assumption rests on the existence of effective data chains or interfile linkages (i.e., that the class schedule file contains those data elements necessary to correctly access the facilities file), some of which may be unknown at the time of implementation.

In summary, information retrieval planning should be conceptually divorced for machine considerations. Initial steps in design should be based on:

1. Recognition of the likelihood of the addition and/or change of interfile linkages.

2. Recognition of the fact that the initial data base (constructed to accommodate the requirements of the existing program library) constitutes an open-ended inventory which will grow, both by the addition of individual records and the addition of new data elements to existing records.

3. Initial retrieval programs should be designed to create the sub-files required by existing programs in order to minimize reprograming and conversion.

Applications Programing.--Discussions concerning the preparation of production programs focused on three general topics: the selection of programing languages; the use of general purpose programs and program generators; and the organization of programing personnel.

One does not need lengthy experience with automatic data processing in order to appreciate the magnitude of the problem of the conversion of programs which frequently occurs when equipment changes are being effected. (58) The high costs of reprograming are often accompanied by: the need for expense allowances which permit programing staff to journey to distant locations where the equipment being ordered is available, so that they may test and debug their programs prior to the installation of hardware; by the need for continued rental of obsolete equipment after the installation of new equipment, in order that production schedules may be met while reprograming is completed; and, more recently, by the purchase of additional hardware which permits the emulation of old equipment, at a greatly reduced efficiency. While there is no universal solution to such problems, the approach most frequently suggested to avoid them is the use of programing languages which, either through tradition or through direct financial support on the part of the users (federal government), offer the highest probability of survival. At this point in time, it appears to the authors that there are only two such languages: FORTRAN which has grown both in sophistication and in acceptance over the last decade; and COBOL, a language initiated and supported by the resources of the federal government. There are other languages currently available which offer promise of achieving the same stature: PL/1, initiated by IBM, is now offered by several equipment manufacturers and software houses; ALGOL has been, and probably will continue to be, offered by selected manufacturers; the BASIC language developed at Dartmouth College by General Electric Corporation also has a history which promises a future. However, until a language has achieved a level of sustained usage which assures its future, its use offers the prospect of massive reprograming efforts.

In recent years, general purpose programs or program generators have become increasingly popular. These software packages may be characterized by relatively simple instructions to the user which permit him to: Format information in accordance with his needs; very simply code arithmetic operations, including several levels of totals; and evaluate input data to the extent that extraneous material may be excluded from processing. (28) The more sophisticated of these programs, usually available only with larger computers, often contain features which are beyond the expertise of even experienced programers in that they tend to optimize themselves and operate at a high degree of machine efficiency. (19) The earliest of this type of program was primarily directed at sorting, and sort program generators have now arrived at a level of efficiency where only the most hardheaded of programers (or the most specialized of data processing installations) still insist on writing their own sorting programs. (To qualify the foregoing, we must point out that certain programer types obtain deep satisfaction from writing sorting programs which accomplish the desired result in less time than that which is afforded by the manufacturer's supplied program; and certain specialty data processing operations, wherein the sequencing of data constitutes the bulk of their production effort, cannot afford the luxury of the waste of as much as twenty-five percent of their available machine time through the use of generalized sorting programs.) We believe that the growth and development of general purpose programs and program generators will parallel that which we have witnessed in sorting programs, recognizing that limitations on their utility will remain.

An entirely different set of considerations apply to the organization of the institution's programing staff: it is vital that the programers concerned with the preparation of reports concerning student affairs, curriculum, facilities, fiscal affairs, library activities, and personnel administration speak the language of individuals having responsibility for these activities; individuals responsible for administering various activities and operations within the institution should have "someone to talk to" regarding their day-to-day problems, so that opportunities for the creative participation of the data system in resolving some of these problems is afforded; and, notwithstanding the fact that the foregoing would suggest that the most effective method of assignment of programing staff would be to place programing personnel within the smallest organizational unit of the institution, programing personnel should be placed sufficiently close to the equipment upon which they rely to stay current in terms of existing and planned operating systems, data base modifications, hardware and software characteristics, and shop procedures. The administrator responsible for the assignment of programing personnel must, therefore, take into consideration not only the area of interest to the programer (student affairs, business management, or academic program), but also the loyalties and levels of sophistication of the individuals: a highly trained machine-oriented programer who is personally acquainted with many shop personnel might well be farmed out to work for a particular unit of the organization, where he will strive to maintain currency with the data processing installation; a programer with a strong accounting background might more effectively be placed within the central offices of the computer installation, as he will most probably retain his contacts with the accounting unit.

In summary, policies relating to the preparation of production programs (applications programing) should be based upon the consideration of the following:

1. Programing languages such as COBOL and FORTRAN should be used to the fullest extent possible.

2. Programing, as such, should be avoided when the requirements of a particular job permit the use of a report-program-generator or other software package to obtain the desired results.

3. The assignment of programing staff should be based upon individual qualifications and situations. Field of specialization, as well as relationship (or proximity) to the user must be considered.

RESULTS

The results of this project are the design parameters of a comprehensive data system for an institution of higher education. The system, consisting of men, machines, programs, and procedures, is user-operator oriented. Although certain electronic and electromechanical devices offer decided operating advantages to an organization implementing such a system, the basic design is not machine dependent.

The foundation of this system is the institution's data base. The design presented here encompasses the structuring of the data base, the materials which should be made available to users and operators in order to facilitate its interpretation and access, and general operating concepts for the storage, retrieval, and maintenance of the data.

More specifically, the results of this project define a mechanism which provides maximum service to the user (responsiveness) and to the operator (flexibility). Included here is a description of the tools which specify the form and content of the data base in terms which are meaningful to users and operators, a method for expanding the data base without requiring major modification of existing operations (storage), a method for the orderly updating of all segments of the data base (file maintenance), and a method for accessing required data (retrieval).

The portion of this report which presents the means by which data is accessed provides for gradual implementation without affecting existing operations and without major equipment expenditures. The object is to establish a systems approach to data management which permits a gradual move to more highly sophisticated information storage and retrieval systems, without changing the initial conceptual framework.

Hardware Requirements.--As has been suggested earlier, this system design is intended to be feasible for implementation on a relatively small scale, magnetic tape oriented computer. Tremendous advantages are, of course, available when a large scale, random access system is employed. The philosophy underlying the decision to orient the design toward a small scale, magnetic tape system is in part due to the current dynamic developmental and economic conditions of the data processing industry, and the fact that much work must be done in developing and refining the data base and its interfile linkage (not to mention related clerical and data acquisition procedures) prior to the point where the magnitude and level of utilization of the data base may justify the use of a large scale computer. (44)

The primary advantages in the use of a large scale, random access computer system are listed below with comments on the adjustments and procedural modifications which are necessary for tape system implementation:

	RANDOM ACCESS	TAPE
1.	Direct access to the data elements is made possible by a combination of hardware and software. (25)	By storing a maximum amount of data in each tape file, the accessing of the data base requires fewer file handling operations.
2.	The direct access capability reduces the amount of sorting required for file maintenance.	By using the shortest possible record length per file, sorting time may be reduced.
3.	File sorting may be accomplished at extremely high speed.	Short records in tape files reduce the time required for sorting.
4.	File maintenance requires the accessing of only one address in the data base per item of change. (25)	By maximizing the data contained in each record in tape files, the number of tapes to be updated is reduced.
5.	Random access software permits the automatic	Programer specified addressing is mini-

addressing of the data base in file maintenance activities. (25)	mized by the use of fewer tape files containing maximum data per record.
6. Meaningful reports may be produced without the need for interim processing. (25)	By minimizing the data per record, the time required for interim (sort) processing is reduced.
7. Because only data to be processed needs to be brought into core, the amount of core required may be minimized.	By minimizing the data contained in each record, a reduction in core requirements is obtained.
8. Back up or duplicate records may be created simultaneously and they are automatically accessible in the event of read failure caused by deterioration of magnetic surfaces. (16)	Back up records may be copied and stored on tape at a minimum cost.

The paradox concerning the length of the record to be employed in magnetic tape systems is discussed later in this report. While careful planning of file organization should reduce both the length of the files which require sorting and the number of files to be updated, even the most sophisticated tape operating procedures fall far short of equaling the advantages of a random access system.

Organization of Data Base

The term "data base " encompasses all of those data elements contained in the records of individual transactions (or other units or cases). As discussed earlier (page 12), information system inputs and outputs may be viewed as a ladder of abstraction, but the key to providing higher levels of management information lies in the existence of the data base <u>and</u> in its organization.

The formal organization or hierarchy of the data base is schematically presented in Figure 3. However, unseen in this diagram are the linkages between the files which must exist in order to permit one file to be accessed (operated on) from another file. The hierarchical organization of the data base is, therefore, only one dimension of the problem of data base organization, and it may be described as follows:

1. The data element is the smallest meaningful unit of information maintained by the system. Student name, identification number, street address, city, state, and zip code of residence are examples of data elements.

2. The logical record is a grouping of related data elements. The above data, grouped by individual student, would be a record.

3. The file contains all of the logical records of a specified type. The file in the above case would be the student address file.

4. The data base is the aggregate of all of the files being maintained.

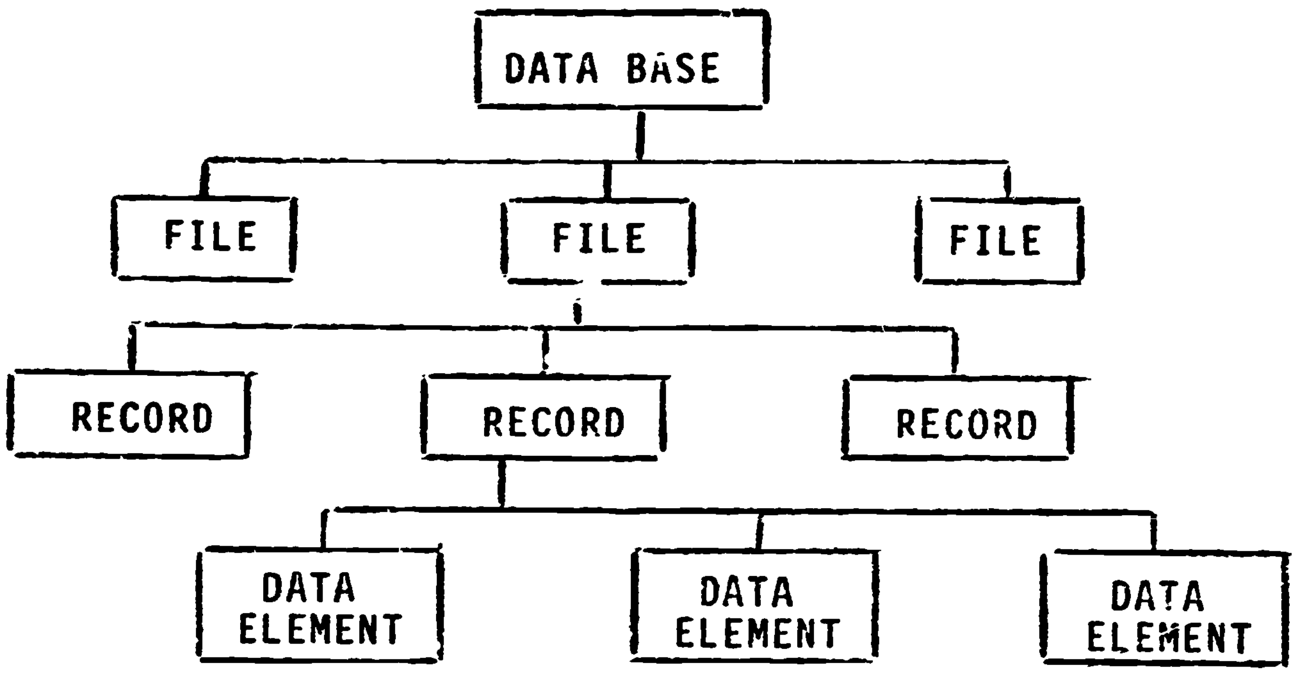

Figure 3. Hierarchy of the Data Base

One may also envision the entire institutional data base as a universe, with each of the major files (student, facilities, faculty/staff, fiscal, curriculum, and library) constituting a set of data.* Each record within each file may then be considered a member of its set and the data elements of which the record is made could be said to define the properties or characteristics of each member. Figure 4 illustrates the presence of Files A and B (as sets) within the universe of the institution's data base.

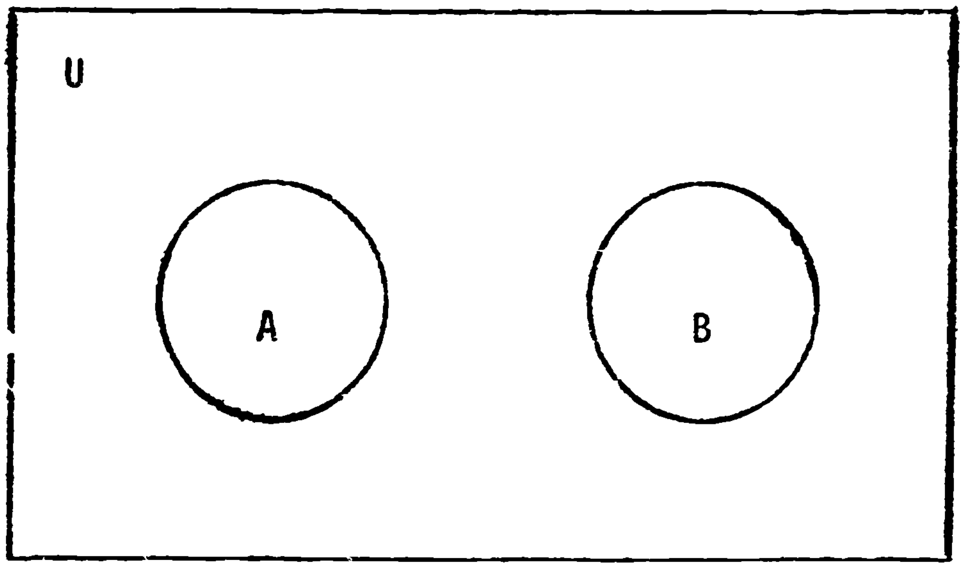

Figure 4. Files A and B as Parts of the Institution's Universe of Data Base

As indicated earlier, the amount of data to be assembled within some of the files is so voluminous, both in numbers of members and in numbers of characteristics, some segmentation will be necessary. Two types of segmentation may occur: division of the file according to certain characteristics (such as student class level); or division of the characteristics relating to each member according to some logical plan (such as student admissions data). Because processing usually involves the interpretation, manipulation,

*Library is set apart from other property inventory files because of the multiplicity of activities which are based upon it.

and summarization of selected characteristics relating to all members of a set, the latter form of segmentation will most frequently be used. Figure 5 illustrates the segmentation of File A into four subfiles, or subsets. Obviously, in order to be able to assemble all of the characteristics of any member of File A, its unique identifier (such as student ID number) must be present in each of the subsets, affixed to his other characteristics.

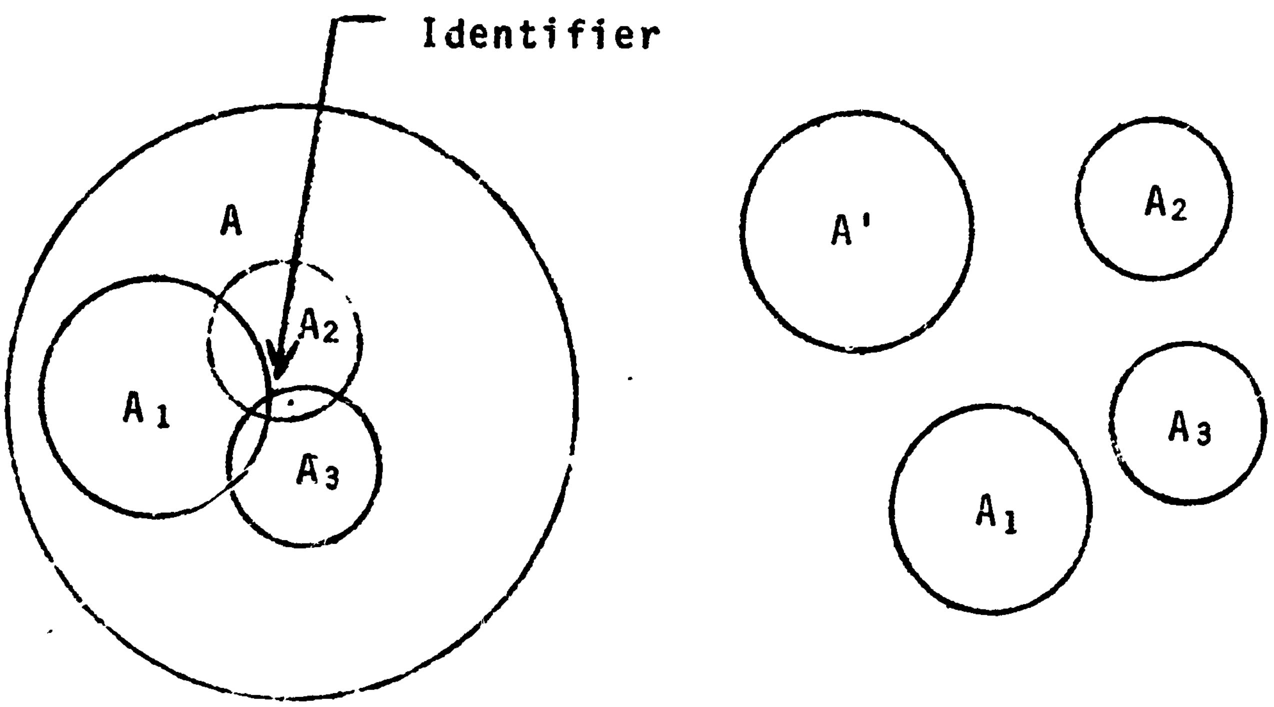

Figure 5. File A Segmented Into Four Subsets

In the early pages of this report, we discussed a traditional problem in the assembly of meaningful management information: The lack of integration of an organization's data base. Figure 4 shows Files A and B existing within the universe of the data base--but, in an unrelated fashion. In order to effectively relate the information contained in File A to that in File B, linking data must exist which permits the members of one set to be related directly to members of the other. The most desirable type of linking data is found when certain characteristics of the members of one file are identifiers of the members of another file (as would be the case when the identification numbers of the instructors of currently enrolled students were retained in the Student Record File, permitting direct access to the Faculty/Staff File). Figure 6 illustrates the linking of Files A and B through data common to both. (In Set Theory, the common data would be termed the "union of A and B," and the unification of the two files would be implied.)

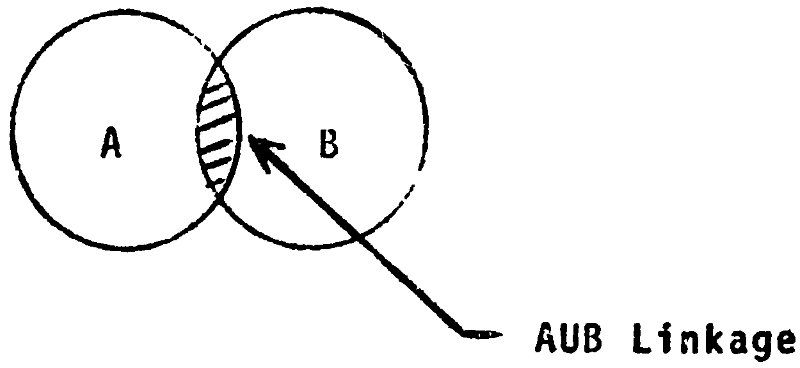

Figure 6. Files A and B Linked by Common Data

<u>Key Files</u>.--Earlier in this report, the semantic problems encountered in discussing various types of data files were mentioned. "Master" files are generally regarded as being files which contain relatively static data, and "working" files are usually designed so as to optimize processing speeds while providing the data base for specific groups of applications programs. Because our operational concepts did not readily lend themselves to the division of the data base into these categories, we determined that the total institutional data base may most effectively be divided into six "key" files according to the nature of the data. We view the characteristics of Key Files to be as follows:

1. The Key Files of the institution possess the entirety of the institutional data base.

2. The Key Files are the points of entry for all new and updating data.

3. The Key Files are never sorted, but reside in predetermined sequences.

4. The specifications (record lengths and formats) of Key Files change from time to time as new data is added.

5. Key Files are processed only by housekeeping programs.

6. Individual Key Files possess all of the linkage data which is necessary for the accessing of related files.

Of particular importance in the above is the fact that the sequence of the Key File must be fixed. Because unusually long record lengths are anticipated and implementation on a small scale, tape oriented computer is being planned, the time required to sort Key Files would be prohibitive. Therefore, it is necessary to limit file sorting to subsets of data (or parts of individual records) retrieved from Key Files (Figure 5.).

Data redundancy in Key Files is limited to that portion of the data base which constitutes the linkages between Key Files. If such linkages are not anticipated and identified initially, they may be added later by introducing data into one Key File which permits accessing another.

As indicated earlier, some of the Key Files will require segmentation. (We currently anticipate a student record in excess of 5,000 characters in length.) The selection of data elements to be grouped into segments of Key Files may be based upon one or both of the following criteria: the logical relationship of the data elements based upon data acquisition and retrieval (or working file) requirements; and the retention schedule established for individual records in the data base.

An example of the first case would be the use of an "admissions" subset of the student record file, based upon both acquisition and processing requirements. An example of the second case would be the use of an "inactive" subset of the student record file, for all students not currently enrolled.

Working Files.--That part of the data base which is processed to provide documents, reports, and analyses, and which is created by the retrieval of specific information from the Key Files is termed the Working File. Ideally, Working Files are designed so as to individually accommodate groups of processing programs. However, their lengths are minimized in order to facilitate file sorting operations.

The assembly of Working Files may require selected retrieval from a single Key File segment, simultaneous selected retrieval from two segments of a Key File, or it may require phased retrieval from two or more linked files. Although the preparation and maintenance of Working Files is covered in more detail later in this report, one further point should be noted here: Data redundancy frequently occurs between Working Files, due to the fact that they are designed, created, and maintained in order to accommodate specific production programs--and existing Working Files often do not possess the data required for newly defined programs. The result is the need to maintain many Working Files, substituting, deleting, and changing data through special purpose programs. A reduction in the number of Working Files, and the related updating requirements, may be achieved by originally creating them with sufficient filler (excess capacity, or space containing no data) within their records to allow for twenty to twenty-five per cent growth.

File Linkage

The following discussion of interfile linkage is limited to the consideration of the Key Files of the system. The Key Files envisioned here are grouped into categories according to the data being retained therein:

1. The Student File (SR) contains all of the data defining student biography, characteristics, academic history, and academic program objectives.

2. The Space-Facilities File (SF) contains all of the data relating to the physical

3. The Personnel File (FS) contains all of the data required for personnel administration, and past and present work assignments.

4. The Fiscal File (BF) contains all of that data required for day-to-day business operations such as purchasing, warehousing, and accounting, as well as conventional General Ledger records and budgetary forecasting data.

5. The Curricular File (CM) contains that part of the data base which defines the contents of the course catalog of the institution, the faculty and facilities resources required for each course, and academic programs which represent current and future enrollment demand.

6. The Library File (LB) contains an inventory of the holdings of the institution, as well as books and periodicals on order and "in process."

Three types of interfile linkages exist in this system: direct, indirect, and multiple. In all cases, linkages are data, apropos to one or more of the Key Files, which may also be readily maintained in Key Files where the existence of such data would not normally be expected to be found, as illustrated in Figure 7.

Direct linkage occurs where data is found in two or more Key Files. For example, the employee ID number basic to the Payroll (BF) File is identical with that maintained in the Personnel (FS) File.

Indirect linkage occurs when an intermediary Key File must be accessed. An example would be the problem of determining the direct cost of instruction, where current student enrollment data must be related to payroll through the FS File, where data defining the part of a professor's salary chargeable to teaching is maintained.

Multiple linkage occurs when identical types of data must be used for different logical purposes. For

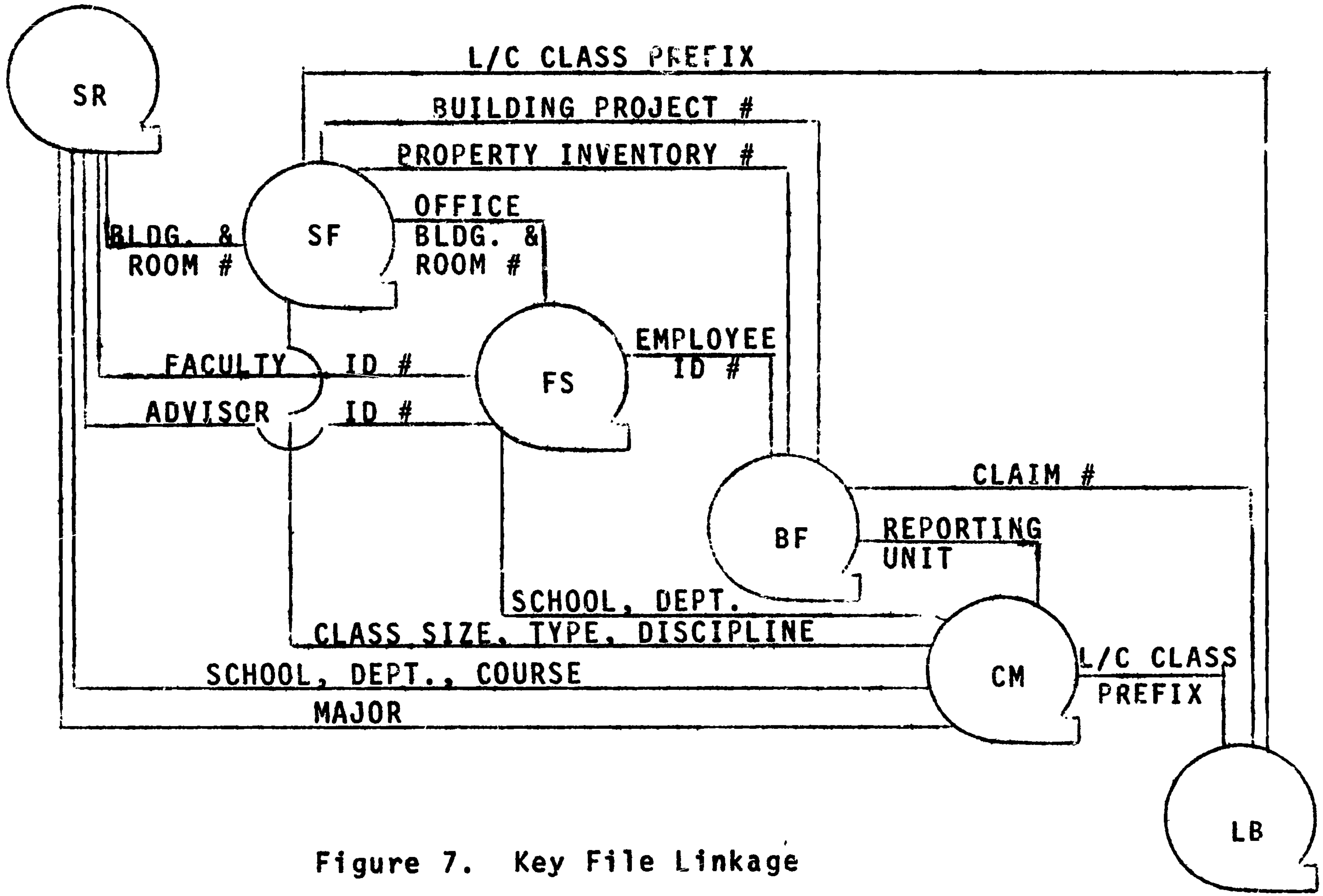

Figure 7. Key File Linkage

example, the maintenance of Faculty ID number in the SR File may be concerned with the identification of the student's academic advisor or the identification of his instructors.

Because the linkages which exist between Key Files frequently reside in different sequences, the accessing of a file will often require the information retrieved from another to be sorted into sequence of the file to be accessed. An example of this would be the data defining the building and room number of the classes in which the students are currently enrolled. If the Student File is sequenced according to student ID number and it was necessary to relate enrollment data to facilities (for a utilization study), the information retrieved from the Student File must be resequenced according to building and room number, in order to access the Facilities File (which would be maintained in building room sequence).

The following interfile linkages are planned for initial implementation in this system.

SR:SF.--The linkage between the Student Record and the Space Facilities files consists of the building and room number of each class in the programs of the students.

SR:FS.--Multiple linkages exist between the Student Record and Personnel files:

1. The identification number of the instructor in each of the student's classes;

2. The identification number of the student's academic advisor.

SR:BF.--The linkage between the Student Record and Fiscal files is indirect. The data defined in the student academic programs (school, department, and course) is used to access the Curricular File where the account code for each organizational unit is retained. This account code (Reporting Unit) then provides the natural data by which the Fiscal File may be accessed.

SR:CM.--Multiple linkages exist between the

Student Record and Curricular Files:

1. The school, department, and course identifications of current student programs;

2. The major objectives of students which may be used to access lists of course in the Curricular File.

SR:LB.--The linkage between the Student Record and Library files is indirect and is probably useful only in a gross quantitative sense. The linkage is established by relating the Student Record to the Curricular File (either in terms of school, department, and course, or in terms of the discipline of his major) in order to access the Library of Congress Classification prefix, which identifies the subject matter of the library's holdings.

SF:FS.--Both direct and indirect linkages exist between the Facilities and Personnel files:

1. The building and room number designation of the office or other work location of the employee;

2. The building and room number defining the instructor's current teaching assignments, which are obtained from student program data residing in the SR File.

SF:BF.--The linkage between the Facilities and Fiscal files (essentially, the institution's General Ledger) is multiple:

1. Building (project) number;

2. Property inventory number.

SF:CM.--The linkage between the Facilities and Curricular files consists of a specification defining optimum class size, type of instruction, and subject matter (or discipline). This specification is the primary tool for the determination of plant capacity and it is a key element in the planning of the institution's class schedule.

SF:LB.--The linkage between the Facilities and Library files consists of the Library of Congress Classification prefix which identifies the space required to organize and house the library's holdings.

FS:BF.--The linkage between the Personnel and Fiscal files is the employee ID number.

FS:CM.--The linkage between the Personnel and Curricular files consists of the school and departmental assignment of the employee.

FS:LB.--The linkage between the Personnel and Library files is indirect. It is obtained by accessing the Curricular File with school and departmental assignment reflected in the Personnel File, in order to obtain the Library of Congress Classification prefix link.

BF:CM.--The linkage between the Fiscal and Curricular files is the reporting unit designation contained in the institution's chart of accounts. (34)

BF:LB.--The linkage between the Fiscal and Library files consists of the order or claim number through which a purchase is executed.

CM:LB.--The linkage between the Curricular and Library files is the Library of Congress Classification prefix which identifies the subject matter of the institution's holdings in terms of the disciplines represented in the college catalog.

User-Operator Tools

In order to provide the administrator and the educational researcher with a convenient tool for determining the availability of data of interest to them and simultaneously provide a vehicle for their communicating specific information requirements to data processing personnel for their action, a set of data base reference materials must be used.

These reference materials are: three different formats of a data inventory; and an assembly of all codes used in the representation of data. The data inventories and code references are briefly discussed in the following:

1. A <u>data catalog</u>, designed for use by administrators and other system users, presents an alphabetic listing of all of the data contained in the system. Where data is retained in coded form, the location of code definition references is indicated. The Data Catalog also presents a unique item number for each data element inventoried. (Figure 8.)

2. An <u>index</u>, designed primarily for use by systems analysts and programers, presents a sequential listing of all elements contained in the data base. The data element item number is the basis for the sequence of the Index. The Index shows all of the files and subfiles in which each data element is regularly maintained. (Figure 9.)

3. <u>A file specification</u>, to be used in the preparation of computer programs which will operate on each file, is also printed. The File Specification includes the data inventory item number, the date that the data was first introduced into the system, the beginning and ending positions of the data element within the file record, the mode of the data, and the data names or labels to be used in programing. (Figure 10.) Also included for each File Specification is a title sheet which defines other characteristics of the file needed in programing.

   The title sheet includes the standard tape label, record and block lengths, the sequence or sequences in which the file is maintained, and the routine update cycle of the file. (Figure 11.)

4. In cases where the detailed nature of the information being maintained by the system cannot readily be determined from the Catalog itself, further references are

D A T A   C A T A L O G   PAGE 1

A L P H A B E T I C A L   L I S T I N G

DESCRIPTION	ITEM NUMBER	CODE REF
ACCOUNT REFERENCE NUMBER	[illegible]	[illegible]
ACT TEST SCORE COMPOSITE	0317	0317
ACT TEST SCORE ENGLISH	0318	0317
ACT TEST SCORE MATHEMATICS	0319	0317
ACT TEST SCORE NATURAL SCIENCES READING	0320	0317
ACT TEST SCORE SOCIAL SCIENCES READING	0321	0317
ADDRESS EMPLOYEE CITY	1109	
ADDRESS EMPLOYEE STREET	1108	
ADDRESS EMPLOYEE STATE	1111	
ADDRESS EMPLOYEE ZIP	1110	
ADDRESS STUDENT CITY (MAILING)	0112	
ADDRESS STUDENT CITY (PERMANENT)	0217	
ADDRESS STUDENT STREET (MAILING)	0104	
ADDRESS STUDENT STREET (PERMANENT)	0216	
ADDRESS STUDENT STATE (MAILING)	0106	
ADDRESS STUDENT STATE (PERMANENT)	0218	
ADDRESS STUDENT ZIP (MAILING)	0107	
ADDRESS STUDENT ZIP (PERMANENT)	0219	
ADEQUACY ROOM PHYSICAL INVENTORY	0929	0929
ADMISSION BASIS FOR	0247	0247
ADMISSION CLASS LEVEL AT	0248	
ADMISSION DATE OF	0251	
ADMISSION DATE OF APPLICATION FOR	0241	
ADMISSION NUMBER APPLICATION FOR	0242	
ALLOTMENT YEAR-TO-DATE	0803	
ALTERNATE FIRST CLASS SIZE/TYPE	0027	0024
ALTERNATE SECOND CLASS SIZE/TYPE	0028	0024
ALTERNATE THIRD CLASS SIZE/TYPE	0029	0024
ANNIVERSARY SALARY ... MERIT	1105	1105
YEAR PUBLICATION	[illegible]	
YEAR-TO-DATE ALLOTMENT	0803	
YEAR-TO-DATE CUMULATIVE EXPENDITURE	0884	
YEAR-TO-DATE ENCUMBERANCE	0804	
YEAR-TO-DATE EXPENDITURE	0805	
YEAR-TO-DATE UNENCUMBERED BALANCE	0886	
YEAR-TO-DATE UNLIQUIDATED ENCUMBERANCE	0885	
ZIP EMPLOYEE ADDRESS	1110	
ZIP STUDENT ADDRESS (MAILING)	0107	
ZIP STUDENT ADDRESS (PERMANENT)	0219	

Figure 8. Data Catalog

DATA INDEX

ITEM NUMBER LISTING

ITEM NUMBER	DESCRIPTION	FILE(S)
0001	SCHOOL CODE	CM0101, CM0106, FS0101, SR0101
0002	DEPARTMENT NUMBER	CM0101, CM0106, FS0104, SR0101, SF0103
0003	SUB-DEPARTMENT NUMBER	CM0101, CM0106, FS0104, SR0101, SF0103
0004	COURSE NUMBER	CM0101, CM0104, SR0101, SR0105, SR0107, SR0109, FS0104, FS0108
0005	OWN DISCIPLINE CODE	CM0101, CM0103, FS0103, BF0102
0006	CSC DISCIPLINE CODE	CM0101, CM0102
0007	COURSE TITLE	CM0101, CM0102, SR0101, SR0105, SR0106, FS0104
0008	FIRST STAFFING CODE	CM0101, CM0106, SF0101, SF0103, SF0107
0009	FIRST UNITS	CM0101, CM0103, CM0106
0766	WORK IN PROG, COURSE NUMBER 4	SR0101, SR0102, SR0105, SF0101
0767	WORK IN PROG, SECTION NUMBER 4	SR0101, SR0102, SR0105, SF0101
0768	WORK IN PROG, SCHOOL 5	SR0101, SR0102, SR0105, SF0101
0769	WORK IN PROG, DEPARTMENT NUMBER 5	SR0101, SR0102, SR0105, SF0101
0770	WORK IN PROG, SUB-DEPARTMENT NUMBER 5	SR0101, SR0102, SR0105, SF0101
0771	WORK IN PROG, COURSE NUMBER 5	SR0101, SR0102, SR0105, SF0101
0772	WORK IN PROG, SECTION NUMBER 5	SR0101, SR0102, SR0105, SF0101
0773	WORK IN PROG, SCHOOL 6	SR0101, SR0102, SR0105, SF0101
0774	WORK IN PROG, DEPARTMENT NUMBER 6	SR0101, SR0102, SR0105, SF0101
0775	WORK IN PROG, SUB-DEPARTMENT NUMBER 6	SR0101, SR0102, SR0105, SF0101
0776	WORK IN PROG, COURSE NUMBER 6	SR0101, SR0102, SR0105, SF0101
0777	WORK IN PROG, SECTION NUMBER 6	SR0101, SR0102, SR0105, SF0101
0801	ACCOUNT NUMBER	BF0101, CM0101, CM0206, CM0209
0802	ACCOUNT NAME	BF0101, BF0107, BF0108, BF0111
0803	ALLOTMENT YEAR-TO-DATE	BF0101, BF0107, BF0108, BF0111
0804	ENCUMBRANCE YEAR-TO-DATE	BF0101, BF0107, BF0108, BF0111
0805	EXPENDITURE YEAR-TO-DATE	BF0101, BF0107, BF0108, BF0111
0806	ACCOUNT REFERENCE NUMBER	BF0101, BF0107
0807	CLAIM SCHEDULE NUMBER (FIRST)	BF0101, LB0101, LB0107, BF0102
0808	CLAIM SCHEDULE NUMBER (SECOND)	BF0101, BF0102, BF0103
1001	ORDER NUMBER, LIBRARY	LB0101, LB0102, LB0104, BF0107
1002	VOLUME NUMBER	LB0101, LB0103, LB0105
1003	COPY NUMBER	LB0101, LB0103, LB0104, LB0105
1004	MAIN ENTRY	LB0101, LB0106, LB0107, LB0109
1005	TITLE	LB0101, LB0106, LB0107, LB0109
1006	AUTHOR(S)	LB0101, LB0106, LB0107, LB0109
1007	PUBLISHER	LB0101, LB0106, LB0107
1008	YEAR PUBLICATION	LB0101, LB0106, LB0107

Figure 9. Data Index

FILE SR0101 **PAGE 1**

ITEM NUMBER	DATE	FILE SEQ	MODE	FIELDS BEG-END	COBOL LABEL	FORTRAN LABEL
0101	18267	0001	IBCD	00010036	STUD-NAME	STNAME
0102	18267	0002	IBIN	00370045	STUD-ID-NUMBER	STIDNO
0103	18267	0003	IBCD	00460046	STUD-SEX	SSEX
0104	18267	0004	IBCD	00470074	STUD-ST ADDRESS	STSTAD
0105	18267	0005	IBCD	00750094	STUD-CITY	STCITY
0106	18267	0006	IBCD	00950099	STUD-STATE	STUDST
0107	18267	0007	IBIN	01000104	STUD-ZIP	STDZIP
0108	18267	0008	IBCD	01050106	DEP-MAJOR	DEPMAJ
0109	18267	0009	IBCD	01070108	INT-DEP-MAJ	[illegible]
0001	18267	0805	IBIN	[illegible]	SCHOOL	SCHOOL
0002	18267	0806	IBIN	36403641	DEPARTMENT-NUMBER	DEPNUM
0003	18267	0807	IBIN	36423643	SUB-DEP-NUMBER	SDEPNO
0004	18267	0808	IBCD	36443647	CRS-NO	CRSNO
0041	18267	0809	IBIN	36483648	SECTION-NUMBER	SECTON
0753	18267	0810	IBIN	36493649	WIP-SCHOOL-2	WPSCH2
0754	18267	0811	IBIN	36503651	WIP-DEPT-NUMBER-2	WPDPN2
0755	18267	0812	IBIN	36523653	WIP-SUB-DEPT-NUMBER-2	WPSDN2
0756	18267	0813	IBCD	36543657	WIP-COURSE-NUMBER-2	WPCNO2
0757	18267	0814	IBIN	36583658	WIP-SECTION-2	WPSCT2
0758	18267	0815	IBIN	36593659	WIP-SCHOOL-3	WPSCH3
0759	18267	0816	IBIN	36603661	WIP-DEPT-NUMBER-3	WPDPN3
0760	18267	0817	IBIN	36623663	WIP-SUB-DEPT-NUMBER-3	WPSDN3
0761	18267	0818	IBCD	36643667	WIP-COURSE-NUMBER-3	WPCNO3
0762	18267	0819	IBIN	36683668	WIP-SECTION-3	WPSCT3
0763	18267	0820	IBIN	36693669	WIP-SCHOOL-4	WPSCH4
0764	18267	0821	IBIN	36703671	WIP-DEPT-NUMBER-4	WPDPN4
0765	18267	0822	IBIN	36723673	WIP-SUB-DEPT-NUMBER-4	WPSDN4
0766	18267	0823	IBCD	36743677	WIP-COURSE-NUMBER-4	WPCNO4
0767	18267	0824	IBIN	36783678	WIP-SECTION-4	WPSCT4
0768	18267	0825	IBIN	36793679	WIP-SCHOOL-5	WPSCH5
0769	18267	0826	IBIN	36803681	WIP-DEPT-NUMBER-5	WPDPN5
0770	18267	0827	IBIN	36823683	WIP-SUB-DEPT-NUMBER-5	WPSDN5
0771	18267	0828	IBCD	36843687	WIP-COURSE-NUMBER-5	WPCNO5
0772	18267	0829	IBIN	36883688	WIP-SECTION-5	WPSCT5
0773	18267	0830	IBIN	36893689	WIP-SCHOOL-6	WPSCH6
0774	18267	0831	IBIN	36903691	WIP-DEPT-NUMBER-6	WPDPN6
0775	18267	0832	IBIN	36923693	WIP-SUB-DEPT-NUMBER-6	WPSDN6
0776	18267	0833	IBCD	36943697	WIP-COURSE-NUMBER-6	WPCNO6
0777	18267	0834	IBIN	36983698	WIP-SECTION-6	WPSCT6

Figure 10. File Specification

FILE SR0105 WORK-IN-PROGRESS

```
SEQUENCE(S):

01 = 0102
02 = 1101, [0001, 0002, 0003, 0004, 0041](1)
      ⋮
     1101, [aaaa, bbbb, cccc, dddd, eeee](N)
03 = [0001, 0002, 0003, 0004, 0041](1)
      ⋮
     [aaaa, bbbb, cccc, dddd, eeee](N)

01 = STUDENT ID NUMBER
02 = FACULTY ID NUMBER,
      [SCHOOL, DEPT, SUBDEPT, COURSE, CLASS SECTION](1)
     FACULTY ID NUMBER,
      [SCHOOL, DEPT, SUBDEPT, COURSE, CLASS SECTION](2)
     FACULTY ID NUMBER,                                  ⋮
      [SCHOOL, DEPT, SUBDEPT, COURSE, CLASS SECTION](N)
03 = [SCHOOL, DEPT, SUBDEPT, COURSE, CLASS SECTION](1)
                                                   ⋮
     [SCHOOL, DEPT, SUBDEPT, COURSE, CLASS SECTION](N)

RECORD LENGTH:  360 CHARACTERS

BLOCK LENGTH:  3601 CHARACTERS

UPDATE CYCLE:  01 = CLOSE OF REGISTRATION
               02 = 5 DAYS AFTER CLOSE OF REGISTRATION
               03 = 10 DAYS AFTER CLOSE OF REGISTRATION
               04 = 15 DAYS AFTER CLOSE OF REGISTRATION
               05 = 20 DAYS AFTER CLOSE OF REGISTRATION
               06 = LAST DAY OF CLASSES

TAPE LABEL:  WIP(TERM)[1](SEQ)[2](UPDATE)[3]

             1 -- SEE REF CODES, ITEM 0091
             2 -- PER ABOVE
             3 -- PER ABOVE
```

Figure 11. File Specification Title Sheet

provided in a manual of reference codes. (Figure 12.)

The three data inventories (Catalog, Index, and File Specification) are produced from a single data inventory record. This record is prepared for each element in the data base for each file in which it is regularly maintained. The data element inventory record conforms to the following format:

Location	Description
001-004	ITEM NUMBER, a unique number assigned to each data element inventoried.
005-009	DAY AND YEAR ENTERED, reflecting the point in time when this element was first introduced into the system. This information will be used to quickly determine the existence of historical information in out-of-date files.
010-013	FIELD NUMBER, reflecting the location (from left to right) of the data element within its record. This information will be used for sequencing in preparation for the printing of the File Specification.
018-025	BEGINNING/ENDING POSITIONS, the number of characters to the right of the beginning of the record where the beginning and ending locations of the data element may be found.
026-048	STANDARD (COBOL) LABEL, to be used in COBOL programing.
049-054	STANDARD (FORTRAN) LABEL, to be used in FORTRAN programing.
055-058	CODE REFERENCE NUMBER, to be used when data is represented in coded rather than natural form. This number may or may not be the same as the inventory item number, depending upon whether or not the same table of codes is used for more than one type of data. For example, when minimum, optimum, and maximum class sizes are to be shown, for

REFERENCE CODES

ITEM 0024: CLASS SIZE/TYPE

THE FOLLOWING CODES INDICATE MINIMUM, PREFERRED, AND MAXIMUM CLASS SIZES FOR LECTURE-DEMONSTRATION AND LABORATORY-EXPERIENCE CLASSES.

GROUP I		MINIMUM	PREFERRED	MAXIMUM
(LECTURE, QUIZ, DISCUSSION, SEMINAR, DEMONSTRATION)	A	5	8	12
	B	8	12	18
	C	15	20	25
	D	18	25	30
	E	20	30	40
	F	30	40	60
	G	40	60	120
	H	40	120+	(Specify)

GROUP II		MINIMUM	PREFERRED	MAXIMUM
(LABORATORY, EXPERIENCE)	A	1	4	4
	B	5	8	8
	C	8	12	12
	D	10	16	16
	E	16	24	24
	F	18	24	30
	G	25	30	36
	H	25	40	60

Figure 12. Reference Codes

both preferred and alternate plans of instruction, the same class size codes may be used.

059-108 DESCRIPTION, an English language description of the data element.

109-114 FILE, showing the file in which the data element is maintained.

The procedure by which the three inventories are produced from this data inventory are described in the following.

Data Catalog.--The Data Catalog, which reflects the data element description, item number, and code references (if any), uses the information contained in positions 001-004, 055-058, and 059-108, respectively. Because the Data Catalog will contain many descriptors which possess more than one significant term, it is necessary to provide a method by which a user may locate a data element through the use of any of its significant terms.*

Key words will be used in sorting and printing the alphabetical Data Catalog. We propose to provide for as many as three key words for cross referencing any data element.

The steps to be taken in the preparation of the Catalog, using the data inventory, are briefly described in the following:

1. The data inventory (in item number sequence on magnetic tape) will be read by the computer and the three fields of data (description, item number, and code references) will be selected for transfer to the output area. Only one output record will be written for each data element item number.

---

*The (KWIC) Keyword-in-Context Indexing System developed by the Service Bureau Corporation offers a technique which, although far exceeding requirements of our system, points out the feasibility of a machine operated cataloging device. (47)

(Remember that a data element which is being maintained in six files would appear in the inventory six times.)

2. Left-hand justify the data description key words. (Our plans for the designation of key words in the inventory record are based upon the use of special symbol, an asterisk, to precede each key word in the description, except for the first, which will be left-hand justified in the description.) An example of this justification procedure follows, using the description of the data element which reflects an employee's merit increase anniversary date:

SALARYΔ*ANNIVERSARYΔDATEΔ*MERITΔINCREASEΔΔΔ...Δ
(Delta [Δ] indicates blank, or space. Initial data description entry above shows first key word in left most position [salary] with subsequent key words preceded by asterisks.)

SALARYΔΔANNIVERSARYΔDATEΔΔMERITΔINCREASEΔΔΔ...Δ
(Initial entry edited for printing. Note suppression of asterisks.)

ANNIVERSARYΔDATEΔΔMERITΔINCREASEΔΔSALARYΔΔΔ...Δ
(Data description after first left-justify movement.)

MERITΔINCREASEΔΔSALARYΔΔANNIVERSARYΔDATEΔΔΔ...Δ
(Data description after second left-justify movement.)

One description record will be outputted onto magnetic tape for each key word in the description.

3. The Data Catalog file, created per the above, will then be sorted on the left-most twenty-four characters of the data description.

4. The sequenced file will then be printed.

We propose to provide all significant users of the data system with copies of the Catalog. Casual users will have access to the Catalog in both the College Library and the Computer Center.

Index.--The data Index, which reflects the data element item number, description, and the file in which the data may be found, uses the information contained in positions 001-004, 059-108, and 109-114, respectively. Because the Index will contain data elements which are present in many files, we propose to use the procedure described below to eliminate the printing of redundant items:

1. The data inventory (in item number sequence on magnetic tape) will be read into the computer and the item number, description, and file designation will be selected and moved to the output area. A counter will be incremented.

2. If the next following data element has the same item number (which will only occur if the data element is contained in a second file), the file indication (positions 109-114) will be moved to the output location next succeeding that which is indicated by the counter. The counter will again be incremented and this iteration continued until a new item number is detected, at which time the counter will be reset to zero.

3. The Index will be printed.

Although some users may be expected to become sufficiently familiar with the data system and programing techniques to effectively utilize the Index, it will essentially be used by systems programing personnel.

File Specification.--The File Specification, which reflects the full particulars on the form and content of each file being regularly maintained, uses all of the information contained in the data element inventory except the description (059-108). The procedure to be used in preparing the File Specification is as follows:

1. The data element inventory file will be

sorted into file/field sequence (109-114 and 010-013, respectively).

2. The file specification, paged by file, will be printed.

3. Title sheets will be added.

As was the case with the Index, the distribution of File Specifications will probably be limited to systems programing personnel.

Classified Information.--Although all data being maintained in the system will be reflected in the Data Catalog, Index, and File Specifications, knowledge of its existence does not necessarily include access to it. Those items of information which are felt to be confidential in nature and which should be made available only to qualified users, will be maintained in selected subfiles to which access will be restricted.

Storage

The problem of storing data in the system may be roughly divided into two categories: random changes, corrections, and updates (such as changes of name or address), which will be processed according to a schedule set to accommodate the work flow of organizational units providing and/or verifying the data; and volume updates which occur in conformance with the work or production schedules of the college (such as registration, late registration, census date, etc.). The first of these is discussed later under the topic of file maintenance. The latter is what we shall call the process of storing new data in the system.

The storage of new data may also be divided into categories: the addition of historical information to individual records, where the established design of such records accommodates such additions; the addition of individual records to established files; and the introduction of data not previously defined to the system.

Storing Historical Information.--When analysis indicates that the assembly of historical information is required, the file which is designated to retain the data may be easily designed to provide space for its storage. (If a fixed record length is used, the maximum number of historical cases or cycles must be predetermined and corresponding fields established.) Once the design of the file permits acquisition of historical information, the storage problem becomes similar to that of file maintenance.*

The establishment of files (and fields) which will retain historical information must be fully detailed in the Data Catalog, Index, and File Specifications, in order that users and programers may understand the structure of the data and its housekeeping requirements; i.e., the difference between the changing of a grade in a student record (from Incomplete to B) must be distinguished from that of the addition of a grade earned in a course.

Storing New Records.--The process of adding records to existing files is so common to data processing personnel that it warrants little treatment here. Essentially, this type of operation involves the creation of a subfile using the records of new cases or individuals, in a format identical to that of the file to which the data will be added, and of merging the new and existing files.

Introducing New Data.--The new-data storage

---

*One technique which facilitates the addition of historical information to files is the inclusion of a "counter" field in each record in an historical file (Figure 13). The function of the counter is to inform the file maintenance program of the number of historical cases or cycles which have been recorded to date. An example of this technique may best be demonstrated by the problem of accumulating the academic records of students, where it is necessary to record the course identification and grade at the end of each term. If, after four terms, a student had completed sixteen courses, the counter (n) would be set at the value of sixteen, and the computer storage program would, when executing the following update cycle, begin storing the new historical information in n + 1 location.

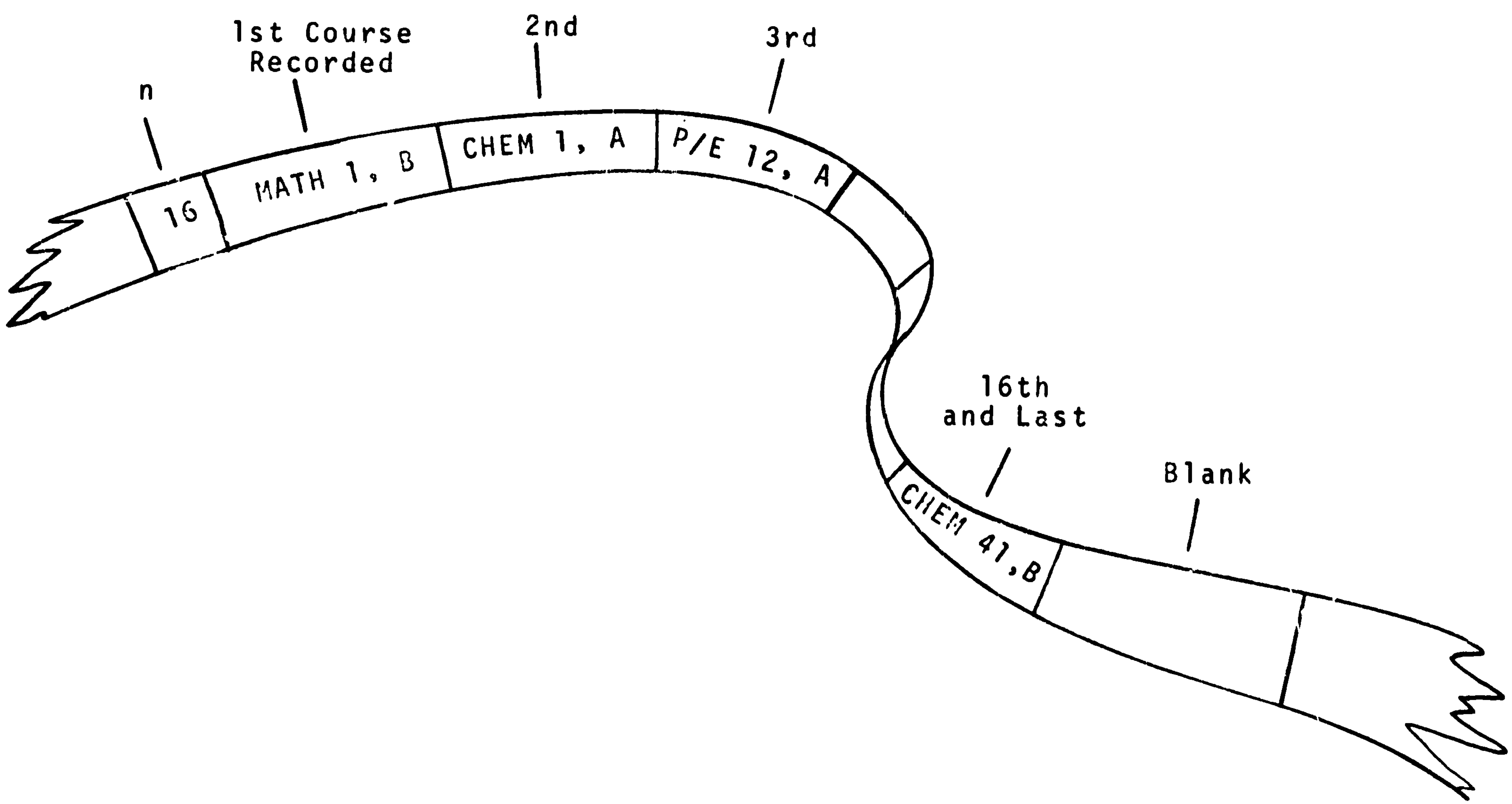

Figure 13. Representation of an Historical Record Containing a "Case Counter" (n)

problem involves the establishment of all information handling procedures required to maintain data which had not previously been defined as part of the system.

At the time that a determination is made to introduce additional information to the data base, systems personnel will analyze the data as follows:

1. Initial data acquisition problems will be evaluated in order to determine whether an effort should be made to capture non-machine readable data already in existence (past history), or whether the data acquisition should be limited to that information which is acquired in the future.

2. The routine input cycles will be defined in order that updating procedures may be identified and scheduled.

3. The appropriate point of entry for the data, together with necessary forms and validation and verification procedures, will be established.

4. The Key Files (or segments thereof) which will retain the data will be identified on the basis of data acquisition and retrieval requirements.

5. Descriptors and other specifications to be used in the Data Catalog, Index, and File Specifications will be defined.

6. Those standard Working Files in which the new data will reside will be defined and the programs which are used to retrieve, or build, the Working Files will be modified accordingly.

7. The housekeeping programs used to maintain the affected Key Files will be modified to accommodate the additional data.

Much of the above would be facilitated in a large-scale, random access system. Third generation software now being introduced provides for many of these steps, provided that the new data is properly related (by linkages) to the existing data base. It is our feeling that the manual effort required for the expansion of the data base in a small-scale, tape oriented computer system may be economically achieved, provided that detailed operating procedures and effective quality control measures are carefully observed.

Information Retrieval

Depending upon the orientation of the individual, the topic of information retrieval may introduce many, extremely varied ideas and notions. Visions of the cathode ray tube mounted in one's desk, with a button which permits him to obtain facts relevant to the problem at hand, may enter some minds. Others (programers) would be thinking in terms of address table lookups, permitting data chains to be generated.

For the purpose of this report, information retrieval is defined as the problem of outputting information based upon a user's requirements and the content and organization of the data base. This output may be obtained at random, by direct access, or it may not; it may relate to single cases, or to the entire universe of the data within the scope of the system. This section, relating to the retrieval of information, will discuss the creation of the Working File which will accommodate (either) groups of standard production programs, or particular special purpose assignments.

The preparation of Working Files which are required for the many day-to-day report requirements of the institution will be deferred and treated under the topic of file maintenance. The steps necessary to be taken in designing procedure for the preparation of such standard Working Files are, however, essentially the same as those which are described in this section.

Information retrieval is discussed in the following in terms of phases of implementation. The step-by-step method to be used in retrieving specified information from the institution's data base will initially require a high degree of programer intervention which will be replaced by automated procedures as the sophistication of users and operators grows, and as equipment capabilities increase.

Implementation Phase One.--The following is a brief step-by-step description of the operating information retrieval system at the outset of implementation. This description assumes the existence of accurate (and complete) data base reference materials:

1. The user defines a problem and specifies the information which he feels should be brought to bear in order to obtain a resolution.

2. The user ascertains the existence of the data required in the preparation of his "problem solving" report through using the Data Catalog (and Code References, if necessary).

3. The user makes his information requirements known to the data processing staff. This may be in the form of (the data definition for) a working file, or a report mock-up.

4. Data processing personnel, through the use of the Index, assemble a list of files for each of the data elements required.

5. By analyzing the files which contain the required data, data processing personnel are able to determine the optimum source (or sources) of the data. This determination is made on the basis of processing time; therefore, the smallest segment of the data base which will yield all of the required data is the one which is selected for processing--unless a larger file will yield the data <u>in the required sequence</u>.

6. If more than one Key File must be accessed for the required data, data processing personnel must also identify, and specify the retrieval of, the necessary interfile linkage.

7. The retrieval parameters are prepared in

accordance with the requirements of the retrieval program.* Pre-punched READ, FORMAT, file OPEN, and DATA DIVISION statements will be available for each file being maintained, in order to: (a) assure error free input; and (b) simplify and speed up programing.

8. If more than one Key File is the source for the required data (Files A and B), each is processed per the above so that all of the data elements and required file linkages have been retrieved from each file, and matching and merging may be effected.**

9. The retrieved information is sequenced according to the extended processing requirements (or applications programs) of the user.

10. The retrieved information is processed and the end result provided to the user.

Implementation Phase Two.--As the implementation of the data system progresses, certain segments of the retrieval procedure will be automated. At this time, it would appear that the first steps which might be programed in order to eliminate workload for the data processing staff would be the following:

1. Step Four (above), consisting of the assembly of the list of files for each required data element, will be a relatively simple procedure to program. It will consist of reading

---

*The retrieval program will also be developed in phases, beginning with the use of special purpose programs for each retrieval problem, and leading to the use of one or more of the direct, random access software packages which are currently available for large-scale systems.

**The problem of consolidating data retrieved from more than one Key File is accomplished through sequencing, matching, and merging. In this case, merging implies the joining of matched (linked) records to form a single record.

the data Index (tape), searching for the required data elements, and then printing the list showing each file which contains that element.

2. Step Five, consisting of determining the optimum source(s) of the data may be accomplished by the system using the same criteria described above: finding the smallest segment (or subfile) of the data base which will yield all necessary data. (File and record sizes stored in the data reference system permit such comparisons to be used as part of the file selection criteria). If the required output sequence is known, it may also be used in making the selection, as file sequences are also stored in the system.

3. Step Six, consisting of the requirement that data processing personnel ascertain and specify the retrieval of necessary interfile linkage may be eliminated in some cases, by storing linkage data in the file reference system.

4. The automation of steps Seven and Eight (above) will be based upon the growing sophistication of both the retrieval and the match/merge programs.

Ultimate Implementation.--Advancements in the development of large-scale information retrieval systems which have occurred during the past few years suggest that major (manufacturer supplied) retrieval packages will soon be available for use with computer systems having large capacity storage devices such as magnetic cartridge, drum, or disc.

Through the use of indirect addressing and automatic table reference and construction, users should soon be able to access data through the use of English language descriptions which, through table lookups, cause data chains to be generated.

Retrieval Program.--The initial development of a generalized information retrieval program will conform to the following objectives:

1. FORMAT (or Data Division) statements for each Key File segment and standard Working File will be maintained by the system, along with READ and OPEN (Procedure Division) routines. Upon selection of files (see below), these statements and routines will be automatically called for use in a generated retrieval program.

2. The user/programer submits retrieval parameters to the program, including

   a) the item numbers of the data elements required in preparing final output;

   b) the sequence required for the final output;

   c) extended processing steps (i.e., special sub-programs); and

   d) output report formating specifications.

3. Although initially done by programing personnel, the evolving system will then select optimum files for retrieval by searching the data Index to determine the file(s) in which the desired data reside. (Assume items A, B, N, and P are requested, along with instructions to sequence on A. The program will search the data Index hoping to find a file containing A, B, N, and P, sequenced on A. If found, retrieval can proceed; if not, then search must continue until files are found which contain the required elements of data <u>and</u> the link between the files, before retrieval may proceed.)

4. The initial step in the retrieval of data will consist of calling the input FORMAT and file OPEN statements for each file to be processed, and of generating a "move" instruction for each item of data to be retrieved from each file. (Assume data elements A, B, N, and P are found to exist in File CM0710, along with many other elements of data, and that the input area

description for this file's data elements is: AIN, BIN, CIN, DIN, KIN, NIN, OIN, PIN, ZIN. The retrieval program will then create a "move" or "substitute" statement for each data element required. Essentially, these will appear as follows: AOUT = AIN, BOUT = BIN, NOUT = NIN, POUT = PIN.) The next phase of the retrieval program generator will analyze the "outputs" reflected in the move statements, and create an appropriate tape write procedure.

5. If the required data is not to be found in a single file, the retrieval program will search the Index in an attempt to find two files which met the following criteria:

   a) contain A, B, N, and P;

   b) possess a link (L) common to both files;

   c) are sequenced on link (L); and,

   d) are the shortest files available meeting requirements A and B.

   If these four conditions are met, the retrieval program selects the appropriate input and file opening statements (for the indicated files) and proceeds to generate the necessary move and output format statements.

6. If no files are found to meet all of the criteria indicated above, the program selects the two "best" files. In this case, files which contain the required data and the linkage, but which require sequencing are "best." (In such a case, the retrieval program will generate sort parameters for both files in order that they may be resequenced on the linkage (L) between the files, preparing them for consolidation via the Match/Merge program.

Match/Merge Program.--When it is necessary to assemble information retrieved from two or more files based upon the matching of interfile linkage, one may be confronted with four distinct logical problems:

1. A one-for-one matching situation where each item retrieved from File A will be matched with a single item retrieved from File B.

2. A one-for-many matching situation where one or more items from File B may match each item retrieved from File A.

3. A many-for-one matching situation where one item retrieved from File B may match many items selected from File A.

4. A one-or-more for one-or-more matching situation where each item retrieved from File A may be matched by one or more items from File B and each item from File B may be matched by one or more items from File A.

Initial implementation of generalized match/merge operations will involve the preparation of two machine programs to accommodate Problems One and Two above. (Because Problem Three may be accommodated by merely changing the logical designation of the files--A to B and B to A--the same match/merge program will satisfy).

Problem Four (one-or-more for one-or-more) will be accommodated by processing Files A and B per Problem Two, storing the output, and then reversing the files to obtain output per Problem Three. Both outputs will then be consolidated.

Following completion of the Match/Merge operation, the data may be sequenced according to the requirements of the users program(s) and processing may be executed as indicated earlier.

File Maintenance

File maintenance consists of the routine updating of the data base. Updating operations may be random changes in individual records (such as changes of address, telephone number, or room assignment) or they may involve large volume inputs of new data, such as registration, course grades, or purchasing and accounting transactions.*

In either case, file maintenance must be accomplished on the basis of a schedule so that all systems users may, at any given time, ascertain the currency or obsolescence of the data base. For example, if student drops and adds are processed once weekly (after five o'clock Friday) and the Dean of Students asks for an enrollment report on Wednesday, he would be informed that three days' drops and adds have not yet been entered into the data base.

Because confusion frequently exists with respect to the obsolescence of various working files, and there is a tendency to update individual subfiles without the use of a general file maintenance plan (in order to have "good" working files available for particular programs), lengthy correction runs are often necessary in the typical data processing installation. The system which is proposed here attempts to avoid this problem by identifying all file update requirements and executing updates according to a specific plan and schedule. The effective control of file updating is obtained through the use of the Key File as the point of entry for all new and update data. By updating the Key Files and then proceeding to create all of the standard Working Files necessary for routine operations, effective controls may be established over file maintenance.

Figure 14 illustrates the updating of Key File A through the File Maintenance system, and the subsequent reduction of the File into subfiles according to the requirements of the Working File system.**

---

*Note that "random changes" in the data base are random in the sense, that they occur at random--they are herein proposed to be processed in batches notwithstanding the fact that a "batch" may contain many different kinds of changes to be made to selected records within a file.

**As has been suggested earlier, the determination of optimum Working File content and update/retrieval cycles must be based upon the input/output

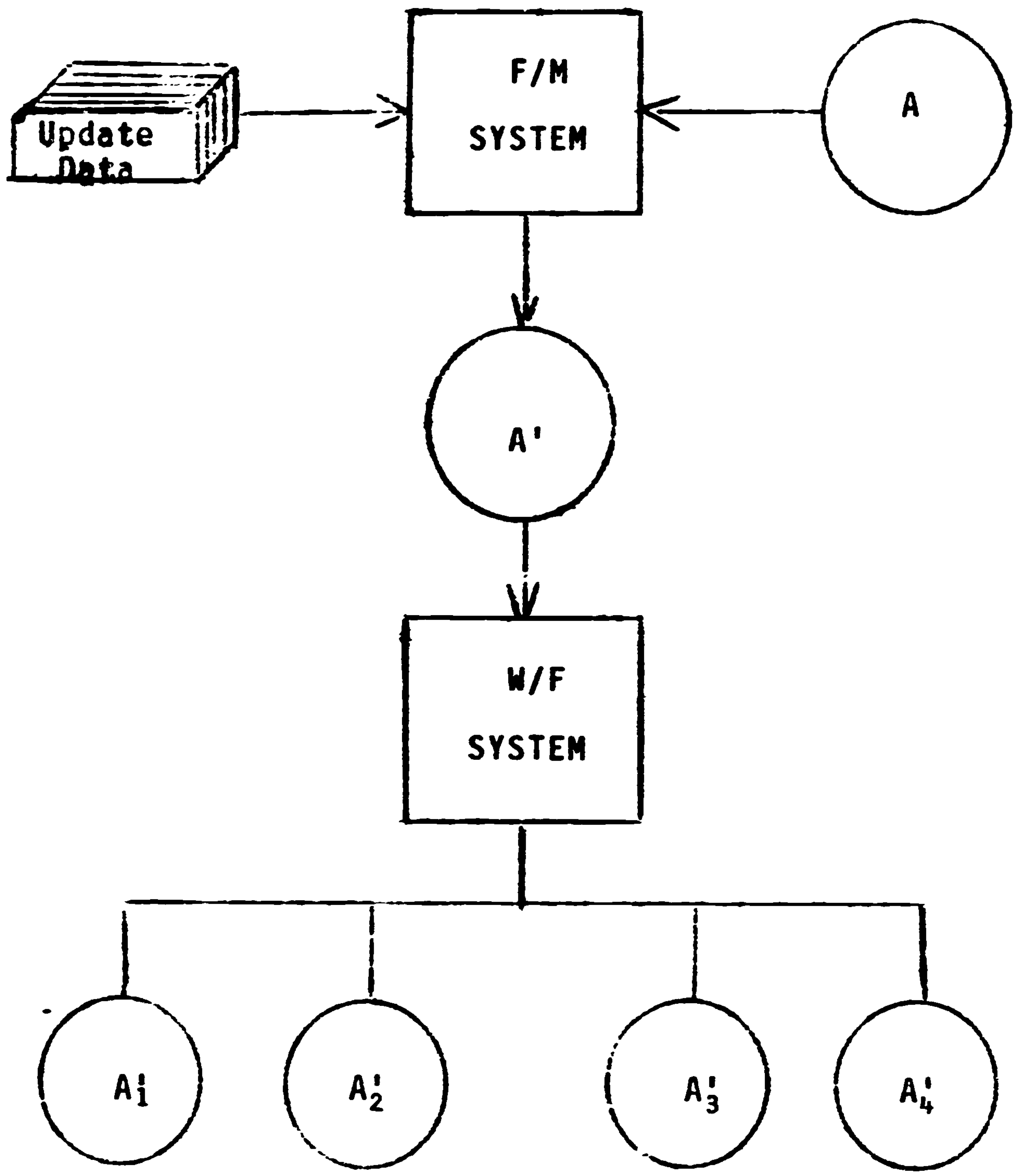

Figure 14. File Updating Through Key Files

Random file updating operations will be of two basic types: one where the exact location of the field to be updated is known; the other where numbers of fields are used to hold similar types of information, such as historical data, and the exact field to be updated is not known. The first case permits the file maintenance activity to be directed at a specific location within a particular record. The second case requires the searching of the fields within a particular record in order to locate the one which is to be updated. Discussions of these two cases follow.

Update 1, Location Known.--By referring to the file specification for that (segment of a) Key File for which updating data exists, the record may be altered by a computer program, as follows:

1. The update data (Figure 16) is introduced into the computer by the program and the term "Update 1" is recognized.

2. The item number of the data which is to be operated on, the data which is to be introduced, and the record identifier are present for interpretation.

3. After searching the file and locating the record to be updated, the new data is inserted into the field specified by the item number and a "before and after" copy of the record is printed.

An example of the input to be used for this type of updating is given in Figure 16, which suggests a fully automated file maintenance system where the program would identify the field to be updated

---

requirements of the institution. By analyzing and inventorying the data required for various types of reports (data element--report cycle), linear programing methods may be employed in solving for the "best" Working File content and preparation schedules, including the necessary sequencing of the files. Figure 15 shows a form used in inventorying the report requirements of the institution which, in turn, permits the analysis of Working File content.

ITEM NUMBER: ___ DESCRIPTION:
1 4

PERIOD AND FREQUENCY											
CALENDAR	ACADEMIC	FISCAL	CONVENIENCE	YEAR	HALF YR	QUARTER	MONTH	WEEK	DAY	NAME OR NUMBER OF REPORT	SEQUENCE (LIST, MOST TO LEAST)
5	6	7	8	9	10	11	12	13	14	15 40	41 80

Figure 15. Data-Element/Report-Cycle Inventory Form

(beginning and ending locations) on the basis of the data element item number and the action to be taken (add, delete, change) on the basis of the Code. Initial implementation of this system will most probably require that the programer or control clerk ascertain and specify such information.

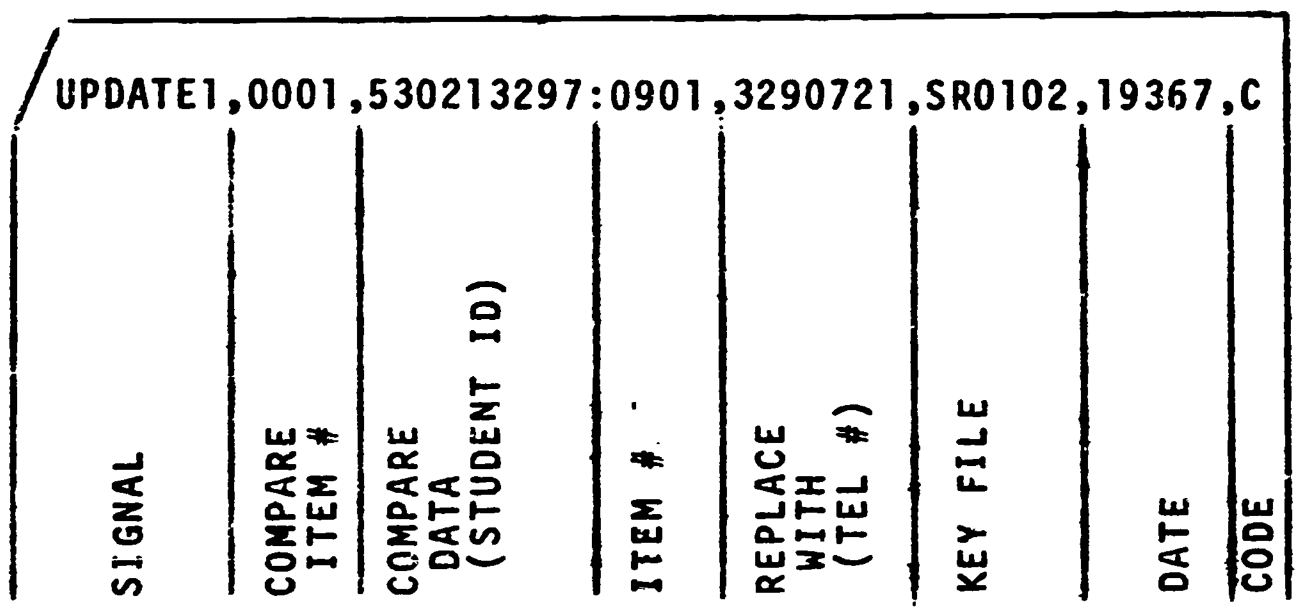

Figure 16. Input for Update 1, Location Known

<u>Update 2, Location Unknown</u>.--In cases where the specific location of a data element requiring change may only be determined through inspection of the appropriate record, provision must be made for the file maintenance program to perform this function. For example, if course grades are added to student records at the end of each term and it becomes necessary to change a particular grade (i.e., from "I" to "B"), the field in the student record which contains the grade to be changed must be determined through a logical procedure. Figure 17 presents an example of a student record which contains grades earned and which require modification. (Grade #20, an Incomplete earned in Math 160, is to be changed to B.)

Description	Actual Data	Beginning Location	Item #
Student ID#	530213297	0001	0001
Course Taken (1)	ART 100	1001	0287
Course Taken (2)	BIO 102	1000	0288
Course Taken (3)	PSYCH100	1017	0289
Course Taken (4)	SPAN 100	1025	0290
Course Taken (19)	ECON 110	1145	0306
Course Taken (20)	MATH 160	1153	0307
Course Taken (21)	MUSIC100	1161	0308
Course Taken (79)	ΔΔΔΔΔΔΔΔ	1625	0366
Course Taken (80)	ΔΔΔΔΔΔΔΔ	1633	0367
Qtr Taken (1)	FALL66	1701	0371
Qtr Taken (2)	FALL66	1707	0372
Qtr Taken (3)	FALL66	1713	0373
Qtr Taken (4)	FALL66	1719	0374
Qtr Taken (19)	WTR 67	1809	0390
Qtr Taken (20)	WTR 67	1815	0391
Qtr Taken (21)	SPR 67	1821	0392
Qtr Taken (79)	ΔΔΔΔΔΔ	2175	0449
Qtr Taken (80)	ΔΔΔΔΔΔ	2181	0450
Grade (1)	B	2201	0460
Grade (2)	A	2202	0461
Grade (3)	C	2203	0462
Grade (4)	D	2204	0463
Grade (19)	A	2219	0418
Grade (20)	I	2220	0419
Grade (21)	C	2221	0420
Grade (79)	Δ	2279	0478
Grade (80)	Δ	2280	0479

Figure 17. Part of Student Record Showing Grades Earned

The procedure which must be followed in locating and updating such data (in this case, the grade) is as follows:

1. The update data (Figure 18) is introduced into the computer by the program and the term "Update 2" is recognized.

2. The program recognizes a) that the first item number specified is to be used in a matching process in order to locate the record to be updated, and b) the data following this item is the data to be matched.

3. The program then recognizes that the following item number, in this case item number 0287, is the beginning of a set of fields to be searched for a match. The number of iterations which are required to be executed in finding the matching data (i.e., the data following the second item number) are counted, and the beginning location of the third item number specified (if any) is established as the sum of the beginning location for the third item number and the product of its field length and the number of iterations required to obtain the match for the second item number.

4. If the data following the third item number matches the data located (per 3 above), the program proceeds to similarly match the fourth item number, if any. The termination of this procedure is established when a colon (:) is encountered.

5. The item number following the colon prescribes the beginning location of the set of fields, one member of which is to be updated. The beginning location of the particular field to be updated is equal to the sum of the beginning location for the set of fields and the product of the field length and the number of iterations required to obtain the match, per three and four above.

6. If the data following the third item number

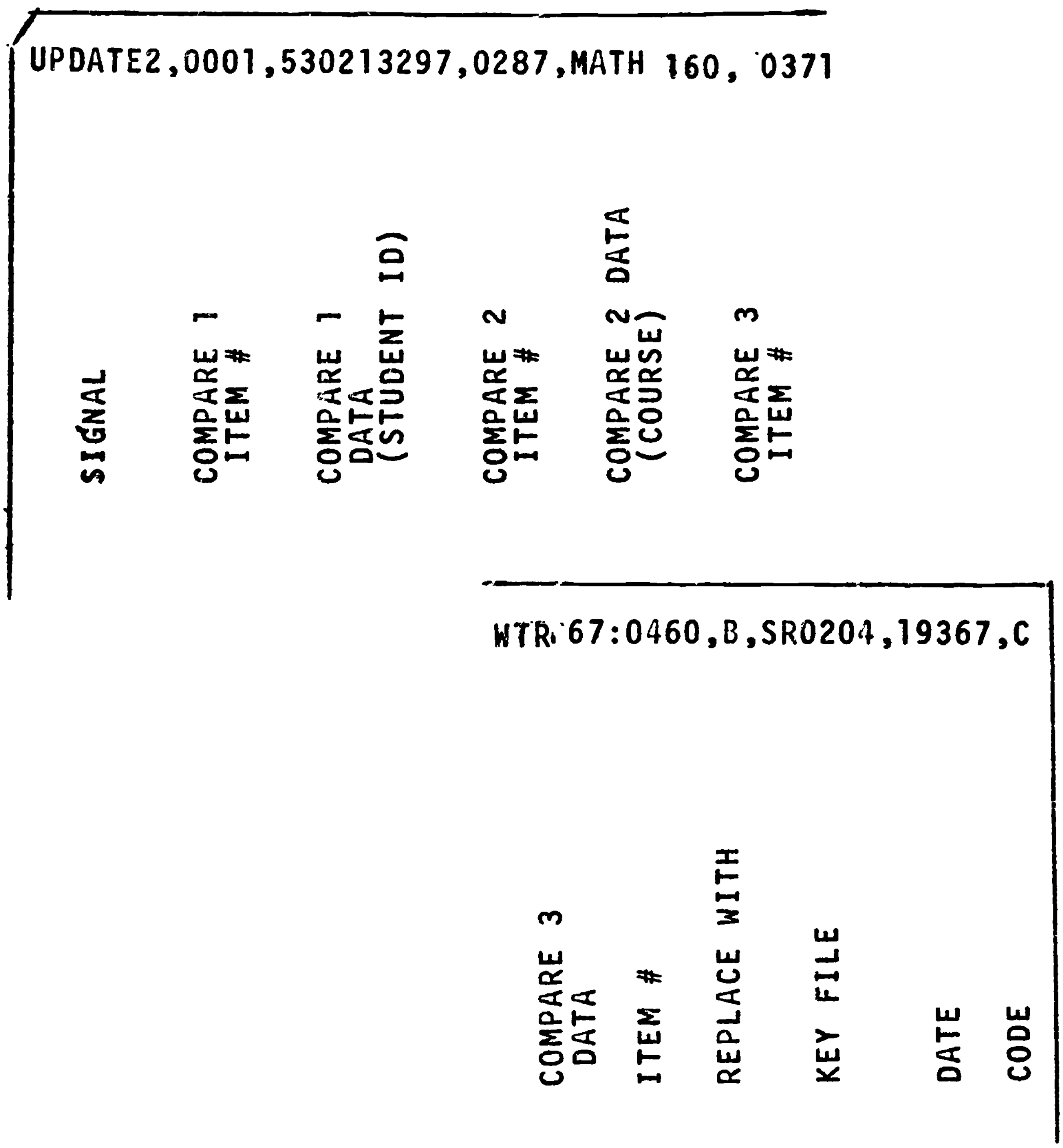

Figure 18. Input for Update 2, Location Unknown

does not match the data in the record, the program loops back to the set of fields first specified (second item number plus iterations) and resumes searching for another match.

Applications Programing

The fullest advantage of a broad base data system is obtained through the interchange of information between related organizations. The key to the effective exchange of information is the compatibility of data base and data base interpretation. (It is not necessary that coded data be expressed in identical form by the cooperating institutions, but only that interpretations of data be related to a common base.)

The authors of this report believe that the future will see an increase in data compatibility (which will result in part from the stabilization of the report requirements of the U. S. Office of Education), and that the analytical treatment of the data base will be enhanced through the exchange of machine programs by cooperating institutions and organizations. In order for machine interpretable data and machine programs to be effectively exchanged between institutions (or between organization units within institutions) it is necessary that:

1. Data base definitions be fully and completely documented;

2. Machine programs be fully and completely documented;

3. Machine programs be prepared in programing languages of a standardized or universal type;

4. Precautions are taken to avoid the use of special features within available programing languages, in order that institutions with equipment installations of lesser capability may be able to utilize the programs prepared by others; and,

5. That insofar as it is possible, programing personnel should be trained in the techniques of converting, or translating, programs from one machine system to another.

Program and Procedure Standards.--In order that other institutions may most effectively utilize the work done by others, the standardization of operating procedures and programing methods must be established. While it is impractical to assume that a great number of cooperating institutions will be able to adopt identical programing conventions and operating procedures (due to differences in equipment and personnel competencies), an important step in this direction is the complete documentation of such procedures. Complete and detailed documentation will permit the personnel of other organizations to quickly identify the work which is necessary to effect conversions.

One step in the adoption of standard procedures is the maintenance of an inventory of all data and documentation. This inventory should describe the location of files, hard copy data, and other important material such as cards and tapes. The maintenance of such an inventory is best accomplished through the assignment of responsibility to a single control point (library clerk) where the storage and cataloging of each source/object/data deck, tape program write-up, and batch of input or output information is effected. A "circulation" control should be established so that information or data may be loaned out when necessary, and its return assured.

Program Documentation.--The following briefly indicates the documentation parameters which are requisite to effective machine program exchange:

1. Authorship, including the institution (and organizational component) preparing the program, the individual(s) writing the program, a meaningful program identification, and the date of publication or release;

2. Machine specifications, defining the minimum requirements for the processing (and compilation) of the program;

3. Operating instructions, including program load, error or other halts, and necessary operator "switch" settings;

4. Library program and/or subroutine requirements, input and output formats, and code interpretations; and,

5. The history of the development of the program and the uses which the output of the program serves.

(When sets of machine programs are required to fulfill the needs of a particular problem, they should be packaged and described as an operating subsystem, in the detail suggested above.)

DISCUSSION

The initial plan for this project was focused on the development of an operational data system for an institution of higher education. Specifically, the procedure involved the definition of the institution's information requirements, a synthesis of the operating system, the development of procedures for the acquisition and processing of data, the subsequent evaluation and extension of the data system (to provide for management level decision-making), and a final testing and evaluation of the system which would permit other similar systems to benefit by the work done here.

A significant deviation from the plan evolved out of the early investigations of this undertaking. The change of focus resulted from many factors: confusion concerning the terms and definitions of institutional information requirements; a constant state-of-change with respect to reporting requirements imposed on institutions of higher education; an unstable operating environment caused by frequent change of equipment, personnel; ill-defined and frequently changing information processing objectives; and the lack of a general plan defining long-term objectives for the data systems (which fostered a lack of integration of the data base).

The above conditions seemed to manifest themselves (in organizations where the utilization of electronic data processing equipment had been established over a period of years) in a distribution of work for data processing personnel which involved as much as seventy per cent of their effort being directed toward the changing of computer programs. (This was necessary to either accommocate changes in the data base or in the institution's report requirements.) In other words, attempts to establish data systems (more properly called subsystems) which would be permanent fixtures and which (once tested and debugged) would accommodate the institution in perpetuity usually failed.

The impact of these findings on this project was a change in focus primarily directed to the development of operating concepts which would permit the development of a durable data management system. The system design must be feasible for implementation on a relatively small scale, but its total scope must be sufficiently inclusive so that large scale implementation is possible without material change in philosophy.

The authors believe that there is no single answer to institutional information system problems. They are most interested in reactions to some of their ideas, and in seeing proofs and refutations of some of their operating concepts. It is believed that the current interest in management information systems for higher education will continue to grow, and that the work presented here will play some part in the development of truly effective systems implementation which will be realized over the next few years.

Qualifications and Limitations.--The bulk of the background information contained in this report was obtained from educational administrators and data processing personnel in the California State College system. Certain planning techniques used by the system have proven to be quite effective over the years. However, the system has fallen behind many other institutions in the utilization of modern technology for the development of management information. Some of the problems described in this report have been effectively overcome in other institutions, but it is felt that the development of a truly effective management information system has yet to be attained, primarily based upon California experience. It may well be that California experience and relatively limited contacts with institutions in other regions have led to an inadequate appraisal of progress in the development of management information systems in higher education and the development of data management systems in general. However, current literature in both of these fields tend to support the contentions in this report.

In addition to the limitations in view mentioned above, the operating concepts presented here are free from dependence upon specific machine considerations. It may be that the full use of a manufacturers hardware and software systems may require the adoption of other operating principles, and that more expedient and economical implementation may be effected through the use of manufacturer prescribed systems.

CONCLUSIONS AND IMPLICATIONS

Recent publications, and attitudes evidenced by many of the educational administrators contacted during the course of this study, have resulted in our conclusion that the development and implementation of management information systems for institutions of higher education will increase significantly in the next few years.

Some administrators, although voicing strong approval of the operating concepts described here, have expressed concern that their institutions were too deeply committed to systems development using a more traditional, segmented approach. It should be noted that such situations do not constitute reasons for the avoidance of comprehensive planning, but, rather, require that plans make provision for system conversion.

The "total" or integrated data system concept presented in this report may become the basis for some systems implementation programs in higher education. Other systems (with similar objectives) may be formulated aroun' hardware-software packages developed and offered by equipment manufacturers. We believe that the most common management information systems of the future will probably grow out of the extension and integration of existing subsystems, and that this will usually occur when an institution is in the process of upgrading its data processing installation, provided an overall plan exists when system conversion is implemented.

The implications of the development and use of a management information system will be felt from many standpoints within an institution. Changes will be evolutionary for the most part, but we believe that many are predictable and should be fully considered when planning the implementation of such a system.

Organization.--Many organization units within institutions of higher education are justified on the basis of the information which they collect and maintain. By establishing a broad base, general purpose information system, the need for many of these independent subsystems will cease to exist. It is suggested, therefore, that the formal organization structure in institutions of higher education may be expected to change as a direct result of the consolidation of existing information systems.

The improvement in intra-institutional communication which will result from a management information system will reduce the degree of autonomy afforded certain organization units, due to the fact that information concerning their operations and effectiveness will be more readily available to higher levels of management. The possibility of sustained inefficiency will be far less likely, as will the opportunity for organization units to operate in contravention to the policies of the institution. In short, management effectiveness (foresight and strength) should be enhanced.

The traditional role of committees in the decision making process may undergo significant change due to the availability of objective information bearing upon the problems at hand. Decision making may increasingly reflect the desire of the institution to formulate policy on the basis of specific (economically utilitarian) objectives. The widespread use of committees in decision making may therefore be reduced, and be limited to those areas where management information remains undeveloped and a more subjective approach obtains.

Staffing.--The consolidation of many of the institution's data collecting, maintaining, and retrieving operations should result in a change in its staffing pattern. The extent to which operating personnel involved in the clerical processes may be reduced through the use of machines will depend upon the findings of continuing management evaluation. It is anticipated, however, that a relative increase in middle and higher level management personnel will occur. A proportionate increase in the number of more highly paid administrative personnel, bringing higher level skills to bear more directly on the operations of the institution, should bring about a change in the character of the institution's staff which should become quite noticeable. Higher level competencies in increased proportion will, among other things, result in increased student contact with individuals holding broader views of the institution, and its capabilities and limitations.

Curriculum.--The availability of analytical program information should tend to minimize needless course proliferation. Ample opportunity for the analysis of program costs and benefits will permit educational administrators to effectively plan the allocation of the institution's resources (the faculty and facilities) to provide for the strongest possible instructional programs. An improvement in the quality of instruction should result from improved facilities planning and faculty recruitment.

Faculty.--The availability of the institution's data base and the existence of tools which permit convenient access to it should be of great value to the institution's faculty. The analysis of the effectiveness of programs and parts of programs may be readily undertaken by the faculty having responsibility for them.

SUMMARY

This report discusses a management information system for higher education decision making. The organization and management of a college or university may be characterized by numbers of relatively autonomous organization units, many of which necessarily maintain their own information systems. Most frequently, these information systems, developed on the basis of specific needs, do not readily interface with each other. Top level management (trustees, presidents, and their key personnel) are therefore unable to readily relate the information maintained by one organizational component to that of another, and they are thusly denied comprehensive information upon which to base important plans and decisions.

The purpose of this report is to prescribe, in lay terms, the characteristics of a data system which serves the needs of the operating organizations, and, at the same time, provides for a comprehensive, integrated data base for the development of management information.

An inventory of the operating information requirements of institutions of higher education was undertaken early in this project. It was found that a single college or university could be expected to maintain identical information within many organization units, and that this redundancy was frequently "justified" on the basis of differences in work schedules. In some cases, the redundancies were not readily apparent, due to the use of different terms and definitions for the same data, and to the use of similar terms and definitions for different types of data. The authors have found excellent source materials setting forth terms and definitions of universal utility to be available. This, combined with the belief that a systems approach to information storage and retrieval, minimizing schedule problems, makes the implementation of a "total" data system feasible and, indeed, overdue.

The necessary characteristics of an effective management information system are flexibility and

responsiveness, combined with the capacities for growth and projective and diagnostic output. This report presents general operating concepts which provide for the fullest possible integration and exploitation of the data base, and which permit the orderly development and growth of an operational data system. The fundamental design principles of the system are as follows:

1. The data base itself may be regarded as a set of files, linked (directly or indirectly) by natural data which are appropriate to two or more files in the set.

2. The maintenance of the data base, be it random changes to be made in isolated cases or wholesale updates, must be effected through a general plan which provides for the accessing of a minimum number of files (according to specific schedules and procedures).

3. The retrieval of information should, insofar as it is possible, be effected through routine, scheduled operations which are based upon the consideration of the total requirements of the institution.

4. The introduction of new data to the data base should similarly be effected in an orderly, scheduled manner, and it should be accomplished in such a way as to minimize changes in existing operating procedures and programs.

5. Educational administrators and researchers are regarded as the users of the data system and, accordingly, they must be provided with tools which will facilitate their understanding of the data base and the procedures which are to be used in its exploitation.

REFERENCES

1. Abell, Victor A. Purdue Academic Student Scheduling: PASS. Lafayette, Indiana: Purdue University. July 1965. 40p.

2. _____; Blakesley, J. F.; Morgan, G. E.; and Sherwood, W. C. A Comprehensive University Scheduling System: CUSS. Lafayette, Indiana: Purdue University. March 1965. 34p.

3. Argyris, Chris. Personality and Organization. New York: Harper and Row, Publishers. 1957. 291p.

4. _______. Understanding Organizational Behavior. The Dorsey Series in Anthropology and Sociology. Homewood, Illinois: The Dorsey Press, Inc. 1960. 179p.

5. Basic Data and Unit Measures Relating to Cost of Instruction. Coordinating Council for Higher Education. Sacramento, California: 1965. Unpaged draft.

6. Bross, Irwin D. J. Design for Decision. New York: The Macmillan Company. 1953. 276p.

7. Calhoun, Jr., John C., and Morrison, Don F. (Texas A & M University). A Report on Total Information Systems for Colleges and Universities. A Supplement to the Final Report on the project To Devise and Test Simplified Adequate Systems of Measuring and Reporting Financial, Manpower, Facilities, Research, and Other Activities in Colleges and Universities. College Station, Texas: 1965. Unpaged draft.

8. California State Department of Education. A Master Plan for Higher Education in California, 1960-1975. A Report prepared for the Liaison Committee of the Regents of the University of California and the California State Board of Education. Sacramento: California State Department of Education. 1960. 230p.

9. California State Department of Education. A Restudy of The Needs of California in Higher Education. A Report prepared for the Liaison Committee of the Regents of the University of California and the California State Board of Education. Sacramento: California State Department of Education. 1955. 473p.

10. Carlson, Walter M. "A Management Information System Designed by Managers," Datamation. May 1967. p. 37-43.

11. Carzo Jr., Rocco, and Yanouzas, John N. Formal Organization: A Systems Approach. Homewood, Illinois: Richard D. Irwin, Inc., and The Dorsey Press. 1967. 22p.

12. Cohen, Irving. "Computers and Communications Symposium," Datamation. March 1967. p. 63-64.

13. American Association of Collegiate Registrars and Admissions Officers. Committee on Data and Definitions in Higher Education. Handbook of Data and Definitions in Higher Education. New York: A. A. C. R. A. O. 1962, 132p.

14. Cowley, W. H. A Holistic Overview of American Colleges and Universities. Stanford, California: Stanford University. 1965. p. 97-103.

15. ______. Some Myths About Professors, Presidents, and Trustees. A paper presented in the Kennecott series. Tucson, Arizona: University of Arizona. 1962. 29p.

16. Edelman, Paul R. "Safeguarding Tape-Stored Data," Datamation. January 1967. p. 46-47.

17. Follett, Mary Parker. "Dynamic Administration," The Collected Papers of Mary Parker Follett. Edited by Henry C. Metcalf and L. Urwick. London: Sir Isaac Pitman & Sons, Ltd. 1941. 320p.

18. French, Wendell (University of Washington). The Academic Manpower Administration System at the

University of Washington: Toward a Generalized Theory. A Supplement to the Final Report on the project To Devise and Test Simplified Adequate Systems of Measuring and Reporting Financial, Manpower, Facilities, Research, and Other Activities in Colleges and Universities. Seattle, Washington: 1965. Unpaged draft.

19. Friedberg, Lou M. "RPG: The Coming of Age," Datamation. June 1967. p. 29-31.

20. Glaser, R. George. "Are You Working On The Right Problem?" Datamation. June 1967. p. 22-25.

21. Head, Robert V. "Management Information Systems: A Critical Appraisal," Datamation. May 1967. p. 22-27.

22. Henle, Reverend R. J., Director and Compiler (St. Louis University). To Devise and Test Simplified Adequate Systems of Measuring and Reporting Financial, Manpower, Facilities, Research, and Other Activities in Colleges and Universities. A Final Report. St. Louis, Missouri: June 1965. Unpaged draft.

23. Hower, Ralph M. and Orth 3rd, Charles D. Managers and Scientists. Boston: Harvard University Press. 1963. 323p.

24. Information Service Guide. The California State Colleges, Office of the Chancellor. Los Angeles, California. Unpaged manual.

25. I-D-S (Integrated Data Storage). General Electric Corporation, Computer Department. Phoenix, Arizona. 6p.

26. International Business Machines, Inc. University-College Information System. White Plains, New York. 15p.

27. Kibbee, Joel M.; Craft, Clifford J.; and Nanus, Burt. Management Games. New York: Reinhold Publishing Corporation. 1961. 347p.

28. Leslie, Harry. "The Report Program Generator,"

Datamation. June 1967. p. 26-28.

29. McDonough, Adrian M., and Garrett, Leonard J. Management Systems: Working Concepts and Practices. The Irwin Series in Management. Homewood, Illinois: Richard D. Irwin, Inc. 1965. 311p.

30. Magee, John F. Production Planning and Inventory Control. New York: McGraw-Hill Company. 1958. 333p.

31. Mahn, Robert E. A Glossary of Terms Used by Registrars and Admissions Officers. Washington, D. C.: 1956. 64p.

32. March, James G., and Simon, Herbert A. Organizations. New York: John Wiley and Sons, Inc. 1958. 262p.

33. Martyn, Kenneth, et al. "A Program Budget For The California State Colleges." California State College at Los Angeles. 27p.

34. Milfs, M. Milo, et al. Report of Special Study Committee on Student Records. Unpublished report. California State College at Palos Verdes. June 1966. 58p.

35. Miller, David W., and Starr, Martin K. Executive Decisions and Operations Research. Englewood Cliffs, New Jersey: Prentice-Hall, Inc. 1960. 446p.

36. Pfiffner, John M. The Supervision of Personnel: Human Relations in the Management of Men. Prentice-Hall Industrial Relations and Personnel Series. Englewood Cliffs, New Jersey: Prentice-Hall, Inc. 1951. 454p.

37. Proceedings of the Conference on College and University Interinstitutional Cooperation. Edited by John J. Wittich. Corning, New York: College Center of Finger Lakes. April 1962. 87p.

38. Reason, Paul L., and Tankard, Jr., George G. Compilers at the U. S. Department of Health, Education, and Welfare, Office of Education. Property Accounting for Local and State School Systems. State Educational Records and Reports Series: Handbook III, Bulletin 1959, No. 22, OE-21019. Washington, D. C.: U. S. Government Printing Office. 1962. 194p.

39. ______; Foster, Emery M.; and Will, Robert F. Compilers at the U. S. Department of Health, Education, and Welfare, Office of Education. The Common Core of State Educational Information. State Educational Records and Reports Series: Handbook I, Bulletin 1953, No. 8. Washington, D. C.: U. S. Government Printing Office. 1960. 116p.

40. ______, and White, Alpheus L. Compilers at the U. S. Department of Health, Education, and Welfare, Office of Education. Financial Accounting for Local and State School Systems: Standard Receipts and Expenditure Accounts. State Educational Records and Reports Series: Handbook II, Bulletin 1957, No. 4, OE-22017. Washington, D. C.: U. S. Government Printing Office. 1962. 235p.

41. Report on Fiscal System Design. Unpublished report. California State College at Palos Verdes. March 1966. 34p.

42. Russell, J., and Doi, James I. Manual for Studies of Space Utilization in Colleges and Universities. Ohio University. Athens, Ohio: American Association of Collegiate Registrars and Admissions Officers. 1957. 130p.

43. Samuelson, Everett V., and Tankard, Jr., George G. Compilers at the U. S. Department of Health, Education, and Welfare, Office of Education. Financial Accounting for School Activities. State Educational Records and Reports Series: Bulletin 1959, No. 21, OE-22018. Washington, D. C.: U. S. Government Printing Office. 1962. 1 ^p.

44. Schneidewind, Norman F. "The Practice of Computer Selection," Datamation. February 1967. p. 22-25.

45. Schultheiss, Louis A.; Culbertson, Don S.; and Heiliger, Edward M. Advanced Data Processing In The University Library. New York: The Scarecrow Press, Inc. 1962. 388p.

46. Selltiz, Claire, et al. Research Methods in Social Relations. Revised. New York: Holt, Rinehart and Winston, Inc. 1959. 622p.

47. The Service Bureau Corporation. "Keyword-in-Context Indexing for the IBM 7094." A users manual. Inglewood, California. May 1967. 22p.

48. Simon, Herbert A. Models of Man. New York: John Wiley and Sons, Inc. 1957. 287p.

49. _____. Administrative Behavior. New York: The Macmillan Company. 1961. 259p.

50. _____. The New Science of Management Decision. New York: Harper and Row. 1960. 50p.

51. _____; Smithburg, Donald W.; and Thompson, Victor A. Public Administration. New York: Alfred A. Knopf, Inc. 1962. 582p.

52. Skillings, H. Hills. Computer Class Scheduling: Experience At The University of Massachusetts. Office of Institutional Studies Information Series, No. 5. Amherst, Massachusetts: University of Massachusetts. December 1962. 12p.

53. Strother, George B., editor. Social Science Approaches to Business Behavior. The Irwin-Dorsey Series in Behavioral Science in Business. Homewood, Illinois: The Dorsey Press, Inc.; and Richard D. Irwin, Inc. 1962. 183p.

54. Von Neumann, John, and Morgenstern, Oskar. Theory of Games and Economic Behavior. Third ed. Princeton, New Jersey: Princeton University Press. 1953. 641p.

55. Werling, Richard. "Action-Oriented Information Systems," Datamation. June 1967. p. 57-65.

56. Wiener, Norbert. Cybernetics. Second edition. New York: MIT Press. 1961. 212p.

57. ______. The Human Use of Human Beings. Boston: Houghton Mifflin Company. 1950. 241p.

58. Williams, Douglass A. "Conversion at Lockheed Missiles and Space," Datamation. January 1967. p. 39-45.