AMRL-TR-66-200

DEVELOPMENT AND APPLICATION OF COMPUTER SOFTWARE TECHNIQUES TO HUMAN FACTORS TASK DATA HANDLING PROBLEMS

K. W. POTTER
A. T. TULLEY
SYSTEM DEVELOPMENT CORPORATION

LAWRENCE E. REED
AEROSPACE MEDICAL RESEARCH LABORATORIES

DECEMBER 1966

JOINT NASA/USAF STUDY

Distribution of this document is unlimited

AEROSPACE MEDICAL RESEARCH LABORATORIES
AEROSPACE MEDICAL DIVISION
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

NOTICES

1000 - February 1967 - C0192-23-488

1687

AMRL-TR-66-200

DEVELOPMENT AND APPLICATION OF COMPUTER SOFTWARE TECHNIQUES TO HUMAN FACTORS TASK DATA HANDLING PROBLEMS

K. W. POTTER

A. T. TULLEY

LAWRENCE E. REED

FOREWORD

This report presents the findings of a research effort sponsored jointly by the United States Air Force and the National Aeronautics and Space Administration. The report was prepared by the Dayton, Ohio facility of System Development Corporation for the Behavioral Sciences Laboratory, Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio. Mr. Fred H. Wise was the principal investigator. The research was conducted under Contract AF 19(628)-3418 during the period of 21 June 1965 through 21 June 1966.

This research was conducted in support of Project 1710, "Human Factors in the Design of Training Systems," and Task 171006, "Personnel, Training, and Manning Factors in the Conception and Design of Aerospace Systems." Dr. Gordon Eckstrand was the Project Scientist and Mr. Melvin Snyder was the Task Scientist. Mr. Lawrence E. Reed served as the contract technical monitor.

The research discussed in this report represents a mid-point in a total effort directed toward the development of techniques for handling human factors task data. A previous feasibility study, conducted by Computer Concepts Inc., and the American Institute for Research is reported in: AMRL-TR-65-131, "The Role of Human Factors Task Data in Aerospace System Design and Development," AMRL-TR-65-206, "The Role of Computers in Handling Aerospace Systems Human Factors Task Data," and AMRL-TR-65-231, "Basic Human Factors Task Data Relationships in Aerospace System Design and Development." Subsequent research will be directed toward computer software and the testing and validation of techniques.

Appreciation is extended to Dr. Stanley Deutsch, National Aeronautics and Space Administration, Washington, D. C., who served as the NASA initiator and monitor, for reviewing the preliminary draft manuscript.

The authors wish to thank the many government and contractor personnel who contributed valuable suggestions throughout the course of this research and who helped review the draft manuscript.

Special thanks are due Mr. Melvin Snyder and Dr. Gordon Eckstrand, who assisted in the review of the manuscript and Mrs. Joan C. Robinette for her suggestions and editorial assistance. Acknowledgement is also made of other System Development Corporation personnel: Mr. Alan A. Anderson and Mr. Robert E. Yates who worked on the research during the period covered, and Mr. Donald E. Blair who edited this report.

This technical report has been reviewed and is approved.

WALTER F. GRETHER, PhD
Technical Director
Behavioral Sciences Laboratory
Aerospace Medical Research Laboratories

ABSTRACT

Research leading to the application of computer software techniques for handling human factors task data generated in support of aerospace system development programs is discussed. It is recognized that data handling techniques must be developed in context with their total operative environment. A concept of an operational data management system for storing, processing, and retrieving human factors task data in a government/contractor environment is discussed and illustrated. This concept is predicated on the assumption that a user-oriented computerized data system will help draw human factors specialists closer to their data. Five problem areas, considered to be fundamental to the development of data handling techniques, were researched. These areas are: (1) analysis of human factors task data, data relationships, and classification schemes, (2) application of vocabulary and thesaurus techniques to increase the effectiveness of communication among man/machine/software functions, (3) application of computer storage and retrieval techniques to human factors task data, (4) application of analytical and simulation techniques to human factors task data, and (5) application of current awareness techniques to provide notifications of data availability.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONT'D)

TABLE OF CONTENTS (CONT'D)

LIST OF ILLUSTRATIONS

LIST OF TABLES

SECTION I

CONCEPT OF AN OPERATIONAL DATA MANAGEMENT SYSTEM

A concept of an operational data management system for storing, processing, and retrieving human factors task data in a government/contractor environment is shown in figure 1. This concept is predicated on the assumption that a user-oriented, computerized data system will help draw human factors specialists closer in their data. A further assumption is that the application of such a system will reduce the problem of data accessibility and allow more effective use of data in the system engineering processes of aerospace system development and operation.

The importance of insuring an early human factors input into the design of systems has long been recognized by both government and industrial specialists in the area. The ever increasing complexity of today's aerospace systems demands that human performance considerations be an integral part of the overall system performance. Both the Air Force and the National Aeronautics and Space Administration (NASA) have instituted programs to insure proper considerations of the human element in system design. The Air Force has established a formalized program known as the "Personnel Subsystem" (AFR 30-8 and AFSCR 80-16)* which directs that system personnel--operator and maintenance--their training, and their support equipment, be regarded as an essential subsystem of the total system. While NASA has not established an overall standard program, the organization provides for the integration of human capabilities into the system design process. Various NASA Centers have established their own requirements. An example is the Marshall Space Flight Center Standard for Human Factors Engineering Program (MSFC-STD-391). This standard sets up the requirement to promote maximum effectiveness and reliability of the man as a system component.

To conceive, define, acquire, and operationally deploy a system is a complex evolutionary process requiring an integrated management approach to maintain performance, configuration, cost, and scheduling control. The detailed requirements and management procedures are contained in various government manuals and regulations (AFSCM 375- series, NASA NPC 500-1). The top section of figure 1 illustrates the rather formal management process for the initial and continued implementation of a personnel subsystem or human factors program.

As indicated above, an integrated management approach is necessary to develop and deploy an aerospace system. One of the principal integrating factors is data. In both the Air Force and NASA, the generation of data occurs throughout all phases in the life cycle of systems. As systems progress through each phase, the content and nature of the data change. For further understanding of the implication that these changes have on the development of a data management system, each phase is reviewed briefly.

The Conceptual Phase is a period of time extending from the determination of broad objectives--usually quite nebulous--to a formal approval of a Program

---

* Titles and dates of regulations, manuals, and exhibits are listed at the end of the References.

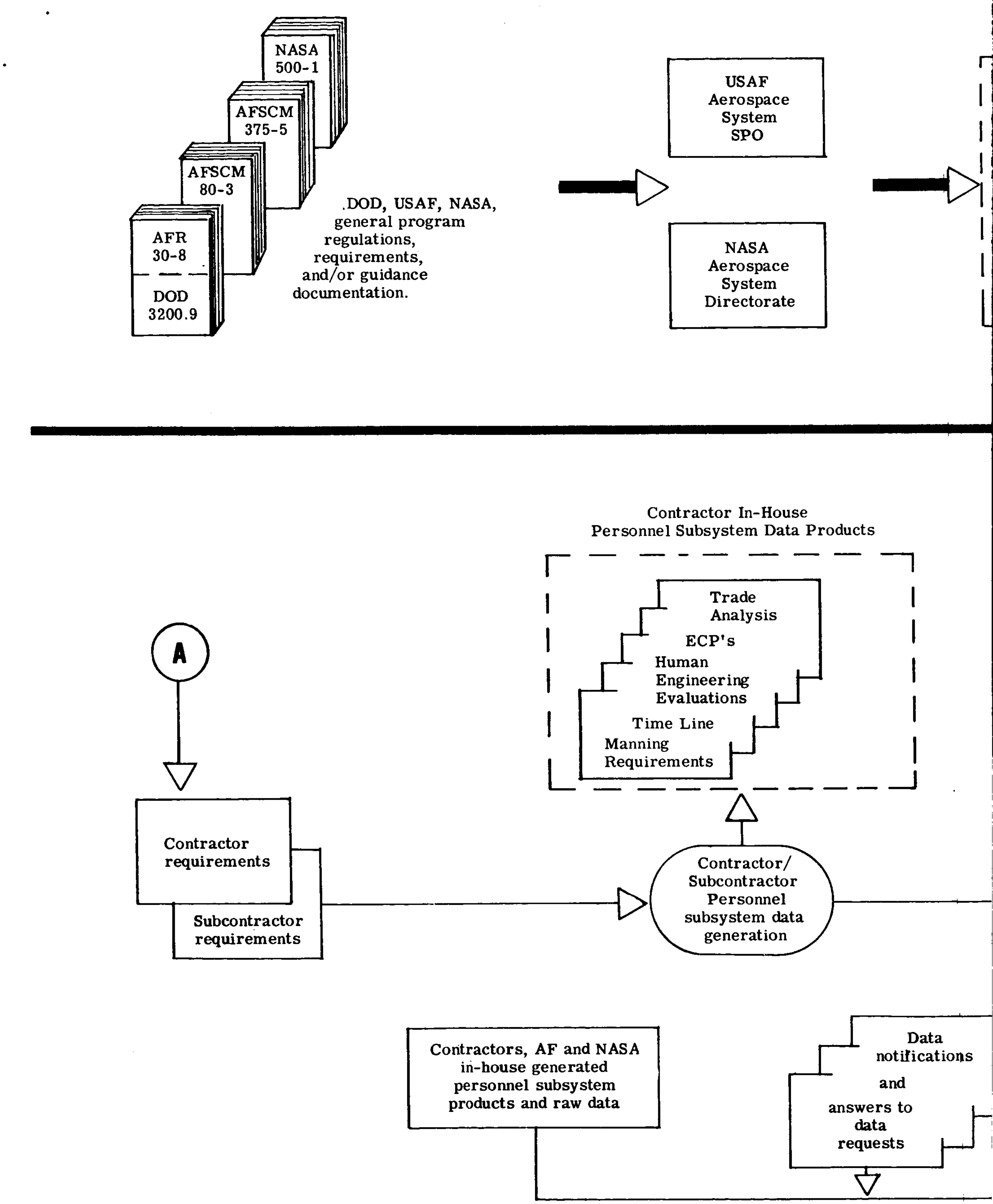
Change Proposal. While the conceptual phase is usally the responsibility of the government, it often requests assistance from contractors. The data generated during this period are usually narrative. Personnel subsystem inputs are concerned with the broad capabilities and limitations of man and his appropriate role in the system functional baseline and gross design requirements. If the mission of a new system requires that a complete study program be initiated, most aspects of the personnel subsystem are given consideration. A rather general task analysis may be performed, predictions are made of the manning requirements for operation and maintenance of the proposed system, and training requirements are identified.

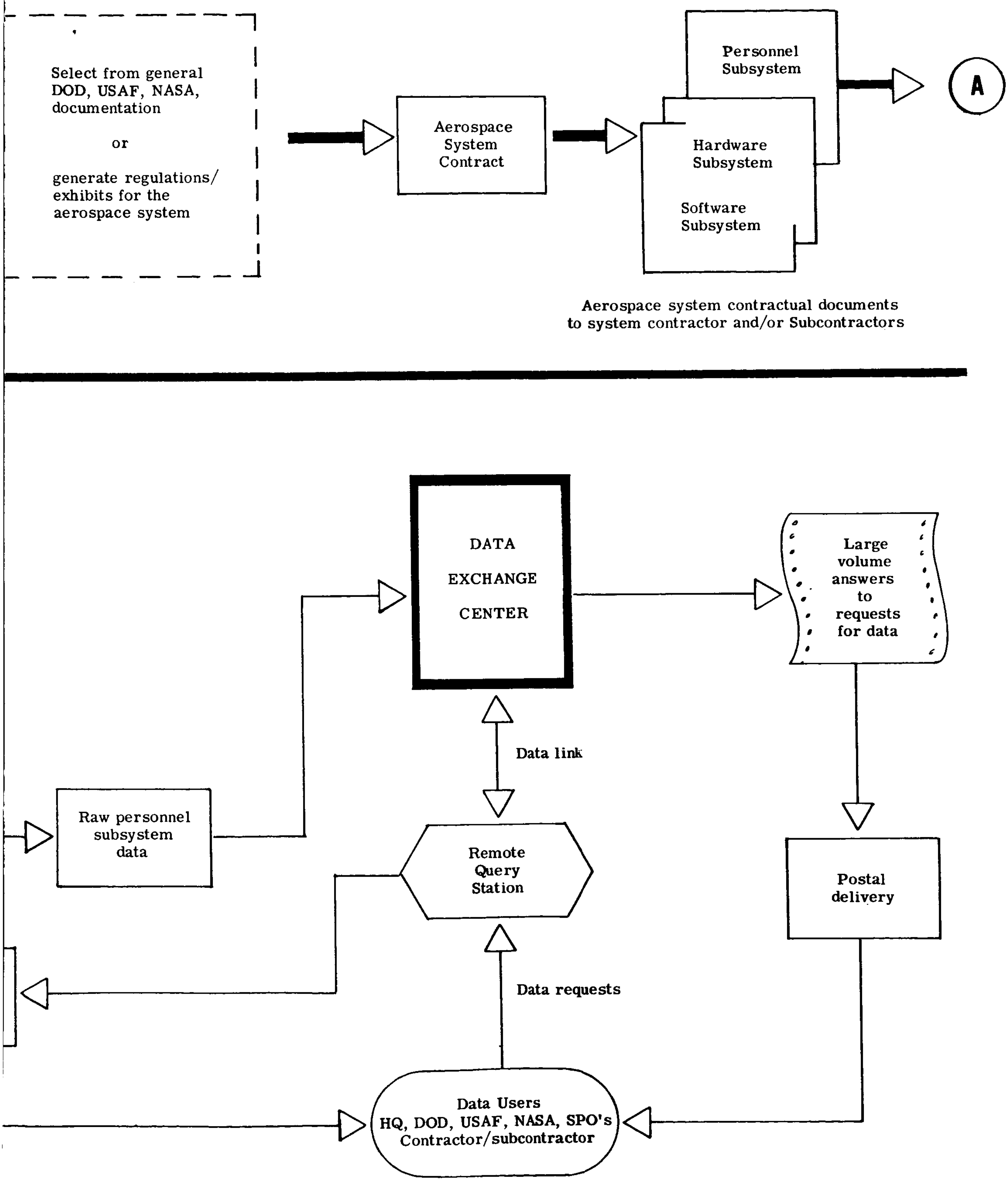
The Definition Phase extends from the approval of the request for competition by higher headquarters to the approval of a proposed "total package" plan. The definition phase is the responsibility of an assigned system managing agency; a USAF System Program Office (SPO) or NASA "Project" Program Office. The SPO cadre are selected at the end of the conceptual phase but are established formally as a management function at the beginning of the definition phase in the System Definition Directive. The SPO draws from a number of regulations, specifications, and manuals, or they may develop special exhibits necessary to define products and methods that the prime contractor and subcontractor are to follow in producing the specified system's hardware and software. The terminal point of the definition phase is the selection of contractors and subcontractors for the initiation of system design and development. During this phase, three general types of personnel subsystem data begin to emerge. These types of data are:

1. Management data required of the contractor to provide the SPO with information necessary for programs and system control
2. Product data that are generated by the contractor and used to support the contractual obligations--documents and data--specified by various system requirements and milestones (AFSCM 375-4, 375-5, and AFSCM/AFLCM 310-1)
3. Test data that are generated to provide a standard against which program progress may be measured, and also to provide a data base upon which the Category I, II, and III testing programs are developed

The Acquisition Phase begins with the awarding of contracts to industry for the development of an aerospace system. The acquisition phase continues until the acceptance by the user of the last operating unit manufactured by the contractor. Early in this phase, most of the effort is expended on the generation of personnel subsystem data. The reduction or reformatting of these data into reports and contractually obligated personnel subsystem products takes place in this phase. This phase is normally divided into development, test, and production. During development, trade-off studies are conducted, personnel tasks are identified in detail, and training and training equipment requirements are documented. Development tests are conducted in this phase on individual components or subsystems to insure maximum operability. The data generated during this period are concerned with evaluation of requirements for personnel and training, human engineering evaluations, and determination of whether the training equipment meets the specified requirements for personnel training.

The Operational Phase is the last phase in a system's life cycle. It begins with the acceptance, by the using organization, of the first operational unit


NASA 500-1
AFSCM 375-5
AFSCM 80-3
AFR 30-8
DOD 3200.9
.DOD, USAF, NASA, general program regulations, requirements, and/or guidance documentation.
USAF Aerospace System SPO
NASA Aerospace System Directorate
Contractor In-House Personnel Subsystem Data Products
Trade Analysis
ECP's
Human Engineering Evaluations
Time Line
Manning Requirements
A
Contractor requirements
Subcontractor requirements
Contractor/ Subcontractor Personnel subsystem data generation
Contractors, AF and NASA in-house generated personnel subsystem products and raw data
Data notifications
and
answers to data requests



4 Figure 1. Operational Data Management System

manufactured by the contractor and continues until the disposition of the system. Category II and III of the Personnel Subsystem Test and Evaluation (PSTE) program are completed and appropriate changes and modifications to the system are accomplished from the test results. The SPO cadre is normally phased out during this period.

Early in the acquisition phase, the bulk of task data is generated and used. The data are highly diverse in that they are generated from, and used by, many different disciplines. The data are also diffuse in that they are applicable to many different problems in system design. A further complicating factor is that the data vary in content and nature within a particular system and across systems. The data may be narrative, highly descriptive, quantitative, pictorial, graphic, or diagrammatic in nature. The same type of data may be described in any of the above mentioned forms, depending on the immediate needs of the human factors specialists. The implication is that any data handling system built to meet the needs of the human factors specialist must be sufficiently flexible to handle the varying types of data.

For purposes of exposition, two groups of personnel are identified as responsible for the personnel subsystem development programs; data generators and data users. There is no clear distinction between the generators and the users. Generators of one kind of data may in turn be users of the same or different kinds of data. For example, training specialists generate training courses and training requirements. In creating those products, they use manning and task or maintenance analysis data which they or others have previously produced. The distinction between generators and users is being made in order to help delineate design requirements for a data handling system. In broad terms, the generators input data into the system and the users receive output data (see figure 1).

During an aerospace system life cycle, generators produce: raw personnel subsystem data which are organized, but unprocessed, into a number of different formats; hard copy data, which include drawings or sketches, layouts, data plots, and graphs; and contractually obligated documents, such as Qualitative and Quantitative Personnel Requirements Information (QQPRI), Training Equipment Planning Information (TEPI), human engineering studies, and Technical Orders (TOs).

Configuration of the Data Exchange Center

As indicated earlier in this section, a need exists to draw the specialist closer to the data generated and used during the various phases of system development and operation. For this discussion, assume that the Data Exchange Center (DEC), illustrated in figure 2, is located at a division of the Air Force Systems Command. Prime system contractors, subcontractors, and government agencies provide the DEC with data generated in support of the Air Force Personnel Subsystem Program. The data, ready for input to the computer, are sent to the DEC by mail. It is envisioned that new data and data updates will be submitted to the center when significant groups of data are developed and not necessarily at design milestones. The DEC contains personnel subsystem data generated in support of several object systems. (In this report, an object system refers to any aerospace system which provides data to the DEC for storage.) This multisystem data store makes it possible

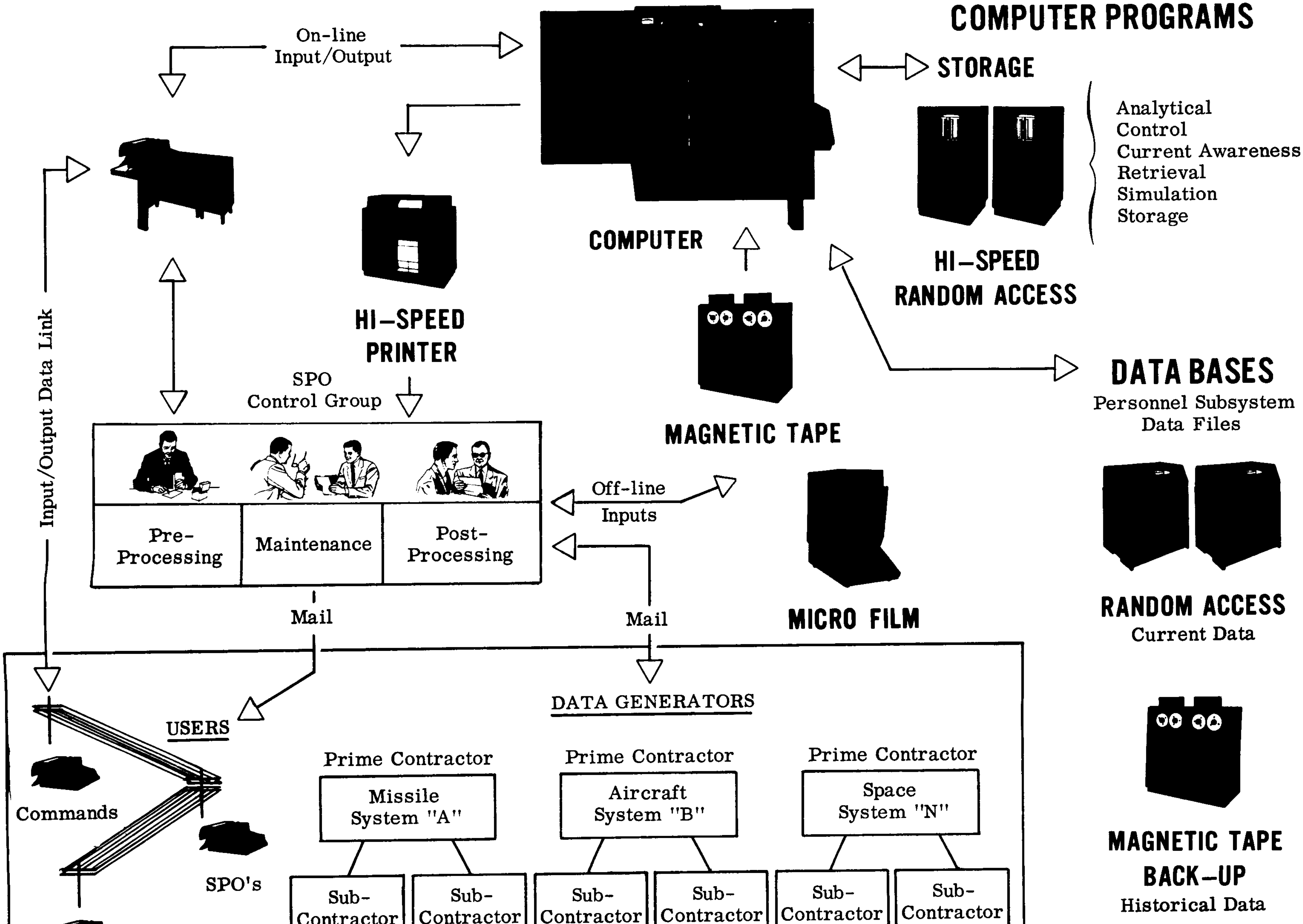


Figure 2.

Data Exchange Center

for the SPO, contractors, and other authorized agencies to access and make comparisons of data from more than one system. For example, suppose that an Air Force agency is developing the requirements for a system which is currently in the conceptual phase. The agency wishes to know if the anticipated manning requirements for the new system are realistic. If off-the-shelf equipment is being contemplated for the new system, the DEC may contain useful and valid information stored during the development and operational phases of other systems. If new equipment is to be developed, the agency has the prerogative of requesting manning information on other equipment with similar requirements. In either case, the data user formulates his request for the desired manning information. The requester has a user's manual available to him with vocabulary and operational procedures guidance. The request is written in a format approaching natural language and sent to the computer by way of an on-line remote query station. After determining if the requester is an authorized user of the DEC, the computer searches the data base for relevant data on the desired systems and performs the sorts which meet the formatting needs of the user. If the selected data are not voluminous, they are transmitted immediately to the user via the remote station; otherwise, they are mailed. Because he is able to request specific facts, the user receives only the information requested and nothing more.

Rapid notifications of current and available data, resulting in decreased reliance on expertise by the specialist, are provided through the application of current awareness techniques. For example, a training specialist at a training center may want to know immediately of any changes in an object system's data that will affect the content of training courses. The training specialist simply establishes an interest profile using object system terminology which indicates his interest in particular data. The profile is placed in the computef at the DEC. During the storing of new data, the computer compares the data with the user's profile and when a match is made, prepares a notification message. The message is transmitted directly to the user via the remote station hook-up. The user interest profiles are maintained and controlled by SPO representatives at the DEC. Changes in the profiles are made rapidly and remain sensitive to the changing interests of those involved in aerospace system development programs.

Two modes of input to the DEC are available. Small amounts of data for storage are transmitted to the computer using the remote query stations. Larger groups of new data, or information not suitable for on-line telecommunication, are mailed to the DEC. These data may be in the form of hard copy documents, punched cards, punched-paper or magnetic tape, illustrations, graphs, charts, etc.

The SPO Control Group, located at the DEC, is an important part of the overall system, and is made up of representatives of each of the SPOs involved in the functioning of the DEC. The group is responsible for collecting, reviewing, verifying, and preparing the input of all incoming data. In addition, the group maintains the data system software and monitors the reproduction and releasability of all outputs. The group is visualized as a semiformal organization that stimulates cross-communication of information relative to common design problems, and provides added control over classified and proprietary information.

The proposed computer is a medium to large-scale machine--core size and processing speed--and performs all arithmetic and logical data operations of searching, sorting, analyzing, etc. It contains those features adaptable to time-sharing and multiprocessing operations for quick response to many on-line users. The computer has a highly flexible growth potential for added storage capabilities. The storage units include high-speed random access bulk storage--disks or drum type--and magnetic tape for backup or off-line storage. The variety of storage units affords greater flexibility in buffering and also allows less current data to be stored in less expensive units. The buffering capability permits simultaneous handling of incoming data with its attendant notification, the processing of requests, and the transfer of current or historical data to either the central processor for a current execution requirement, or to a buffer unit in preparation for the next requirement.

A library of hard copy materials--graphs, pictures, and documents--is included in the DEC. This library contains, in addition to the hard copy received with input data, the physical materials directly related to the data bank--program listings, punched cards, diagnostic routines, and program descriptions. To reduce the physical size or volume, and perhaps make the library more readily accessible, hard copy--pictures, flows, and reports--is stored on microfilm. An index of the microfilm contents is inserted into a master index stored in computer data bank. Thus, as data are retrieved from computer storage, cross references are included for retrieval of supplementary data from the microfilm storage.

The remote query stations--keyboard entry devices such as teletype--are directly linked to the SPO/contractor facilities and the DEC. Generally, any number of units may be included in the network with use of the data transmission lines based upon a time-shared design. On-line units are also located in the DEC, enabling the SPO Control Group to monitor all traffic as well as to access the computer for their own purposes.

A wide variety of outputs such as synthesized reports, summaries, and results of analytical and simulation runs are available from the system. In all cases, the outputs are in natural language and ready for immediate use.

The data base structure and content of the operational system primarily involve the storage of behavioral data at the human action level. Input data are comprised of those data items that support and better describe the action. Thus, the human action becomes a pivot around which many of the data classes are referenced. Complex human behaviors, i.e., behaviors composed of more than one action and requiring a combination of supporting data--hardware descriptions, time parameters, and skill levels--increase the problems of interpreting the qualitative state of the data. The qualitative data, such as the written performance descriptions and equipment nomenclature, generally require mental evaluation to analyze the content. The quantitative data, however, contain measurable and highly structured values. These values can be directly subjected to analysis procedures stored in the computer. The final evaluation of the results still rests with the user.

The operational system includes a large data bank and services many users. The system's capability to reduce, tabulate, and synthesize data can be used extensively. The results of such processing may be very significant to various users. Therefore, the system will retain the results of such analytic operations, including simulation modeling, in historical data files. Lengthy results are stored external to the computer in documents or microfilm. The location and other index information for these external files are stored in the computer and provided on request.

PRECEDING PAGE BLANK NOT FILMED.

SECTION II

RESEARCH PROBLEM AND APPROACH

PROBLEM

The operational system concept, described in Section I, was developed following the first year of a four-year effort. That initial effort was dedicated to the study of: (1) current practices, and (2) the feasibility of using computer techniques for handling human factors task information (Hannah, et al, 1965; Whiteman, 1965; Hannah and Reed, 1965). It was based on the premise that the complexity of modern aerospace systems has led to the generation of increasingly large and unwieldy amounts of human factors data. This generation of data has: (1) decreased the impact of data on system design and development decisions, (2) increased the reliance on expertise when existing data are either not known to exist or are inaccessible, (3) led to the inadvertent duplication of data, and (4) led to the scattering of costly information. An assumption was made that these problems would be partially alleviated through the use of an automated system for storage, processing, and retrieval of aerospace system data.

It was recognized early in the overall research program that techniques for handling data must be developed in context with their total operative environment. The design and development of data handling techniques are contingent upon the nature of the data to be handled and the needs of the data generators and users. Thus, the first year effort was directed toward the following objectives:

- Identifying the representative groups of technical and professional specialists involved in the generation and use of human factors data
- Identifying the types and classes of data generated and used by the government and contractors in system design, development, test, and operation
- Identifying relationships among data categories, characteristics with regard to the creation and use of the data, phases in which the data are generated and used, and final data products
- Assessing the types of current and potential uses of computers for handling system data
- Assessing the current and desired data retrieval times

On the basis of information gathered from generators and users of human factors information, Hannah et al (1965) recommended that a computerized system for handling human factors information be capable of performing the following functions:

- <u>Supply data, including task analysis, manning and training requirements, for any part of a system which has been duplicated in past systems or on an experimental basis</u> This recommendation simply means that a need exists to provide data on past systems for design and development decisions in new systems.

- Provide rapid access to information concerning facilities, training aids, aerospace ground equipment, and trained personnel necessary to design, develop, operate, and maintain a system This function should be available at any time during the life cycle of a system.

- Simulate any proposed system, or portion thereof, at any time during the system life cycle and at various levels of detail

- Be amenable to frequent updating in order to provide the most recent information to the user

- Generate replies to specific queries in the event that the required data is available in handbooks or when the latest data are required Also, the data system should provide the user with basic analytical tools to aid in the solution of system-specific problems.

- Provide data, throughout system development, in support of personnel subsystem requirements, such as QQPRI, TEPI, and TOs

OPERATIONAL SYSTEM REQUIREMENTS FOR HANDLING DATA

An extensive review of the recommendations, information, and experiences, resulting from the first year research effort, led to the determination of fourteen requirements for the data handling system. The requirements provided the guidelines for the research discussed in the remainder of this report. These requirements are:

- The system must be oriented to user requirements. It must satisfy the needs of aerospace scientists and engineers as well as serve management needs for data.

- The system must provide for the storage, updating, and retrieval of human factors task data.* The data should be indexed to permit retrieval based on several reference points.

---

* Human factors task data, for the purpose of this research, are defined as including qualitative and quantitative data about human performance of operator and maintenance personnel. These data will emphasize the behavioral and man-machine system data of human engineering, human learning, and training. For example: (1) the demands that the system, man, or the situation make upon one another e.g., the working environment, time criticality, and performance accuracy; (2) the discrete task information such as expected or required task and skill parameters for fixed and/or variable task procedures; and (3) the applications of skills within system mission segments and time base. Skills pertain to such functions as detecting and processing information, monitoring, communicating with and directing machines or humans, command or decision making, feedback, and self-alignment or adjustment. Human factors task data are further defined as not including: (1) data found in technical guides, manuals, standards, "cookbooks," etc.; (2) system management oriented data such as required reports, available or desired training devices, schedules, and contractor and SPO responsibilities.

- The data system must be responsive to current Air Force and NASA system and data management concepts. Compatibility, rather than dependability, with the system engineering process is emphasized.

- The system must provide simplicity of use by accepting and outputting data in a form approaching user terminology. This means that all inputs and outputs must be immediately interpretable by the user. This includes all data, whether they are qualitative or quantitative in nature.

- The user of the data system must have easy access to the stored data through the use of a user-oriented query language. The terminology interpreted by the system must be compatible with the language employed by the user in his system-specific activity.

- Provisions must be made for external storage of data that cannot be coded economically for computer storage. Where applicable, cross-indexed data should be stored in the computer for referencing information filed externally to the computer, e.g., documents, pictures, and graphs.

- The data bank structure must be flexible enough to allow for future expansion and inclusion of additional data elements--categories of information. This flexibility will allow for changes in data concepts and new system requirements and avoid major changes in the structure of the data base.

- The data bank must be capable of frequent updating while retaining selected data for such uses as design trend analysis. The system must be capable of purging unwanted historical information. The updating capability of the system should allow for the storage of information generated in support of on-going phases of an aerospace system life cycle.

- The data system must be capable of retrieving similar information generated in support of different aerospace systems. This capability will allow maximum use of data in making design decisions for new systems.

- The data system must be capable of selectively retrieving data elements by qualifying them with other data elements. The user is permitted to receive only the data he needs and nothing else.

- The data system must have the capability of protecting data having security classification and/or proprietary status.

- The data system must provide the capability of processing retrieved data through the use of analytic programs and simulation models with a minimum of human intervention.

- The system must provide the user freedom in specifying the format of outputs.

- The system must provide for current awareness notifications to qualified users, in response to interest profiles. A notification is defined as a statement that data meeting the requirements of a qualified user now exist in storage.

EVOLUTION OF THE RESEARCH

The current research effort is oriented to investigating problems and conducting experiments using computerized techniques for improving current practices of handling human factors task data. Experimentation involves the manual testing and trial program operations performed during the design and development stages of the pilot research program*. The results of the experiments will help determine the configuration and performance of the computer programs, and other pilot components such as vocabulary, and profiles. The experiments will apply existing state-of-the-art computer systems technology. The application of the techniques employed will result in a pilot configuration which presents solutions to the overall problem--improving data handling procedures in Air Force and NASA system programs.

The approach described below provides the orienting principal theme for activating and controlling the direction of the several research activities. Seven stages of evolution are depicted in figure 3. The contingency of the stages are such that research and development carried out in each stage are dependent upon, and conditioned by, the results of the previous stages.

This report covers the work completed for stages one through four. Completion of the research covered in the fourth stage, and the work to be done in stages five through seven, are to be carried out in follow-on efforts.

First Stage - Concept of an Operational Data Management System

Research into the improvement of data handling procedures should be conducted with an application in mind. An application is the computer-driven data management system described in Section I of this report. It advances the concept of automating Air Force and NASA human factors data in aerospace system development programs. The concept presents an overview of the entire data system and it provides a method by which the various areas of the research can be guided.

Second Stage - Concept of a Research Pilot

A concept of the computer software functions in the research pilot further examplifies the inner workings of the operational system concept. The pilot concept narrows down the research and development aspects of this program. It also focuses more clearly on the functional features to be investigated, i.e., user procedures, program techniques, and characteristics of the system. This concept is presented in Section III of this report.

Third Stage - Select Research Areas

The essential research activities to be performed are identified in this stage. The concepts provided in stages one and two, and the system requirements (see pages 12 and 13) form the basis for selecting those areas of system techniques to be investigated. The selection of significant and fundamental subject areas is made with the intention of augmenting the research with other available technology. For example, investigation into analytical techniques can draw from existing applications even though they may not be currently used in personnel subsystem problems. The areas that have been selected are identified in Section IV.

---

* Research pilot program refers to a study program, henceforth called the research pilot, or simply, the pilot.

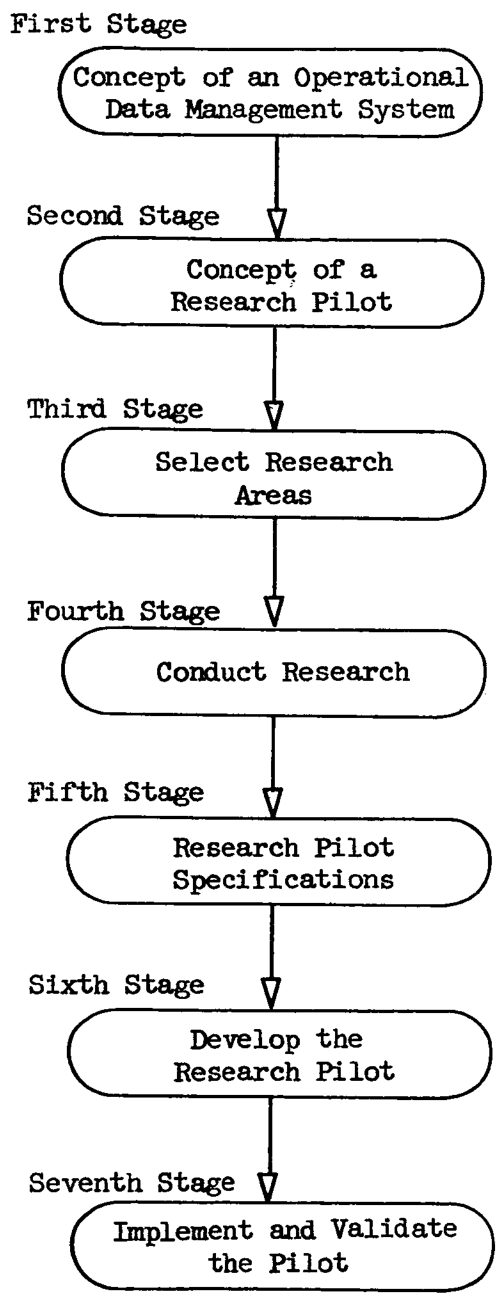


Figure 3. Stages of Pilot Evolution

Fourth Stage - Conduct Research

The activity leading up to formulating an initial, yet detailed, pilot design takes place during this stage. This activity includes describing particular problems and objectives for each research area, as well as planning and carrying out the objectives of the research. The results gained during this stage will provide the basis for the research pilot design. The nature of the operational system concept and its man/machine functions will also become more clearly described as problems are explored. Modifications to the concept will be made during the course of the study.

Fifth Stage - Research Pilot Specifications

The fifth step in the research program provides the design for developing and testing the pilot software and related procedures. A set of design specifications will be written after the research results are studied and conclusions are drawn regarding the necessary software, detailed characteristics of the data that will be used, and other design inputs. The design specifications--including the vocabulary, profiles, and computer programs--will describe the techniques, functional interface, and the approach to be used in pilot development.

Sixth Stage - Develop the Research Pilot

This stage of the research covers the period during which the pilot is developed and tested. Development procedures will include the generation of a test data base, developing and checking out the computer programs, and modifying both according to test results. Vocabularies and user interest profiles which have been generated are included as developmental projects.

The entire activity of subjecting pilot components to test procedures is experimental. Interim test results will cause changes to the pilot design. For example, attempts to process representative user queries may indicate the need to reclassify data categories in order to respond to more efficient retrieval requirements, or even to more easily formulated user data requests.

Seventh Stage - Implement and Validate the Pilot

The final stage of the research program will determine the feasibility for building an operational system. A program of field testing and verification of the pilot will be carried out during this seventh stage.

The pilot will be subjected to realistic situations in order to validate the techniques applied during its operations. These situations will involve using the test data base, made up with as near real data as possible, as a means of running "live" queries. This implies that human factors specialists, engaged in present aerospace systems development programs, will be used to formulate queries, including those processing functions available. This approach will provide objective evaluation for establishing the feasibility of operational system capabilities such as demonstrated by the pilot.

SECTION III

CONCEPT OF A RESEARCH PILOT

The pilot concept is developed, basically, to focus more clearly on the operational aspects of the pilot software functions. The configuration and techniques of the pilot software are based on principles of information systems design and also the operational requirements (see pages 12 and 13). This concept provides a starting point for the investigation of the pilot and helps to identify those areas of the research program that will be studied in detail. The actual design of the software is a matter for the research to determine. Depicted in figure 4 is a diagram of the conceptualized flow of the major computer program operations. A narrative description of the workings of each function is given below. Each subtopic--e.g., Editing--is directly related to the diagram blocks.

The functions are grouped into three processing cycles--storage, current awareness, and retrieval. Analytic functions are included in the retrieval cycle. A fourth major area, system control, is described but is not a part of the diagram.

STORAGE CYCLE

Editing

Data that are input to the pilot system pass through the SPO Control Group for review and then are directed into the computer for subsequent storage. The task data are normally the raw, or unrefined data that are generated by system contractors during the life cycle of aerospace systems. Manning allocations, training requirements, equipment designators and characteristics, performance descriptions, and time allocations are types of input data.

The data are received from the contractor agencies in coded form, i.e., indexed according to the guidelines laid out in the system's handbook. This handbook--most often referred to as a user's manual--contains instructions for indexing, operating, retrieving, etc. The SPO preprocessing unit reviews the data, checks for security and other required identification, and feeds the data into the computer. The initial program function--Editing--checks the data against system standards to determine if they are acceptable for storage. The checking process will accept or reject the data according to criteria controls stored in the program. These controls include the system descriptions for data formats, element and item location, character make-up, and recognizable terminology. This function also provides a back-up review for the checks that determine if the data are sent from a qualified source, and if appropriate security levels are maintained.

Prepare for Storage

Data sent to the system for storage are usually organized and described according to the individual generator's requirements. A particular generator may need to express newly created data, e.g., specialty codes, in terms of the training requirements involved. The particular data arrangement, expressing training concepts, will not meet the requirements for all users. A user of the same data may need to form his request with regard to the mission involved or other system functions. This means that the data system must be able to assign the various elements of data to a framework that allows retrieval of data using more than one prime reference point. Therefore, the incoming data will need to be rearranged and cross-linked to correspond to the organization of the data in the various data bases, which are structured to meet the various requirements. This activity will be done by the computer using stored schematics, which are auxiliary files that contain data linkage, pointers to hierarchical levels, and other data structure information. The organization takes into account all reference points which allow data items to be retrieved with respect to other data items. One of the intentions of the storage scheme is to eliminate as much duplication in the files as is feasible. This reduces the size of the data bank and facilitates search time during retrieval. For example, where task descriptions involve identical equipment identifiers, only one file entry for the equipment data is necessary if that entry can be referenced to each appropriate task. One of the features of this structure is providing dynamic cross-indexing among data classes to meet changing user requirements. Should users change a requirement for accessing data, the change is carried out by modifying the data linkage involved, rather than by restructuring whole files or data bases. This ability to modify data relationships is a significant characteristic of data management systems. It is considered an important facet of the pilot concept.

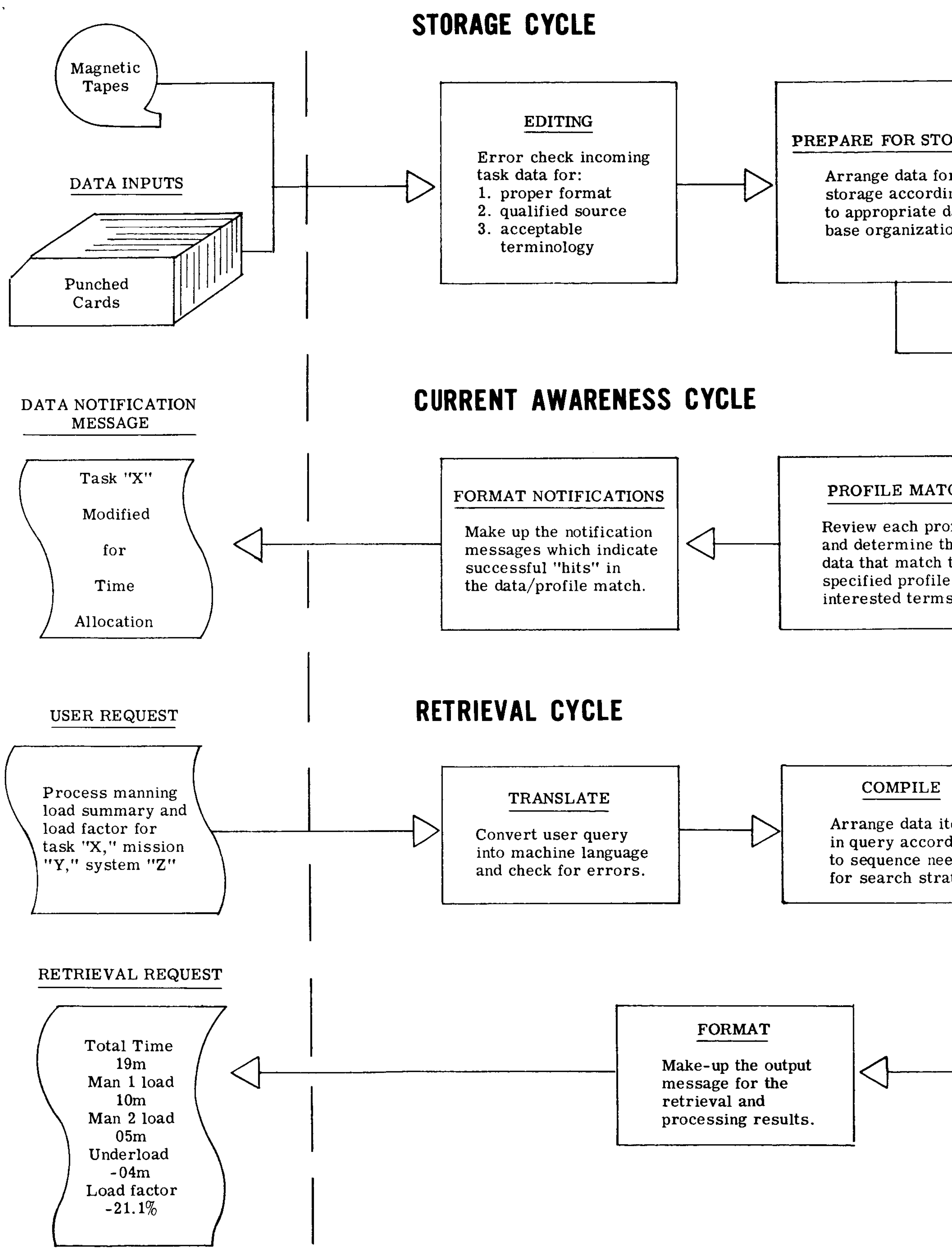
Storing

The task data are merged by the storing function into the appropriate data base files in accordance with the arrangement of the data provided by the Prepare for Storage function. The data files make up the total store of object system data as they reside in the computer storage devices. Each object system's data are stored as a separate data base. Data are stored by classes within each base--e.g., mission information is stored as one group, and performance information is subdivided into another group. In addition to the object system data and auxiliary index files, other data are stored, such as user profile terms--see "Current Awareness" (below)--and system "dictionaries." A dictionary is a parameterized file that provides automatic control--data editing, formatting, file characteristics, etc.

The system uses modular storage units that handle a high data growth rate. The ability to access specific data items is augmented by the use of storage devices that provide a capability for direct-path search strategies, thus minimizing time consuming serial retrieval runs. Random-access storage equipment--disks, drums, and bulk core-memories--provides this capability.

CURRENT AWARENESS CYCLE

A system requirement (see page 13) is that interested users be notified of incoming data. The Current Awareness cycle meets this requirement by assembling and matching new data against user profiles. The profiles contain lists of user interest terms, taken from the system vocabulary, which express the various users' interests in particular data. A profile can indicate interest in a class of data as well as specific items. The organization of the profiles is similar to the organization of the data base. Therefore, the prearrangement of data for system files in the prepare for storage function directly supports the current awareness cycle.



STORAGE CYCLE
Magnetic Tapes
DATA INPUTS
Punched Cards
EDITING
Error check incoming task data for:
1. proper format
2. qualified source
3. acceptable terminology
PREPARE FOR STO
Arrange data fo
storage accordi
to appropriate d
base organizatio
CURRENT AWARENESS CYCLE
DATA NOTIFICATION MESSAGE
Task "X" Modified for Time Allocation
FORMAT NOTIFICATIONS
Make up the notification messages which indicate successful "hits" in the data/profile match.
PROFILE MAT
Review each pro
and determine th
data that match t
specified profile
interested terms
RETRIEVAL CYCLE
USER REQUEST
Process manning load summary and load factor for task "X," mission "Y," system "Z"
TRANSLATE
Convert user query into machine language and check for errors.
COMPILE
Arrange data it
in query accord
to sequence nee
for search stra
RETRIEVAL REQUEST
Total Time 19m
Man 1 load 10m
Man 2 load 05m
Underload -04m
Load factor -21.1%
FORMAT
Make-up the output message for the retrieval and processing results.

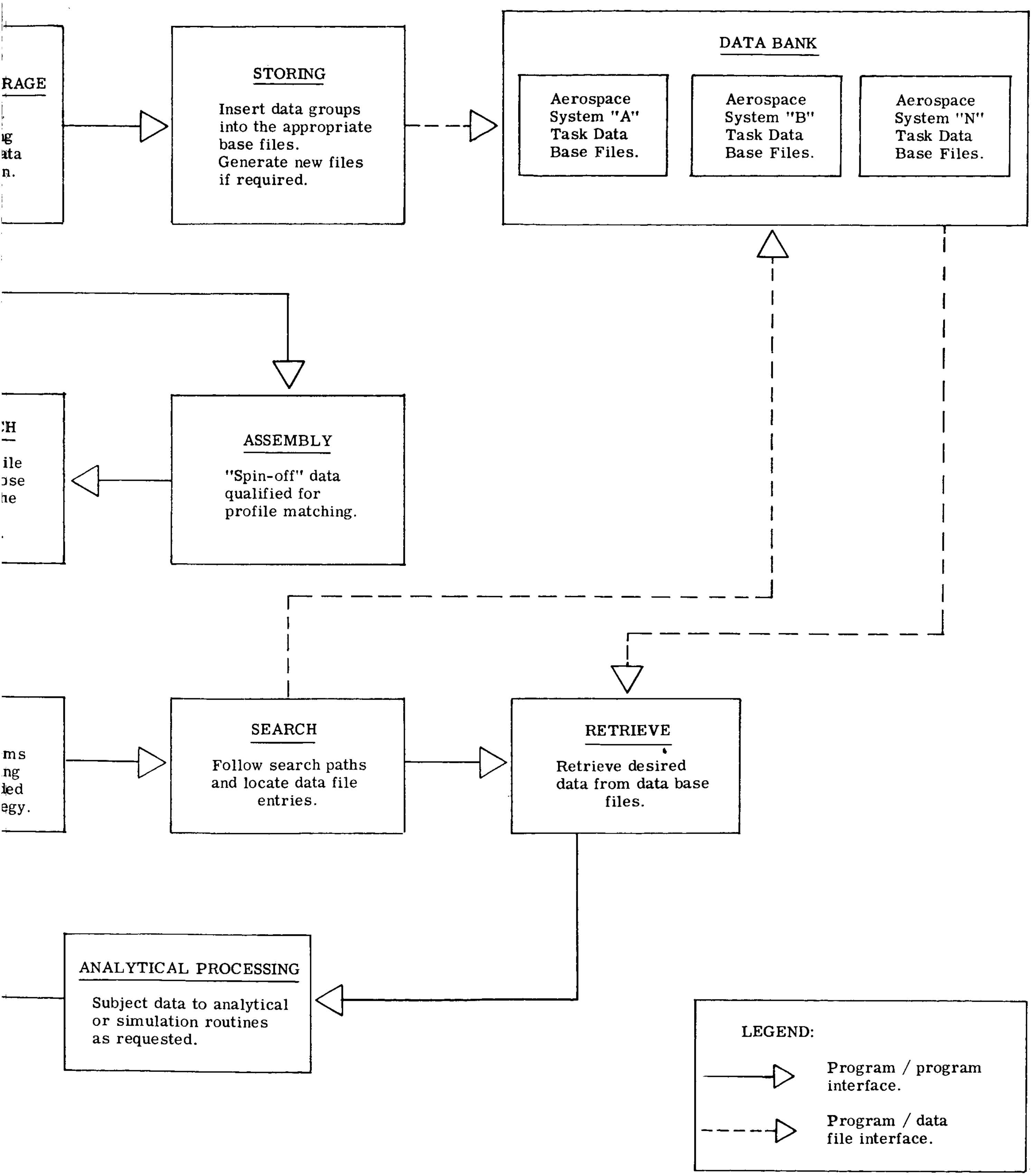


Figure 4. Functional Flow of Major Pilot Programs

Profile Match

This function matches data identifiers with the individual profile terms. The identifiers are tags for data items and data groups. As incoming data are matched against the user interest terms, notifications are formed for those profiles that contain the matched terms. A further check is made by the program to determine if security requirements are met. Those notifications not passing the test are dropped from further processing.

Format Notifications

This function formats the notifications into individual user messages, and sends them via the remote transmission lines. A message consists of: (1) user's profile identity; (2) the notification of the matching data identifiers; (3) security classification of the actual data; and (4) data and user's location.

RETRIEVAL CYCLE

The user-orientation of the system allows a person to directly access the data store without having to consult a system specialist to formulate his retrieval request. A user's manual is sufficient guidance for communicating with the data system.

The system data are controlled by a vocabulary that is tailored for human factors specialists and other users of the system. It contains terminology that expresses the data and their various concepts as they are used in the object systems. A thesaurus technique which classifies and associates vocabulary terms, provides the user with the capability to quickly reference the concepts he needs to form his request for data. This vocabulary is a part of the user's manual.

Figure 4 shows an example of a request message being input from a remote query station. The data request sheet indicates what the user might type in for direct transmission to the computer. His identification, source, and the date would be included in the message and used by the computer for making the reply to the request.

In the example, the requester asks for a task manning load summary and the load factor for task X. He provides the desired object system identifier Z, and other qualifying information. The input is converted on-line into machine format and sent through the retrieval cycle. Following the execution of the request, the result is sent on-line to the requester console where it is printed out.

The retrieval cycle is made up of two subsystems which interface so directly that they operate as one cycle. The first subsystem retrieves data from the storage files; the other does all required processing of the retrieved data, using analytic and simulation modeling programs.

The functions of the retrieval cycle are explained next.

Translate

The translation function converts the request into machine language. The function also performs error-checks for incorrect terminology, unqualified users, and improperly formatted requests.

Compile

The request is organized into a "job operation." The data identifiers are hierarchically arranged into search keys according to the logical associations of the data being requested. The hierarchical arrangement for the sample request in figure 4 is given below. The numbers indicate the level of specificity required in the search strategy. Thus, a qualifying technique is implied whereby the search is first keyed on system Z, which qualifies the specific mission. Mission Y qualifies the task, and so on, until the desired "time" value is located.

(1) System Z

  (2) Mission Phase Y

    (2) Task X

      (4) Manning allocation value

      (4) Time allocation value

If there are special qualifying factors involved, they also become part of the job operation. For example, in a request for a job load summary, which is associated with different time variables, the relative start time can be used as a special qualifier to select the precise allocation information desired.

Directions for processing the data are also a part of the job operation. In the example, in figure 4, the request for the summary and load factor causes the automatic selection and execution of the required analytical functions. The execution sequence of programs--translate routine, compile routine, load summary routine, and load factor routine--is implied in the directions provided in the request.

Search

The search function locates the particular file entries in the data base which contain the desired data. These routines use the stored auxiliary files that provide the cross-indexing to all data base files. Search paths are set up that bypass irrelevant entries to speed up search time. The selection of the paths are determined by the arrangement of the search keys identified in the "compile" function.

Retrieve

The retrieval routines, which pull the requested data from the storage files, function as direct companion routines to the search routines. After the search routines have located the desired file entry, the particular values are retrieved and set up for analytical processing and/or final formatting for the retrieval reply. The decision whether to do processing or not is based upon whether only raw data are desired--i.e., a request that does not require interim processing--or if a data product is asked for, e.g., manning load summary.

This set of routines also determines whether or not a request can be satisfied. If not, the request is voided and the user is informed of the failure. Several factors can lead to a voided request, such as data requested that are not in storage, and security or proprietary restrictions.

Analytical Processing

A pool of analytical routines is available to the user to further assist him in his need to effectively and efficiently handle task data. The research will determine the type and function of the routines which can be implemented in the system and used by the human factors analyst. Mathematical simulation models, a function of this subsystem, are selected and operated according to the specifications provided by the user in his request. These routines process the data supplied by the retrieval routines, or can even work on supplementary data supplied in the request. The user has several options available, such as creating special routines by using on-line program building techniques. He can request that particular routines be executed or he may simply ask for a product, in which case the routines are selected for him.

Format

This function arranges the retrieval result into a specified user format that meets the requirements of each user. If special instructions are provided in the request for arranging the data, they are carried out by this function in the message formatting. The data values are tagged and the requester's identification, location, and date of message are included in the output message.

CONTROL CYCLE

The concept of the system describes a capability of executing a variety of computer program functions and maintaining near-simultaneous responses to requests. In an operational environment, where many users and generators are tied into the system, methods for maintaining an efficient system operation are needed. The system control cycle permits the other system cycles--storage, current awareness, and retrieval--to operate as an integrated unit. Two functions of the control cycle are: (1) program control, and (2) manual control.

Program Control

Automatic program control is carried out by supervisory programs that monitor and control the performance of each cycle as they are executed in the computer. These programs assure that the sharing of the computer by the different cycles is done efficiently. The cycles operate on a priority basis, which cause more important functions to precede lesser functions. Data requests, for example, can take precedence over the storing of new data. In this way, users can access the data bank as the need arises, rather than having to wait until the computer completes a cycle operation. A time-sharing capability is built into the system which allows several users to enter requests on-line at the same time, again without waiting for the completion of processing of an individual user's data request. Many current generation computers are being designed and built to handle a time-shared and program-shared system, thus enhancing the capability of serving a number of users utilizing common program operations.

Manual Control

The system is designed to relieve the user and generator from being directly involved with the execution of the programs. But certain decisions cannot be made automatically by the programs. Unforeseen processing priorities will occur and the need will arise to add or delete object systems or user profiles, etc. Also, and most important, provisions are necessary to assure that communication networks and system maintenance activities are supervised. These supervisory functions are handled by the SPO Control Group responsible for the operation of the entire data system.

SUMMARY OF THE PILOT CONCEPT

The type of pilot software conceptualized is a large, user-oriented, factual retrieval and processing system. The system uses a controlled vocabulary to govern the diversity of data terminology, increase user-generator-computer interface, and act as a focal point for system growth, utility, and other management considerations. The vocabulary also increases the capability of handling task data in a single computer store for several aerospace systems.

The system is user-oriented. This means the user can communicate directly with the computer and is able to access the data bank with little specialized training or use of system specialists. Each user receives quick response to his request, even though he shares the computer with others. The data are indexed in such a way that the user can make requests that meet his requirements without being restricted to a predetermined "line of thinking" that may not match his particular data needs.

The system provides factual retrieval. The system is able to store, and make available upon request, data facts within several levels of specificity. Task data are comprised of many differing and individual data items. Each item has its own concept and yet different concepts are formed when items are merged. These varying concepts form the basis for factual data. A user is able to refer to the specific information he wants. This eases the burden of time and increases his work capability. It is this characteristic which is most valuable to people who use data management systems. When users are forced to wade through a mass of data to locate a relatively small fact, they are discouraged and rely more heavily upon their subjective thinking.

The system has a processing capability. The human factors specialists, human engineers, training specialists, procurement analysts, SPO monitors, and other users of the system are kept informed of new data by user profile matching techniques. There are also a number of analytical techniques available to them for producing data products.

SECTION IV

SELECTION OF THE RESEARCH AREAS

The concept of the operational system is quite broad. The complete system involves hardware, communication networks, control points, maintenance and operating functions, and advanced system software including computer programs, data vocabulary, indexing procedures, and user manuals. Human functions are also involved--users, generators, operators, and controllers. The design and development of such a system requires a wide range of activities and considerations.

The design of an information system is based upon its functions, that is, what it is intended to do. The system requirements (see pages 12 and 13) lay out the conditions for handling task data in a computer system. Included in the requirements are stipulations that affect the manner in which the conditions are met--user orientation, security regulations, consideration of future applications, factual selectivity, standardization, and management needs.

The efforts put forth in this study are focused on five areas of the research. These five areas are considered fundamental, from the standpoint of studying the problems and proposing solutions, and in conducting tests to support the research. Cognizance of other research and development activities in the general field of information systems is also a factor in the selection of areas for research.

The five areas represent distinct topics for study. Each area has been assigned individual problems to investigate. There also exists a great interdependency among the areas. Problems in one research area can very likely have an impact on decisions made in other areas. Examples of this interdependency were identified early in the research program. Some examples are:

- The manner in which system users wish to identify and reference data influences the vocabulary terminology used and the storage and retrieval techniques needed.

- The organization and content of the vocabulary will directly impact with the design of user profiles.

- The determination of the nature of the data, including its qualitative and quantitative characteristics, is basic to the application of analytic tools, as well as being basic to applying factual retrieval and current awareness.

The five areas selected and discussed in the following sections of the report are presented below. The diagram in figure 5 shows the interface of each area in the fourth stage of the research program.

- Analysis of Human Factors Task Data, Data Relationships, and Classification Schemes

  One of the most important considerations in applying information system techniques to data problems is the determination of the characteristics

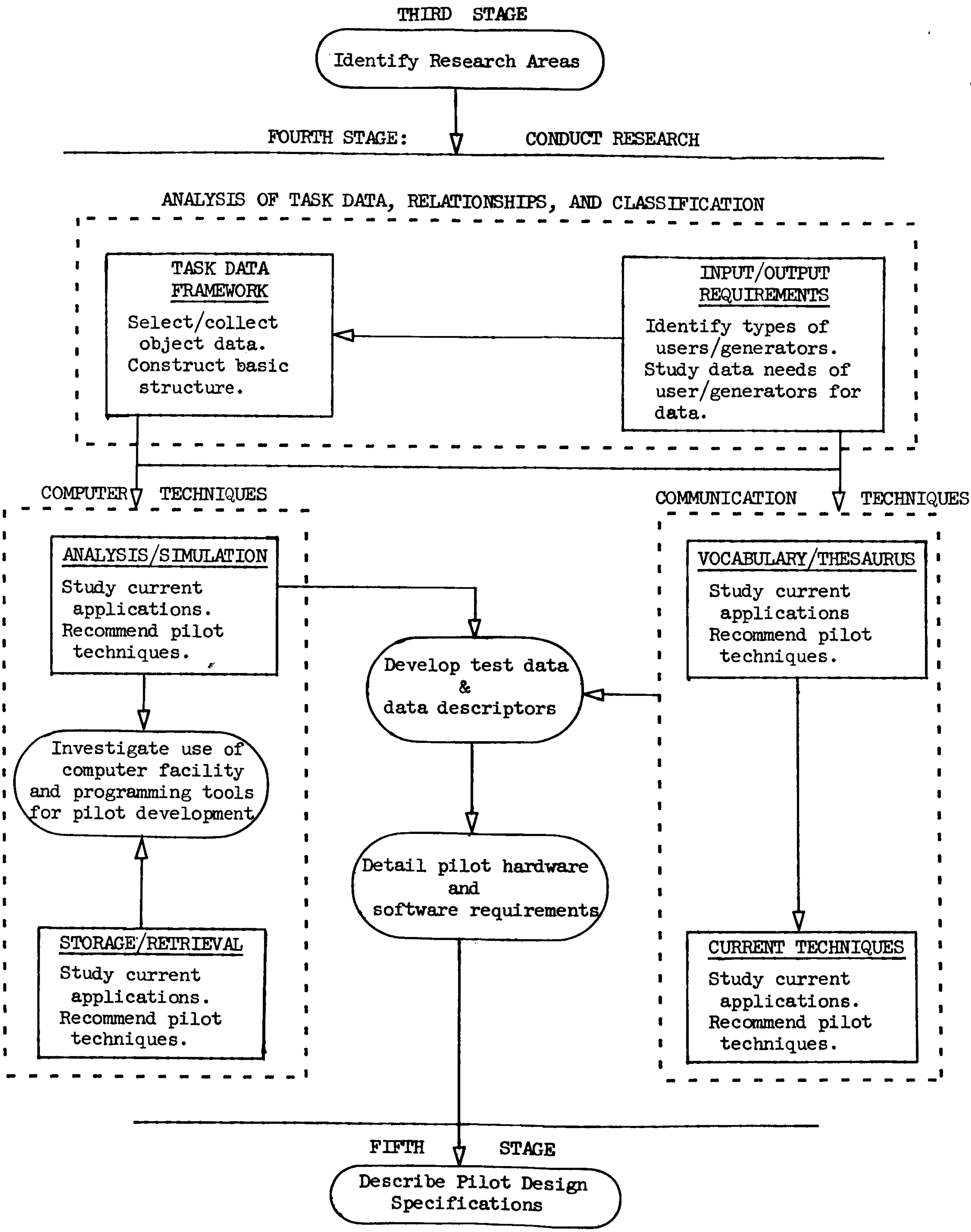


Figure 5. Interface of Research Areas for Stage Four

of the data to be handled by the system. The factors involved in gaining a complete understanding of the data include its diversity, application, environment, content, life cycle, and significant phases in its generation and usage. The personnel involved in data handling are also important factors--their problems, response times, tools, scheduling, and products. Therefore, an analysis of the data, and the people directly concerned with it, is needed (see Section V).

- Vocabulary and Thesaurus Techniques Applied to Human Factors Task Data

  A capability that increases the effectiveness of communication among man/machine/software functions is highly desirable in computer systems. Vocabularies in today's systems cannot be used as off-the-shelf components; they must be tailored to the environment of the system. In this research program, a special consideration, standardization, is apparent in studying task data terminology. Not only is the problem apparent for treating multisystem data, but in providing for multiusers. A data vocabulary is considered essential to this information system and requires careful study in its applications (see Section VI).

- Computer Storing and Retrieval of Human Factors Task Data

  Particular problems, identified in the first part of this research program (Hannah et al, 1965), that are of concern to users of task data are the handling of large amounts of data, dealing with scattered sources, and drawing from previous and current systems' experience. The recommendations (see pages 11 and 12) call for a store of information, mutually available to those who require such information, and in a form that expedites the use of that information. These recommendations lead to the conclusion that a factual storage and retrieval capability is needed which performs on an up-to-date basis (see Section VII).

- Analytical and Simulation Modeling Techniques Applied to Human Factors Task Data

  Techniques employed by analysts in refining task data into useful products are, many times, the result of scientific analysis and modeling procedures. Such procedures are highly amenable to computer applications. It follows that if raw quantitative data are easily accessible in a computer store, that various processing techniques can be immediately and directly applied to refine the data into the required products (see Section VIII).

- Current Awareness Techniques Applied to Human Factors Task Data

  A special problem pointed out in the Hannah et al study (1965) involves the inability of analysts to keep up with the fast pace of data generation. This is particularly true when the data are scattered, or if channels providing awareness are inefficient. Inefficiency is almost always present when several separate organizations are involved. The requirement that the user of task data be immediately aware of its generation, when it is pertinent to his interests, can be facilitated by the functioning of a common data store. The problems of assuring awareness can be lessened by setting up a major automated control point that acts as a disseminator, which provides notifications of pertinent data to interested people. The operational system can act as the control point (see Section IX).

SECTION V

ANALYSIS OF HUMAN FACTORS TASK DATA, DATA RELATIONSHIPS, AND CLASSIFICATION SCHEMES

PROBLEM

In the absence of comprehensive, universally accepted descriptions, it is difficult and, at best, arbitrary to conduct a research analysis on all personnel subsystem data. In addition, the volume of personnel subsystem data, even on a single system, is unwieldy and would be difficult to analyze. Fortunately, all personnel subsystem data need not be used in this study. Hannah et al (1965) concluded that task analysis not only forms the bulk of personnel subsystem data but it also influences the decisions made in the generation of many personnel subsystem products such as human engineering studies, QQPRI, and TEPI. Furthermore, the type--numerical and narrative--formatting structure--graphs, pictorial, and fixed formats--are representative of the total personnel subsystem data generated in support of systems. Based on these conclusions, task data are selected as the sample information for the research.

Most human factors analysts are acquainted with a variety of task analysis methods. Each contractor has considerable freedom in the selection and implementation of task analysis procedures and data generation/recording formats. The results of task analysis are used "in-house" as part of the developmental feedback loop to engineering such as design, test, and fabrication; and to other functional groups as either input or informational data. These results may also be used as a major support in the preparation of such contract end items as QQPRI, TEPI, and PSTE.

The usual considerations regarding task analysis differ somewhat in their approach to the current research of an automated task data handling capability. Task analysts must often rely on their own expertise in data collection for man/machine design decision making. This is primarily due to the time constraints caused by the "doctrine of concurrency." This research is directed towards the development of a capability which relies heavily on common data storage and indexing techniques. These techniques would make it possible to form relationships between data from different sources and from different points in the object system's life cycle, i.e., from the concept stage through the operational stage. The need for a common method of labeling and defining data is apparent in the current research effort. The data from various object system sources vary in content and identity. A scheme for identifying the data in common terms is a necessity. At present there is no precedent for a user-oriented information system that includes the number of diverse types of object systems and establishes information handling on a factual rather than a document level. The complexities of this effort are first encountered in the attempt to define and classify groups of human factors task data.

Human factors task data terminology is generated and used within an interdisciplinary environment. The terminology consists of many variations of terms from the pure, social, and applied sciences. Such a wide variation in expression creates problems in developing effective man/machine information techniques. Task data are presented in various forms and formats including Air Force or NASA task analysis forms, contractor prepared task analysis forms, graphs, photographs, and engineering drawings. Thus, the development of techniques to recognize, classify, and prepare data from several systems and sources for a computer-based task data system is a major problem confronting the research.

Task data form a complex collection of performance observations or simulations of man and man, man and machine, or machine and machine interactions and interfaces. Each collection of observations has a unique set of dependencies, any one of which may be totally contained within the observation itself. The interdependency contains all performance stimuli and reactions within the perceptual boundaries of the observer. The cross-dependency may be expanded beyond the boundaries of the observer, including other's observations. The complexity of task data dependencies is further compounded by a "ripple effect." The ripple effect is initiated by a change made to a performance. The change requires adjustments by all affected participants, man and machines, and spreads to all cross-dependent observations. These cross-dependent tasks may, because of the change, initiate changes to the original observation or to other cross-dependent observations. For example, a change in the time required for a given task may result in changes to training requirements, skill levels required, subsequent performance of other related tasks, and special equipment required. The complexity of task data, compounded by the concept of dependencies and the ripple effect has created several problems in task data handling procedures. An analyst is never certain that the data on hand are the most current, or if a design change has been approved and effected. Many other complexities of data are not so readily apparent and will likely present problems during the course of the research.

OBJECTIVES AND APPROACH

The overall objective of analyzing the data, establishing data relationships, and developing classification schemes is to provide a quantity of defined, structured data--the experimental data pool for pilot development. The data pool will be the main source of test data for the other research areas. The lexicographer will have a common source of words and terms for vocabulary and profile building. The software designer will have a preliminary structure to apply techniques and design rules for file generation and retrieval logic. The analysis area will draw on the pool for inputs to apply processing techniques. Several specific objectives to accomplish the overall objective are given below:

- Develop a method of identifying, collecting, and grouping similar task data from selected object systems The object systems should represent different types of aerospace systems.

- Provide data that satisfy the quantitative and qualitative research requirements of the other related research areas This objective will lead to the establishment of an experimental data pool.

- Develop a technique, possibly using a standardized collection form, to permit orderly extraction of the data to be used in the research

- Develop procedures for examining the extracted data to determine if the standardized format preserves the intent of the data generator

- Analyze the data to determine if a common organization exists that permits the application of taxonomic techniques to data categorization and classification This objective will allow a deeper, closer inspection of the initial investigation by Hannah and Reed (1965), on task data.

- Establish one or more task taxonomies for the purpose of systematically structuring task data into their factual characteristics

- Develop procedures to test the validity of the categorization and classification scheme when applied to both generator and user requirements

- Identify potential users of a data system and determine their requirements for accessing task data on a factual basis

The approach employed in this research area will satisfy the above objectives primarily in the order they are listed

METHODS AND RESULTS

Aerospace contractors and human factors specialists have created a variety of formats for the collection of human factors task data. After collection, the data items are combined to produce a number of outputs to satisfy the requirements of both in-house data users and contract end items. The satisfaction of the user requirements for data, i.e., the specific information desired is of pivotal importance in this research.

Manual data handling operations, which is the method used in many current aerospace programs, do not always require formalized data classification and indexing techniques. However, any automatic processing system requires some method of data identification. Data identifiers, or tags, are used throughout computer operations. Numerous schemes have been developed to classify and index data for both automated and manual handling techniques. An examination of several existing techniques was made to determine if any of those techniques can be applied to the research pilot.

The concepts derived from this examination of techniques are next used in the development of theoretical boundaries within which task data are defined, classified, and indexed. When real object system data are collected, the theory is tested against the real data. Trade-offs are made by first testing the requirements of users, and then testing generators' requirements. The final product is a preliminary input/output structure for the experimental data pool.

Examination of Task Data Generation and Reporting Techniques

Technical reports from a recent related study of task data generation and reporting techniques were examined first. Next, several task analysis techniques developed in support of Air Force, NASA, Army, and Navy weapon systems were examined. Finally, a collection of system and task analysis techniques developed by industry and research institutes, academic and government, was studied. The following discussion presents a review of the techniques studied.

Whiteman (1965) reported that eight categories of human factors task data formats exist:

1. Functional flow diagrams
2. Highly formatted forms
3. Tables
4. Equations
5. Freely formatted forms
6. Blueprints, sketches, etc.
7. Combinations of 1 through 6
8. Completely narrative forms

A conclusion is advanced that the first five categories could easily be converted for computer handling while the last three categories would present a problem for computer handling.

In a second report (Hannah et al, 1965), a further examination of human factors task data was made. The data factors examined are as follows: content, relations with personnel, impact on decisions, design, development, operation, and interface with contractors. The authors, with considerable detail, construct representative human factors data networks. Finally, the authors group all human factors data into six data groups:

Group I - Hardware Identifiers
Group II - Function Identifiers
Group III - Data Origin Classifiers
Group IV - Personnel Identifiers
Group V - Task Description Identifiers (Behavior)
Group VI - Human Factors Variables

A third report (Hannah and Reed, 1965) illustrated, in a series of eleven charts, the generation, use, and flow of human factors task data in aerospace system design and development. The emphasis is on the parts of human factors task data that are required for personnel subsystem products for either formal or informal contract end items.

The following paragraphs summarize some of the techniques developed in support of Air Force, NASA, Army, and Navy aerospace systems.

Several forms were developed for collecting and describing task procedures in weapon systems. Standard Air Force procedures are used for the Minuteman Weapon System (BSD Exhibit 62-90) in the collection of information to satisfy human factors task data requirements. This procedure consists of: Form B, System Analysis; Form C, Equipment Maintenance Analysis; Form C_1, Maintenance Function Analysis; and Figure A, Equipment Description and Parts Lists. A series of forms are used for the Atlas Weapon System that results in a complete Task/Maintenance Analysis, Time-Line Analysis, and Personnel Equipment Task Summary (PETS). The PETS is automated for updating and printout. Lockheed Missiles and Space Company, in their Human Engineering Program Plan for the Ranger/Agena D Program, developed five formats as follows: (1) Human Engineering Equipment Review Form--a narrative form that describes an item of equipment to the component level and indicates human interaction problems, recommends solutions, and provides a record of the action taken; (2) Task Analysis Work Sheet--a highly formatted check list based on requirements generated by the program; (3) Decision Action Diagram--a functional flow type of diagram that is oriented toward an analysis of human procedures and control/display requirements; (4) Human Engineering Check List--a highly formatted check list based on the requirements of MIL-STD-803; and (5) Preliminary Human Engineering Equipment Action List--a free format summary of the information contained in the other forms. It provides space for recommendations of other human factors or personnel subsystem requirements such as training, training equipment, handbooks or manuals, and procedural aids.

The Task Equipment Analysis for the Ground Air Missile, GAM77, developed by North American Aviation, Inc., consisted of the following four parts: (1) narrative description of each operating phase; (2) a functional block flow diagram of the operational segments, normal and emergency, for each phase; (3) a time-line analysis with narration and a segmented time-line; (4) a detailed Task Equipment Analysis.

The Maintenance Engineering Analysis Record (MEAR) system developed for the U. S. Navy Bureau of Weapons (BUWEPS Exhibit XWR-30) was used by General Dynamics, Fort Worth, for the F-111. This system produces maintenance analysis data in describing man/machine maintenance interfaces as recorded by Maintenance Requirements, Maintenance Tasks, and Personnel Planning Data. Automation Services Corporation (1965) conducted a study of the TITAN II Basic Data Pool and the related Air Force data systems developed for the Ballistic Systems Division. The objective was to analyze the data pool for a probable restructuring to enhance its effectiveness for an operational weapon system environment. Among the inputs to the data pool were the products generated in compliance with AFBM Exhibit 60-26A, Personnel Subsystem Development Functional Analysis and AFBM Exhibit 60-50A, Instructions for Implementing MIL-D-9412C Maintenance Analysis. The outputs were: Air Force Form C, Equipment Maintenance Analysis Data Sheet; Air Force Form C_1, Maintenance Function Analysis; and a fund of human factors data essentially oriented toward maintenance and logistic requirements analysis.

Shapero and Bates (1959), and Schaffer and Shapero (1964) introduced the concept of a complete weapon systems analysis using a structured model, the Systems Analysis and Integration Model (SAIM). SAIM is a descriptive matrix model that classifies the elements of a weapon system into three categories: (1) those determining the nature and form of the system; (2) those comprising the system; and (3) those integrating the parts of the system. A matrix is then constructed using the specific parts of the three elements. Two levels of matrix construction are indicated, system and subsystem, and primary data connections are identified. Subsequent connections can then be identified by reentering the matrixes. SAIM uses a fixed formatted form.

Reed et al (1963) discussed a small computer system for storing task data generated in the conceptual phase of system development. The technique introduces a detailed category scheme for identifying, categorizing, and coding task information. A standardized task analysis format is used for recording the categorized data. The system includes programs for updating and retrieving task information by specific category.

Kurke (1961), and Channell and Tolcott (1960), working separately, described a very similar method of presenting task-equipment analysis. The method employs an enriched flow chart technique providing for decision points, decision information requirements, the man/machine, machine/machine, and man/man interfaces. The end product is the primary difference between these studies. Kurke's objective is an operational sequence diagram, manual or automated, while Channell and Tolcott's objective is an optimum arrangement of men and equipment for military systems or industrial production operations.

Smith (1965) compared the features of four methods of task analysis. The methods studied are those developed by R. B. Miller (1953), R. C. Demaree (1961), J. D. Folley (1964), and M. P. Willis (1961). The Miller study is an application of task analysis procedures to a commercial item of equipment, the A. B. Dick Memograph 445 (Miller, 1953). Task analysis is described as an eight step procedure and may be applied in either of two presentation categories, narrative or free format. Each is applicable to systems or subsystems. Demaree's method is very similar to Miller's but the results are reported in tabular rather than matrix format. Folley's four step method initially identifies tasks and relationships among tasks; then, using functional task descriptions and behavioral detail descriptions, it further defines a task time chart by describing activities within tasks, relationships among activities, and the psychological characteristics of activities. Willis creates some fifty design problems; display, operator output, and feedback. The categories include, as appropriate, thirteen learning principles.

The results of the examination of task data techniques indicate that existing data generation and reporting techniques do not satisfy the data indexing methods needed for multisystem input capability and data item retrieval at the "factual" level. The existing techniques reviewed are rejected for any of four reasons: (1) the technique is directed toward the collection and processing of a limited area of human factors data, and would be ineffective if expanded to handle a wider spectrum of data; (2) the orientation of the technique is toward a single system and would be difficult to adapt to multisystem problems; (3) the techniques are document specific and do not provide for expansion to a data item level orientation; and (4) they are too complex or theoretical to be used in computer systems. Further research for the development of new techniques for data collection, organization, and reporting is required. The following five conclusions are presented for continued study: (1) task data are composed of miscellaneous groupings of heterogeneous and homogeneous categories of human behavior and machines; (2) differences exist in methods of generating task data, but the basic approach is to record the interactions between men and machines during the performance of an activity; (3) different methods exist for presenting task data outputs, but the basic uses are satisfied by the organization of task data items; (4) new formats for describing and standardizing multisystem task data inputs are needed; and (5) the manipulation of human factors task data must be largely standardized to satisfy user needs for data from other systems.

Analysis of Data Classification and Indexing Schemes

Some of the terms used in the discussion which follows may be taken out of their normal context; therefore, the following definitions are given for clarity:

Classification - the basic organization of data including rules for categorization and indexing

Categorization - the theoretical base by which data are grouped and interrelated

Indexing - the application of categorization to the data content, i.e., assigning one or more terms or tags to an item of data to identify it

Vocabulary - a collected list of human factors words or terms extracted from research input data--the research vocabulary will not contain words and terms that encompass the entire human factors spectrum of activities

Thesaurus - a set of rules applied to control vocabulary building and maintenance

Classification and indexing schemes are included in all information processing operations. In a small manual system, such as a student's notebook or home record collection, the scheme is usually informally developed by each user to meet the individual's requirements. For more voluminous and complex data collections, as in a public library, a formalized classification and indexing system is used to provide standardized access to users and filing points for maintenance of the collection. The Dewey Decimal System is an example. Whenever a manual operation is automated, part of the design process is the description of a scheme for classifying and indexing data. The scheme is used for both computer and manual references.

Current human factors information retrieval systems are normally related to document retrieval. Many problems involved in the classification and indexing of data, on a factual basis will differ from those posed by document storage and retrieval. These problems are due to the level of detail needed to describe data concepts. However, several classification and indexing schemes exist for document or bibliographic systems, and similarity does exist between document and factual data systems. Many of these schemes have been in existence for some time and have undergone a great deal of testing. Therefore, those techniques employed in document data bases will be studied in detail. A discussion of these techniques and a few applications are presented here.

One of the simplest approaches to organizing information for easy access is alphabetical indexing. The information is simply arranged alphabetically within logically grouped categories. This permits the use of the standard search technique of alphabetizing for information in each category.

Williams (1965) described a variety of indexing schemes and their application. One technique, enriched coordinate indexing, was developed by the Computer Department of the General Electric Company in 1959. The technique is an indexing system which permits reconstruction from the index of the basic text. It is also a "fact" or "information" content retrieval system. The technique provides specific rules for indexers, permitting controlled and uniform indexing to any level of specificity.

Rubinoff (1965) described a procedure for developing a microthesaurus based on a relatively small number, 50 to 100, of index terms. The microthesaurus is an organized list of standardized index terms--single words, pairs of words, or phrases--that described significant concepts in a technical area of specialization. The indexing procedure requires both a file-lexicon and subject lexicon.

Janning (1963, 1965, and 1966) described a coordinate indexing retrieval system that was developed at the University of Dayton for the Air Force Materials Laboratory. This system is composed of a dual dictionary and deep indexing of documents using links and roles. The use of roles in the system was determined to be more economically unfeasible than helpful and was discontinued. The final system index is a controlled vocabulary using thesaurus principles such as scope notes. Using the technique of controlled vocabulary indexing, speed is increased without sacrificing vocabulary flexibility or response to user demands.

Ronco (1960) reported on the establishment of the Human Engineering Information and Analysis Service at Tufts University. The methods adopted for organizing the human factors literature centers about a topical outline. The outline has some fifteen major headings with numerous subheadings and sub-subheadings to fill out the major areas. The outline is a logically related set of descriptions that provides structure to the field of human factors engineering and permits a systematic search of the literature to a relatively detailed level. By using contiguous and cross referenced headings, a searcher gains some knowledge of unfamiliar areas even as he conducts his search.

Keenan (1965) described a method of classifying the varieties of human performance in a nonmachine system by defining five parameters of performance and interactions, and three levels of detail. The five parameters are:

(P)	Individual Personnel Functioning
(PxEp)	Inter-Personnel Interaction
(PxEeq)	Personnel - Equipment Interaction
(PxEac)	Personnel - Ambient Condition Interaction
(PxA)	Personnel - System Activity Interaction

The levels of detail are:

Gross - corresponding to mission or executive requirements
Major - corresponding to system or subsystem requirements
Specific - corresponding to individual items of man-machine requirements

Three levels of data tags are also defined:

1. Object System Basic Identifiers
2. Non-Object System Basic Identifiers
3. Data Identifiers (applicable to both 1 and 2)

Keenan's parameters are contained in the data identifiers level and the levels of detail are specified in both the object system and non-object system basic identifiers. The data identifiers are structured as follows:

Data Identifiers

Performance Name	Man type and Quantity Required	Event P Time	PxEp	PxEeq	PxEac	PxA	PxU	Event Criticality	Training Requirements & Equipment

These structures do not provide a solution to the classification and indexing problem because task data are more heterogeneous than homogeneous. However, the structure is used to help classify the initial data groups. These groups of data are: (1) general information--descriptors of the specific worksheets, and conceptual or overview data for the specific object system; (2) preliminary design--sequential organization of system functions, hardware and software design specifications, and man/machine functional descriptions; and (3) development--detailed information relative to specific hardware/software end item performance, operator and maintenance requirements, man/machine interfaces, and man-loading analysis. However, it must be emphasized that the development cycle is an iterative process. Detailed data is developed through successive refinements from the general while some general information is developed through successive broadenings of data detail.

Ullmann (1966) provided a comprehensive summary of a number of classification and indexing techniques. The following paragraphs are from his report on techniques used in document classification and indexing.

A. <u>Subject Headings</u>

One of the approaches to the organization of information and to facilitate easy retrieval when the information is wanted, is the Subject Heading. Subject headings are used by all major libraries in the United States. They are not a hierarchical scheme but, in a sense, they are the continuation of such a scheme providing a convenient way to subdivide information into small units. Subject headings are based on the subject content of the material to be indexed, but by necessity they are for the most part used generically rather than specifically since it would be uneconomical to have a unique subject heading for each item cataloged. Subject headings should be logically constructed, uniform in depth, and mutually exclusive. Assignment of headings depends upon the point of view of the indexer. Since the depth of his knowledge of the subject of the item cataloged determines the pertinency of the subject heading assigned, the indexer can only hope that he has selected the right heading and that the user will serach for the information under the headings that the indexer has selected. To overcome this dilemma, elaborate cross-references have been developed to aid in retrieval of the information. "See" references are made from unused terms to the subject heading chosen for use in the catalog; or "see" references may make reference from a specific to a more generic term. "See Also" references are made from a related comprehensive term to specific subject headings, or to accepted synonyms.

1. Direct Subject Heading

The Direct Subject Heading is a word or phrase denoting the general content of an item of information. It may be followed by other words or phrases serving as subdivisions . . . .

2. Inverted or Dictionary Type Subject Heading

The Inverted Subject Heading, in contrast to the Direct Subject Heading, is arranged by main noun only. It, therefore, lists all related headings under the same noun providing an instant assessment of the information at hand. It reduces the onerous task of having to consult numerous entries in the catalog in order to obtain a complete set of information. It also diminishes the need for elaborate cross references . . . .

B. Keyterm Indexing

In the last few years a different approach has been made to overcome the lack of a common language between originator and user. This approach is referred to as keyterm indexing. It consists of analyzing the information to determine the appropriate terms (keyterms and descriptors) or vocabulary representing the subject content.

To retrieve the information, these terms may be freely associated or coordinated by logical operations of sum, product, or complement into combinations representing the content of the information desired . . . .

The Keyterm method of indexing claims many advantages over conventional indexing systems. One of the main advantages is its flexibility and adjustability to particular needs, conditions, and requirements. A second advantage is the reduction or virtual elimination of obstacles of communications between the originator, indexer, and user of a particular item of information since the terms used to index the information are selected from the information itself and the only judgment exercised by the indexer is the determination of which keywords contained in the information are significant indexing terms. This conclusion is based upon the assumption that the originator of a scientific or technical language and, therefore, the terms contained in the paper and utilized for indexing are common to both the originator and the user. One added feature of the keyterm system is its adaptability as a tool for high speed digital storage and retrieval systems.

1. Coordinate Indexing

a. Uniterms

In early 1952 a contract was negotiated between the Armed Services Technical Information Agency (ASTIA), now Defense Documentation Center (DDC), and Dr. Mortimer Taube, Documentation Incorporated, Washington, D. C., to make a study of existing indexing systems including all known classification schemes. He was to evaluate their adequacy for the control of the vast collection of documents of ASTIA and, if necessary, develop a new system that would satisfy ASTIA's requirement for indexing its heterogeneous collection. This research by Dr. Taube and his associates resulted in the development of the Uniterm System of Coordinate Indexing. As all new ideas, this system at first drew quite some controversy, but since has proven successful for many activities, both commercial and governmental.

Each term in the system is a filing term and hence an access point to information. There are no subdivisions which means that each term in the system has the same value as any other term which eliminates "see also" references from general to specific subjects. In the case of synonyms, a "see" reference may be made from the least common or used form to the most common form, e.g.,

Devices
  See Equipment
Apparatus
  See Equipment . . . .

The original Uniterm concept differentiated between "free" and "bound" terms. Those found in combination with only one other term within a given closed collection or system were considered "bound" whereas, terms which may be combined with many other terms were called "free."

b. Descriptors

Another form of keyterms are the Descriptors. Descriptors are inclusive terms for a particular field. It is claimed that a small set of descriptors for each discipline is sufficient. If the proposition that descriptors follow the idea of of set inclusion is accepted, they are not as expressive of the content of any single piece of information as are Uniterms. Any need for greater depth or specificity will require an associated increase in the number of descriptors . . . .

c. Topic Tags or Precoordinated Terms

In order to increase the preciseness of meaning and concept, and to reduce the number of coordinations required to retrieve information, System Development Corporation has developed the concept of Topic Tags. There are pre-coordinated terms somewhat larger in scope than Uniterms.

National Security
  Civil Defense
  Internal Security
    Security Classification

The terms are less inclusive than Descriptors but more inclusive than "bound" Uniterms . . . .

2. Rotational Indexing: Keyword-In-Context

Utilizing the well-established technique used in the compilation of concordances of important works of literature, the Keyword-In-Context concept developed by the late Hans Peter Luhn, lists selected keywords together with the surrounding words which act as modifiers. By this means, the user obtains a more specific idea in what connection the keyword has been used. The added information conveyed by this method helps to bridge more readily the gap of communication between user and originator of information. However, it must be cautioned that the technique is not an indexing method by itself but is used primarily for computer-generated printed bibliographical indexes and is only mentioned here as an example of the unique use of keywords . . . .

C. Expanded Keyterm Indexing

One of the main problems associated with using Uniterms, Descriptors, or Topic Tags is the occurrence of noise commonly referred to as "false drops." To overcome this difficulty, methods have been developed to increase the pertinency of keyterms by adding various appendages denoting semantic and syntactic interrelationships. This process sometimes is called "Concept Coordination." The assumption is that this will facilitate the determination of relevance of keyterms to a given item of information and, in addition, permit the retrieval of peripheral information. Among these appendages are:

1. Roles

Roles may be used as action indicators, i.e., who is doing what to whom, or simply to indicate relationships of terms to each other, or to explain the context in which a term is being used.

2. Links or Interfixes

Links are claimed to be especially useful in the field of chemistry to indicate conceptual relationships, or may be used to show sequence, or to link various parts of phrases, sentences together, or simply to modify another word.

3. Prefixes

In some systems, prefixes in the form of mnemonic codes are used to reduce ambiguity and to indicate generic levels which can be utilized to retrieve the information.

4. Suffixes

Numerical suffixes are a means of distinguishing one concept from other similar concepts. DoD Instruction 3200.8, dated 18 February 1964,* suggests that similar numeric suffixes may be used to indicate "the relative weight of the subject of the key word in describing the technical content of the report. The 'weight factors' shall range from 0 (zero) to 3 (three), with the highest weight assigned to suffix numeral '3.'"

D. Semantic Factors

Perhaps the most extreme form of an attempt to instill more meaning into words to facilitate retrieval, reduce ambiguity, and assure relevance is the "Encoded Abstract." This is a standardized telegraphic abstract composed of segments called "Semantic Codes." The basic units for building "Semantic Codes" are the "Semantic Factors" which are said to "symbolize the concept represented by a word." The "Semantic Codes" contain most of the elements discussed in the preceding paragraph such as roles, infixes and suffixes, but in addition use telegraphic punctuation to denote syllable, subphrase, phrase, etc.

The investigation of the indexing schemes and classification techniques did not identify a technique that can be directly applied in building an experimental data pool. Dr. Paul G. Ronco, Director of Human Engineering Information Analysis Service, Tufts University, was asked to study some of the problems of classifying and indexing human factors data on a factual level. A brief paraphrased summary statement of Dr. Ronco's advice follows:

> Since the subject data to be included in the experimental data pool is so diverse and diffuse, no one single indexing technique could possibly be expected to provide a solution. The research

* The information (quoted by Ullmann, 1966) is also contained in Air Force Regulation 80-29, Army Regulation 70-31, and Secretary of the Navy Instruction 3900.29.

should break with the tradition of using a single classification or indexing technique for the system and assign an indexing technique that best operates with the type of data included in each category.

In summary, there is not an immediate solution to the problem. However, several techniques are identified that can be applied to problems that best fit each technique. The research conducted in constructing a classification and indexing scheme for the experimental data pool is discussed in the following paragraphs.

Development of the Task Data Structure

The examination of documented techniques, described above, indicates both the existence of a variety of task analysis formats and the significant role of task data in technical reporting procedures. These reports will range from data generated specifically for human factors specialists to broader management information, both contractor and government, to contract end items. A concept of data generation (input) and user (output) can be determined by using both the documented techniques and through the advice given by people involved in current systems. Using the generation requirements, judgments regarding data identification and classification are made before user requirements are applied. The first need, is to develop some type of organizational framework that involves more than one system. The framework will identify, define, and categorize task data to a level detailed enough for computer input, and analysis. The framework will also be compatible with the user's vocabulary and current awareness requirements.

Two alternatives exists for initiating the data framework design. One alternative is to devise a data structure using formats and theoretical information obtained from the examined documents. The second alternative is to devise the data structure using actual data worksheets and formats in current and past aerospace systems. The first approach will use theoretical inputs as object system data. Two reasons for this approach are: (1) it will take time to select object systems and collect data; and (2) if only historical systems data were used, the data structure would tend to be fixed, since its origin was from operational systems rather than from theory.

Data Framework

The basic design of the object system data base postulates a logical data framework for the system. This framework consists of stratified levels which are amenable to classification into subdivisions or cells. The number of levels and cells required is flexible and contingent on data criterion. The research will not be able to determine in detail how potential users will want to access task data. These final decisions are made by users actually using the data bank. Therefore, the need exists to provide a framework with as much flexibility as possible. The data criterion is considered satisfied, and therefore, a "data item defined," when the basic information state can be retrieved in terms of optional frames-of-reference. A user's frame-of-reference is the thought process in which he groups and relates data facts. For example, in one instance the user relates performance with time; in another he relates performance with mission. He is also thinking in levels of detail, that is, performance on a group level, or performance on an individual level. Various users will have various frames-of-reference. The framework will need to correspond to these various frames-of-reference. The system might contain a capability for allowing each user to define his own unique frame-of-reference in addition to standard frames-of-reference which are more common to all users.

The framework is based on the premise that data can be logically organized into a hierarchical structure, permitting the data to be identified in levels which become progressively more detailed. An example of six levels are:

- Type
  - Class
    - Category
      - Item
        - Element
          - Unit

If necessary, additional levels are added to satisfy the data criterion. The labels assigned to the various levels are arbitrary. The content of the levels, however, fit general categories of data. Examples of the framework applied to task data are presented in table I.

The framework indicated a need for a more precise method of categorizing information groups. It identifies levels, but it does not assure that the content of the levels are specified to prevent overlap. An alternative procedure of identifying data was taken. The term "element" is redefined as a data category that fits a more generic class of information; lower levels are designated as items.

The procedure developed to arrive at a more detailed data identification scheme is based on the derivation of data elements. A data element is a label used to represent a generic class which contains any number of subordinate data. For example, a data element and its data values might be "Part Number" and "MX19432-Y," or "Task" and "disconnect hydraulic actuator."

As an exploratory attempt to substantiate the data element concept, a group of source documents was selected. The documents were not necessarily a representative sample, but they did direct the research to the specific area of task data generated through the use of task analysis. Although a large number of documents fit this description, the number used was kept small enough to permit maximum use of the documents selected for the sample. The documents selected were: AFSCM 80-3, Hannah and Reed (1965), Hannah et al (1965), Atlas Series F Integrated Task Index, Reed et al (1963), and AFBSD Exhibit 65-14. Each data element referenced in one of the source documents was transferred to an index card. The card file, when completed, contained all data elements described in the source documents. The data elements and associated sources were compiled into a list (see table II, a partial list for illustration). Next, a minimum number of independent elements were derived, which accounted for all of the data elements contained in the original list. This was accomplished by using the definitions of the elements provided in the source documents and applying an iterative judgmental process of element combination and redefinition. The resulting list of data elements was the first research trial in categorizing human factors data. There was some

Table I. Initial Data Framework

TYPE	CLASS	CATEGORY	ITEM	ELEMENT	UNIT
Personnel Subsystem	Anthropometry				
	Operability Design				
	Task/Equipment Analysis	Identification			
		Equipment			
		Task	Task/Subtask Number		
			Description	Action Verb	
				Object Noun	
			Location		
			Communication	Human to Human	Type
					Method
					Rate
					Duration
				Human to Machine	
				Machine to Human	
			Time	Time-line Graph Number	
				Total Time to Complete	Time per person
				Time Criticality	
			Support Tools & Aids Required		

Table II. Sources and Data Elements

1	2	3	4	5	6	DATA ELEMENTS
	X	X				Phase
	X					System Development Phase
	X	X				Phase Segment
		X		X		Function
	X					Maneuver-Assignment
		X				Job
X					X	Job Operation
	X	X			X	Task
X						Duty Statement
X						Task Sequence Number
X						Task Statement
X						Task Data
		X				Task Elements
	X					Sequence Number
		X				Indentures
	X	X		X		System
X						Model
X	X	X		X		Subsystem

question about the validity of the data elements contained in this first categorization because of the inexact method of sample source selection. Therefore, the data elements were compared with the fifty-one categories recommended by Hannah et al (1965). Again, through subjective judgments, a reevaluation was accomplished which resulted in a second list of categories. This second list contained thirty-two data elements and was accepted as a point of departure for future research in the definition of task data. These thirty-two elements, their definitions, and probable parameters are contained in Appendix I.

The original thirty-two data elements were derived from the outputs of data generators. A question arose as to whether a data structure would be organized according to generator or user requirements. It was decided that the data structure, wherever possible, will primarily reflect the needs of the user. The functions of an operational data system is designed to satisfy it's users. A test of the ability of the thirty-two data elements to satisfy user requirements was then established. An attempt was made to illustrate the relationships between data elements based on input and user requirements based on output.

The relationships are illustrated in a matrix of input/output relationships developed from the thirty-two data elements (see figure 6). Ten users of human factors data were given descriptions of the thirty-two data elements and were requested to make judgments of input/output relationships. These judgments were recorded in accordance with the following scale: 0 = no relationship exists, for example, between a performance prerequisite and a mission phase; 1 = an indirect relationship exists between the input/output aspects of the two data elements, e.g., there may be only an implied relationship between the hardware subsystem and the number of personnel required for operation or maintenance; 2 = a direct relationship, that is, a one to one correlation exists between the input/output aspects of two data elements, e.g., the number of personnel required for an operation or maintenance performance and the developmental status, mock-up or production, for example, of the hardware item. The nearest whole number of the mean was recorded on the matrix. The judgments were achieved by a subjective content value rank-ordering of the thirty-two data items.

The research efforts described above represent the achievement of a preliminary research step in the identification, definition, and categorization of task data. Evidence collected indicates that task data are amenable to separation into identifiable categories of similar or related data. These data may then be generally defined and parameters postulated from a theoretical framework. Additionally, it appears possible to satisfy most user requirements from data that have been generated by a number of different users.

Analysis of Data Elements in Object Systems

A basic concept of a task data structure is established. The next step in the research process replaces data elements in the theoretical framework with actual data elements. To accomplish this aim, data and technical reports from three object systems will be used.

The selection of object systems was based on the following criteria:

INPUT FACTORS		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Object System	1																																
Mission	2	2																															
Mission Phase	3	1	2																														
System	4	1	1	2																													
Subsystem	5	0	1	1	2																												
Component	6	0	0	0	1	2																											
Part	7	0	0	0	1	1	2																										
Hardware Status	8	0	0	0	0	0	2	2																									
Source of Data	9	2	2	2	2	2	2	2	2																								
Reference	10	2	2	2	2	2	2	2	2	2																							
Security	11	2	2	2	2	2	2	2	2	2	2																						
Performance Level	12	0	1	1	1	2	2	2	2	2	2	2																					
Performance Desc.	13	0	1	1	2	2	2	2	2	2	2	2	2																				
Performance No.	14	1	1	1	1	1	1	2	2	2	2	2	1	2																			
Performance Prereq.	15	1	0	0	1	1	1	2	2	2	2	2	2	2	2																		
Personnel Classif.	16	0	0	0	1	1	2	2	2	2	2	2	2	2	1	1																	
Number of Personnel	17	0	0	0	1	1	2	2	2	2	2	2	1	2	1	1	2																
Performance Loc.	18	0	1	1	2	2	2	2	2	2	2	2	2	2	1	0	1	1															
Environment	19	2	1	2	2	2	2	2	2	2	2	2	2	2	0	1	1	1	2														
Communication	20	0	0	0	1	1	2	2	2	2	2	2	1	2	0	0	1	1	1	0													
Tools, etc.	21	0	0	0	1	1	2	2	2	2	2	2	1	2	0	0	0	1	1	1	0												
Frequency of Perf.	22	0	0	0	1	1	2	2	2	2	2	2	1	2	1	1	0	1	0	0	0	0											
Time	23	0	0	1	1	1	2	2	2	2	2	2	0	2	0	1	0	0	1	1	0	1	1										
Criticality	24	1	1	1	1	1	2	2	2	2	2	2	1	2	0	1	1	1	1	1	0	1	1	2									
Hazards	25	1	1	1	1	1	2	2	2	2	2	2	1	2	0	1	1	1	1	2	0	1	0	0	1								
Human Output	26	0	0	0	0	1	2	2	2	2	2	2	0	2	0	1	1	0	0	0	0	0	1	0	1	0							
Machine Output	27	0	0	0	0	1	2	2	2	2	2	2	0	2	0	1	1	0	0	0	0	0	1	0	1	0	2						
Knowledge	28	0	0	0	0	0	1	2	2	2	2	2	2	2	0	1	1	1	0	0	0	0	0	0	1	1	1	1					
Skill	29	0	0	0	0	0	1	2	2	2	2	0	2	2	0	1	1	1	0	0	0	1	0	0	1	1	1	1	2				
Difficulty	30	0	0	0	0	0	1	1	2	2	2	2	0	2	0	0	0	1	1	1	0	1	2	1	1	1	1	1	1	1			
Performance of Prob.	31	0	0	0	0	0	1	1	2	2	2	2	0	2	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	2		
Reliability of Perf.	32	0	0	0	0	0	1	1	2	2	2	2	0	2	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	2	

Legend:

0 = No Relationship

1 = Indirect Relationship

2 = Direct Relationship

Figure 6. Data Element Output Relationships

- The systems should be in their early stages of development in order that the evolutionary process of data generation usage can be given consideration.

- The systems must contain a personnel subsystem or life sciences program.

- The systems must contain programs that require the generation of task information.

- The systems should be representative of those developed for the government, e.g., space systems, aircraft systems, and missile systems.

The systems selected for this research program are:

- C5A Heavy Transport Aircraft
- Airborne Launch Control Center (ALCC)
- SV-1C Saturn V Booster, Stage 1C

Data collection forms for C5A and ALCC were reviewed to determine the type and content of data to be generated for those systems. The forms are based primarily on classical task analysis methods. Following the identification of data elements, they are compared to determine if a basis of standardization exists in the elements for the two object systems.

A second objective of the review is to compare the degree of compatibility between the theoretical elements and the deduced elements from the input forms. The methods used to accomplish these objectives are as follows:

- <u>Collect lists and descriptions of the raw task data items for each system</u> Where this information is not available, probable definitions are developed.

- <u>Develop a structure for the data items related to C5A</u> The structure must be meaningful with regard to data usage. Repeat this step for ALCC and SV-1C.

- <u>Explore the possibility of combining the three object system data structures in order to provide one common structure</u>

The analysis conducted, using the C5A Flight Crew Task Analysis input form, resulted in the sixteen data elements below. Some subelement data are also included:

1. Crew Member
2. Phase of the Flight Mission
3. Identification of Sheet of Related Charts
4. Date of the Analysis
5. Revision Notes and Date of Revision
6. Reference Line Number
7. Function Element
8. Task
9. Equipment and Location
10. Equipment Characteristics:
    Visibility
    Readability

Reachability
Manipulability

11. Frequency of Use (Equipment and Task)
12. Task Characteristics:
    Difficulty
    Criticality
    Training Requirements
13. Functional Relationship
14. Time Breakdown:

Vision	Feet
External	Cognitive
Internal	Audio
L. Hand	Verbal
R. Hand	

15. Time Budget:
    Time Constraint
    Time Started
    Time Completed
    Clock Time
    Overload
16. Tolerance, Hazard, and Remarks

Using the C5A categories as a comparison standard, the data collection forms for the Minuteman Weapon System's ALCC were analyzed. The objective of the analysis is to determine if the data elements are compatible i.e., standard, for the two object systems. The result of the analysis shows that the data elements defined appear to be completely compatible with the addition of one data element. The element is:

17. Special Tools and Equipment

Figure 7 shows a dependency relationship among the seventeen data elements listed above. Since all forms analyzed were blank, compatibility of data content between systems cannot be assured. However, it is postulated that standardization is possible and a completely different set of data elements will not be required for each object system.

Further Analysis of Data Elements

An attempt was made to resolve the differences between the original organization of elements into thirty-two categories and the second organization into seventeen categories. Output relationships of both sets of categories were tested using a simple query language to form sample queries (see Appendix II). The language is a tool used to subjectively test the validity of the assigned relationships for output purposes. Some judgments in interpretation were needed because analysis was performed using formats rather than actual data. The first few queries proved the set of thirty-two categories to be harder to handle in the attempts to provide answers to the sample queries. One major difficulty encountered was that the data elements are presently defined into system-specific categories, and it could not be assured that the relationships leading to the query answers were not also system-specific. The problem of data being system-specific would violate one system objective, that is, the ability to mutually use systems data in response to users' queries. The difficulty forced a reevaluation of the entire problem

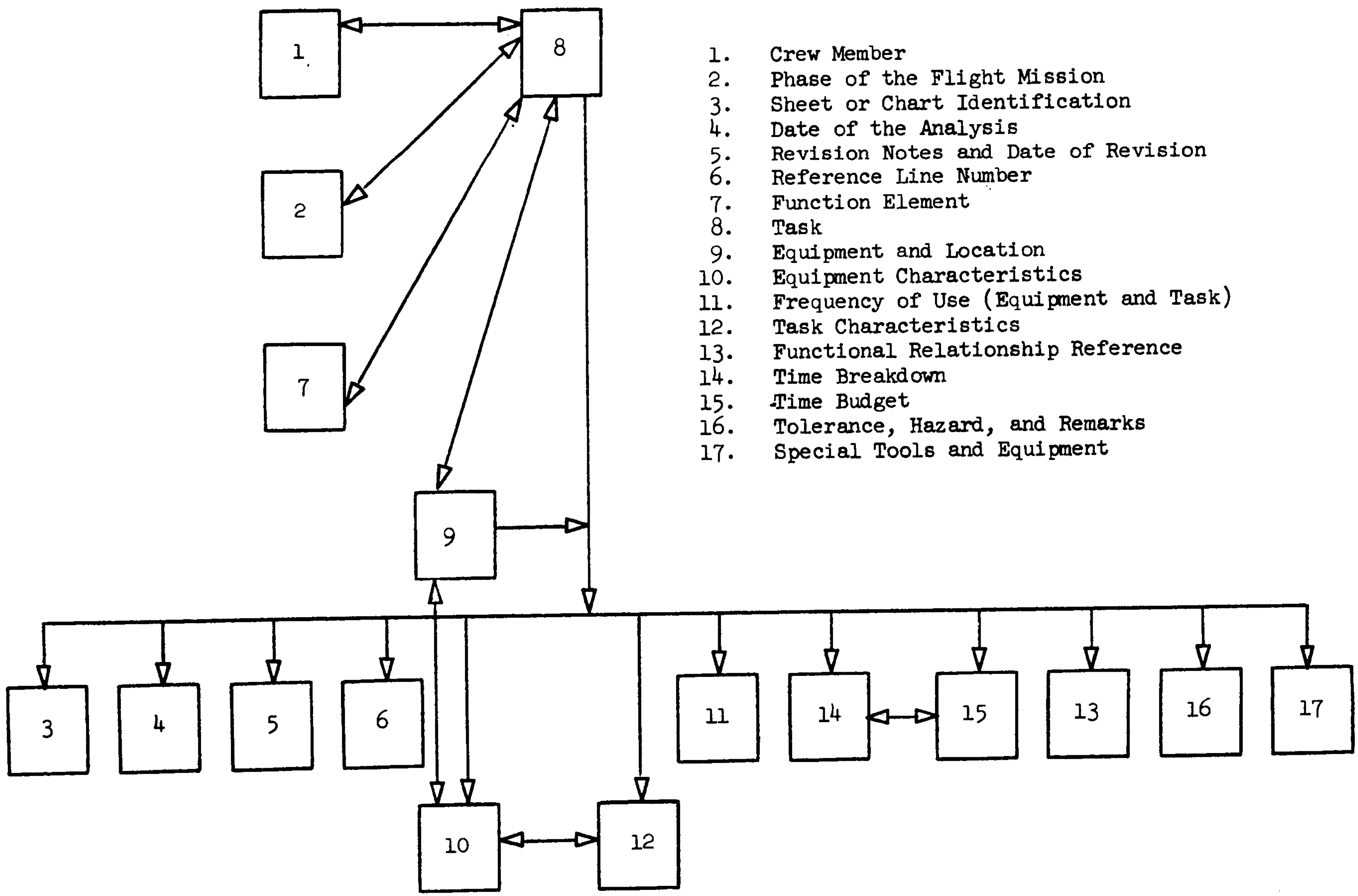


Figure 7. Initial Structure for ALCC/C5A Element Relationships

of data definition. A reassessment of the original approaches of developing a task data structure was necessary. One approach was to develop a data structure from the theoretical elements derived from the literature search; the second approach was to develop a data structure using only object system data. The present approach uses both structures developed, i.e., the thirty-two data elements and the seventeen data elements. This approach will result in a set of general but completely defined data elements that will accommodate a wide variety of data items. These data elements will be the common pivotal points around which an experimental data pool is developed.

A list of ten data elements (see table III) was derived by comparing the original thirty-two elements with the elements found on the preliminary C5A and ALCC data input forms. This list does not represent all possible data elements for all object systems, nor is it likely to contain all the data elements that will be utilized in the experimental data pool. It does, however, provide a basis for determining illustrative data relationships needed to proceed with the design logic of storage and retrieval techniques. Thus, the research described above is directed toward the satisfaction of the system requirements that the system must provide for the storage, indexing, and retrieval of human factors task data. The data bank structure must be sufficiently flexible to allow for future expansion and inclusion of additional data elements.

Data Element Relationship

The data elements, when examined for basic input relationships, tended to fit naturally into a hierarchical structure. The basic relationships, i.e., class dependencies, among the elements are illustrated in figure 8. These relationships are as follows:

- Object system is the prime element, that is, all other elements are structured under it.
- Three elements--mission information, system information, and performance descriptions--are directly related to the object system element. Performance description is related directly to mission information and system information.
- The hardware characteristics element is directly related to system information and performance descriptions.
- Performance characteristics, personnel descriptions, and time information are directly related to performance descriptions.
- The two remaining elements, source identifiers and remarks, are directly related to each of the other nine elements.

Each of the ten data elements represent a major grouping of data. A fixed set of subelements, or data items, for each element will be included as each object system's data are further analyzed. Some items may be amenable to standardization, but flexibility in the definition and categorization of data elements must be provided in the design of the data structure. The design must not be of very rigid structure. Rather, it should be open ended to

Table III. Definition of Ten Data Elements

Data Element	Definition
1. Object System	The designator of a specific aerospace system
2. Mission Information	A specific operational maintenance profile or profile segment for the specified object system
3. System Information	Specific data relating to the hardware and software required to accomplish the specified mission or segment
4. Performance Description	Specific data relating to the level of detail to be included in the related performance descriptions
5. Performance Characteristics	Specific data relating to the man/machine, and man/man interfaces and duties required to accomplish the specified mission or segment.
6. Hardware Characteristics	Specific data regarding the human engineering characteristics of the hardware required to accomplish the specified mission or segment
7. Personnel Description	The job title and/or Air Force specialty code of personnel required in the specified performance --any special skills or knowledge required of the performer are also noted
8. Time Information	Specific data regarding performance or mission related time values
9. Remarks	Miscellaneous comments and remarks necessary to explain any material contained in other data elements
10. Source Identifiers	Specific data regarding the origin and author, date of completion or revision, references used by the generators, and security or proprietary restrictions

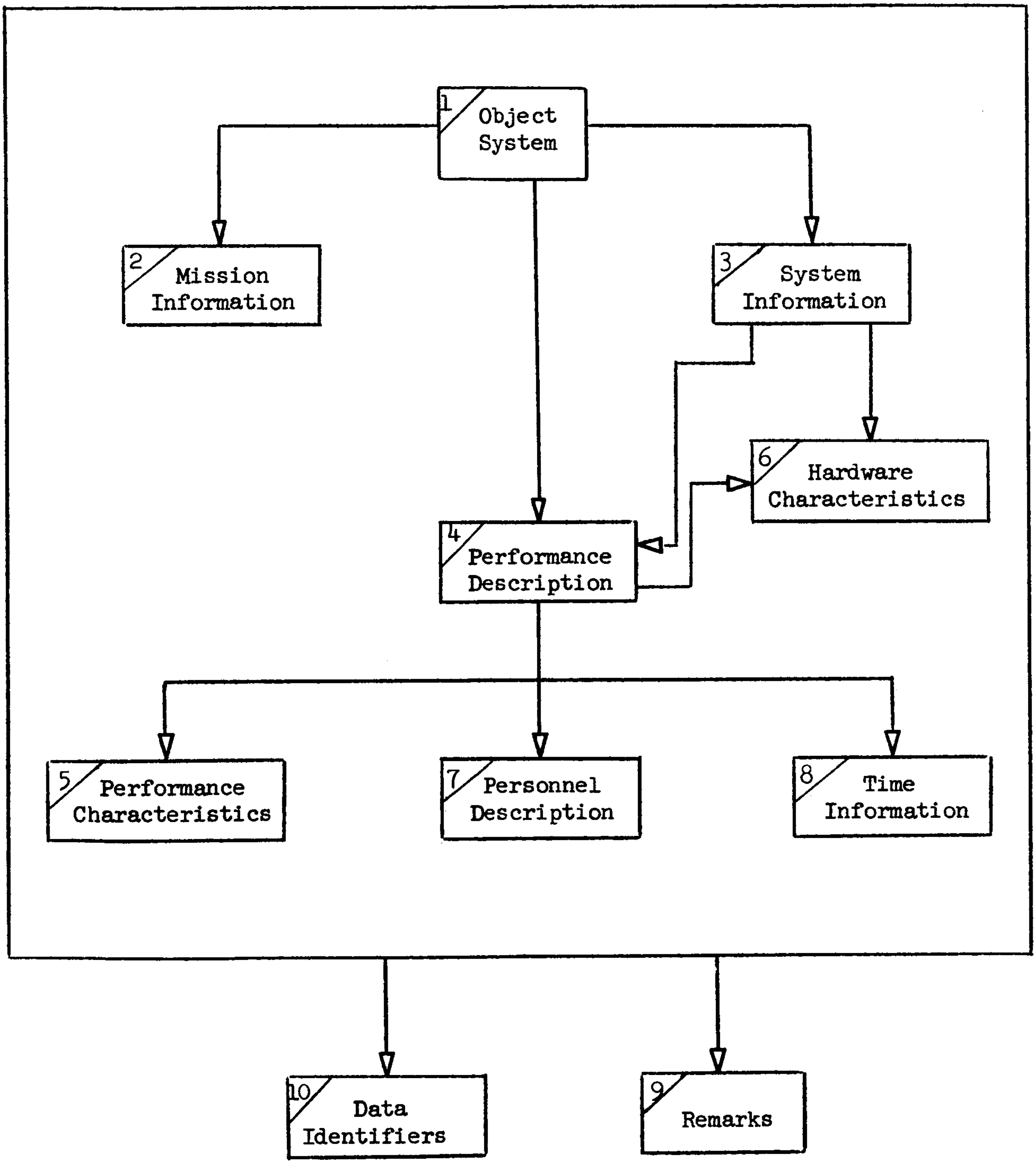


Figure 8. Hierarchical Structure of Input Relationships

permit addition, deletion, or combinations of data element categories to accommodate variance in the various object systems' data. These variances may be in data content, data classes, system-specific data, data organization, etc.

The relationships discussed above represent the most basic structure of the data. A more detailed structure is needed, that will take into consideration the relationships among data items. These relationships are the cues by which a user would refer to, or access, the various combinations of data within the ten elements. The list of elements will be reviewed for all possible output relationships which are meaningful to a user. These output relationships will be recorded and, if necessary, reevaluated. The resultant list should include all of the basic relationships, excluding nonmeaningful relationships such as equating source information with time information.

A preliminary organization of the ten data elements, with their assigned output relationships, is shown graphically in figure 9. For the data structure, at the element level, the object system is the entry point. Table IV illustrates the basic data sets representing user frames-of-reference. These frames-of-reference provide the relationships among major classes, i.e., data elements.

Data Element Frame-of-Reference

The use and meaning of the frame-of-reference are explained through examples. The frame-of-reference may be used separately as an individual path through the data structure to arrive at desired data, or may be used in combination with one or more frames-of-reference.

Two examples are provided to explain the concept of single frame-of-reference usage:

1. A human factors specialist, preparing a task analysis on a new system, needs to assign performance characteristics to a task involved in a certain mission. The specialist wishes to review all characteristics related to similar tasks in another system currently in operation.

2. A training specialist at the Air Training Command (ATC), in preparation for establishing course requirements for a position (AFSC), wishes to retrieve the task performance descriptions for the specific personnel description (AFSC) in a certain mission for an object system currently under development.

Consider the frame-of-reference, example 1 above. In this example, a requester wishes all of the performance characteristics (Element 5) related to a particular set of performance descriptions, (Element 4) written for a specific mission, (Element 2) and for a certain object system (Element 1). Referring to the data structure, the following steps show the user's frame-of-reference identified in this example.

- The specific object system is directly related to a set of mission information.

- The specific mission, as a subset of the mission set, is directly related to a set of performance descriptions.

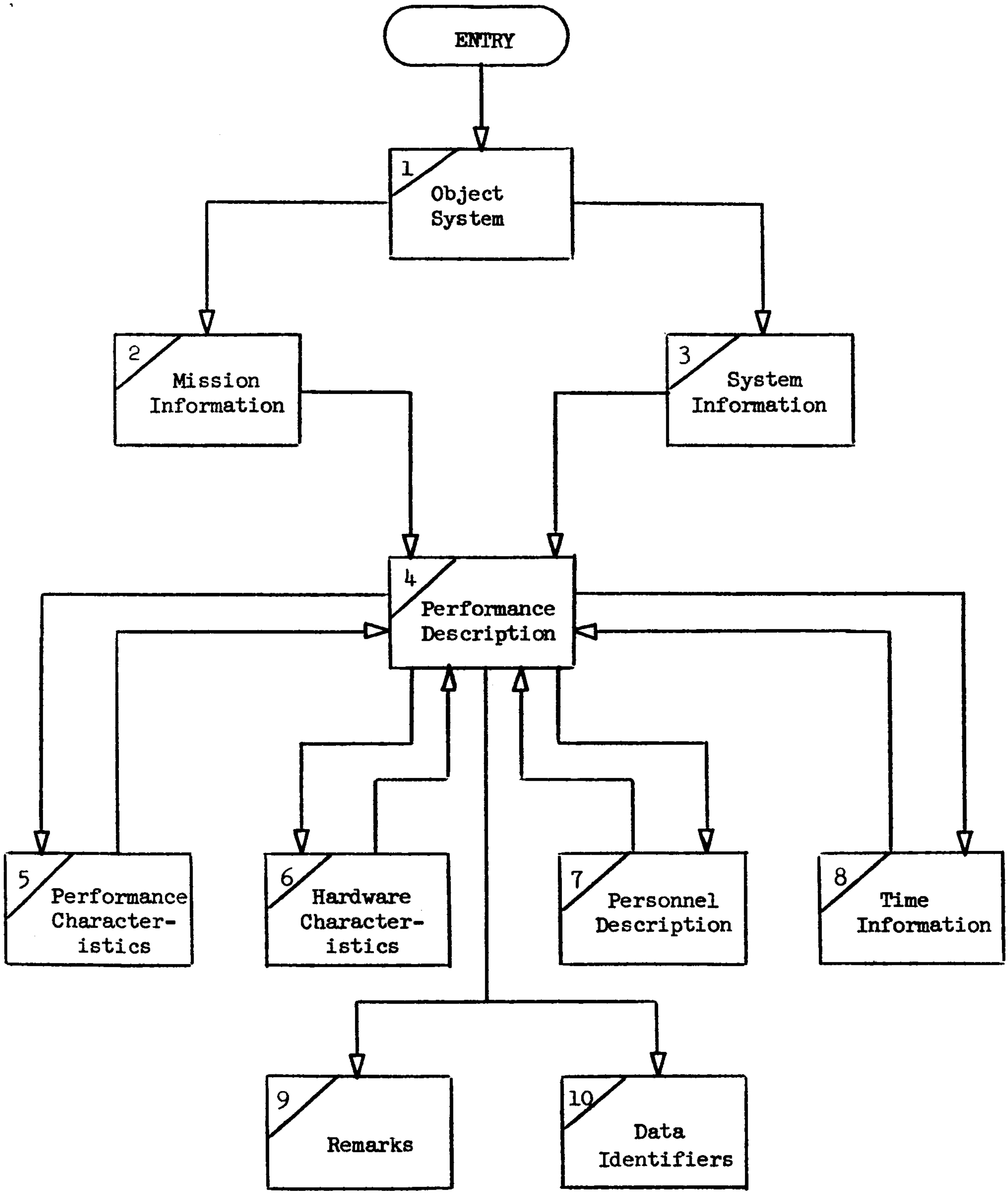


Figure 9. Data Element Output Relationships

Table IV.
Hierarchical Structure of Internal Relationships for the Ten Data Elements

1-2	1-3
1-2-4	1-3-4
1-2-4-5	1-3-4-5
1-2-4-6	1-3-4-6
1-2-4-7	1-3-4-7
1-2-4-8	1-3-4-8
1-2-4-9	1-3-4-9
1-2-4-10	1-3-4-10
1-2-4-5-4	1-3-4-5-4
1-2-4-6-4	1-3-4-6-4
1-2-4-7-4	1-3-4-7-4
1-2-4-8-4	1-3-4-8-4

1 - Object System
2 - Mission Information
3 - System Information
4 - Performance Description
5 - Performance Characteristics
6 - Hardware Characteristics
7 - Personnel Description
8 - Time Information
9 - Remarks
10 - Source Identifiers

- The particular performance descriptions referenced by the requester are used as a senior set of information by which its related performance characteristics are subordinated.

- The lowest order in this particular hierarchy is the subordinate set identified in step 3 and consist of the information desired by the user. The logical connection of the elements for this particular request is given in the following sequence: (1 - 2 - 4 - 5).

The frame-of-reference in example 2 (p. 54) is (1 - 2 - 4 - 7 - 4). In this example, the user wants access to all of the performance descriptions that have a specific personnel description assigned to the performance descriptions are written for a particular mission, and for a specific object system. This example utilizes the following steps:

1. The specific object system is directly related to a set of mission information.

2. The specific mission is then used to identify its subordinate set of performance descriptions.

3. The subordinate set identified in step 2 now becomes a senior set directly related to its subordinate set of personnel descriptions.

4. The set identified in step 3, the specific personnel descriptions, in turn relates to a subset of the set of performance descriptions identified in step 2.

5. The subset resulting from step 4 consists of the information desired by the user.

Use of Combined Frames-of-Reference: An example of combining frames-of-reference is as follows; a member of a System Program Office wishes to know the frequency of use of an item of equipment required for the performance of a specific task, by a specific AFSC, for a particular mission phase of an object system. To access this information, the following basic frames-of-reference would be used:

1. (1 - 2 - 4 - 7 - 4) Object System
   - Mission Information
     - Performance Descriptions
       - Personnel Descriptions
         - Tasks

2. (1 - 2 - 4 - 6 - 4) Object System
   - Mission Information
     - Performance Descriptions
       - Hardware Characteristics
         - Tasks

3. (1 - 2 - 4 - 5) Object System
   - Mission Information
     - Performance Descriptions
       - Performance Characteristics

In each set of element associations above, the last element is the prime information. In set 1, those tasks with the desired Personnel Descriptions are referenced. In set 2, tasks with certain Hardware Characteristics are located. In set 3, the pertinent Performance Characteristics are located. These frames-of-reference would be combined as follows:

- (1 - 2 - 4 - 7 - 4) combined with (1 - 2 - 4 - 6 - 4) results in (1 - 2 - 4 - 7 - 4 - 6 - 4), which specifies a particular set of tasks with both the qualifications of certain Personnel Descriptions and Hardware Characteristics.

Those tasks with the above qualifications are now used to qualify the Performance Characteristics (Element 5), thus detailing the reference to only the data desired.

- (1 - 2 - 4 - 7 - 4 - 6 - 4) combined with (1 - 2 - 4 - 5) results in the combined frames-of-reference (1 - 2 - 4 - 7 - 4 - 6 - 4 - 5).

A summary of the study of element frames-of-reference is given below:

- There are ten data elements, with each element containing a varying number of data items.
- The data elements are interrelated as shown in the data structure (see figure 9).
- Twenty-four basic frames-of-reference are contained in the data structure (see table 4).
- The frames-of-reference identify the logic used to access specific data.
- The frames-of-reference may be used individually or in combination to access a desired set of data.

Data Item Analysis

Actual object system data will be inserted into the experimental data pool. The data must be further analyzed, and data items isolated for assignments to various elements. The purpose of this data item analysis is to: (1) extract and define the character (e.g., alpha, numeric, symbolic) and make-up (e.g., three numeric characters, maximum of forty-five alphanumeric characters, ten lines of narrative, two symbolic characters) of each data item; (2) identify the contextual meaning of each item and classify them according to the appropriate scheme; (3) examine the data items and determine their probable output relationships with other items in the same data element and with data items of other data elements; and (4) provide indexing procedures for the assignment of the items to element categories.

To assist in the mutual extraction and organization of data, from different systems, a common data collection form was needed. A review of the data collection formats of USAF, NASA, and contractors was made to determine if an existing format could be used as a standard for the research. The forms examined did not provide the flexibility required for common classification. Therefore, a standard input data form was developed to be used for the recording of data. This form, illustrated with ALCC data, is shown in Appendix IV. The research is engaged in extracting data from technical documentation and describing task data for the three object systems involved in the research. An instructional sheet, giving instructions for extracting data from ALCC documents, is given in Appendix III.

CONCLUSIONS AND RECOMMENDATIONS

The principal effort during this research period was the establishment of a method for defining and classifying human factors task data. The results of the research provide a logical base from which an experimental data pool can be established. This data pool, and its structure, is intended to supply the necessary requirements for design and development activities in the companion research areas, i.e., storage and retrieval techniques. The structure of the data pool is represented by a task taxonomy which specifies both the general and specific classes of data and defines their relationships. Further research is needed to complete the taxonomy and establish an experimental data pool.

Three remaining activities which must be completed during this research are: (1) determine the extent of the data to be extracted in order to set the size of the data pool; (2) establish the structures--items and subitems--for each element category; and (3) develop a detailed data pool including total structure and assigned data values for three systems. A discussion of the needed research is given in the following paragraphs.

Analysis of Human Factors Task Data and Data Relationships

Further investigation into the content of data items is required to determine the extent to which items can be standardized for application to any object system. A method of selecting output frames-of-reference for data elements has been developed. Investigation must be conducted to determine if the same method is applicable to data items. If it is not, research must be conducted to determine a method of extending the hierarchical arrangement of items to specific levels. The data item input frames-of-reference for ALCC data requires continued research. Analysis of the input data and definition of data relationships must be initiated for the remaining object systems, Saturn V-1C and C5A.

Human Factors Task Data Classification Schemes

Upon completion of the ALCC task taxonomy, investigation as to the type(s) of indexing to be applied to the data categories within the taxonomy must be initiated. The objective of this taxonomy is to establish a standard capability to access several systems. Flexibility may include substructures to account for system-specific data, i.e., data that do not fit the overall organization or that cannot be classified according to a mutual scheme. A substructure will involve the identification of individual sets of system-specific data so that the functions of storage, retrieval, and analysis can be performed on these data.

Research Testing

Research testing involves testing the validity of the taxonomy. It is a test program for the entire research. It includes establishing detailed user/generator requirements, and finding differences in size and make-up of the data pool and the operational data bank. Research testing is necessary for the completion of Stage 4 (see Section IV).

The objective of the research in the data analysis area is the establishment of a data pool that is adequate in content and organization. It must satisfy the requirements for efficient factual storage and retrieval, study of analytical techniques, and current awareness. It should also be adequate to initiate work on vocabulary. The taxonomy provides the initial design for indexing and organizing vocabulary.

The task taxonomy will be submitted to current generators/users of task data for evaluation to insure that the extraction process does not alter the meaning of data or the intent of original data generators. This verification activity will also supply output frames-of-reference of possible users of a data system. Answers are needed to such questions as: Given this kind of data base, what probable use would you make of it? What format would you expect for your output? Were there any additional data combinations required? A further application of a query capability is required for the manual testing of the task taxonomy to provide a vehicle for establishing a range of potential user frames-of-reference. One alternative is to develop an automated query capability to test and modify the experimental data pool structure according to the user requirements established for test criteria.

SECTION VI

VOCABULARY AND THESAURUS TECHNIQUES APPLIED TO HUMAN FACTORS TASK DATA

PROBLEM

The experimental data pool encompasses task data from a number of object systems. One objective of the data pool is to provide the method for making comparisons between and among object systems. Therefore, the vocabulary of human factors terms must be organized to permit not only the storage, retrieval, and output of individual object system data, but also cross-indexing or cross-correlation among and within object systems.

The creation and use of task data terminology has not been a controlled evolution. Human factors literature is characterized by rapid growth, heterogeneous data, and a catholic approach to the use of human factors terminology. The heterogeneity of human factors terminology is attributable to its originating in several scientific and engineering fields, i.e., engineering--all of its disciplines--psychology, physiology, medicine, anthropometry, and many others. As a result, redundant and ambiguous terms exist and special meanings are often used. Individual analysts--contractor and governmental agencies--have created specialized jargon influenced by their academic training and work experiences. This imposes a problem of interpreting and translating terminology and integrating the terms into a comprehensive, broadly useful vocabulary and indexing technique.

Requirements stemming from Project LEX, a Navy sponsored DoD-wide technical thesaurus developmental activity and the Committee on Scientific and Technical Information of the Federal Council on Science and Technology, Sub-Panel on Cataloging and Indexing, are considered in the vocabulary/thesaurus activity.

Specific Problems

Others have attempted to produce a standard human factors vocabulary. None of these attempts have met with any general acceptance. Therefore, this research activity investigates alternative methods of vocabulary building that do not require formalization.

General, document oriented vocabularies exist which contain human factors data. The Tufts University, Defense Documentation Center, National Aeronautics and Space Administration, and the National Library of Medicine have all developed such vocabularies for use within their systems, which can assist in structuring a preliminary task data vocabulary. These vocabularies are organized on either a general bibliographic level or for only one segment of task data. The rationale for cross-linking concepts, forming syntactical inferences, and retaining semantic context for the multidisciplinary terminology on all levels of specificity does not exist. Techniques used in building documentary vocabularies may prove to be applicable to data vocabulary construction. If this does not provide an acceptable solution, some original research will be necessary to develop new techniques for data vocabulary development.

The terminology available in the experimental data pool will be treated as representative of human factors task data. However, it is only a sample of task data--a subset of the total personnel subsystem terminology. If all words and terms presently used within personnel subsystem programs were compiled into one vocabulary, the resultant would be an excessively large and probably unwieldy lexicon. The size alone would cause a breakdown in application, without considering its extreme complexity. In addition, it would require an inordinate amount of computer storage. Methods of controlling vocabulary development must be imposed. This approach, however, poses a problem of its own. If a standard and controlled personnel subsystem vocabulary can be developed, will it receive general acceptance and use?

OBJECTIVES AND APPROACH

The overall objectives are to investigate and develop a human factors task vocabulary for the research pilot. This vocabulary effort and its mutual activities, such as the study of indexing and thesaurus techniques, is intended to determine the feasibility of:

- Standardizing terminology and definitions of task data

- Studying and using classification and other taxonomic concepts for vocabulary and thesaurus construction The results of the efforts in the companion area--Analysis of Human Factors Task Data, Data Relationships, and Classification Schemes (Section V)--will provide the concepts necessary to initiate the development of vocabulary techniques.

- Utilizing these techniques as a basis for building a future human factors task data management system

Specific objectives are:

- Compare the methods, techniques, and related problems for both freely-indexed and controlled-indexed arrangements of terms

- Study vocabulary development principles for locating and maintaining pertinently indexed arrangements of terms

- Investigate the techniques and related problems of thesaurus principles and determine the application of such techniques to the construction of a vocabulary

- Prepare lists of task data terminology using the data prepared for each object system represented in the experimental data pool

An approach to the vocabulary problem can be expressed as a linear extension or modification of the techniques used in document retrieval systems. A vocabulary provides the data descriptions which serve the requester as search keys. These descriptors are presented in arrangements corresponding to the various methods of referencing information e.g., by subject, author, topic, and alphabetically. The descriptors also serve as reference points for the storage scheme. Document retrieval techniques will be analyzed. This analysis will determine the probability of modification to provide the greater detail necessary for a data retrieval system. This detail must be provided in the areas of requesters' data concepts and search keys and computer program reference points. Modifications to these techniques do not involve a linear extension but require a very careful analysis to insure that the extension has not caused a cavitation or breakdown of the technique.

Methods and Results

The most striking phenomenon in contemporary, scientific, and technological research and development is the rate at which new ideas are generated and new terms are generated to communicate ideas. Emphasis and importance also ebb and flow. The indexing vocabulary needs to keep up with these changes as well as permit improvements based on searching experience.

The constraints imposed by a small vocabulary on the searcher are removed by encouraging the development of a very large dictionary of referral terms. This dictionary should, ideally, carry every term that may occur to the searcher in connection with a query. It should be capable of an unambiguous interpretation of the term.

A comprehensive vocabulary is needed in order to:

- <u>Achieve system efficiency</u> Since each new term, in effect, sets up a new filing point, an uncontrolled vocabulary would result in a continuously growing set of files that would soon become unworkable and unmanageable. Total efforts of the system would be consumed in servicing the files, rather than in giving service to the user.

- <u>Insure that each specific concept has only one frame-of-reference, and corresponding search path, in the files</u> Information can be retrieved only if all references to a specific concept can be compiled together.

- <u>Minimize grammatical, syntactical, and semantic problems</u> Word spelling and grammatical form must be standardized. Since the computer does not recognize nearly-alike words, the same term must appear in the same way every time for successful automated data notification and retrieval.

  EXAMPLE: Surface-to-air missile must not appear as:

  surface to air missile
  surface to air missiles
  SAM
  or any other variation

  The concept represented by a term must remain constant

  EXAMPLE: Pitch cannot be used to represent acoustical pitch on time, aeronautical pitch another time, or gear pitch at other times.

- <u>Increase consistency of communication</u> A list of standard terms will increase consistency among various users and generators by displaying an identical set of authorized terms. Proper generic and hierarchical levels will be indicated by terms used in the list.

- Increase retrieval effectiveness Retrieval will be more effective if a concept is represented by one term rather than by different terms scattered throughout the file.

A particular problem is that of generic level. The generic level of each term in the vocabulary must be that which will give a useful degree of discrimination without excessive fragmentation of concepts. Terms of a higher generic level simplify indexing, and result in a high rate of recall; however, data relevance will be lower. That is, since the term represents a broad concept, many specific concepts are subsumed under it. When this broad term is utilized to fill a request relating to one of the subsumed specific concepts, many "facts" will be retrieved. Many of these facts are irrelevant because they relate to all subsumed data concepts of the generic term and not specifically to the data term.

More specific terms can achieve more precise indexing. However, terms of a more specific generic level increase indexing difficulties and retrieval problems. A vocabulary containing many specific terms would be much larger than one containing more generic terms. The searcher would be forced to make more precise decisions among highly specific terms, and consistency in forming frames-of-reference would be harder to achieve. Retrieval would be more painstaking, since it must be as precise as the indexing.

The generic level of each term must be determined pragmatically by the anticipated uniqueness and volume of fragmented data rather than by an esthetic or theoretical formula.

A lesser, but nevertheless very real, problem is that of hierarchical arrangement of terms. A basic hierarchical structure is being investigated for its usefulness of grouping generic levels of data according to a succession of related orders of ideas. The same is tied directly to the amount of cross-indexing desired. Specificity is a function of the required level of detail contained in a data set. If the uniqueness of a specific class of data can be assured, it remains to associate the class in an ordinal scheme, involving all classes. The scheme would position each class according to the concept in which the class, or its contents, is referred. As discussed in Section V, the element of Performance Characteristics is assigned to a lower level and thus becomes subordinate to Performance Descriptions. However, the element Performance Characteristics can also assume a higher position. This is due to cross-referencing. The terms of Performance Descriptions are now subsumed under their characteristics. The problem is further intensified with the consideration of unique data within each generic level. Each level of specificity then assumes a geometrically increasing amount of detail.

As a hypothetical example, action verbs such as TURN ON and START might appear in a structure as follows:

<u>OPERATIONAL TASKS</u>

<u>ACTION VERBS</u>

ACTIVATE

START
TURN ON
ETC.

ADJUST

LOOSEN
SET
TIGHTEN
TURN
ETC.

Study of Existing Techniques

The problems involved in the construction of a data vocabulary are believed to be similar in scope, complexity, and in implementation to a documentation vocabulary. Therefore, vocabularies constructed for human factors data, automated, or semiautomated systems were reviewed. This provided familiarity with document or data vocabulary construction efforts.

Ronco (1960), Devoe and Saul (1959), and Ronco, Devoe, and Saul (1963) describe the establishment of a manual human factors document retrieval system centered about a topical outline. The outline has fifteen major headings, numerous subheadings and sub-subheadings. The outline provides a logically related set of descriptors that permit a systematic search to a relatively detailed level. Through contiguous and cross-referenced headings, the user of the system gains some knowledge even as he conducts his search.

Janning (1963, 1965, and 1966) describes, in considerable detail, the operations that were performed at the University of Dayton in the establishment and modification through vocabulary control, of a coordinate indexing retireval system for the Air Force Materials Laboratory (AFML).

Drew, Summit, Tanaka, and Whitney (1966) report on an experimental on-line reference retrieval system which uses a coordinate search strategy. The Automatic Data Acquisition (ADA) system requires that the user search a file of information categories, select the inquiry card, and initiate the system. If no category card exists, another descriptor is selected and input. The ADA system then assists the user in further searching.

Evans (1959) describes the marginal punched card bibliographic system used by the Human Factors Group at Douglas Aircraft Company. The index vocabulary was divided into twenty-six major subject headings. A glossary of terminology was then developed for each major subject. The glossary of vocabularies was next examined to assure relevancy and reduce redundancy. The vocabulary, arranged alphabetically, became the Detail Subject Index. From the Detail Subject Index, reference cards were punched.

The Foreign Technology Division of the Air Force Systems Command has developed a Centralized Information Reference and Control (CIRC) System. The CIRC System uses a thesaurus to provide a standard language for content description of input items. The thesaurus is established as the authority for content description--vocabulary--in all CIRC System information processing.

McMillan (1964) describes the procedures that were taken in developing an interim vocabulary for the CIRC system. The vocabulary was alphabetically arranged into thirty-three specialized subject areas. The terms were later arranged in a hierarchical structure within each subject area.

The review of current practices provided relevant applications of two vocabulary development techniques: (1) the subject area arrangement developed for CIRC, and (2) the term cluster arrangement for AFML. The subject area arrangement uses two listings. One is an alphabetical listing of scientific terms extracted from document search terms and validated by users. The second is an arrangement by sociopolitical blocks of countries. In the term cluster arrangement, the vocabulary terms are arranged alphabetically by subgroup-clusters of materials properties, i.e., tensile strength, and initial research into human factors data vocabulary building.

Vocabulary Building

The problem of reducing a comprehensive vocabulary to a manageable, controlled indexing vocabulary, containing all the concepts needed for storing and retrieving task data, would require a large effort in itself. Since the experimental data pool vocabulary will be relatively small, as compared with a total vocabulary requirements for an operational system, it is not necessary to establish a complete cross-referencing scheme. Cross-indexing may be accomplished by using term "codes," "see also," or other similar techniques.

Human factors task data, as explained in Section V of this report, are generated in support of any number of aerospace system development programs, and prepared by a variety of analysts. To provide a manageable standardized data source, the concept of an expanded preferred term must be enforced. The application of the preferred term concept differs slightly from prior applications of the concept in that it must also encompass the object acted upon.

For example: A toggle switch may be:

activated
set
pushed
hit
tripped

A pushbutton may be:

pushed
activated
hit
tripped

The preferred term concept must not only encompass the best possible definition for the term, but must also be tied to a specific human task performance activity. A preliminary list of verbs, key words, and synonyms, in common use by human factors specialists, was developed from the ALCC data (see Appendix VI). This list is being used as data for further analysis into vocabulary building. The task taxonomy (experimental data pool) is expected to include a task description format of verbs, objects, and modifiers.

Thesaurus Principles

Adequate control over an indexed vocabulary is often desirable to keep term proliferation to a minimum. Authoritative lists are established and appropriate cross-reference category lists are published showing the category headings approved for use. Additions of terms to the approved lists are made upon approval by a thesaurus committee. Suggested new categories--elements, and items--are submitted by system users.

One of the most widely used tools for vocabulary control is the thesaurus. This instrument is designed to alleviate the proliferation of vocabulary by means of synonym control. It allows the indexer to assign terms freely, but establishes "preferred terms" to facilitate retrieval. The "preferred terms" are nothing more than the synonymous terms most frequently used in the data or agreed upon to be used. "Structured" thesauri may be developed by superimposing a hierarchical scheme.

To keep the thesaurus a flexible tool, responsive to change in technology and concepts, new terms are added as required. These new terms may be empirically derived through free indexing. Periodically, the thesaurus is updated through the inclusion of new terms. However, no term should be incorporated without ascertaining its status, i.e., whether the new term is a synonym, or an independent keyterm. Usually a reviewing body is established to consider new terms and pass on their inclusion in the thesaurus as an accepted or "official" term. This review helps to maintain consistency, to resolve differences in interpretation, and to provide continuity.

Periodically, the thesaurus is purged of terms that become obsolete through lack of usage. As the state-of-the-art progresses, a certain number of terms become inactive as retrieval items. It is advisable to eliminate inactive terms to keep the size of the thesaurus within manageable limits and avoid obsolescence. Use of terms by requesters can be established by statistical means during computer operation. For instance, the number of requests against a term can be evaluated and the terms not used can be purged from the vocabulary.

Use of Links and Roles for Multiterm Relationships

A process used to associate one term or set of terms with other terms to denote semantic and syntactical interrelationships involves the technique of assigning link and role indicators. Links are appended to terms to indicate that a relationship exists; roles are used to show contextual meaning of the term. For task analysis data, it may be possible to utilize this technique for relating a specific term, an action verb for example, to a task within an activity for a particular mission phase. It can also be linked to several other tasks. But as the number of links grow, the amount of relevancy diminishes. A role can be used to facilitate the degree of relevancy. However, care must be taken because of ambiguous definitions assigned to role indicators. This is particularly true when information of a multidisciplinary nature is involved. Research indicates that both multidiscipline and system-specific requirements need to be carefully considered here. If the complexities of the multiterm relationships can be reduced to a set of generic rules, then roles and links may be of value in showing these meanings and relationships, and useful in computer based file structures for storage and retrieval purposes. While the use of roles and links has not met all expectations in other applications (Janning, 1965), it should be considered in this application.

Techniques for Establishing Conventions for Constructing Terminology and Phrases

A set of construction conventions and restrictions to the terminology (terms and phrases) is required for the research vocabulary. Following are some of the conventions which must be considered:

> The dictates of machineable vocabularies specify a maximum fixed-length term. Investigation into the use of vocabulary terms by human factors specialists is required. This investigation will determine the exactness of current terminology and assist in the establishment of these term limits.
>
> The use of hyphens to group compound factors which conjunctively are of interest, such as "motor-arm response time " or "skin-heat response reaction time," may necessitate a more complicated constraint on spelling simplification.
>
> The rationale of no-plural formats will be weighed against customary practices which generally employ plural forms. Prepositions, participles, and conjunctions can be eliminated unless they are necessary to the meaning or context of the term or phrase.

CONCLUSIONS AND RECOMMENDATIONS

The work accomplished in this research area was primarily oriented to investigating existing systematics in vocabulary application to storage and retrieval systems. The bulk of the existing techniques are directed to both manual and automated document handling procedures. There is also a heavy emphasis on the consideration of the user--user-orientation--in the development of most of the vocabularies.

The effort in the study of vocabularies was accomplished concurrently with the data study discussed in Section V. The results and approach in the relationships and classification schemes were reached as an extension of the requirements for vocabulary design. Thus, many of the results of the data definition research apply to this area as well.

The results of the study indicate the feasibility of applying to data problems those techniques used in building document vocabularies. There has been little indication of nonusability of techniques, with the probable exception of "links and roles." However, as more detailed analysis in the factual aspects--data items--of task data is completed, special problems with vocabulary building techniques may become more evident. For example, a term defined during data analysis may require a redefinition for use during analysis research. Which definition should be included in the vocabulary? If both are required, how can they be differentiated?

A difficulty with the human factors task vocabulary arises from a juncture of soft and hard scientific disciplines. Such terms as stability, behavior, performance, and response are shared but interpreted differently by the psychologist and physical scientist. The degree of this interaction may lead to the requirement of "scope notes" on such a high percentage of terms that a dictionary will also need to be prepared. The dictionary would be a listing of ambiguous terms and definitions. These ambiguous terms would be indicated in non-structured listings of the vocabulary.

Precision in the terminology is a large factor in developing a vocabulary. The precise terminology would clarify all that a concept includes, and all that it excludes. This exclusion is frequently an assumed factor in vocabulary work, but may require more explicit handling to satisfy users.

To insure a good product in any undertaking, it is desirable to establish safeguards. Quality control is especially important for information processing. This applies equally for senior personnel as for trainees. A system of checking, although costly, is imperative for an efficient information system.

One of the checking functions researched should be for completeness testing methods to assure that the information is fully represented by the terms or headings assigned. It is also important for future retrieval that generic as well as specific terms be applied. Paradoxically, the completeness check should insure that not too many indexing entities have been assigned to any item of information to forestall excessive noise in the total information system.

It is important to safeguard consistency of indexing. This applies especially for serial inputs, that is, data concerning an activity but being generated in sequence and perhaps by different sources. Although many terms assigned may vary, there always should be a core of indexing terms which will be the same for each item of information and, in effect, tieing the various individual reports together.

The application of the vocabulary will have an important influence on the format requirements. Since the system requirements encompass data management aspects, consideration must be given to some internal accounting mechanism tied in with content factors. This capability is highly amenable to making the products suitable for machine processing.

The emphasis in this research was on vocabulary building. However, exploration of thesaurus principles was also initiated. With only a finite vocabulary of human factors terms available from the experimental data pool, it is postulated that the research requirements (see Section II) can be satisfied with a rigidly controlled vocabulary. Thesaurus building principles (ONR-25, Project LEX) will be used as the vocabulary control. Thus, a rigidly controlled vocabulary will provide the capability of demonstrating the use and value of a human factors data thesaurus.

The approach, which involves the following activities, will be used to test the validity and utility of the vocabulary:

- Interviews conducted with vocabulary/thesaurus experts to assist in technique selection and application
- Interviews conducted with both potential data system generators and users and also human factors experts to ascertain the use and acceptance value of the developing vocabulary/thesaurus
- Generate a parallel effort, manual and automated, which would enable a given group of developers or potential users a flow of data constituted by two modes of dissemination acting on the same inputs
- Establish a review group of experienced thesaurus personnel to evaluate the vocabulary and determine the applications for a thesaurus in a human factors data processing system

SECTION VII

COMPUTER STORING AND RETRIEVING OF HUMAN FACTORS TASK DATA

PROBLEM

The storage and retrieval techniques used in information systems are usually considered fundamental in attaining efficiency in system operations. This is particularly true if the size of the data store is large, and response to input and output is important. Two important factors that must be considered in studying storage and retrieval problems are the computer hardware being used and the requirements of the people who use the system. A third factor, perhaps having prime importance, is understanding the data involved--the characteristics, volume and flow rates, formats, and the different ways the concepts and values are expressed.

Several aspects are involved in studying the application of storage and retrieval techniques for handling task data. The most basic of these is determining an effective data bank design. The design describes the structure of the data files and the methods for accessing the files for storage and retrieval, and is the most important problem being researched in this area. The design must consider changes in user requirements. In the development of aerospace systems such changes occur frequently.

Other related problems of concern to this study involve the implementation of storage and retrieval techniques. These problems include such contingencies as: (1) the use of an on-line retrieval capability, (2) machine-translation of data requests formulated in user-oriented language, and (3) consideration of the requirements of providing data to analytical processing programs. These problems also involve such factors as determining appropriate computer characteristics and programming techniques, such as program languages. This study will need to consider what types of computer equipment are used in aerospace system development, and what problems are presented if an operational system is built for one or more applications.

OBJECTIVES AND APPROACH

Early in the research period, it was recognized that an investigation of these system functions--storing, updating and retrieval--was dependent upon the results of the other research areas; i.e., analysis of data, vocabulary, and analytical techniques. The knowledge gained about how users will reference data, the kind of responses needed to satisfy user requests, and contents of the data will all impact directly and heavily on the investigation of this area. The objectives below are intended to meet the need for a concurrent effort.

- Explore system techniques being used in both aerospace task data procedures and in commercial information systems designed to handle information similar to task data The intent of this objective is to provide the research with an understanding of the present applications of storage and retrieval functions, particularly as they apply to human factors task data. This knowledge will be used in the design stage of the research pilot, either in directly utilizing existing techniques or in determining the need for modification of techniques.

- Identify techniques which are to be studied in detail for subsequent pilot development A more intensive investigation will be done for those techniques and software functions which are directly germane to the functions of the pilot.

- Identify one or more computer facilities for use in the development and testing of the pilot software The facility should include appropriate developmental tools to assist in the building and debugging of the programs. A programming language must be selected for use in the coding of the pilot programs.

  The investigation of available computers is intended to supplement the investigation of systems The type of computers available and their software capabilities are factors in selecting programs that can be used in the research pilot development. Language limitations, input and output characteristics, and storage capabilities must be considered.

METHODS AND RESULTS

Investigation of Information Systems and Techniques

Many types of information processing systems exist. The concept described in this report proposes a number of features which characterize the type of system to be investigated. Thus, the concept sets a standard for the investigation of existing systems and techniques. The investigation identifies, for further detailed research, systems that support personnel subsystem procedures or provide techniques that may be applicable to improving these procedures.

The following questions form the basis for investigating techniques of information systems:

Does the program design involve a technique for storing diffuse, heterogeneous data similar to task data, and also provide a multilevel file organization that is relevant to factual retrieval?

Does the program involve techniques for searching in-depth for subsets of data in a manner that by-passes an inappropriate data item thus introducing efficient retrieval strategies?

Is the retrieval function responsive to user-oriented retrieval requests? Does a user-oriented vocabulary play a significant role in the storage and retrieval strategy?

If a technique appears appropriate for task data, is it one that can be applied directly in its present form or does it require modification?

Are adequate documentation or other information sources available for further review?

What techniques of an advanced nature are used in system operations? These include time-sharing, and near-English queries.

Discussed below are systems that were studied in detail. A further breakdown of the systems is given in Appendix VII along with references to field trips and documentation.

The Formatted File System (FFS) was developed for use on IBM 7090/7094 computers. It is an operational system used primarily in the Intelligence Data Handling System (IDHS). It has a file maintenance capability which handles both numerical and alphabetical data. The system files are tape-oriented, although the latest model (Mod 7) uses disks for supplementary storage. A user of the system may specify several retrieval options for high selectivity; "and, or, not, equal," are just a few of the options available. A variable formatting subsystem in Mod 7 provides the user with any number of alternatives of organizing data output. The system is programmed in assembly language and is quite large. Approximately thirty programs are involved.

RECAP--Reliability Experience Correlation and Analysis Program--is a modified version of the Formatted File System. The system is being designed and developed for Ballistic Systems Division, AFSC, by Aerospace Corporation. It is being built to handle aerospace systems reliability and maintainability data, such as that contained in component and parts test data, maintenance action reports, and reliability summaries. The system allows a user to specify several retrieval options, including ranking, sorting, and conditional data parameters. Graphic output is also available for off-line printing. The system is being modified to provide modularity of program functions. That is, special programs may be fitted into the system to support unique users needs for data analysis processing. Since the investigation of that system, RECAP has been redesignated System Effectiveness Correlation Analysis Program (SECAP). Also, modifications to its design are planned, or are being made.

North American Aviation is using a system--the Apollo Task Data Bank--which processes task description data. The data include task functions, task elements, time-line information and equipment designators. The system provides time-line computations and allocates tasks for specified missions. A consistency check technique built into the system determines if certain actions are "legal." For example, a check is made to determine if a task performance involving a switch setting requires a prior reset action. The system is programmed in FORTRAN for IBM 7094 computers. The data files are tape-oriented although disk storage is utilized.

RCDMS--Reliability Center Data Management System--is being developed at Rome Air Development Center by Auerbach Corporation (Sable, 1964). The system handles reliability data dealing with equipment such as diodes and transistors, for reliability analysts. The system is a generalized data management system and is designed to handle a variety of data problems. The retrieval function, for example, responds to on-line queries. The storage logic utilizes a form of tree-structures and cross-file data linkages, involving a variety of indexes. A feature of the system allows for setting up a logical file structure without modifying the physical file structure. This means that a user may specify a data organization that fits his unique needs, but the basic file structure remains static. Selective retrieval options, as well as elementary analytical processing, are available to the user. The system is written in JOVIAL, a high order programming language, for the UNIVAC 1218 time-shared computer at RADC.

TDMS--Time Shared Data Management System--is a general-purpose system being developed by the System Development Corporation. It is designed to handle large volumes of mixed data. It includes a retrieval function that responds to user-oriented queries as well as to processing requirements of other programs of the system. The file organization is emphasized in the system which permits data to be classified on several levels. A technique of automatically associating hierarchical data relationships is built into the file design. In addition, a concordance index is included to increase the number of reference points. These techniques provide the capability of retrieving, by use of classes of data or values of data, for reference points. Elementary analytical routines are included as part of the system. The query capability uses a nonprogrammer language and numerous selective retrieval options including a report generator for formatting retrieval replies. The system is being programmed in JOVIAL for the IBM 360 models 65 and 67. It will operate in a time-shared mode providing remote on-line input and output services.

TDMS will replace LUCID (Grant, 1965) an existing time-shared retrieval system implemented on the IBM AN/FSQ-32 computer. Advanced Research Projects Agency (ARPA) is supporting the use of the Time-Shared System (TSS) at SDC. LUCID, Language Used to Communicate Information System Design, employs a concordance file index that provides direct linkage to data base files for all data categories and, in addition, for unique values of data. The concordance lists all values and cross-links these values with a list of the data categories; both lists are linked to the base files containing the actual data.

All systems discussed handle large volumes of data. With the exception of the Apollo system, they are used for retrieving specific units of data. The Apollo Data Bank is more specifically used to produce refined or simulated results, based on execution of selected routines, which use prespecified units of data stored in the data base files.

Two of the systems, RCDMS and TDMS, are employing techniques in their storage and retrieval operations which orient the system to users not necessarily knowledgeable of the system file structure. They provide on-line interaction between users and the system with time-shared, remote query consoles. RCDMS and TDMS use random access storage devices, and are designing file structures to take advantage of random-access data selection techniques. The selection techniques include auxiliary file indexes to pinpoint data base file groups as well as to structure the data base files according to primary search strategies. Both systems are written in JOVIAL.

RECAP is a formatted file system used for reliability data problems. Some modifications are being made in the development of RECAP. For example, two functions being modified are floating point calculations and executing nonsystem routines. FFS, like TDMS and RCDMS, is a highly selective retrieval system. FFS uses tape-oriented data files; however, it uses a serial search strategy. The user must be knowledgeable of the structure of the files to allow effective retrieval. FFS employs a cross-file index technique. The technique allows a user to reference data in individual files and cause the values to be used as selection criteria for retrieving in other files. This is done by the system in a single retrieval process. FFS can automatically create additional data files using data which were retrieved from a set of existing files. The user designates the structure of the new file in his retrieval request. Lengthy or complex requests, involving several data categories, data associations, and processing instructions, can be stored by the system. These stored requests are then referenced by name and processed. FFS is written in an assembly language oriented specifically to the IBM 7090/7094 computer.

The Apollo Data Bank is specially designed for performance data processing, involving techniques primarily concerned with analytical processes. Such processes involve sequencing task statements, modifying task elements, and developing mission performance specifications. These specifications involve time constraints, and performance and equipment check-off procedures. This system is not designed for storage and retrieval applications. However, due to the systems orientation to performance data and the techniques employed to describe and process data involving tasks, task elements, time variables, and equipment description; a further study of this system is needed.

The RCDMS and TDMS systems are considered directly applicable to the pilot study. Many of the capabilities described in the operational system concept correlate with many of the techniques employed in those systems. Generally, these capabilities are: time-sharing, mixed data storage, user-oriented retrieval, remote query interaction, random-access storage, and multifile indexing. Another important aspect of these systems is that they are general-purpose information systems. The importance of being general-purpose, which is usually a system design objective, lies in the adaptability to various data problems. Also, general-purpose techniques are used so that single systems handle different data problems concurrently. Completely varied data bases are handled in a common manner.

The storage and retrieval techniques that are being considered for the pilot are discussed in the following paragraphs. Many of the techniques described are used in the RCDMS and TDMS systems.

Investigation of Pilot Techniques

The investigation of the techniques, which can be applied to the pilot, is tied directly to the development of the task taxonomy and vocabulary techniques. The results of the studies in the current awareness and analytical processing areas also have an impact in the design and application of storage and retrieval techniques, but to a lesser degree. This lesser impact is due to the nature of the interaction of those two areas with storage and retrieval operations. Current awareness is almost an independent operation, with its prime interface existing between profile structure and the organization of incoming data. Analytical functions require that retrieved data have a highly structured framework, but an interim formatting routine to restructure values will perform as the interface between analytical and retrieval functions.

The taxonomy describes all classifications of data and cross-relates facts among data groups. The structure of the data base will need to follow the taxonomy framework. User requirements for referencing data facts are also a significant factor in setting up a storage and retrieval strategy. Detailed requirements for establishing and maintaining reference points are included in the indexing techniques used in constructing the vocabulary.

Storage Techniques

The structure of the data base files is an important aspect to storage techniques. The structure is dependent upon the need to access the data contained in the files. In some systems, simple structures are adequate. For example, bibliographic data are often organized by subject or title, and associated data such as date of publication, author, and keywords are all linked or grouped with a primary reference point, e.g., subject. Searches are made using the subject as the prime search key and associated data are thus retrieved with reference to this primary key.

More complex structures are needed; however, when the requirements to retrieve data involve a higher level of selectivity. A greater degree of depth is needed if many subassociations are implied in the data. This is the case of handling task data on a factual basis. The data can be subdivided into several levels of detail.

Performance descriptions, discussed in Section V, are divided into three levels--Position, Job, and Task--and each level contains further categories, e.g., verb, object, and noun. In addition, each description can be potentially associated to several categories of data, such as time, equipment, performance characteristics, and mission. Within each of these associated categories are subcategories of data. Even these subcategories may need to be directly associated with categories not a part of the particular hierarchy in which the subcategories are classified.

Hierarchies of data can be handled in both physical and logical arrangements of data in storage files. A physical arrangement involves using storage equipment characteristics, such as tape reels, tape files and records; or disk units, disk channels, and tracks. Logical arrangements, the most detailed of the two, involves the grouping of data in fields and sets so that the respective location of one entry implies its association with other entries.

Embedding entries is a technique that can be used for storing hierarchies of data. An array of entries is formed in the data structure so that the location of a group of similar data, stored in a particular file entry, provides the reference point to the location of other data that are associated with the particular file entry, e.g., job descriptions involved in the mission phases.

An illustration of a logical array is given in figure 10. In the illustration, the entry containing position description C in the position description array is in the same logical position as its associated entry C in the

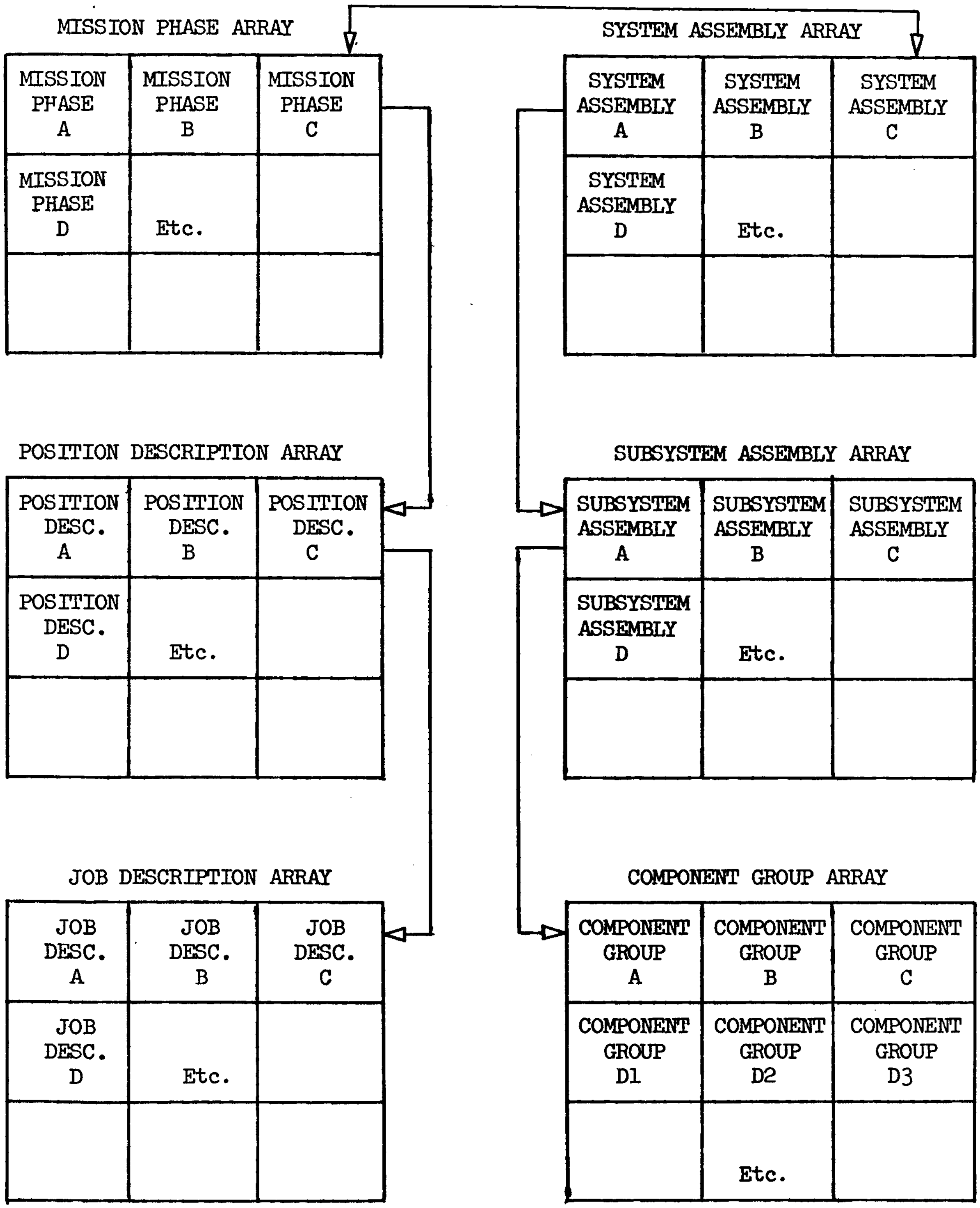


Figure 10. Logical Organization of Data Groups

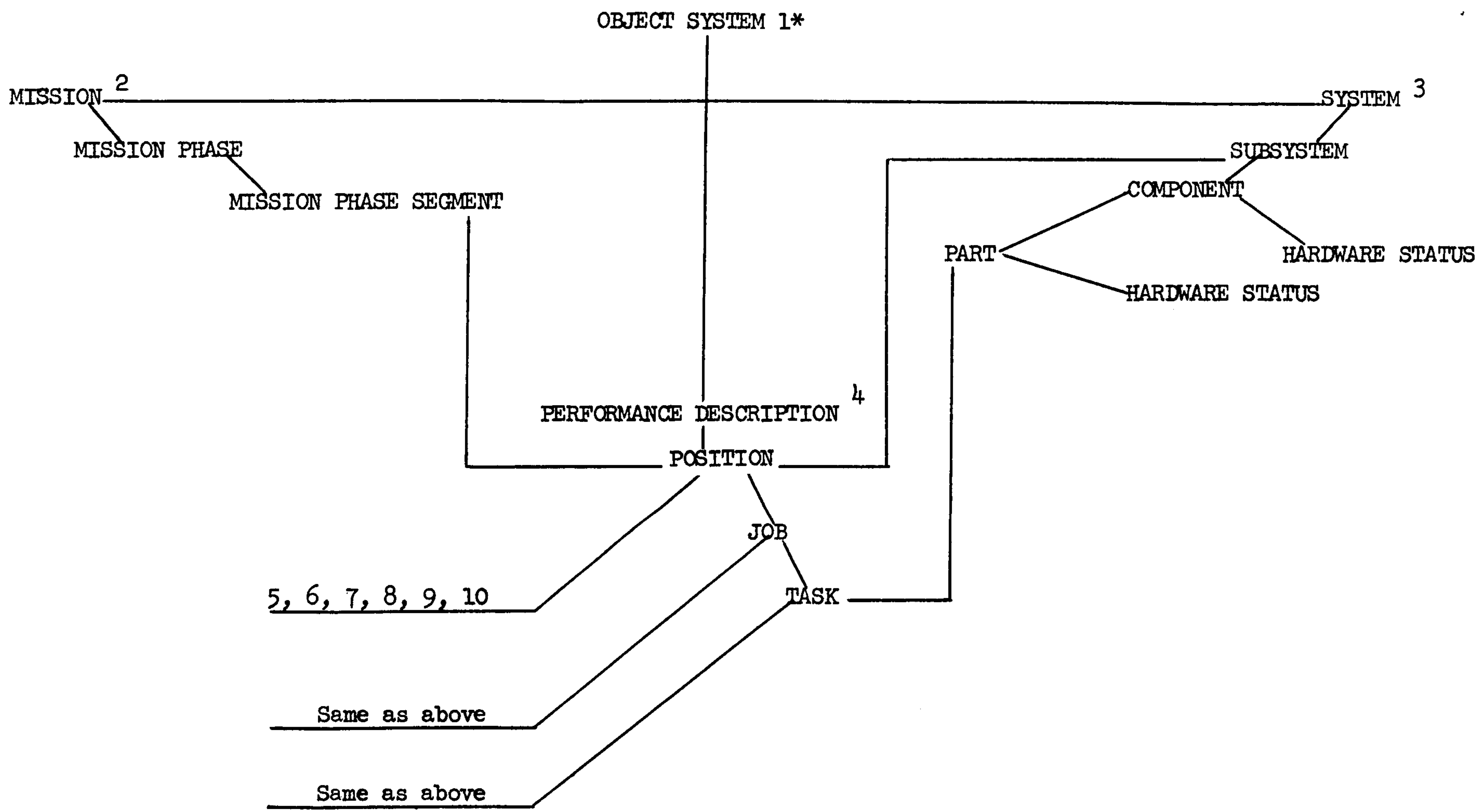
mission phase array. The same arrangement is followed in the job description array, where the job C entry involved with position description C is similarly located. More complex arrays can be organized to provide for several levels of detail. For example, if several groups of components are involved in subsystem D, then a string of subordinate entries are arranged in connection with the prime entry--see component group D_1, D_2, etc., in the component group array in figure 10.

The horizontal arrangement of the arrays implies a subordinate grouping of data categories. However, any array can be considered on the same level as an associate array. Information concerning the entries in one array can be arranged so that a particular entry is logically connected to an entry in another array. For example, in figure 10, mission phases are in a different hierarchy as system assemblies. However, by logically connecting mission phase C with system assembly C, either entry will point to associated data in the other hierarchy. The system array, then, would be associated with subordinate arrays that contain subsystem entries, etc. This type of data base structure would help satisfy the classification scheme described in Section V. Figure 10 illustrates the scheme under investigation.

Further cross-linking of data entries is necessary if the user wishes to retrieve data with reference to other data that are not arranged in the same hierarchy or logically located. Figure 11 shows three hierarchies--mission, system, and performance--with object system as the prime entry. The line connecting task to part, for example, indicates a cross-link across hierarchies for parts that are associated with tasks, and vice versa.

A technique to cross-link data groups in storage involves the use of auxiliary indexes. Task data in the data base files would be organized according to the most often used search paths or hierarchies and implies a fixed structure. For example, such a structure may cause performance descriptions to always be grouped with performance times. Thus, the data base provides for direct association of the performance descriptions with their associated times. Assume that it also provides for direct association of entries of components and equipment characteristics. However, if user A needs to retrieve performance descriptions with reference to the equipment characteristics involved, a duplicate grouping is necessary unless a cross-reference is provided. Such cross-references are stored in auxiliary file indexes. These cross-references are index pointers that are used to select alternate paths to search for data with respect to other data. Thus, the string of task entries in figure 11 that are connected to a certain job need not be logically arranged precisely as the string of part entries. An auxiliary file completes the linkage necessary to associate one entry with another. Too many cross-links burden the system with excessive look-up for data searches. The number of cross-links, then, must be kept to a minimum to provide a trade-off between providing "over-retrieval," and "under-retrieval." Both RCDMS and TDMS employ auxiliary file indexes to provide multi-reference points.

Another alternative is to organize the data into duplicate or supplementary files arranged according to the different reference requirements. All data probably would not need to be duplicated. Some data, for example, will be referenced similarly for all users. A natural hierarchical arrangement of systems assemblies--such as a set of part data subordinate to a set of component data which is subordinate to subsystem data--is an example of an



* Element number from the ten data elements. See Appendix V for all element descriptions.

This illustration is an example of relationships within an ordered data structure. The structure represents associations among prime elements at various levels. It also shows that relatively few elements are directly tied together, but the capability of referencing all data selectively exists within the organization. A further and similar organization of subelements (5, 6, 7, 8, 9, 10) is needed to complete the hierarchy.

Figure 11. Data Element Output Relationships

organization that can reflect mutual access points for several users. This type of supplementary storage becomes too bulky if a large volume of data is stored and various access points are desired. The data may be referenced efficiently, but the cost of storage equipment needed to contain the duplication of data files overrides the efficiency of indexing.

A technique used in document retrieval systems for indexing information is the inverted index. The inverted index allows for a generic arrangement of index filing points, whereby the filing points for a "family" of data items, such as specialty codes, are grouped under each key descriptor or code number for that family or group.

The inverted technique requires as many indexes as there are cross-associations. If a certain specialty code is to be associated with categories of task descriptions, special skills, system assemblies, etc., each file location of the specialty code is listed under those specific categories. The complexity of association increases if each of the various items of information is to be associated with all other items. In a large store of task data having many thousands of subsets of distinctive facts, complete association using inverted indexes would compound the problem beyond efficient machine handling capability. The storage of indexes would be many times the size of the actual data store. However, inverted indexes are efficient for shallow retrieval, i.e., retrieval concerning relatively few major categories. Inverted indexes should be considered in an operational data system for those files containing bibliographic references indexed to the store of documents, charts, etc.

Retrieval Techniques

Retrieval techniques are, in many ways, identical to storage techniques. The strategies used to organize and index data within the files for storing data are essentially the same strategies used for searching and referencing data from those files. However, another facet of retrieval exists; the combinations and conditions of the data that are retrieved. Those techniques needed to make the decisions regarding final selection of data items with reference to other items are discussed here.

Many retrieval systems provide for high selectivity using various retrieval logics. The application of a retrieval logic allows a user to indicate, in his input request, conditional elements which are interpreted by the retrieval program and used to limit or combine data results. For example, suppose a request for maintenance task descriptions is made concerning a hydraulic tank and pump unit, but the requester is interested in only those tasks which require the concurrent performance of more than one person. The entry for task descriptions is connected in storage to the entry containing hardware, i.e., tank and pump unit. Using the direct association, the descriptions are located. However, further selectivity is needed. This is accomplished by checking the manning allocation for the number of assigned personnel, determining the relative start and stop times of the performance, and providing the result for each person involved in the tasks.

The techniques normally used for retrieval combinations involve subjecting data values to symbolic logic. A Boolean technique used in many systems is one of interpreting logical properties of statements; for example, or; and;

not; equal; not equal. These properties serve as logical functions which assign one or more elements of a retrieval statement to other elements. In formulating a data request, a user will connect the data descriptors with the logical function according to the combination or selectivity desired. Example: Retrieve tasks <u>with</u> hazardous index <u>equal</u> to 3 <u>for</u> job activate missile engines <u>and</u> job activate missile guidance system. In the example, the logical function <u>with</u> assigns a relationship between tasks and hazard--hazard is a subcategory of the element performance characteristics. Further, the function equal to provides for a degree of hazard to be specified. A value of three could stand for an abnormal hazard condition. The function <u>for</u> assigns the prime elements of the statement, task, and job. <u>And</u> assigns a must condition for locating job descriptions, i.e., both jobs must be used in the retrieval strategy.

Data parameters, such as the number "3" in the prior example, can be included in connection with any data having some measurable characteristics. The parameters are either numeric or alphabetic. If, for example, the index for hazards is defined in narrative terms--e.g., gaseous, electrical, extreme, minimal--then those terms are used as the qualifying factors.

Another condition exists for retrieved data. If a small number or individual facts are retrieved, the format of the information is usually a simple listing of the data. But, in the case where several or complex information items are retrieved, a special arrangement of presentation may be necessary. In document systems, a ranking technique is used; for example, bibliographic data are presented by year of publication in ascending order. A system that handles personnel data could arrange employee information by years of service or academic degree. Fundamentally, common sorting techniques are used which key on terms and values of primary, secondary, etc., levels. Examples of this technique applied to task data would be to present certain task descriptions by their relative start times, by arranging specialty codes in descending order, or listing equipment by the mission phase.

An important function of retrieval is the language used to formulate retrieval requests. A user's request consists of one or more statements, each serving to direct the retrieval function in responding to the total need for data. The parts or elements of a statement are generally identified below with examples:

- Data parameters, e.g., <u>launch</u> (mission); <u>support adaptor-S12/A</u> (equipment nomenclature)
- Data descriptors, e.g., <u>tasks</u>; <u>subsystems</u> (tags for data categories)
- User identifiers, e.g., Performance Evaluation Group; South Aviation Corporation
- Processing directions, e.g., <u>perform</u> manning summary; <u>retrieve</u> Apollo LEM hydraulic components (directs the computer to execute various routines)
- Logical functions, e.g., <u>and</u>; <u>equal</u>; <u>for</u> (connects elements of a retrieval statement)

A fundamental problem in designing a retrieval language is making the tradeoff between man and machine communication requirements. Users desire to express requests in natural language--English syntax--whereas computer logic does not permit the use of abstract expressions so common in human communications. There are on-going attempts to solve the problems of advancing man/computer communication. Kellogg (1966) has approached the problem with a research system accepting nonprocedural requests for fact retrieval. "Baseball," a research program developed by Lincoln Laboratory, uses a relatively small, but flexible, natural query language to communicate English requests for facts dealing with baseball data stored in computer files.

The degree to which English language forms can be used in retrieval requests is difficult to determine. This research program will rely on results developed in other research programs to provide the capability for using syntactic rules rather than highly coded expressions or relationships. However, the word forms to be included in the system vocabulary will alleviate a great deal of burden from the user. That is, the vocabulary terminology will be as near to actual data expressions as possible, and other elements of a retrieval statement will approximate user-oriented terms. The request will still need to be formed according to rules acceptable to the request translation program, but sufficient flexibility can be built into the rules so that users feel they have freedom to format requests.

Programming Languages

There is an ever increasing use of higher-order languages in developing computer systems. Higher-order languages allow the programmer to state his instructions by using self-explanatory English words and other symbols easily recognized by programmers. These languages are usually tailored for general purpose applications, that is, programs developed in a higher-order language can be run on various computers. Reprogramming, caused by transferring a system to other computers, is minimized by using translator programs that convert standardized higher-order program statements into the basic machine instruction sets. In addition, program maintenance efforts are facilitated by the comprehensiveness of the higher-level program statements. Two procedure-oriented languages, FORTRAN and COBOL, were recommended by Whiteman (1965) for programming an operational data system. Both languages incorporate a wide range of language requirements, e.g., file generation and maintenance algorithms, input and output MACROS, and matrix indexing. They are used in programming several different computer types and are applied to many different system applications. A third language, JOVIAL, is also being considered. It provides a broad spectrum of capabilities including scientific processing combined with detailed file management algorithms. JOVIAL incorporates procedures for optimum packing and manipulation of mixed data terms.

Assembly languages are equipment oriented and require more programmer time for building programs. These languages are generally based on a one to one ratio between a program statement and a computer instruction. However, instruction sets are usually provided where a routine of instructions are built into the program with a single symbolic notation. These languages do provide the capability of building greater efficiency into execution times for programs due to the ability to code precisely the instructions that best fit a problem, or the characteristics of a particular computer. However, it is this preciseness that complicates the maintenance of a program or reprogramming it to fit other computers.

A trade-off between optimum efficiency for "run times" and the programming effort spent in developing and maintaining computer programs is a prime factor in selecting a language. Optimum program efficiency is a major consideration for the operational system, but higher-order language programs have demonstrated sufficient program efficiency in current information systems, as well as provided for easier development and maintenance.

The selection of a programming language for the pilot is dependent upon whether pilot programs are written or if an existing system is used. If the latter course is taken, the need to select a language is obviated.

Selection of Computers for Pilot Development

An appropriate computer facility is needed to develop and check out the pilot software as well as to be available during the final stage of the research program, i.e., the verification and field testing of the pilot. The same facility is not necessary for both development and verification. However, if two facilities are used, they must have similar components and characteristics to eliminate redesign or reprogramming. The major considerations in selecting a facility are listed below.

- The machine must be general purpose. It should incorporate both high processing capability and rapid input and output characteristics. It should be capable of running large general purpose programs.

- The storage equipment should include magnetic tape for permanent storage and magnetic drums or disk for random-access capability. A large core memory is needed to facilitate extensive processing operations, such as sorting and large file searching, while providing for lengthy programs.

- The computer should be openly available for use. There should be no unreasonable restrictions that prevent easy access to the machine.

- Program developmental tools should be available to assist the programming effort. These tools include compilers which meet program language needs and diagnostic routines.

- The facility should exhibit as many characteristics of the operational system as possible. Thus, pilot techniques can be patterned in an environment approximating the operational system.

Two facilities were selected for consideration for pilot operations.

1. Research and Technology Division (RTD), Wright-Patterson Air Force Base.

   One of the RTD facilities uses an IBM 7094/7044 Direct Couple Computer System. The storage units include a large core memory and tape and disk equipment. The direct couple complex requires that data base files be tape-oriented, i.e., serial tape entries. All input/output transfers and program executions are controlled by a stored executive subsystem, which is oriented to batch processing of program operations. There are no storage and retrieval systems available; however, the Formatted File System is being considered as an addition to the RTD software library.

2. System Development Corporation (SDC), Santa Monica, California.

   The Command Research Laboratory at SDC uses an IBM AN/FSQ-32 computer. This computer, supported by the Advanced Research Projects Agency, is used in a time-shared mode with user interaction consoles located at remote sites. The equipment configuration includes a large core memory, random-access storage, disks and drums, and magnetic tape storage units. Several storage and retrieval systems exist, including document, personnel data, and a general purpose data management system, LUCID.

   The Q-32 computer will be replaced in the near future with an IBM 360 Mod 67. This computer will also operate in a time-shared mode using remote query consoles. The TDMS system will be implemented on this computer.

The capabilities of the two existing computer complexes at RTD and SDC are similar in many respects. Both are large scale machines with extensive storage available. Higher-order programming language compilers are used and adequate programming checkout tools are available at both facilities. The Q-32 facility provides an on-line programming technique, TINT (Brewer, 1965), by which a user may build and operate programs simultaneously. The technique employs a simplified language set that is oriented to users that are not programming specialists. An on-line development of programs within the scope of the research pilot would be difficult and expensive using remote consoles. However, small auxiliary routines can be built and operated in conjunction with the regular pilot programs.

CONCLUSIONS AND RECOMMENDATIONS

The investigation of information systems techniques indicates a wide range of capabilities in storing and retrieving data. There are applications in engineering, personnel, intelligence, materials, criminology, etc. In some cases, special purpose systems are employed which are tailored to specific, current requirements. In other cases, general purpose systems are used and tailored to broader, more flexible applications.

Techniques used in the structuring of data files, program applications, and retrieval strategies all vary with each system. The factors involved in the use of those techniques are dependent upon the computer equipment, response requirements for input and output "flow" rates, volume of data, and functional requirements such as retrieval, and analytical. A prime factor also involved is the nature of the data to be handled, i.e., whether it is highly structured numerical data, textual data, graphical, or combinations of these.

A fundamental conclusion from the investigation is that the state-of-the-art for information systems provides sufficient software techniques to handle task data on a factual level. The major requirements then, are to define the data adequately, express their characteristics, such as classes, and relationships, and determine what demands will be made against the data, i.e., user requirements. These requirements primarily constitute the objectives of the study carried out in the data analysis area discussed in Section V of this report.

The results of data analysis and classification techniques show that task data, when observed on a factual basis, contain many classes and are interdependent with respect to most categories. However, the data appear to fit a manageable organization, i.e., one that is adaptable to information retrieval systems. Systems technology used in the TDMS and RCDMS systems can handle the majority of the problems presented by automating task data procedures. Such problems are: the heterogeneity of the data, the many index points needed to reference individual facts, the frequency of user requests, and the rapid response needed to service human factors specialists.

One problem that has not been investigated sufficiently is the potential size of a task data bank. Also, the rate that data would flow into an operational system is important to the design and execution of a system. A proper balance is necessary between the rate of storage and the output rate which involves new data notification, basic retrieval, and analytical processing. It is possible that the need to store large volumes of data on a frequent basis may overload such systems as TDMS or RCDMS. This problem needs further investigation. The potential flow rates of both input and output are required in order to determine a proper design for the operational system.

Exploration of existing and planned information systems should continue. The state-of-the-art of computer technology, both hardware and software, will continue to be advanced. For example, an advancement in the machine translation of natural language queries would have a great impact on the utilization of an operational system.

The use of an existing, general-purpose information system is strongly recommended for the development of techniques for storing and retrieving task data. The effort required to design and develop new programs is a costly one and may or may not result in an effective payoff. The recommendation to use an existing system for pilot experiments is based on the availability of a system that provides sufficient task data handling techniques for research. It is not necessary for the system employed to be the ultimate in handling capability, or even to possess all the characteristics conceived for an operational system. The techniques that will be recommended for a proposed operational system can be deduced from practical experimental application. Therefore, both LUCID and TDMS should be used as pilot systems. LUCID can be used to initiate the framework of the experimental data pool, described in Section V, and for the study of on-line retrieval techniques. TDMS should be used as the system for completing the research on information retrieval for task data. TDMS is capable of handling a data structure consisting of many classes and levels of data. TDMS also can be used to demonstrate, in a field environment, the techniques of on-line, time-shared retrieval and processing. The Q-32 computer facility is also recommended as the computer facility for pilot development. It provides the widest range of capabilities for the continuation of the research program. Existing data management systems, random access storage, time-sharing, remote user consoles, and on-line programming are the major reasons for the selection. In addition, a greater capability will be provided as one of the newest generation of computers becomes available--the IBM 360. This computer will provide the research with the opportunity to investigate the capability of advanced hardware in helping to solve task data problems.

SECTION VIII

ANALYTIC AND SIMULATION TECHNIQUES APPLIED TO HUMAN FACTORS TASK DATA

PROBLEM

General Problem

The importance of analysis and product simulation, as tools for facilitating the design and development of aerospace system, is well recognized by both the government and their contractors. Irrespective of contract obligations, many contractors rely heavily on computerized simulation for forecasting reliability and the effectiveness of design changes. Hannah et al (1965) reports that necessary design changes have been made as a result of computer simulation in early development cycles. Such changes would have been impossible under other circumstances.

The application of simulation and analysis to personnel subsystem programs during system development cycles is illustrated in figure 12. Computer simulation in the conceptual, definition and early acquisition phases facilitates the estimation of costs, development steps, man-machine trade-offs, design of training devices, etc. As systems move into production and operation, the need for simulation modeling decreases. Where simulation continues during this period the payoff, in design change recommendations, decreases. The system, for all practical purposes, is built. Testing programs, such as PSTE, that are initiated during the latter periods of the development cycles, rely on statistical and other analytic processes. It is during this time period that a human factors specialist has the opportunity to evaluate the effectiveness of his early estimates--some of which were based on simulation modeling--through the application of significance testing on PSTE data.

The information in figure 12 is based on a generalized concept of the uses of simulation and analysis tools in system development. In practice, quantitative techniques are applied to system problems in varying degrees from system to system and from time to time. Results from questionnaires, distributed by Hannah et al (1965) to representative government and contractors agencies, show that program and personnel subsystem managers are using computers for decision making functions such as cost estimates. Department heads are using computers for general simulation purposes. Although nonmanagerial personnel are making little use of computers, approximately 75 percent of their responses to the questionnaire indicated that they could make use of computers for simulation and analysis purposes. Several factors are assumed to be involved in this lack of use by nonmanagers: (1) adequate tools may be lacking, (2) access to a computer may be difficult, and (3) the compressed developmental schedules are such that the specialist is forced to rely on expertise, rather than subject his time to the lengthy and complex process of simulation modeling.

The research on simulation and analysis techniques addresses itself to the data system requirements listed in Section II and is intended to answer the following questions:

- Do current simulation and analysis techniques lend themselves to pooling into a generalized data handling system?

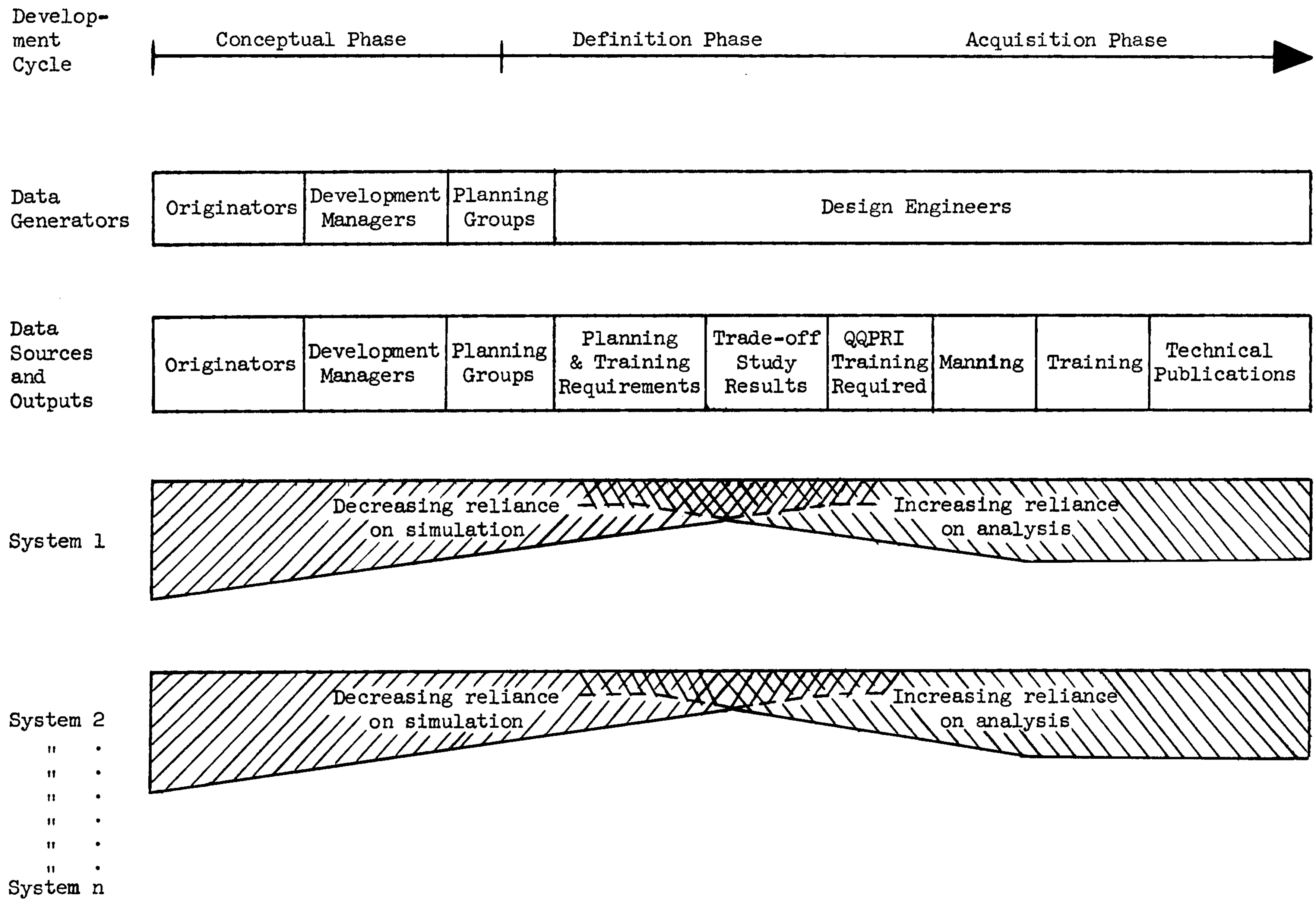


Figure 12. Application of Simulation and Analysis During Object System Development Cycles

- What procedures must be followed in the creation of a multisystem data base that lends itself to quantitative operations?

- Can a data system provide ready access to simulation and analysis tools for both managers and nonmanagers?

Specific Problems

Redefinition of Data for Analysis

Preliminary decisions, resulting from a review of the general problem, defined several specific subproblems: definition of analysis content and analysis units for the experimental data pool, selection of measurement characteristics, generation of data, analysis and simulation techniques and applications, and interfaces with other research efforts.

- Definition of Analysis Content and Analysis Units A general definition of human factors task data is contained in Section V. The efforts required in this research provide for a reevaluation of these general definitions to extract the data contents that are amenable to analysis and simulation. Once these basic categories have been identified, an analysis is required to identify the quantitative and nonquantitative contents--analysis units. This analysis includes the identification of dependent and independent variables within each analysis unit and among analysis units. Coordination with the data and vocabulary research efforts is required.

- Measurement Characteristics Each analysis unit must be analyzed to identify the measurement characteristics contained therein. The identification of these measurement characteristics assists in the specification of the mathematical class and distinguishes the contents of the class, within each object system's data, for each analysis unit. An understanding of the measurement characteristics assists in the selection for research on analysis and simulation routines.

- Generation of the Data The organization and definition of data for analysis and simulation is especially rigorous. The stored data must be organized so that any individual data, whether quantitative or nonquantitative, is capable of retrieval. Care must be exercised in the software design to insure that none of the characteristics of the data units are mutilated or destroyed during the retrieval process. The resolution of this problem can only be met through close coordination with the research software design effort and an exact definition of the analysis and simulation analysis units.

- Analysis and Simulation Techniques and Applications The selection of analysis and simulation techniques and application decisions will depend on a number of decisions made during the conceptualization of the operational data system and the definition of the content of the research. Additional restrictions or requirements will depend on research performed to provide answers to the following: What is the nature of the generalized techniques and applications and how does this differ from the nature of system specific techniques and applications? What kinds of analysis and simulation techniques will have the greatest initial payoff to human factors specialists? What kinds of analysis and simulation are prime necessities for human factors specialists; are these the initial demonstration items; if not, what are the maximum return trade-offs? What is the scope of the initial applications?

Specification of User Requirements

Anticipated users of the data system within both government and industry extend from lower top management to engineering and human factors specialists. The requirements of such a wide spectrum of users for operable analysis and simulation techniques extends beyond the present state-of-the-art. Research into these requirements to identify break points and trade-off points is required. When user requirements have been defined, techniques that satisfy the greatest number of the user requirements must be identified.

Current trends in the application of analysis and simulation, during aerospace system design and development, have resulted in the creation of a number of object system or operating system/subsystem specific techniques. For example, a simulation technique (Inaba, 1963) was developed to simulate the characteristics of the maintenance functions for the inertial guidance system--an operating subsystem--for the TITAN II--an object system. Considerable difficulty, possibly complete reprogramming, would be encountered in using this model either with another operating subsystem or with an inertial guidance system in another object system. Indications are that the converse is also true because a current generalized maintenance function simulation model would not develop the necessary level of specificity required by TITAN II. A problem exists in defining the trade-off point which will provide the maximum number of users the maximum levels of application and specificity.

Data Creation or Synthesis

During the investigation of techniques, certain universal analysis or simulation techniques were selected for further critical review. They were selected because they appeared to satisfy a number of user requirements. A data problem is anticipated in that these universal techniques may require data items that are not available from the experimental data pool. The problem is essentially twofold: (1) Since the research is exploratory, should consideration be given to the foundation of a rational developmental baseline for the operational system, or is it adequate to demonstrate that analysis and simulation processing, to some very limited degree, is possible using a multiobject system experimental data pool? (2) Since certain necessary data items are missing, what is the best source of missing data? Should this require another object system for the experimental data pool? Are these data obtainable, but not retrieved from the available supply of object system data? Is it acceptable to create a "synthetic object system?" This system includes only the data items required by specific research objectives, but not available from the research data pool. This synthetic object system is a miscellaneous collection of addressable files.

OBJECTIVES AND APPROACH

The general objectives of this research area are:

- Isolate, define, and categorize data for analysis and simulation
- Identify analysis and simulation techniques in current object systems, and identify user requirements
- Apply selected analysis and simulation techniques to the experimental data pools to demonstrate operability within a multifaceted environment
- Establish congruent interfaces with relevant data handling system research efforts

The scope of this research effort involves the selection, development, and validation of analysis and simulation techniques and applications. The methodology allows the application of priority considerations in order to optimize the output from this research. General problems are to be delineated, but only specific aspects of the general problems are to be attacked in order that relatively more definitive outputs might be obtained. Specifically, attention will be focused on applications which represent a single hierarchical level, phase or time frame, and installation.

Methods and Results

Two research avenues were explored. The first was to investigate the availability of current analysis and simulation techniques--off-the-shelf--and evaluate their applicability to a multiobject system data base. The data necessary would need to be selected for the experimental data pool. The second was to investigate the data types currently available from the experimental data pool. The selection of analysis and simulation techniques appropriate to the data types and user requirements would follow.

The implementation of the first alternative necessitated a review of currently available analysis and simulation techniques. The techniques selected are applicable during various phases of an aerospace system life cycle development. The availability of techniques proved to be so voluminous that only a sample of promising techniques was investigated.

Types of Analysis Techniques

Analysis sources were grouped into four general categories: descriptive, general, forecast, and validation.

Descriptive Techniques

Woodson and Conover (1964) summarized the levels of systems analysis, the purpose of the analysis, and their applicable techniques:

LEVEL OF ANALYSIS	PURPOSE	APPLICABLE TECHNIQUE
System	To determine effectiveness of system in performing a specified mission	Operations-research methods
Subsystem	To determine best way of meeting a specified requirement of the mission	System analysis Integration matrix

Function	To determine best combination of components required to make up subsystem	Man-machine system analysis Function analysis
Task	To determine best allocation of man's capabilities to perform required functions	Task analysis Time-line analysis Logic models Information theory
Subtask	To determine best method of utilizing man's capabilities to perform the assigned tasks	Operator load analysis Operator sequence diagrams Decision theory Information flow analysis
Element	To determine best method of utilizing man's capabilities to perform assigned subtasks	Time and motion analysis Elemental task analysis

In their summary of questionnaire results, Hannah et al (1965) indicated both present and potential use or requirements for analytical techniques.

General Techniques

System or application specific and omnibus or general purpose techniques are grouped together for evaluation: The factor analytic approach was proposed for predicting system maintainability (Topmiller, 1964). The "components-of-variance" model is suggested. Questionnaire data were used against AFM 66-1 criterion data. Eight orthogonal maintainability design factors were obtained in the analysis. A user-oriented general analytical technique (BOMM) is described (Bullard et al, 1966) in which arithmetic operations on time series are provided for. Provisions are made for a variety of data formats, automatic removal of gross errors, a variety of arithmetic operations, and does not require a professional knowledge of programming. A user-oriented, time-shared, set of routines (TRACE) for analysis, classification, and evaluation is reported by Moore et al (1965). This technique is probably compatible with operational system requirements. A program (SYSTRAN) has been written in FORTRAN IV for the IBM 7094 which provides for computations in higher mathematics, such as Fourier transforms and complex algebra (La Jeunesse et al, 1965). A manual is provided which discusses program aspects that are important for effective usage.

A general listing of life sciences programs (Hammid et al, 1965) included information storage and retrieval, cross-tabulation of data, tests of statistical hypotheses, regression analysis, multivariate analysis, time series analysis, analysis of variance and covariance, mathematical diagnosis, linear programming, simulation, numerical techniques, and matrix algebra. This listing emphasizes the fact that a variety of techniques for analysis and simulation are available. The major problem involves the determination of applications to specific requirements in the object system life cycle. The BMD

(Biomedical computer programs) represented a widely used set of analytical techniques (Dixon, 1965). Six classes of programs are included, i.e., Description and Tabulation, Multivariate Analysis, Regression Analysis, Special Programs, Time Series Analysis, and Variance Analysis.

Forecasting Techniques

One of the known requirements of a human factors task data system will be the compilation of manning requirements. One analytic tool that is used in the analysis of manning requirements is forecast analysis. Several applications using this technique were evaluated.

The Computerized Advance Personnel Requirements and Inventory (CAPRI) system, developed for the Bureau of Naval Personnel is discussed at length in two volumes. Volume 1 (Pulscak et al, 1965) is concerned with the description and operation of the system, and Volume 2 (Byer et al, 1963) presents the computer program specifications. Volume 1 reports on the four phases into which the design and operation effort was divided. Phase I was a feasibility study, conducted by the Navy, to determine if personnel and training requirements could be computerized. Phase II was the development of a prototype system. Phase III was a pilot test phase and Phase IV was a training program conducted for operational personnel. The CAPRI system is one which uses data similar to that proposed for the operational system, but limited to personnel and training requirements, is manipulated by a computerized program.

Barton et al (1964) developed a queuing technique for computing manning requirements. Queuing tables (Purvis et al, 1964), which were developed to be directly applicable to characteristics of military systems having quantitative spares, personnel, component failure rates, and operational readiness performance requirements. These tables were then field tested (Purvis et al, 1965) on the F105D fire control system and the C141 system to establish the validity and reliability of a technique of mathematical modeling for predicting manning requirements for weapons systems. Losee et al (1961) reported on another manpower forecasting technique. This technique uses Task Equipment Analysis as the input data and develops a manning estimate consisting of position descriptions and numbers of men. Early training information is deduced directly from the Task Equipment Analyses, while prediction of logistic support requirements is available as a by-product of the method. Plans are presented for an extension of the method into other requirements such as estimation of man hours for activities not amenable to direct task analysis, the effects of environment on manning requirements, and the estimation of maintenance activity frequency rates.

Validation Techniques

The final group of techniques that were reviewed contained both those used for concept validation and those used for the validation of other analytical techniques.

Alexander and Cooperband (1965) described a method using statistical decision theory analysis to provide information in support of functions allocation decisions.

Three tests to determine the utilization of human factors information by system designers are described by Meister et al (1965) as an experimental vehicle for improving cross-disciplinary communications. Additionally, the report presents a theory of design activity in behavioral terms.

A novel approach to the definition and classification of data reported by Coombs (1964) reduces data to a geometric formulation. "Data may be viewed as relations between points in a space." This geometric approach developes unique classifications for data types and new scaling models, and describes some similarities among data and models from different areas in psychology.

Bottenberg and Ward (1965) described the application of multiple linear regression techniques to problems from the fields of behavioral sciences. These techniques lend themselves very readily to binary-coded information and are less restrictive than multivariate correlation techniques.

Van Buskirk and Huebner (1962) discussed a method for determining a system's performance--later called system effectiveness--through a study of equipment and human failure data. The resultant is a theoretical technique for monitoring performance and reliability during system development as a measure of the ability of the system to conform to its design objectives.

OMNITAB (Hilsenrath et al, 1965) is a computer program developed for, and employed by the National Bureau of Standards for the statistical and numerical analysis of experimental data. This program is of particular interest because of its language, and a concise instruction set for a wide class of problems. This last feature is of particular relevance in this study because the computer applications to remote stations is very similar to those proposed for the operational system.

Types of Simulation Techniques

Simulation techniques were grouped into three general categories for evaluation: abstract model synthesis, realistic model generation, and system operating model generation.

Abstract Model Synthesis

Abstract model synthesis uses parameters of selected dimensions from the human factors source data bank to synthesize systems. A method for synthesizing a simulation was described by Kirby (1966). This method is based on successive approximation. Hogan (1966) described a general simulation routine in which he emphasizes the practicality of the event-store type of simulation model. The International Business Machines Corporation has developed a series of General Purpose Systems Simulator Programs (GPSS II 1963 and GPSS III 1966). GPSS II is designed to assist in studying the design of systems. Statistical and sampling techniques may be introduced and a variety of computer printouts requested. GPSS III is a more powerful program that permits the introduction of relatively complex models of systems.

Realistic Model Generation

The realistic model is based on a computerized system performance using data from a selected system engineering process. This system performance is to be used as a "root model." The model builder uses this approach when a proposed system is substantially similar to an existing system for which system engineering data are computerized. Logical constructs are added, modified, removed, or replaced. A discussion of the present digital simulation theories of human organizational systems was presented by Lackner (1964). The author discusses only general system theoretical development but suggests a scheme for modeling processes, and producing a well-organized simulation model that will overcome these shortcomings. Finally, the author introduces the notion of using the mathematical concept of a calculus of change as an approach to the general problem of developing system theory.

System Operating Model Generation

Techniques have been developed for simulating system operator models. A specific simulation technique was designed (Howard and Inaba, 1963) which simulate the characteristics of the maintenance functions for the TITAN II Inertial Guidance System. Four principles guide the development of the model. They are: (1) representative of the actual system, (2) means of considering personnel performance errors, (3) use of currently available data, and (4) capability of being developed in time for the Category II testing program. Further instructions are provided for implementing the Category II testing and evaluation (Inaba, 1963).

An operating simulator for the LEM Mission (McLaughlin et al, 1963) was described in terms of its specification requirements. The document is very explicit in the details of design, fabrication, testing, and preparation for delivery. While a simulator, as such, is beyond the scope of this research effort, the description of some simulation techniques utilized in designing some of the simulator's systems seems to have a very relevant bearing on the problems anticipated for a task data system.

The method of developing a simulation model for statistical validation of an operational system was presented by Gafarian and Walsh (1966). The authors advocate the development of simulation models of important system properties. The results of these simulations can then be analyzed statistically, accepting those properties of the system that yield the best results, and rejecting those in which the simulation model displays too much disagreement with their representative values. The statistical requirements are for nonparametric methods. This technique seems to provide an applicable tool to data system requirements and further investigation should be conducted.

While many analysis and simulation techniques have been successfully applied to specific system problems, the pooling of these techniques into a data handling system is fraught with problems. The selection or development of applicable techniques for a wide range of systems and data types must be preceded by the establishment of a common data base line. Human factors data, by their very nature, are diffuse and ill defined. Even the simplest quantitative applications, such as the summation of human activities, cannot be conducted without the loss of important data during retrieval. While the thesaurus techniques discussed in Section VI will provide the necessary control over the vocabulary for retrieval, the same procedures must be followed from the analysis and simulation point of view. Cross-communication must be maintained between the establishment of an adequate vocabulary and the generation of a common base line for the application of quantitative techniques to the experimental data pool.

Establishment of a Data Baseline

The establishment of a base line was the first requirement in the process of isolating data units for analysis and simulation. The experimental data pool contains the following data units: data indicant, nominal, interval, ordinal, and ratio scales. Other data units are qualitative indicants such as effects, assessments, or correlates of a man/machine system interface. The initial problem was to convert indicants into standard sets of quantitative units. The following definitions apply.

Data indicant	-	A verbal expression, or combinations of numerical digits or symbols which apply to structured, meaningful, factual information
Nominal scale	-	An ordered, but unrestricted, assignment of values as labels or types
Interval scale	-	Values which are equidistant and, therefore, amenable to infinite uniform fractionation
Ordinal scale	-	Values which are not equidistant but contain a determination of greater or less
Ratio scale	-	Values which express a determination of equality of ratios, generated from index numbers, parameters, statistical numbers, or numerical estimates

The classification of task data into ten elements (Section V) provided the required level of specification of data elements into analysis and simulation requirements. There were five steps employed in this procedure: (1) examine each element and the preliminary item content to determine the existence of analytical content; (2) classify or describe the content in analysis units, data indicants or measurement units; (3) determine if the analysis unit contains the level of exactness--rigor--required for definition, expression, and determination of class boundaries; (4) identify dissimilar analysis units contained in presently defined individual data elements, or items, as well as similar analysis units contained in dissimilar data elements, or items; and (5) identify the basic values and quantity characteristics contained in each analysis unit. This procedure was applied to the ten data elements presented in Section V. Preliminary results are illustrated in Appendix VIII.

The development of a method which permits a rigorous definition of task data for processing requires the identification of mathematical procedures for satisfaction of user requirements. The user requirements and applications presently identified are:

- Simple single step mathematical applications such as addition and subtraction For example, in object system 1--an operational system--a need may exist to determine the amount of time, during flight operations, the command pilot spends using an equipment item. The satisfaction of this need requires the retrieval of the tasks performed during flight operations using the equipment item. To this data is added times indicated during which the command pilot used the equipment.

- Multiple step mathematical applications, such as the determination of means, ratios, and solutions to single equations For example, determine the mean time required to accomplish certain preflight inspection tasks by a hydraulic mechanic with a skill level of 7. The satisfaction of this request requires that the total task times, whether expressed in a time-line or in discrete numbers, be added and the total divided by the number of tasks.

- Complex operations such as analysis of variance and simulation modeling For example, estimate the maintenance activity manning requirements along with spares and consumables required for turn around for object system 10--a system in the definition phase. The satisfaction of this request would involve analytic routines for estimating manning requirements and simulation techniques for predicting logistic requirements under various conditions.

The evaluative method and applications of the redefined data to mathematical procedures as described above, are conceptual developments. When data elements and data items are specifically defined, redefinition can be implemented.

CONCLUSIONS AND RECOMMENDATIONS

The redefinition of data for analysis and simulation purposes prior to the selection of application techniques is of great relevance in this research. Continued effort is required to specifically describe the analysis content for data elements and items, and to more rigorously define analysis units.

In conjunction with the redefinition effort, coordination with users is required to assure that their needs and requirements for analysis and simulation techniques can be satisfied.

These two areas of continued effort result in a third requirement. This requirement is the coordination of the research in vocabulary/thesaurus development with user profile content and design. This coordination is required to assure that the data system user is able to interpret outputs and coordinate meanings and concepts of expressions produced by the analysis and simulation techniques.

Research into the possibility of introducing a simulation language should be investigated. A number of simulation languages are in existence, for example, SIMSCRIPT. Others are being developed. Coordination of this aspect of the research with the software development activity is required.

SECTION IX

CURRENT AWARENESS TECHNIQUES APPLIED TO HUMAN FACTORS TASK DATA

PROBLEM

Documents and data regarding a human's capabilities and functions in a man/machine system are being produced by the scientific community at a prodigious rate. The digestion and application of these data to a specific man/machine system presents one set of problems that is beyond the scope of this research. The collection, integration, and production of data about a specific man/machine system are the functions with which this research is concerned. Current awareness is the vehicle which provides notification of the receipt and availability of new or updated object system data to data system users. Dr. Paul Ronco (1960) discussed the problem from a document system point of view.

> One of the most pressing needs of modern science is that of more adequate communication within the scientific community. We are now in an era which can be characterized as one in which science and technology are experiencing accelerated growth. This growth, desirable as it is, has, in turn, created serious problems, not the least of which is an inadvertent frustration regarding dissemination of scientific information within every discipline is becoming a momentous task. No longer is it possible for the individual researcher to keep account of the staggering volume and wide dispersion of existing and cumulative technical information in a given field. While he may be certain that work relevant to his own interest has been or is being done, he often encounters tremendous difficulty in locating it. As a result, scientists are spending increasingly and irritatingly large amounts of time searching for information. The real problem seems to be not so much that there is undue publication of information, although this is a contributing factor, but that information publication has been extended far beyond our present ability to make use of it.

A specific problem concerning awareness, is one of restricting the demands of users of data to realistic boundaries. These boundaries should encompass the type, content, and frequency of notification that a system can reasonably produce and from which a user can derive the maximum utility. The techniques for profile design and construction have not yet become an integral part of information systems. CIRC, a bibliographic system developed for the Foreign Technology Division, Air Force Systems Command (McMillan, 1964), is the only information system investigated that has a formal notification function. Therefore, it is anticipated that some original research will be required if the bibliographic techniques are not adaptable to the processing of data.

OBJECTIVES AND APPROACH

The purpose of this research area is to explore the rationale and selection criteria for developing a user profile. This effort involves three objectives:

- Investigate the usefulness of applying task data terminology to describe various levels of technical interest for different representative discipline groups and individuals who use task data

- Investigate techniques for building and maintaining computerized profile lists of user interest terms using terminology selected from and controlled by the task data pilot vocabulary

- Investigate the system interface points between the current awareness and storage functions. Once these interface points have been isolated, the specification of requirements can be initiated

This approach is based on the premise that a highly structured organization of profile terms is required to prevent undue detail in matching--reviewing--input data with individual terms. This premise further indicates that the basic structure of the vocabulary and the parallelism of that structure in the data files and index points, must be applied to the design of interest profiles. The rationale used in the pilot vocabulary will be essentially the same as used in designing profiles.

METHODS AND RESULTS

A current awareness function is a highly desirable feature of the system. This function provides the user with notification that data entering the system for storage matches his particular area of interest--disciplinary, functional, etc. If he is interested in details, or receiving the data, he uses the retrieval function.

Two components make up the current awareness capability. One component is a profile or list of words and phrases--data terms--that specifies a user's technical interest in task data. Because the vocabulary is the control point for defining and describing the data in storage, the profile must also be controlled in essence, by the vocabulary. Construction and maintenance of the profiles are carried out via system vocabulary rules and guidelines. The second component is the computer programs which provide the automatic screening, matching, qualifying, and preparation of data and the subsequent notification.

For the pilot research, current awareness techniques will be investigated for their capability of providing further interface between the task data generated and system users. The research results will serve partially as a validation of the vocabulary concept if task data and users' interests can indeed be expressed in similar terminology amenable to computer processing.

There are some current users of profile techniques and current awareness functions. The development in these systems is significant. The Foreign Technology Division in the Thesaurus for Centralized Information Reference and Control (CIRC) incorporates an elaborate profile structure and a current awareness notification. The vocabulary terms utilized are rather extensive and may be qualified to various levels of specificity.

Doyle (1966) described the operation of a profile technique in an experimental computer program system for the statistical classification of a large number of documents. Profiles are incorporated as a method of achieving selective dissemination of the documents.

Tomlinson (1965) discussed a practical scheme for document classification for staff members of a personnel research laboratory. The scheme involves the development of clustering indexing terms--pairs of topics having the highest probability of being marked together--into field-of-interest categories or profiles. The clustering is accomplished by a computer program.

An article appearing in a trade journal (Olson, 1966) emphasized the constant requirement for new information by scientists and engineers. The article discusses IBM's approach toward satisfying this requirement.

Parker (1966) approached the problem of information exchange from another point of view. He recommends psychological research to provide adequate specifications of user needs for system designers, and adequate criteria for the evaluation of systems.

Murray (1966) discussed a number of techniques that can be employed to provide technical information to the scientist and engineer. While the author does not postulate an operational system, he does describe several equipment items which are in existence or are being developed that will speed the processing and dissemination of scientific and technical information.

Young (1966) discussed the selective dissemination of information (SDI) system developed for NASA by IBM. An IBM 7090/94 computer program compares user interest profiles, constructed from the Subject Authority List, with reports announced in the Scientific and Technical Aerospace Reports and the International Aerospace Abstracts. Users are provided with announcements of selected literature citations.

The research conducted in other systems has reinforced the requirement for a new data notification system tied to user's profiles (see Section II). The research also provides two alternative techniques for the development of profiles. These two techniques are: (1) clustering indexing terms, and (2) user selected profile terms from the vocabulary. One problem is that no current research describes the use of data as the subject matter for new data dissemination. The research has dealt only with documentary bibliographies. Since these were the only techniques available, research was initiated with the possibility of modifying these techniques for use in a data system.

User profiles are a comparatively recent addition to information processing. They are valuable as a basis for selective announcement of acquisition and for controlled dissemination of information. The main source for building a profile is the user, as the profiles are a reflection of his data interests.

To develop a profile, each user is contacted and instructed in creating a list of keyterms which will express his particular needs. This task may be facilitated by means of skeletal profiles, constructed on the basis of previous interests expressed by the user from functional organization tables or from established structured thesauri. Upon completion of the profile, the profile is discussed with the user to assure completeness and uniformity.

There are two types of profiles; those which are developed for individuals and those which are used for organization groups. The latter is the most desirable from the standpoint of the system. This is because of the amount of "overlap" or redundancy existing in individual profiles that have similar interests. For purposes of discussion, references to a user profile implies either type.

Profiles, by definition, are completely responsive to the user's needs. The user is completely at liberty to make changes, refinements, additions, and deletions to the individual profile from the human factors task data vocabulary. Periodically, even though no changes have been made by the user, he is contacted to assure that his profile is up-to-date.

Profiles will be constructed by all levels of system users, government and contractor systems management, SPOs, or personnel subsystem groups. The profile is developed from a vocabulary term which describes the user's field of interest or data requirement. If unrestricted, the profile term may cause excessive volumes of data to be included in the notification. Thus, the field of interest term is further qualified by descriptors which successively narrow the retrieval field. For example; the term "aircraft" would provide volumes of data, a first qualifier "fighter" eliminates all other classes--bombers, cargo, passenger, and reconnaissance--a second qualifier "100" eliminates all other fighters, and a third qualifier "G" eliminates all other models. The completed profile would be: aircraft - F100-G.

Three conditions are considered in the functioning of current awareness. The effects on system economy and satisfaction will determine the selection of one of these alternatives.

1. Notify the user that data matching his interest have been received by the system

2. Provide the user with actual data as received by the system

3. Provide the user the option of either receiving data notification or the desired data The option will be controlled by decision indicators built into his profile.

A further aspect of the current awareness function, timeliness, is closely tied to the function of notification. The frequency with which data are screened and matched with profile terms can be handled as data enters the data bank, or at periodic intervals. The latter involves "spinning-off" data as they are stored and batching them for subsequent matching. The interval is, of course, dependent upon the needs of the user. The choice of frequency primarily depends on software operating procedures initiating the current awareness function.

A definition of what constitutes new data is required. Once the characteristics of the data are established, the boundaries of new data will need to be set. Each data item need not be examined to achieve awareness.

CONCLUSIONS AND RECOMMENDATIONS

The scope of the effort in this research area is constrained by its dependency on other research--task data definition, vocabulary development, and data storage design. The description of current awareness and the identification of problem areas, defined in this section, has given only an approach for further research. It is anticipated that the design of profiles can be initiated concurrently with vocabulary design. There is essentially one variable to be considered in the mutual design of profiles and the vocabulary; level of specificity.

The boundaries of new data discussed above are an important factor to the design of profile specificity. A distinction must be drawn between the need to access specific data and the need to be aware of their existence. It is quite possible that inclusive generic levels established for the vocabulary will provide the necessary boundaries.

There does not appear to be an existing capability other than bibliographic techniques which might be used to conduct live experiments. Should it be necessary to run such experiments, using computer programs, they must be developed specifically for the job. A determination of the feasibility of utilizing data profiles can be demonstrated through the validation of the vocabulary. A manual method could be developed to construct test profiles and make matches against data in the experimental data pool. The results of these matches would then be subjected to validation procedures by probable users.

SECTION X

GENERAL CONCLUSIONS AND RECOMMENDATIONS

The research described in this report has investigated the feasibility of using automated techniques to help solve task data problems. A continuation of this research is planned for the purpose of developing a research pilot. The pilot will apply selected computer techniques to demonstrate methods for handling actual data. Five major research areas have been identified to study and develop techniques pertinent to the problems of computerizing task data. The results of the research performed in each of these areas are reported in Section V through IX, along with conclusions and recommendations relevant to each area. The overall conclusions, drawn from the total effort, with recommendations for future work are given below.

CONCLUSIONS

A detailed study of task data has been the major effort of this research. The study emphasizes the establishment of a taxonomic scheme by which task data from various aerospace systems can be categorized. The scheme also identifies the relationships among those categories. This scheme will be used as the basis for computerizing the data and making it possible for factual characteristics to be handled as specific entities or groups.

The initial attempts to form a scheme, now called a task taxonomy, were concerned with identifying an extremely large amount of data items having complex relationships. A relationship is the fitting of two or more data items together to form a common meaning. The complexity and extensiveness of relationships of task data appeared to be true because of the numerous interdependencies that characterize human performance, especially when performance is considered with its environment, i.e., hardware, location, and time.

Attempts to classify human factors information in other research programs supported the initial idea that task data are made up of a proliferation of classes. In some of these attempts, data were taken and fractionated until the smallest meaningful item was identified, resulting in a bulky organization of classes. Many classes were ambiguous because of other classes into which data would fit. These attempts to classify human factors information were not, however, primarily concerned with organizing data for computer storage and retrieval. Because the techniques of storage and retrieval, along with analytical processing, is a primary concern of this research program a different approach to the problem of classifying human factors data is taken. That approach is to classify task data in a generic structure, then provide for subclasses, or more specific data, according to the requirements for using the data. This approach involves studying the problems of users and then determining the level of detail needed to access individual data according to the user's needs. The objective is to limit data organization to a minimum degree of specificity in order to efficiently access a large data store. In addition, a vocabulary will be used to orient the data system to various users and provide an interface between natural data terminology and computer conditioned terminology.

The results of this research indicate that more than one type of object system's data can be subsumed under a general scheme (see figure 11). The scheme limits the data structure to a basic hierarchical arrangement which controls computer retrieval strategies. A strategy for storing and retrieving data according to such a structure is described in Section VII (see figure 10). Further detail is still needed to assign subclasses of data to the major elements. The breakdown of elements into subcategories will require that the lower levels in the structure be clearly defined. That is, the retrieval of specific items will be made accurately by the computer only where categories are distinct and conform to an explicit order.

The investigation into existing techniques of storing and retrieving factual data has shown a broad application in handling task data. However, there are practical limits involved in using computers. The establishment of those limits for a task data system are difficult to determine. An effectiveness study, to establish trade-offs between the worth of the system and its cost and capabilities must be conducted. The following factors are involved: the volume of data to be stored, the growth rate, frequency and extensiveness of user/computer interaction, and, whether new hardware and software components are required or existing components can be used. Several conclusions regarding these factors can be drawn from the current results. The examination of computer systems and techniques, in conjunction with the detailed study of task data, gives a strong indication that existing software techniques are well suited to task data management. The study revealed that actual software systems can be used at least partially in storing, updating, and retrieving task data. However, the concept of an operational system involves a very wide range of functions. It is unique for a data management system, such as the operational concept, to include so many advanced techniques and capabilities, e.g., on-line and remote retrieval, time-sharing, current awareness, analytical processing, and user-oriented vocabularies.

Because of the inclusiveness of the operational system's capabilities, it is extremely doubtful that any existing data system can be used without a great deal of modification. Adding analytical programs to a system is not a particular problem as most systems are designed for modularity, i.e., have the ability to add on programs as processing requirements are increased. The functions of current awareness, the interface of an extensive system vocabulary, and the need to maintain a data base made up of data from separate, divergent aerospace systems, are the major reasons why an operational system must be designed and developed as a new system. However, many segments of the system can be made up of existing software. The on-line user interaction and time-sharing techniques, storage file structures, and retrieval and programming languages are examples of software that can readily be used in an operational mode.

The investigation of vocabulary techniques identifies several methods for building an indexing capability. The special problems involved in expressing task data terminology cannot be fully identified until a basic structure is established. The vocabulary will provide for a controlled assignment of data into categories, and unambiguous indexing so that the narrowing of a concept can be easily followed to derive a fact. The narrowing process is for both user and computer. Just as a person considers a general category in an encyclopedia before searching for specific information, the user will normally relate general concepts, mutual to many users, then narrow his frame-of-reference to pertinent items. The computer will, likewise, start its search at a generic level and follow a search sequence to locate a specific data item. However, where a computer can rapidly scan large numbers of entries, a user should be required to scan only a minimum number. The vocabulary, therefore, will provide for quick, easily recognized index points and cross-references that guide users to desired terms. Because many users are involved, standard terminology is considered highly desirable. Standardization not only simplifies computer file structure, but increases communication among users. A major problem in building the task data vocabulary is to establish the degree of standardization for various object systems' data. This standardization process is necessary in order to limit the extent of divergency currently used in task terminology, and make it easier to reference terms for computer retrieval and processing.

The same problems described above apply also to profile techniques. The results of the investigation indicate that techniques used for vocabulary usage can be similarly used in building profiles. One major problem is to determine the boundaries of awareness for factual data. This research compared problems involved in notifications for documents with those involved in factual notification. A result was that there was little to compare because factual current awareness is virtually nonexistent. That is, there is no established state-of-the-art for factual awareness. The technique for notifying users of incoming task data will either need to be patterned after document techniques or it will need to be designed. This research approaches the problem by using document techniques. However, a prime difference between document and factual data is recognized. Document data are much more generic. Even using generic subject concepts, some document profiles are large and unwieldy for computer efficiency. Factual data are even more extensive by their very nature of having all levels of information detail involved. There is a strong need to establish boundaries of awareness. These boundaries will define the extent of detail which will make up an interest profile. The results of the data study indicate that levels of detail may vary for different systems, depending on the need of each system to specify data content. Therefore, these awareness boundaries may need to be variable in order to coincide with the needs of users in the various systems. Profiles cannot include reference to all levels of data or they will be too large and cause extensive computer processing. A particular level in a data hierarchy, for example, the fourth level in a seven level hierarchy, can be selected as the maximum degree of specificity to be used in a profile. The vocabulary terms which express the meaning of data on that level, or above, are then used as profile terms. This limiting of profile terminology will prevent "over-awareness," that is, flooding a user with so many notifications of new facts that he ignores them altogether.

Investigation into analytical and simulation techniques has identified a large source of techniques, and many have potential application in the behavioral sciences. However, analytical processing is an infrequent use of computers in current aerospace systems. Normal analytical techniques require data to be organized to suit the input conditions of each technique. System personnel are not usually trained in setting up data for computer processing. Also, a technique used in one system may need to be reprogrammed to fit the computer used in another contractor's facility. Computer time is often difficult to obtain and is scheduled according to system priorities. This research approaches the problem by examining task data for their fundamental characteristics regarding analytical properties. This approach is taken to break down task data to determine properties mutual to various systems data and to help identify the applications, or outputs, to which common routines can be applied. More sophisticated or complex techniques will be investigated as a framework is developed. The task data used in this study indicate only limited use of analytical routines can be made unless more quantitative data are available in the experimental data pool. Such data as time, manning allocation, task sequences, and scalars are present. A greater degree of quantitative data are not usually included in task data because generators seldom have the time or are required to describe the data in the necessary detail.

A technique that appears to be useful to analytical processing is on-line programming. This technique allows nonprogramming people to select existing routines, or simultaneously write and operate a new routine using English language terms and notations. The computer provides immediate response and guidance during the program writing and operation. A user can access stored data and use them in the program operations. To help the user better reference data with regard to analytical processing, the vocabulary will describe such concepts as ratios, scales, and variables in addition to the data descriptors which provide qualitative meanings.

Further research is needed in all areas to complete the fourth stage of the research program and to provide the necessary information to design the research pilot. The major need is to establish an experimental data pool, including the taxonomy and the analytical data framework. The pool will be the central point of the pilot and all activities of the research, will revolve about the pool. Listed below are recommendations for completing the fourth stage.

RECOMMENDATIONS

User requirements are mentioned extensively in this report. These requirements have a major impact on deciding:

- Relative positions and degree of specificity of data categories in the data structure
- The formats and rules of grammar to be used in making retrieval requests
- The degree of standardization in the vocabulary
- The type and output of analytical processing tools
- Current awareness boundaries limiting factual data notification to a manageable and useful level
- The completeness and utility of the data content in the experimental data pool

A representative group of potential users should be consulted regarding the points listed above. The objective of these contacts with users is to define, in detail, how a typical user wishes to access data, how he wants to receive it, and what differences exist among classes of users, i.e., what user variables are to be built into an operational system.

A comprehensive list of data requests needs to be formulated and used as a guideline in setting up a data structure. Several types of requests should be established ranging from simple queries to requests dealing with complex data relationships including processing implications. These request types will help specify system design requirements. These types include arranging categories of requests by: frames-of-reference, output formats, specificity--generic and hierarchical levels--and types of analytical processing required.

A basic stroage and retrieval capability should be used in the early stages of development of the experimental data pool. The selection of techniques for constructing the vocabulary and profiles will also be greatly facilitated by using automated methods for testing and examining various data structures.

The research area dealing with storage and retrieval systems, Section VII, recommends using existing systems for pilot development and verification. The recommended systems are LUCID and TDMS, with TDMS being implemented in the latter stages of the research program. A further recommendation is made here with regard to the total research program. LUCID should be made available to the research program as soon as a preliminary data pool can be assembled. Only basic categories and relationships are necessary to develop an initial data pool for LUCID operations. The use of a retrieval technique, such as LUCID, providing direct user interaction with the computer will greatly assist in investigating and collecting user requirements. Actual retrieval demonstrations will help potential users better judge what their requirements are. Realistic demonstrations are particularly needed in those cases where potential users are not familiar with on-line selective retrieval methods.

User-orientation is a characteristic that covers nearly every aspect of the operational data system. A user-oriented system implies that system functions directly interfacing with users are designed to communicate on or near the same level as the user. The primary interface between users and the operational system proposed in this report is the vocabulary. The vocabulary will be used directly by users of the system similarly as any person presently references a dictionary. The system vocabulary is a user-oriented vocabulary. That is, the vocabulary contains terminology and references that are natural to typical users.

Task data terminology should be established prior to building a pilot vocabulary. Terms and multiterms extracted from representative data can be matched or associated with various data classes. This activity of developing pilot terminology needs to be done in conjunction with developing the data structure and establishing user requirements.

The terms should express data content as well as represent the structure of the data. Computer storage files will be more efficient if they can be designed similarly to the vocabulary. If the designs differ sufficiently, an extensive language translation will be needed to convert terms and term associations into file language and file organization. However, the primary consideration here is that the vocabulary reflect, as nearly as possible, natural terminology and grammar rules that are used in describing task data.

APPENDIX I

GLOSSARY OF INPUT DATA FOR RESEARCH PILOT STUDY

The following is a glossary of data elements and subelements (data items) which formed the initial scope and identification data for the research. It should be noted that the definitions of the data elements and items are considered tentative. The parameters are simply an attempt to give coding characteristics further definition.

TERMS USED IN THE GLOSSARY

Record - A term used to identify all of the elements for a single described performance

Element - A term used to identify a logical class of data

Item - A term used to identify a subset of data within an element--in some cases, there is only one item in an element--where this is true, an item is not identified as such, but is considered to be the same as the element

DATA ELEMENTS, ITEMS, AND PROBABLE PARAMETERS

1. Object System - Any combination of alphanumeric characters, up to a maximum of twenty-five, which designates a specific, total aerospace system

2. Mission - Any combination of alphanumeric characters, up to a maximum of twenty-five, which designates a specific operational profile for the specified object system

3. Mission Phase - Any combination of alphanumeric characters, up to a maximum of twenty-five, which designates a specific segment of the identified mission

4. System - Any combination of alphanumeric characters, up to a maximum of fifteen, which designates a major functional subdivision (consisting of related elements of man/hardware/software) of the specified object system

5. Subsystem - Any combination of alphanumeric characters, up to a maximum of fifteen, which designates a logical subdivision (hardware/software orineted) of the specified system

6. Component - Any combination of alphanumeric characters, up to a maximum of fifteen, which designates an identifiable self-contained unit which performs a specific function necessary to the proper operation of the specified subsystem

7. Part - Any combination of alphanumeric characters, up to a maximum of fifteen, which designates a particular hardware/software item within the specified component--this element may have up to ten entries for any one described performance

8. Hardware Status - A maximum of fifteen alphabetic characters which designates the developmental status of the specified component

9. Data Source - This element contains the following data items:

   Organization - A maximum of ten alphabetic characters which designates the organization responsible for the data being submitted

   Author - A maximum of twenty-five alphabetic characters, consisting of one, two, or three initials plus last name

   Date - A six digit code designating the month/day/year (dd/dd/dd) on which data is submitted

   Revision - A two digit code designating the revision number of data being submitted

10. Reference - This element refers to reference information (charts and documents) and contains the following data items:

    Related Data Element(s) - Up to sixty-four digits or the alphabetic characters "ALL" designating the specific data elements to which the specified reference pertains

    Identification - Any combination of alphanumeric characters, up to a maximum of fifteen, which designates the name and/or identification number of the reference

    Location - Any combination of alphanumeric characters, up to a maximum of twenty-five which designates where the specified reference is physically located

    This element may be repeated up to ten times

11. Security/Proprietary - One or two alphabetic characters which designate the security classification and/or industrial proprietary status of the data--the legal values for this element are: C, S, T, P, SP, TP; where C = confidential, S = secret, T - top secret, and P = proprietary to the organization specified in the data source element

12. Performance Level - A single alphabetic character which designates the level of specificity to which the performance description pertains--the legal values for this element are: P, J. T, E; where P = position, J = job, T = task, and E = task element

13. Performance Description - This element consists of the following data items:

    Verb - A maximum of fifteen alphabetic characters which designate the action portion of the performance

    Object - A maximum of twenty alphabetic characters which designate the object of the specified action

Modifier - A maximum of fifteen alphabetic characters which designate noun, adverb, and/or adjective used to modify the specified verb and/or object--a maximum of three modifiers may be used for any one verb or object

14. Performance Number - A maximum of fifteen digits which designates a specific identifying number of the described performance

15. Performance Prerequisites - A maximum of fifteen digits which designates the number (see Element 14) of the performance that must be accomplished in order to make possible the successful accomplishment of the described performance--this element may be repeated a maximum of three times

16. Personnel Classification - A maximum of ten alphanumeric characters, in any combination, which designates the type of personnel required to accomplish the described performance (AFSC number will be a common entry in this data element)--this element may be repeated a maximum of ten times

17. Number of Personnel - A two digit code which designates the actual number of personnel required to accomplish the described performance

18. Performance Location - Any combination of alphanumeric characters, up to a maximum of twenty-five, which designates the physical location at which the described performance is accomplished

19. Environment - Any combination of alphanumeric characters, up to a maximum of thirty, which designates a critical factor of the environment associated with the described performance--this element may be repeated a maximum of ten times

20. Communication - This element is composed of those items which provide a description of the transmission of information from one human to another in relation to the described performance

Personnel - Any combination of alphanumeric characters, up to a maximum of ten, which designates the specific personnel involved in the communication process

Method - One or two alphabetic characters which designate the method used in the communication process--the legal values for this item are: OD, OI, W, and G; where OD = oral-direct, OI = oral-indirect, W = written, and G = gesture

Rate - Four digits and one alphabetic character which designates the number of times per unit of time (dd/dd/l) the communication process occurs:

dd/dd/l = number of times
dd/dd/l = number of time units
dd/dd/l = the specific time unit; H = hours, M = minutes, S = seconds

Duration - Six numeric and three alphabetic characters (ddHddMddS) which designate the length of time the communication process takes:

ddHddMdds - number of hours
ddHddMdds - number of minutes
ddHddMddS - number of seconds

21. Tools and Equipment - Any combination of alphanumeric characters, up to a maximum of twenty-five, which designates the specific tools, equipment, fixtures, or supplies that are required to accomplish the described performance--this element may be repeated a maximum of ten times

22. Performance Frequency - Four digits and one alphabetic character (dd/dd/l) which designate the number of times, per unit of time, the described performance occurs (see Item "Rate" in Element 20)

23. Time - This element consists of the following two items:

    Total Time - Six numeric and three alphabetic characters (ddHddMddS) which designate the time the described performance takes (see Item "Duration" in Element 20)

    Incremental Time - Twelve numeric and six alphabetic characters ddHddMddS/ddHddMddS) which designate the start/stop times of the described performance relative to the next higher level described performance

24. Criticality - A single digit code which designates the possible effects which would arise from the failure to accomplish the described performance

25. Hazards - Any combination of alphanumeric characters, up to a maximum of twenty-five, which designates a possible source of physical or psychological injury which may be encountered in the described performance--this element may be repeated a maximum of ten times

26. Human Output (man/machine interaction) - Any combination of alphanumeric characters, up to a maximum of twenty-five, which designates a control output which a man must provide in order to accomplish the described performance--this element may be repeated a maximum of fifteen times

27. Machine Output (machine/man interaction) - Any combination of alphanumeric characters, up to a maximum of twenty-five, which designates an output, from a machine, which a man must perceive in order to accomplish the described performance--this element may be repeated a maximum of fifteen times

28. Knowledge Requirements - Any combination of alphanumeric characters, up to a maximum of fifteen which designates a specific aptitude required to accomplish the described performance--this element may be repeated a maximum of five times

30. Difficulty - A single digit code which designates the complexity of the described performance

31. Human Performance Error - Two digits which designate the estimated probability of the described performance resulting in failure due to human error

32. Reliability of Equipment Performance - Two digits which designate the estimated probability of the described performance resulting in error due to equipment failure

APPENDIX II

SAMPLE QUERY CAPABILITY

COMMANDS	RETRIEVE - Request for a specific item of data SUM - Request for a summation of the referenced item of data
LOGICAL CONNECTORS:	WHERE - indicates qualifying condition AND - both adjoining terms must be used OR - one of the adjoining terms must be used
LOGICAL OPERATORS:	LS - (less) LQ - (less or equal) NQ - (not equal) EQ - (equal) GQ - (greater or equal) GR - (greater) EXISTS- (has a value) FAILS - (has no value)

SAMPLE QUERIES:

1. RETRIEVE CREW MEMBER WHERE TASK EXISTS AND CLOCK TIME GQ 15M (minutes)
2. SUM TASK WHERE CREW MEMBER EQ PILOT AND DIFFICULTY EQ 4 *
3. SHOW CREW MEMBER AND FUNCTION ELEMENT AND TASK WHERE TOLERANCE HAZARD EXISTS AND TASK CHARACTERISTICS FAIL

*A difficulty index is assumed, for example: 1 = minor, 2 = normal, 3 = considerable, 4 = extreme. Similar indexes were used for criticality, etc.

APPENDIX III

GUIDELINES FOR DATA EXTRACTION (ALCC)

General

The fundamental task to be performed is to extract the data from the source materials and record it on the forms provided. The information should be recorded exactly as it appears in the source material. Few judgments will be necessary on your part to complete the sixteen items on the form.

Specific Instructions

Number forms sequentially beginning with 001. Place the number in space labeled Index Number.

Block 1 - Will always be "ALCC-AVE "
Block 2 - Will always be 1/25/66
Block 3 - Write Unclassified
Block 4 - Will always be "Boeing"
Block 5 - Document title will always be "ALCC Operational Task Analysis and Timeliness "
Block 6 - Data are found on line #30 of the Task Information Summary (TIS)
Block 7 - Data are found on line #11 of TIS
Block 8 - Data are found on line #1 of TIS
Block 9 - Data are found on lines #3 and 4 of TIS
Block 10 - Extract data from lines #18 and 19 of TIS

a. Copy each different coded equipment designator from the time-line of the corresponding task time-line. (Examples of coded designators are circles in red on page 2 of Section 2 (Time-lines).)

b. Each TIS has a corresponding TAWS (Task Analysis Work Sheet). This TAWS is found on the page following the TIS. Extract from paragraph 8 of the TAWS, any controls or equipment that appears with a check (✓) mark. Supplemental equipment information may appear in paragraph 2 of the TAWS; if so, include this information.

NOTE: Tools are not equipment. We are concerned here with equipment components and subassemblies: We are not concerned with conditions under which they are used. (See Block 16 remarks.) Do not duplicate equipments or controls.

Block 11 - Will always be blank
Block 12 - Although the section appears early in the form, it should be completed only after all other sections have been completed. (Proceed to Block 13.)

a. Miscellaneous comments necessary to explain any of the other items

b. Specifically - if paragraph #6 (Special Handling) of TAWS is checked (✓) <u>considerable</u> - record from paragraph #2 of TAWS, the corresponding comment regarding what that special handling consists of.

c. Record any comments from paragraph #5 of TAWS regarding Consequence of Deviation here.

d. Any other pertinent information which you feel should be called to our attention regarding this "task."

Block 13 - Copy each step in sequence from Task Time-line of corresponding task. (Example is shown on page 2 of Section 2, Time-lines.)

NOTE: <u>Do not include time information</u>. It is recorded in another block.

Block 14 - a. See line #8 of TIS for location.
b. See line #10 of TIS for frequency.
c. See line #17 of TIS for criticality/difficulty.
d. See line #22 - 25 of TIS for Training Requirements.
e. See paragraph 2 of TAWS for special tools - if none, write <u>none</u>.
f. See paragraph 7 of TAWS for Safety Precautions.

<u>NOTE</u>: For this item, please label the information <u>a</u> - <u>b</u> - <u>c</u> - etc.

Block 15 - See line 12 of TIS.
Block 16 - See lines 14-15-16 of TIS. Please give time, title, i.e., elapsed time, and numerical value.

At this point, complete Block 12, a - d.

Aerospace Medical Research Laboratories Human Factors Task Data		Index No. 023
1) Object System: ALCC - AVE	2) Date/Revision: 1/25/66	3) Security/ Proprietary Classification: Unclassified
4) Originating Organization: BOEING	5) Author/Document: ALCC Operational Task Analysis and Timelines	
6) Reference: Drawings, flows,TAWS	7) Type of Performance: Operations	
8) Function/Task, Name/Number: (T) Send PLC - B Command 2.4.2.7	9) Mission Information: Perform Airborne Operations Perform 494L Post Attack Operations	

10) System Information

F16 A, 1999, 1997, 1996, 1988 ALCC System
1-Lp, 2, 1-LP/2, RO/1/2, 1/3, 1/2
Push buttons, toggle switches, selector switches, thumbwheel selector

11) Hardware Characteristics:

None

12) Remarks (Indicate Specific Referent Block or Subject):

Consequence of Deviation

Wrong target

Form
AMRL-0 89
Aug 66

Index No: 023

AMRL Human Factors Task Data

13) Performance Description:

Rotate ADDRESS thumbwheels to insert address of WING/SQUAD/FLT/LF
Check ALIGN toggle switch & change to MRT if required
Rotate PLC-B TARGET thumbwheel to appropriate target
Check DELAY toggle switch & throw to HOLD position if required
Rotate COMMANDS selector to PLC-B position
Confirm STA 1 selection agrees with message (check PROGRAM display)
Throw RADIO PERMIT toggle switch to ON position
Receive status: RADIO PERMIT display ON
Depress & release message INITIATE button
Receive status: IN PROCESS display ON
Receive status: TRANSMITTING display ON
Detect RF output & coordinate with STA 2
Receive status: MINIMUM COMPLETE display ON
Coordinate "beyond range" with pilot navigator
Depress & release STOP button
Receive status; IN PROCESS, TRANSMITTING & MINIMUM COMPLETE displays OFF
Coordinate with STA 2
Throw RADIO PERMIT toggle to OFF position
Receive status: RADIO PERMIT display OFF
Coordinate, return all switches to normal, & log information

14) Performance Characteristics:

Location	Aircraft
Frequency	3
Crit/Diff.	X-C
Tng. Reqmts.	Operator
Spec. Tools	None
Safety Precautions	None

15) Personnel Description:

1416, 3016, 29372C, 1045H

16) Time Information:

Time (1) 1.2, (2) 0.8
(RO) 0.1 (3) 0.1
Elapsed Time 1.3
Total Time 2.2

Form
AMRL-0 89
Aug 66

APPENDIX V

ALCC INPUT DATA ITEM DESCRIPTION AND RELATIONSHIPS

An analysis of the ALCC involved the accomplishment of the following tasks:

- Assign data to previously defined data elements (Table III, page 52)
- Deduce the definition of the data item content
- Identify data parameters
- Outline input data item relationships

A summary of preliminary results for the first three tasks is presented in chart format.

The accomplishment of the fourth task is illustrated in two formats. Relationships contained in data elements two, three, and four are presented graphically (see pages 130 - 132). Data elements five through ten are presented by means of a narrative description (pages 127-129). This analysis is preliminary and associates only those data items whose relationships are obvious. Neither the illustrations nor the narratives are intended to be an exhaustive analysis of the relationships that can exist between and among data items. The analysis may never be completed because of unique relationships that may be developed by a user to satisfy a newly defined requirement. Additionally, there may be relationships that exist but become obvious only through manipulation of the data items themselves.

DATA ELEMENTS	DATA ITEM	DEFINITION CONTENT	PARAMETERS	RELATIONSHIPS
1. Object System		ALCC	4 alpha characters	The entry point for all data
2. Mission Information	2.1 Mission	Launch Minuteman missiles from airborne launch control center	20 alpha characters	See figure 13
	2.2 Phase	Preflight operations Airborne operations Postflight operations	25 alpha characters	
	2.3 Segment	Obtain operational readiness Preattack operations Postflight operations	30 alphanumeric characters	
3. System Information	3.1 System	Statements regarding the system required for the performance being observed, e.g., aircraft, ALCC, 494L, SAC/LCF	30 alphanumeric characters	See figure 14
	3.2 Subsystem	Statements regarding the aircraft and ALCC subsystems required for the performance being observed, e.g., aircraft power supply, ALCC station 7, ALCC commo	20 alphanumeric characters	
	3.3 Components	Statements regarding the various components of subsystems required for the performance, e.g., power panel (ALCC), address panel, power panel (aircraft), data processing panel	20 alphanumeric characters	

DATA ELEMENTS	DATA ITEM	DEFINITION CONTENT	PARAMETERS	RELATIONSHIPS
3. (Continued) System Information	3.4 Part	Statements regarding the various parts of components, assemblies, or subassemblies, required for the performance, e.g., toggle switch, thumbwheels, headset and microphone	20 alphanumeric characters	
4. Performance Description	4.1 Level	A descriptor that specifies the detail--ranging from gross to detailed--of the performance being observed. The levels are: position, job, and task.	8 alpha characters	See figure 15
	4.2 Description	Performance goal or objective will contain the following:		
		<u>Verb</u> The action or behavioral verb that describes the activity of the performance	15 alpha characters	
		<u>Object</u> The object noun that describes the locus of the performance	15 alpha characters	
		<u>Modifiers</u> Various adjectives, adverbs, pronouns, and nouns that are used to further define the object of the performance	50 alphanumeric characters	
	4.3 Procedural Steps	Statements that describe the sequential steps the operator performs--man/machine interactions--to accomplish the performance objective--form of presentation will be: verb, object, modifiers	300 alphanumeric characters	

DATA ELEMENT	DATA ITEM	DEFINITION CONTENT	PARAMETERS	RELATIONSHIPS
5. Performance Characteristics	5.1 Location	Statements regarding the location of the performance being observed, e.g., aircraft power supply panel, ALCC Station 1 This category may often be identical to category 3--System Information		See Page 127 para. 1
	5.2 Frequency	Statements regarding the number of times the performance is required per operation, e.g., once per flight, unscheduled (will occur as required by the mission), three per operation	15 alphanumeric characters	
	5.3 Difficulty	An entry from one of the following codes: G - equipment design makes task performance difficult P - skill or knowledge required makes personnel selection and training difficult, and a high probability of performance failure is anticipated X - other	1 alpha character	
	5.4 Criticality	An entry from one of the following codes: A - performance failure will cause an abort resulting in equipment damage or personnel injury C - performance failure will allow operation to be carried out but not within planned limits D - performance failure will allow operation to be carried out but delayed beyond operational tolerances		

DATA ELEMENT	DATA ITEM	DEFINITION CONTENT	PARAMETERS	RELATIONSHIPS
5. (Continued) Performance Characteristics	5.4 Criticality (Continued)	E - performance failure will not affect successful operation S - task is applicable to supporting activities and will not materially affect operations	1 alpha character	
	5.5 Hazards	Statements regarding hazardous conditions in equipment or operating environment--one or more of the following entries may be used: mechanical, electric, pneumatic, hydraulic, and explosive	15 alpha characters	
	5.6 Training Requirements	Statements regarding the training necessary for operating and maintenance personnel	50 alpha characters	
	5.7 Special Tools/ Equipment	Statements regarding the necessity and the type of special tools and equipment required for successful performance	60 alphanumeric characters	
6. Hardware Characteristics	6.1 Accessibility	Statements relative to equipment accessibility for operability or maintainability	150 alphanumeric characters	See Page 127 para. 2
	6.2 Visibility	Statements relative to equipment visibility or legibility for operability or maintainability	150 alphanumeric characters	
	6.3 Manipula- bility	Statements relative to equipment manipulability for operability or maintainability	150 alphanumeric characters	

DATA ELEMENT	DATA ITEM	DEFINITION CONTENT	PARAMETERS	RELATIONSHIPS
6. (Continued) Hardware Characteristics	6.4 Equipment Status	Statements regarding the developmental state of the equipment--these statements will usually include the following conditions: mock-up, breadboard, prototype, and production	150 alphanumeric characters	
7. Personnel Description	7.1 Type	A descriptive name--job title--of the type of operator required for the performance, and, if applicable, the AFSC or other descriptors Additional job titles and AFSC's will be added for any helper personnel required	50 alphanumeric characters 50 alphanumeric characters each	See Page 127 para. 3.
	7.2 Number	The total number of operators and helpers required for accomplishment of the performance being described	3 numeric characters	
	7.3 Special Skills/ Knowledge Required	Statements regarding special skills and knowledge required of the operator of the performance	70 alphanumeric characters	
	7.4 New Skills/ Knowledge	Statements that describe any skills and knowledge required for the performance that are not presently included in the job description or AFSC	50 alphanumeric characters	
8. Time Information	8.1 Time	Time required for job or task performance, expressed in minutes	3 numeric characters	See Page 128 para. 4.

DATA ELEMENT	DATA ITEM	DEFINITION CONTENT	PARAMETERS	RELATIONSHIPS
8. (Continued) Time Information	8.2 Total Time	The accumulation of task element time increments to produce the position time for the mission segment, expressed in minutes	3 numeric characters	
	8.3 Time Constraints	Values relative to any time constraints that may affect successful mission performance, expressed in minutes	3 numeric characters	
9. Remarks		Miscellaneous remarks necessary to explain or supplement any information contained in other data elements	250 alphanumeric characters	See Page 128 para. 5
10. Input Identifiers	10.1 Author	The name of the person or group responsible for the input	40 alpha characters	See Page 129 para. 6
	10.2 Organization	The name of the prime organization submitting the input	16 alpha characters	
	10.3 Date	Preparation date for the input	20 alphanumeric characters	
	10.4 Revision	The revision code designated by the contractor or sub-contractor that specifies revisions to data inputs	25 alphanumeric characters	
	10.5 Security/ Proprietary	An entry from the following codes that indicates the security or proprietary status of the input T - Top secret S - Secret C - Confidential U - Unclassified A - Atomic energy data P - Proprietary	2 alpha characters	

DATA ELEMENTS	DATA ITEM	DEFINITION CONTENT	PARAMETERS	RELATIONSHIPS
10. (Continued) Input Identifiers	10.5 Security/ Proprietary	The letter A will be used with other codes where applicable--E.G., TA - Top Secret atomic energy data		
	10.6 Type of Performance	A descriptor that identifies the type of performance being observed--the types of performance are: operational, maintenance, routine, continuous, and discrete	50 alpha characters	
	10.7 References	The references include: the documents, charts, and flows, used by the analyst in the preparation of his analysis reference to a previous or subsequent TEA that is required as a prerequisite performance for the current TEA	80 alphanumeric characters 80 alphanumeric characters	

ALCC INPUT DATA ITEM RELATIONSHIPS

This discussion is not intended to provide an exhaustive analysis of the relationships that exist between and among data items. This analysis probably will never be completely exhausted because of individual relationships that may be developed by specific users. Additionally, there may be relationships that exist, but will become obvious only through manipulation of the data items themselves. This analysis is considered preliminary and will associate only those data items whose relationships are obvious.

1. Performance Characteristics Items

The data items of data element 5--Performance Characteristics--are normally related to each other. However, they are more closely related to the data elements; System Information--element 3, Performance Description--element 4, Hardware Characteristics--element 6, and Personnel Description--element 7. Since these relationships have been or will be described with the other elements, they will not be repeated here.

2. Hardware Characteristics Items.

The items of element 6, Hardware Characteristics, may be related to each other, but the major relationship is to the equipment being used in the performance. These data items define the performance characteristics of the equipment specified by the items of the element System Information. Accessibility indicates whether or not the equipment required for the performance is readily at hand. It answers two questions: (1) Is the equipment where it is required for the operator? (2) Can the equipment be easily exposed for failure diagnosis by the maintainer? Visibility indicates that the displays and control labels are within the visual limits of the operator at his normal work station or that test points, parts, subassemblies, and assemblies of components are adequately labeled and the labels are visible to the maintainer. Manipulable implies that there is adequate space to operate the controls required in the performance by the operator, or that there is adequate space for both the maintainers hands and tools required. Status indicates the developmental condition of the equipment used by the operator in the performance being observed.

3. Personnel Description Items

The data items of data element, Personnel Description, are not related to each other except that an inherent relationship exists between Type and Number. However, these items are directly related to the data items of Performance Description and Performance Characteristic items. Where the performance is accomplished, the frequency, the criticality hazards, and difficulties involved will determine preliminary types and numbers of personnel required for successful accomplishment. The type of training required will often depend on the special skills or knowledge required which in turn may be dependent upon training. As the level of performance becomes more detailed, i.e., position to task element, the requirements for the type of personnel can also become more detailed. The simple accumulation of data will not supply the facts for the items in this element. These facts must be evaluated.

4. Time Information Items

The items of element 8, Time Information, are three separate measures of time, directly related to each other. The start/stop times are measurements of estimates made for each task element recorded on the time-line analysis. These element times are accumulated for the total task and then recorded on the task information summary format as time. However, when the task consists of task elements performed by different operators, the recording on the task information summary format is twofold: (1) the accumulated start/stop times for each operator as segments of time, and (2) the accumulation of the clock time that has elapsed between the start and completion of the task, i.e., elapsed time.

Total time also has a twofold concept. First, the accumulation of the times for task elements are recorded as a total time for the task. Note: If two operators times are accumulated, the total time will differ from the elapsed time. Secondly, the accumulation of the task times will result in a total time for each ALCC operation and the accumulation of operation times will result in an ALCC mission total time. These times appear only on the time-line summary sheet.

Time constraints are those time values that specify limitations or constraints on a performance. For example, the performance must be accomplished within a specified time limitation, (2.3 minutes); the task performance must not be initiated before a specified time measured from the beginning of an operation; or the task performance must be initiated before a specific time measured from the end of an operation.

5. Remarks Items

The data items of data element 9, Remarks, usually are not related with each other within the data element, but are related only to the data item of the element that generated a need for the remark. The generation of a remark data item may be attributable to one of two sources:

a. When performing the analysis, the analyst noted some special characteristic that would be detrimental or hazardous to men and/or equipment in the performance of the specified mission. The area is checked on the analysis form and a special comment or remark made to support the analysts findings. These remarks may lead to special training requirements, special procedures, special tools or equipment or recommendations for design or redesign.

b. When performing the analysis, the analyst noted some special characteristic not covered by the specific areas on the analysis form. These data are then included as a separate remark and is preceded by a descriptive subject indicator.

6. Input Identifier Items

All of the data items of data element 10, Input Identifiers are, by their very nature, directly interrelated and are also related to data items of other elements. For example, an analyst is responsible for the system, subsystem, component or part being analyzed for performance. The date is the time the analyst completed or revised the form and the revision code indicates whether it is the first analysis or a revision. The security classification is dictated by the appropriate security guideline for the object system involved, and the references are the materials utilized by the analyst in the performance of his function, etc. All of these data items are directly dependent upon each other. However, the user may be only interested in any one of these data items, or in any combination of them. The use of these data items cannot be always specified beforehand. These data items are indirectly related to each item of each other element. For example, the analysis of performance recorded is further reflected in the references of any documentation used by the analyst in preparation for observing the performance or during the performance.

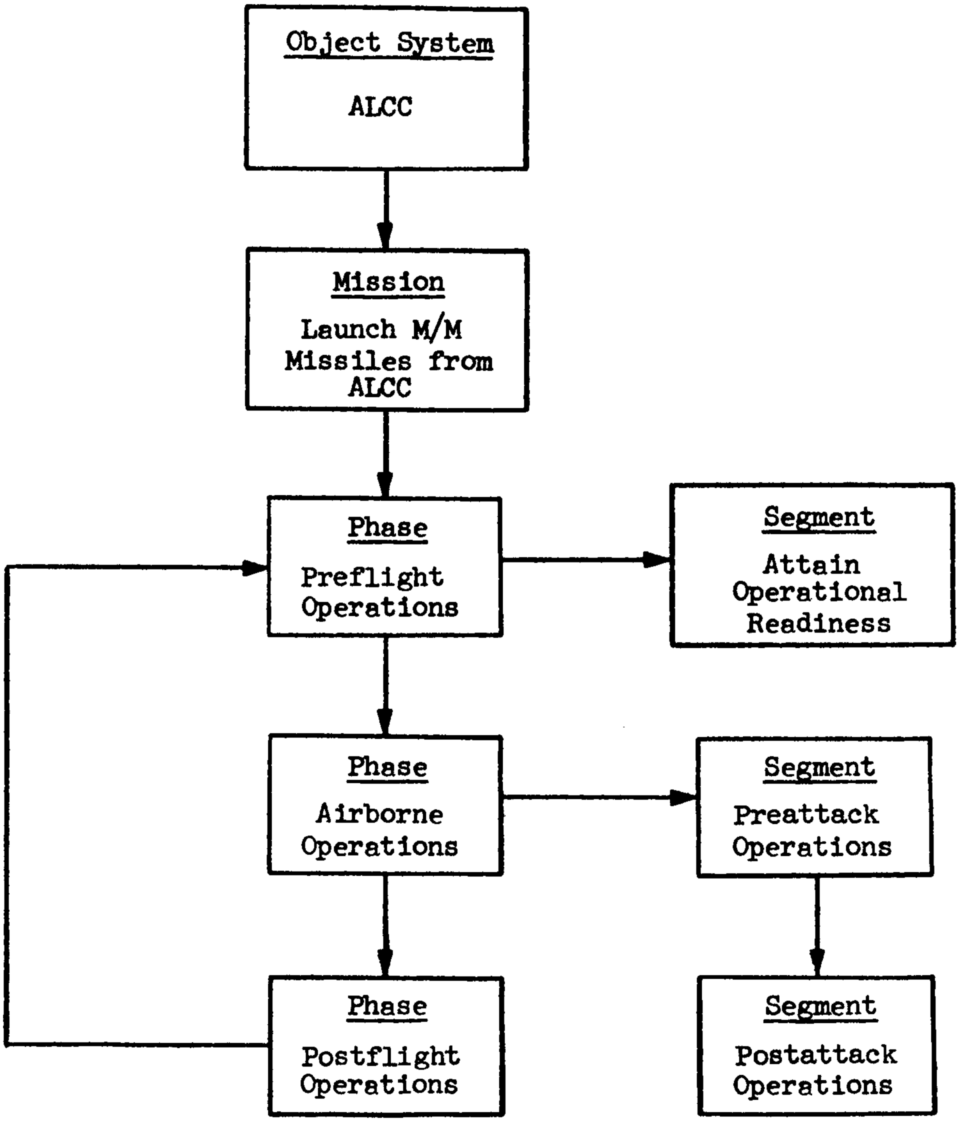


Figure 13. Mission Information Relationships

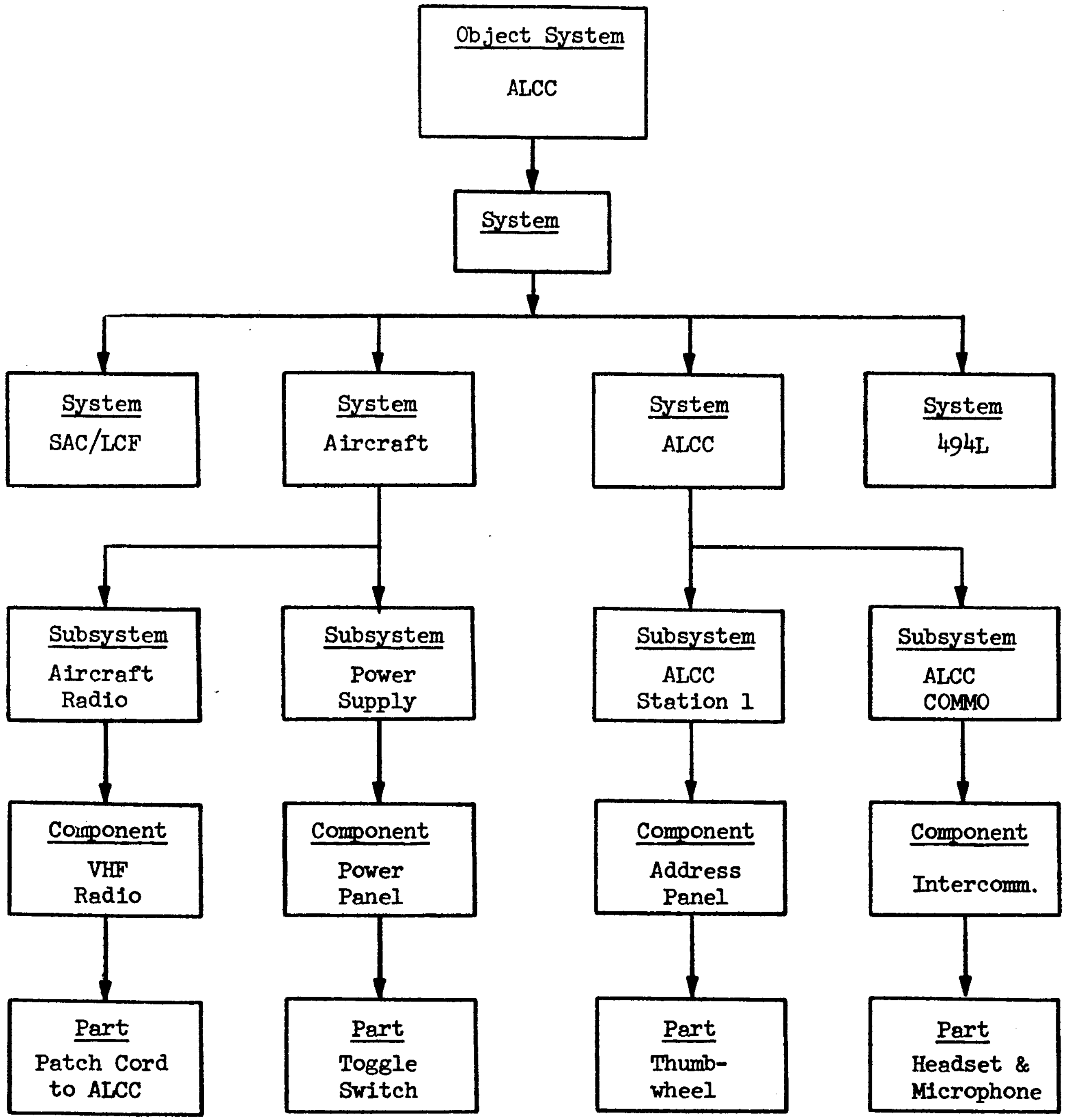


Figure 14. System Information Relationships

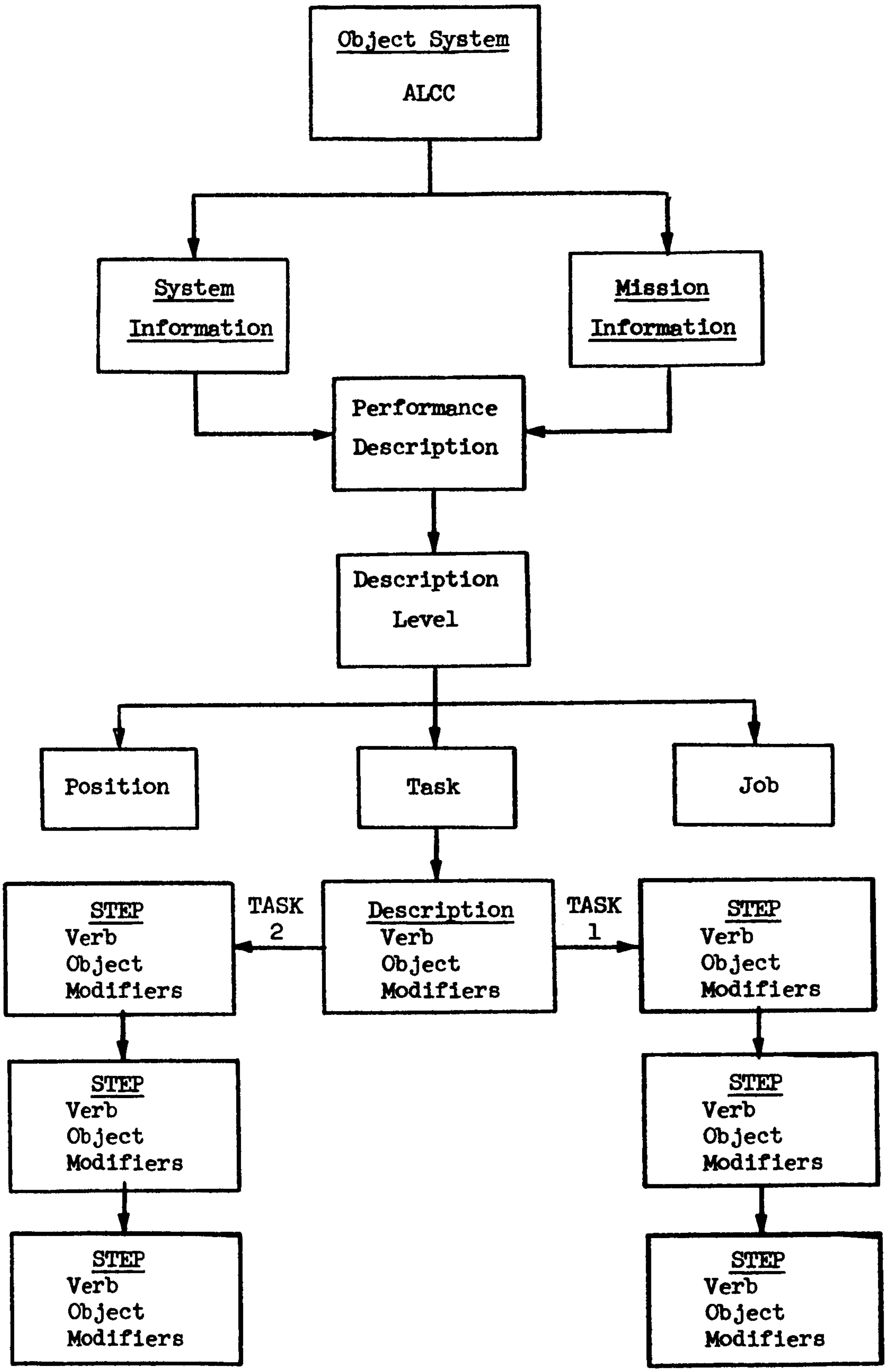


Figure 15. Performance Description Relationships

APPENDIX VI

PRELIMINARY TASK ACTIVITY VERB LIST

1. accompany
2. accomplish
3. acknowledge
4. activate *
5. actuate (4)
6. adjust (9)
7. advise (21)
8. analyze *
9. align *
10. ascertain
11. assemble (14)
12. assume
13. assure
14. attach *
15. attack
16. attain
17. attempt
18. balance (115)
19. bleed
20. brief (21)
21. call *
22. center (115)
23. check *
24. chock *
25. clean
26. clear
27. close *
28. command
29. communicate (21)
30. compare
31. complete
32. compress
33. compute
34. condition
35. conduct
36. confirm
37. connect (14)
38. consult (21)
39. control
40. convert
41. coordinate (21)
42. crimp
43. cut out
44. deactivate *
45. decrypt
46. de-energize (44)
47. deliver
48. depress (130)
49. depressurize
50. detect
51. determine (23)
52. dewrinkle (182)
53. direct
54. disassemble (44)
55. discard
56. disconnect (44)
57. disengage (44)
58. dispatch
59. display *
60. draw
61. enable (4)
62. encode
63. energize (4)
64. engage (14)
65. ensphere
66. enter
67. erase (155)
68. establish
69. evaluate (8)
70. exercise
71. extend (4)
72. fasten
73. feed
74. fill
75. finish
76. gain
77. generate (33)
78. give
79. ground
80. handle
81. hoist
82. identify
83. illuminate
84. increment
85. inform (21)
86. inhibit
87. initiate (4)
88. insert
89. inspect
90. install

- Numbers in parentheses indicate probable synonyms.
- Asterisks indicate probable key words or "preferred terms."

91. instruct
92. insure
93. interconnect (14)
94. interrogate (21)
95. isolate
96. key
97. lace
98. launch
99. lie
100. load
101. localize
102. locate
103. lock
104. log *
105. loosen
106. maintain
107. manage
108. mark
109. mate (14)
110. measure
111. monitor
112. mount
113. note (140)
114. notify (21)
115. null
116. observe
117. obtain
118. open *
119. operate
120. package
121. patch
122. perform
123. place
124. play back
125. plug
126. position (9)
127. post (104)
128. prepare (14)
129. pressurize
130. press *
131. probe
132. proceed
133. provide
134. pull *
135. punch
136. push *
137. raise
138. read
139. reassemble
140. receive
141. recheck
142. record
143. reduce
144. re-enter
145. reinstall
146. release
147. remove
148. repair
149. repeat
150. replace
151. replenish
152. report
153. request (21)
154. require
155. reset *
156. response
157. restore
158. resume
159. retract
160. return
161. reverse
162. rewind
163. retate (194)
164. scan
165. schedule
166. seal
167. secure
168. select
169. send
170. separate
171. sequence
172. set *
173. set up (14)
174. ship
175. shut
176. slide
177. solder
178. stabilize (115)
179. stop
180. store
181. stow
182. straighten *
183. submit
184. supply
185. switch
186. synchronize
187. test *
188. throw
189. torque
190. transfer
191. transmit
192. transport
193. travel (1)
194. turn *
195. uncap (117)
196. unlock (117)
197. unpack
198. unscrew
199. update
200. use
201. verify
202. watch
203. wait
204. zero (115)

KEY WORD DEFINITIONS

Words in parentheses are key word antonyms.

Activate (deactivate) - to make an equipment item active--work--or cause to engage in a task activity

Analyze - to make decisions regarding the operation or maintenance of a man/machine system function

Align - to line up, track, position, or precisely adjust an equipment item

Attach (detach) - to join two or more equipment items mechanically or electrically, fit or put together

Call - to give or receive information between individuals--face to face-- via intercom, or via radio

Deactivate (activate) - to make an equipment item incapable of activity--work--or cause to cease some task activity

Detach (attach) - to separate two or more equipment items or take apart

Display - to cause a task condition--operation or maintenance--to become visible

Log - to record in writing or on tape

Null - to center a meter, or to void an error

Reset (set) - to reestablish a prior condition in an equipment item of the man/machine system

Set (reset) - to establish an initial condition by manual manipulation in an equipment item of the man/machine system

Wait - to remain inactive until an event has occurred or a point in time is reached or a specified quantity of time has elapsed

APPENDIX VII

REVIEW OF INFORMATION SYSTEMS

RELIABILITY EXPERIENCE CORRELATION AND ANALYSIS PROGRAM

Data Base Content: The data base consists of system reliability data. The data consist of summaries from such data groups as: component and part test data; category I, II, and III test data; cost data maintenance action reports; operational failure data; and, reliability Engineering Change Proposal summaries.

Input: Data are input via punched cards and tapes.

Input to the system includes file generation instructions, data storage, and retrieval requests. New files are structured according to formats described on card decks. The index to the data files is formed from those inputs. Updating of the files, inserting, changing or deleting data in the data base, is accomplished using formats that follow the file generation instructions.

Retrieval requests are also formulated on card decks. The cards contain instructions for retrieval, such as sorting or summing, and stipulate the type of data desired.

Output: Various outputs, formats, and types are available, ranging from specific data requests to data summaries. Some of the products are: status of achieved reliability; critical problem analysis; trade-off evaluations; evaluations; reliability improvement program analysis; data for special studies and evaluation.

The output products include graphical and tabular printouts from a 1401 computer. Analytical products involve summing, averaging, and determining means.

New history files, resulting from data pulled from existing files and structured according to instructions in the retrieval request, can be output.

Storage: The data are stored as separate data bases on magnetic tape; that is, they are not integrated with other systems' data. The bases can be retained, as historical files, for prediction of future data.

Retrieval:	A comprehensive retrieval function is available to the user. Several options exist which allow the user to select only the data desired. These options include conditional operators--equal, greater than, and less than--in conjunction with connectors, data field name, and parameters. Formatting options are also available via sorting and ranking, i.e., key under a control item. The retrieval language is highly formatted using programming-like terms and symbols. A lexicon is provided to guide the user to appropriate data fields. A cross-file retrieval technique is used to correlate data entries in various files. This technique allows a user to specify a category of data in a file to be used in further selection of data in other files. Another capability of the system is the storage of lists of complex retrieval instructions and the processing of those lists upon request of a user. The user specifies the list by name.
Control:	Standard IBM operating routines and procedures are used to control the operation of the system. These include the standard operating system (SOS), IBSYS and IBJOB, etc. A modular approach to program structure allows the control, through a priority mode, to accept nonsystem programs for individual operations. There are no security procedures or program routines to handle classified data. The discretion of the use of the data is left up to the user or generator.
Data Framework:	The IBM 7090/94 Formatted File System, FFS, forms the basic system design. This design is based on clustering similar data categories about a generic subject category. These clusters range from complete tape files down to single bits and items of information. Logical fields and sets of data are formed to further describe data categories. The basic framework hinges on serial search techniques for magnetic tape files.
New Data Notification:	There are no techniques available to notify users of the system that data is available in their interest area.
Simulation Modeling:	There is no modeling capability available. Some investigation is underway to look into predictive models.

Analysis: There are elementary analytical techniques available. More sophisticated techniques are being studied such as: multiple regression; correlation, simple and multivariant; Bayesian estimators; confidence limits; and variance and covariance.

Standard analysis routines provide summing, mean and average derivations.

User Orientation: Users of the system are primarily concerned with the use of reliability data during the entire life cycle of aerospace systems.

Various uses of the products of the system for analysts, engineers and management are: monitoring reliability program status; flagging critical problem areas early in a program; measuring contractor performance against incentives; defining reliability improvement programs; making trade-off studies; determining cost effectiveness, e.g., reliability versus maintenance costs.

A lexicon and users guide are provided for untrained users.

Miscellaneous: Programming is done in assembly language for the IBM 7090/7094 computer. Because the status of the system is still developmental, a further review at a future time is needed to complete the study.

All of the program features described in the RECAP system are standard with FFS. Some modifications to standard techniques include increasing the use of floating point calculations and providing for auxiliary programs to be plugged into normal executions.

References: Information on RECAP and FFS was gathered from the following: FFS Generation, Modification, and Updating, 1966; FFS Retrieval Techniques, 1966; Field Trip,* Defense Intelligence Agency, Washington, D. C., April 1966.

AIR FORCE MATERIALS LABORATORY MATERIALS DOCUMENT RETRIEVAL SYSTEM

Data Base Content: The data bank contains references to documents on materials including adhesives, ceramics, cermets, coatings, fuels, lubricants, electrical and electronic materials, fibrous materials, metals, oils, plastics, polymers, and various types of manufacturing procedures and methods.

* Information about field trips is given in Appendix IX.

Input: Data indexing consists of extracting the pertinent document identification data, preparing stylized statements reflecting the important ideas, selecting key words from the stylized statements, and assigning appropriate links and roles to each term. After bibliographic data are copied on cards, the 2 links--reserved for data of document and name of originating organization--are entered on the first two data lines. "A" links are then used for the terms from the first stylized statements. There are 15 roles used. Free indexing is used and followed by a correlation with the vocabulary tape and an error tabulation printout. Punched cards are used to prepare all input to the computer including the thesaurus and dictionary.

Original index cards by accession number and editing cards are retained and stored to permit manual searches.

Output: Printouts show all searches and results. Invalid searches show a tabulation of terms which do not correlate with the thesaurus. These are processed for possible update of the thesaurus and the dual-dictionary.

Storage: The index data are stored to permit either manual or machine search. The storage includes: document card, index cards, vocabulary cards, word access number tape, master word tapes, thesaurus tape, and generic tape. The word-access number tape contains the library of the system.

Retrieval: The search program reads search information from cards containing word code numbers which are sorted in numerical sequence. The sorted input is run against the word master tape. The output is actual valid words on special tape. Valid word numbers are then run against the word-access number tape. All valid document numbers are output on separate tape for printing. Retrieval can be done using both links and roles, links only, roles only, or neither. Links consist of alphabetics while roles are numerics.

Data Framework: Code numbers are used instead of natural language. An inverted coordinate index with document accession number on coded terms is the method of selecting retrieval paths.

The vocabulary and thesaurus structure include the following references:

1) Prime word or phrase
2) See also
3) Post on
4) Generic to
5) Seen from
6) Related to

Control: There is no automatic control built into the programs. Security control is not a factor in this system.

User Orientation: Users are provided "error" and "abort" messages. Searcher is free to use a maximum of 40 words, and to apply links and roles if he so desires. The searcher may use the manual file to obtain more data about the document before making a hard-copy retrieval from shelves. Copies of the thesaurus and dual-dictionary are available to assist searchers in formulating requests.

References: Information on AFML Retrieval System was gathered in a field trip, USAF Materials Laboratory, Technical Information Center, University of Dayton Research Institute, Dayton, Ohio, January 1966.

APOLLO TASK DATA BANK

Data Base Content: The bank contains astronaut performance data for the Command Module (NAA) and the Lunar Excursion Module (Grumman). The data are stored in three levels; function, task, and task element. Time-line parameters are stored by level. A description or function is stored only once, with an associated code number for reference.

Input: Update cards are used to change, add, or modify any item in the function and task banks. An editing process monitors task and element action verbs for standard, acceptable terms.

Procedure cards indicate various sequences of functions and tasks by code numbers. These cards trigger a mission phase development run. Certain options exist such as omission of items, and development of a time-line.

Output: All outputs are direct machine printouts. A run processing update cards will print out the complete content of whichever bank is being updated. The printout is formatted, but without heading information.

Mission phase development runs provide several optional printouts, all utilizing function and task procedure cards. Flight Crew operational time-lines, prespecified function, task and element sequences, astronaut and flight crew performance specifications, start and stop constraints are included as results of the runs.

Storage: The contents of the data bank are stored on magnetic tapes. The bank consists of four master files; function and task bank files. Each file has two subfiles, command module data, and lunar excursion module data. Standard tape storage techniques are used.

Retrieval: No selective retrieval techniques or methods are used in the system. A request for a prespecified mission run is made by manually selecting the program operation and related bank file, and feeding in procedure cards. Certain data are retrieved for processing, but only in the sense that they were specified by the procedure cards for internal use.

Analysis: There is no standard "off-the-shelf" analytical processing done in the system. A discontinuous analysis technique has been applied to determine time-line data. Link analysis processing is available, but is used infrequently due to the need for modification and upgrading.

Simulation Modeling: There is no modeling capability available in the system. Some man-machine modeling is done using data provided by the printouts, but only as indirect nonsystem operations.

New Data Notifications: There are no techniques available to notify users of the system that data are currently available in their interest area.

Data Framework: A standard list of action verbs is used to formulate task descriptions and task elements. This list is provided to data generators in an Interface Control Document, (ICD).

Function and task descriptions are assigned control numbers for reference in processing requests and in updating the banks.

These descriptions and associated components; time, identifying code, etc., are stored by standard tape file procedures--reels, files, records. Serial search paths are followed to locate appropriate records.

Control: There are no automatic or built-in control methods to monitor computer operations; therefore, only manual procedures are used.

Security control techniques are not considered in the system. All control procedures are the responsibility of the individual users.

User Orientation: The prime users of the system, and generators of data, are human factors analysts working on the command module and lunar excursion module performance specifications. Other interested user personnel include operational time-line analysts and system engineers.

A small set of informal documentation is provided to users to assist them in formulating inputs, task and element descriptions, and to assist in reading formatted printouts.

Some uses of the system are the interactive process of mission phase performance specifications, including time-line analysis, determining load constraints on an astronaut, and providing consistency checks e.g., the task "position switch "A" to on" may be declared illegal due to the failure to reset the switch prior to the action.

A generator is assisted in the formulation of a task or task element description by an automatic check mode when either bank is updated. Action verbs are rejected as illegal if they do not match a term in the official, standard list of verbs.

Miscellaneous: The FORTRAN programming language has been used to develop the nine programs.

References: Information on Apollo Data Bank was gathered in a field trip. North American Aviation, Inc., Downey, California, November 1965.

TIME-SHARED DATA MANAGEMENT SYSTEM

Data Base Content: TDMS employs the same general file management techniques to any number of data bases, regardless of content. The contents and structure of each data base are completely user controlled, and can be expressed in terms and relationships that best describe the information from the users viewpoint. Thus, the system is able to take advantage of any inherent data relationship that may exist with respect to the intended use of the data.

Input:	Storage input to TDMS is in two steps: (1) definition of the data base itself, in terms of logical data components, and (2) assignment of values to the elements for each entry in the data base. Definitions and data are specified via a user-oriented language which can be entered on-line with reactive guidance from TDMS. Data can also be input as punched card decks. The definition step permits building in legality and format checks that can be used to verify element values as they are input to the data base.
	The data entries can be mixed. Alphanumeric, symbolic, and textual are the major data types. On-line updating is allowed, whereby any number of entries may be modified in storage.
Output:	TDMS permits a user to outline the format of data retrieved from a data base. Output is either on-line via an interactive console or by line printer. The retrieval results are expressed in nonprogrammer terms.
Storage:	Each TDMS data base is stored as an entity. A large storage capability--tape, disk, and data cell--is available.
Retrieval:	A TDMS data base can be queried via remote interactive console. Retrieval requests can be formatted using AND/OR logic, and specify varying degrees of output, ranging from single values to entire entries in the data base.
	The format of the retrieval output is controlled by the user.
Analysis:	Standard arithmetic capabilities are included in the system in order to update values in the data base and to qualify elements for retrieval output. Qualifiers such as sum, maximum, minimum, count, average, and standard deviation are provided to permit formulation of meaningful retrieval requests from a user's standpoint.
Simulation:	No simulation techniques exist.
New Data Notification:	No automatic dissemination mechanism exists.
Data Framework:	TDMS data base entries are self-defining, permitting efficient data base storage. Entries are stored in value lists, with a separate list of unique values for each element in the data base. Cross-index files provide association between

individual entries by item values. A technique referred to as "repeating groups" enables the system to efficiently access information by taking advantage of user-defined hierarchies that naturally occur in the data. Random-access search strategies are used permitting search paths to begin at the point defined by the group level indicated by the data element specified in the request.

Control: TDMS operates within the framework of the SDC Time Sharing System. All input and output transfers are made by the system and automatic control over program halts is built in the executive program.

User Orientation: Since TDMS is a generalized system, it is quite naturally intended to service a wide variety of users. Since each group of users, however, defines and references each data base in terms of its own needs, the system appears to have a tailor-made aspect for each user.

Communication with TDMS is accomplished in a user-oriented language via on-line interacting console. A reactive guidance dialogue is maintained between program and user, assisting the user in the use of the system. A user needs to have little knowledge of the operation of the program or of the computer.

References: Information on TDMS was gathered from the following Kribs, C. A., 1966; Hopkins, J. S., 1966; Field Trip, System Development Corporation, Santa Monica, California, November, 1965.

APPENDIX VIII

REVIEW OF DATA AVAILABLE FOR ANALYSIS AND SIMULATION APPLICATIONS

DATA ELEMENT[1]	DATA ITEM[2]	ANALYSIS CONTENT[3]	ANALYSIS UNITS[4]	REMARKS[5]
1.0 Object System		None Descriptive identifier of an aerospace system		
2.0 Mission Information	2.1 Mission	None Descriptive identifier of major object system design objective		Requires a translator to convert object system missions to a common reference base for data accumulation
	2.2 Phase	None Qualifier of mission descriptor to detail major identifiable operations		Requires a translator to convert various mission phases to a common reference base for data accumulation
	2.3 Segment	None Qualifier of mission phase descriptor to delineate mutually exclusive individual parts		Requires a translator to convert various mission segments to a common reference base for data accumulation

1 Data Element - See Section V and Appendix IV for definitions

2 Data Item - See Section V and Appendix IV for definitions

3 Analysis Content - Descriptions of the data types that are available, within the stipulated categories of data elements and items, and identification of those types that are amenable to analytic or simulatory manipulation

4 Analysis Units - Identification of the quantitative and nonquantitative analytic units contained within the data content

5 Remarks - Qualifiers, cautions, or references that are applicable to the specific data categories

DATA ELEMENT	DATA ITEM	ANALYSIS CONTENT	ANALYSIS UNITS	REMARKS
3.0 System Information	3.1 System	None Descriptive identifier of major equipment functional groups required in support of the attainment of successful accomplishments of object system design objectives		Requires a translator to convert system descriptors for various object systems to a common reference base for data accumulation
	3.2 Subsystem	None Qualifier of system descriptors to identify the individual functional equipment groups required to accomplish a performance		Require a translator to equate subsystem descriptors for various object systems to a common reference base for data accumulation
	3.3 Component	None Qualifiers of subsystem descriptors to identify the independent and dependent equipment combinations required to accomplish a performance. Components will usually be identified by a Federal Stock (FSN)		Require a translator to equate nonFSN component descriptors to a common reference base for cross object system performance comparison
	3.4 Part	None Qualifiers of subsystem or component descriptors to identify the specific equipment item required to accomplish a performance. All parts will be identified by (continued)		Requires a translator to equate nonFSN part descriptors to a common reference base for inter- or intraobject system performance comparison

DATA ELEMENT	DATA ITEM	ANALYSIS CONTENT	ANALYSIS UNITS	REMARKS
3.0 (cont.) System Information	3.4 Part	a FSN. However, during initial development a contractor's part number will be used pending the assignment of an FSN		
4.0 Performance Description	4.1 Level	Indicator of retrievable details of performance activity from experimental data pool. Reflexive to Mission (2.0) and System (3.0) Information Performance goal or objective	Three category units: a. Position equates with system (3.1) b. Job equates with subsystem (3.2) c. Task equates with component (3.3) and part (3.4)	Each unit is expansive in descending order and contractive in ascending order
	4.2 Description	4.2.1 Verb - behavioral descriptive non-behavioral Indicator of performance action required, of man/machine system, to accomplish Mission (2.0) or System (3.0) requirements	Verb list contained in vocabulary	Verb used may differ in similar contexts in different OSs due to semantic choice of generators. Vocabulary must provide links to resolve semantic differences
		4.2.2 Object - Usually simple or compound noun that describe the (continued)	Object descriptor list contained in vocabulary	Equipment descriptions are controlled by FSN permitting rapid cross-reference within or between object system (continued)

DATA ELEMENT	DATA ITEM	ANALYSIS CONTENT	ANALYSIS UNITS	REMARKS
4.0 (cont.) Performance Description	4.2 Description	locus, man and/or machine, of the performance activity		components or parts. Higher order references may require vocabulary correlations for reference. Operator description requirements, if not standardized AFSC coded, may present semantic problems to be resolved during vocabulary development
		4.2.3 Modifiers - (Adjectives, adverbs, nouns and pronouns). Qualifiers of object (4.2.2) used to specify the object of the performance	Qualifier list contained in vocabulary	Semantic problems of similar descriptors used for different objects within or between object systems require vocabulary correlation
	4.3 Procedural Steps	Combinations of verb-object-modifier sets required to describe the individual sequential operations of the man/machine system required to accomplish the level of performance goal described in 4.2	Vocabulary lists	
		4.3.1 Verb (see 4.2.1)		
		4.3.2 Object (see 4.2.2)		

DATA ELEMENT	DATA ITEM	ANALYSIS CONTENT	ANALYSIS UNITS	REMARKS
4.0 (cont.) Performance Description	4.3 Procedural Steps	4.3.3 Modifier (see 4.2.3) Additional specific values of performance. Requirements and/or tolerances permitted will be included to provide structure to the level of specificity required	Interval, ordinal, ratio scales, individual alpha or numeric values, or statistically calculated values may be used as required	No problem is apparent within each individual level. A problem may develop when referencing within or between object systems due to generators' choice of nonstandard modifier values
5.0 Performance Characteristics	5.1 Location	Specific information identifying the location of the performance being described. Similar performance activities may be required at different places within the aerospace system envelope. Locators may apply to the entire spectrum of system information (3.0)	Simple or compound nouns or numerical designators for a specific location within the object system	Requires a translator to equate inter- or intra-object system locators to a common reference base for cross-system performance comparison
	5.2 Frequency	A designator specifying the number of times a performance is required per Mission Information, (3.0) designator, or per unit of time	An alphanumeric equation usually expressed as a ratio scale	Requires a translator to reduce object system specific designator to a common reference base for cross-system comparison
	5.3 Difficulty	Usually an object system specific (nonstandard) coded entry. The code is assigned, by the data (continued)	An ordinal scale containing n number of values expressed as a rank order presentation	Difficulties are encountered in attempts to relate difficulty scale values developed from (continued)

DATA ELEMENT	DATA ITEM	ANALYSIS CONTENT	ANALYSIS UNITS	REMARKS
5.0 (cont.) Performance Characteristics	5.3 Difficulty	generator, on a subjective judgment basis after an assessment of the performance. Difficulties may be generated by: equipment design, personnel limitations, mission information constraints, etc.		diverse baselines. A correlation index is required for each object system and an integrator for the data system
	5.4 Criticality	Usually an object system specific (nonstandard) coded entry. The code is assigned, by the data generator, on the basis of a subjective judgment. Criticality factors may be generated by: mission success or failure, levels of equipment degradation, personnel/equipment hazards, performance constraints, etc.	An ordinal scale containing n number of values, expressed as a rank order presentation	Difficulties are encountered in attempts to relate criticality scale values developed from diverse baselines. A correlation index is required for each object system and an integrator for the data system
	5.5 Hazards	Statements of an object system specific coded entry describing hazardous conditions. The conditions emanate from object system equipment or operating environmental conditions. Hazardous conditions result from either equipment design and operational requirements or as a result of equipment failure or human error	Statements or an ordinal scale containing n number of values	No problem is apparent in using hazard statements for cross-system performance comparison. Difficulties are encountered in attempts to relate cross-system comparison using a hazard scale. The scale values are developed from diverse baselines thus, requiring an object system correlation index and a data system integrator

DATA ELEMENT	DATA ITEM	ANALYSIS CONTENT	ANALYSIS UNITS	REMARKS
5.0 (cont.) Performance Character-istics	5.6 Training Require-ments	Objective statements describing object-system operator or maintenance personnel training requirements	Statements describing individual training requirements	
	5.7 Special Tools/ Equipment	Descriptive statements regarding requirements for special tools and/or equipment required for performance completion. The type, description, and part number - if available - will also be included	Statements describing special tools, and/or equipment requirements types, and part number related to specific performance	
6.0 Hardware Character-istics	6.1 Accessi-bility	Descriptive statements or object system specific codes regarding an equipment item's accessibility for operation and/or maintenance during a performance	Statements or an ordinal scale containing n number of values	No problem is apparent in using phrase statements for cross-system performance comparison. Difficulties are encountered in attempts to relate cross-system comparison using an ordinal scale. The scale values are developed from diverse system specific bases. Therefore, both an object system correlation index and a data system integrator are necessary
	6.2 Visibility	Descriptive statements or object system specific codes regarding an equipment item's visibility or legibility for operation and/or maintenance during a performance		
	6.3 Manipul-ability	Descriptive statements or an object system specific code regarding an equipment item's manipulability for operation and/or maintenance during a performance		

DATA CONTENT	DATA ITEM	ANALYSIS CONTENT	ANALYSIS UNITS	REMARKS
6.0 (cont.) Hardware Characteristics	6.4 Equipment Status	Descriptive phrases regarding the development state of the equipment used in the performance. These conditions include: mock-up, breadboard, prototype, and production	Nouns - simple and compound	
7.0 Personnel Description	7.1 Type	Descriptive name (job title) of the type of operator/maintenance individual required for the performance; and, if available, the AFSC or other numerical designator. Additional job titles and AFSC's are included for helper/supervisory personnel required	Statement and numerical designator	A correlation index is required to equate job titles and/or numerical designators for each government agency. (AF and NASA). A data system integrator is required to permit cross-system references
	7.2 Number	A value signifying the total number of personnel required for a specific performance (operator, maintenance, helper, supervisor)	Nominal Scale	
	7.3 Special Skills/ Knowledge	Descriptive statements regarding requirements for performance, specific special skills, and knowledge. Reflexive of statements contained in training requirements (5.6)	Statements	

DATA CONTENT	DATA ITEM	ANALYSIS CONTENT	ANALYSIS UNITS	REMARKS
7.0 (cont.) Personnel Description	7.4 New Skills/ Knowledge	Descriptive statements relative to an analytical requirement for new skills or knowledge developed from specific performance requirements. Related to statements contained in training requirements (5.6)	Statements	
8.0 Time Information	8.1 Performance Time	Time values associated with individual performance requirements	Interval Scale	Caution must be exercised to insure that the values included are man values for a particular performance. Fractional man time (hand, foot, etc.)
	8.1.1 Fractional Performance Time	Time values associated with fractional (hand, foot, arm, leg, etc.) performance measurements	Interval Scale	
	8.2 Total Time	Time values associated with the accumulation of performance time (8.1) from a start to stop reference	Interval Scale	
	8.2.1 Elapsed Time	Time values associated with the accumulation of total individual performance times for a specific performance activity	Interval Scale	

DATA ELEMENTS	DATA ITEM	ANALYSIS CONTENT	ANALYSIS UNITS	REMARKS
8.0 (cont.) Time Information	8.3 Time Constraints	Time values associated with constraints that affect the performance accomplishment or initiation, e.g., the performance must be accomplished within X time units, or the performance must stay at Y time units from the start of the mission segment	Interval Scale	
9.0 Remarks		None, except as constraints to other analytical descriptions. Related to the originating element or item by index number		
10.0 Input Identifiers	10.1 Author	None Identification of data generator		
	10.2 Organization	None Identification of data generator contractor/ subcontractor		Provides a search indicator for the object system
	10.3 Date	None Identification of date of preparation of data generation		Provides an indicator of the relevant timeliness of the data
	10.4 Revision	None Identification of the revision of a performance by a data generator		A limitation exists in that the causative reason for the revision is not usually provided

DATA ELEMENTS	DATA ITEM	ANALYSIS CONTENT	ANALYSIS UNITS	REMARKS
10.0 (cont.) Input Identifiers	10.5 Security/ Proprietary	A coded designator of the security/proprietary classification of the performance and for the document.	Ordinal scale containing the following values: T - Top Secret S - Secret C - Confidential U - Unclassified A or RD - Atomic Energy Date P - Contractor proprietary	Codes A, RD, and P may be added to other codes, e.g., TA Top Secret or TRD Atomic Energy Data or Top Secret Restricted Data
	10.6 Type of Performance	None Descriptor that identifies the general classification of the performance, e.g., operational, maintenance, routine, and continuous		
	10.7 References	None Designators of the references used by the performance data generator. Designators of required predecessor or successor performances.		

APPENDIX IX

Field Trips

Air Force Logistics Command (Hq.)
Wright-Patterson Air Force Base
August 1965
Discuss implementation and content of TITAN II Data Pool

Ballistic Systems Division
Minuteman SPO - Airborne Launch Control Center (ALCC) system
November 1965
Discuss task data requirements (exhibits) for ALCC

Ballistic Systems Division
Norton Air Force Base
July 1965
Discuss Minuteman and TITAN II personnel subsystem data problems

Discuss use of current systems for research

Ballistic Systems Division
Norton Air Force Base
November 1965
Investigate the techniques used in the RECAP system

Boeing Company
Development Division
Seattle, Washington
March 1966
Discuss the collection of task data and associated data handling problems with the Minuteman/ALCC project

Defense Intelligence Agency (DIA)
Automatic Data Processing Systems Center (ADPS)
Washington, D. C.
April 1966
Study the Formatted File System (FFS) and other government computer systems

Douglas Aircraft Company
Saturn Division
Huntington Beach, California
March 1966
Discuss the collection of task data and associated data handling problems with the Saturn project

Flight Test Center (Air Force)
Edwards Air Force Base
Lancaster, California
February 1966
Discuss personnel subsystem problems in Category II testing

Lockheed Georgia
Personnel Subsystem Engineering Department
Marietta, Georgia
March 1966
Discuss data problems associated with the design and development of the C5A Transport Aircraft

Marshall Space Flight Center
U. S. Army Missile Command
Huntsville, Alabama
April 1966
Discuss availability of Saturn task data

Discuss data classification and management problems with U. S. Army Engineering Data Center

NASA
Manned Spacecraft Center
Houston, Texas
July 1965
Discuss utilization of NASA systems for research purposes

North American Aviation, Inc.
Apollo Project
Downey, California
November 1965
Investigate the techniques used in the Apollo task analysis data bank

North American Aviation, Inc.
Columbus, Ohio
May 1966
Discuss the definition of the population of analysis and simulation techniques which apply to task analysis data

System Development Corporation
Santa Monica, California
November 1965
Investigate research in natural language queries.

Investigate techniques in LUCID and TDMS information systems

Air Force Materials Laboratory
Technical Information Center
University of Dayton Research Institute
January 1966
Investigate techniques and methodologies used in development of a classification system for materials document storage and retrieval

REFERENCES

Alexander, Lawrence T., and Cooperband, Alvin S., 1965. SP-2168, A Method for System Task Analysis Using Statistical Decision Theory, System Development Corporation, Santa Monica, California.

Automation Services Corporation, 1965. Utilization Study of the TITAN II Basic Data Pool, BSD-TR-65-194, Ballistic Systems Division, Norton Air Force Base, California, pp viii + 88.

Ballistic Systems Division, 1965. Integrated Task Index, Form WS 107A-1, Atlas Series F, Volume 1, BSD Report AP60-0424G, Ballistic Systems Division, Norton Air Force Base, California.

Barton, H. R.; Purvis, R. E.; Stuart, J. E.; and Mallory, W. K., 1964. A Queuing Model for Determining System Manning and Related Support Requirements, AMRL-TDR-64-2 (AD 434 803), Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio.

Bottenberg, Robert, and Ward, H. Jr., 1965. Applied Multiple Linear Regression, PRL-TDR-63-6, 6570th Personnel Research Laboratory, Aerospace Medical Division, Air Force Systems Command, Lackland Air Force Base, Texas, pp viii + 139.

Brewer, R., Q-32 Time-Sharing System User's Guide TINT: An On-Line Program System, 1965. TM-2708/201/00, System Development Corporation, Santa Monica, California, pp 1 + 34.

Bullard, E. C.; Oglebay, F. E.; Munk, W. H.; and Miller, G. R., 1966. A User's Guide to BOMM: A System of Programs for the Analysis of Time Series (AD 629 877), Institute of Geophysics and Planetary Physics, University of California, LaJolla, California, pp ix + 108.

Byer, R. J.; Hardy, R. W.; Lynch, E. J., Jr.; Morse, A. R., Jr.; Pulsak, M. W. and Spitzer, M., 1963. The CAPRI System for Naval Manpower Planning and Control, Vol. II, Computer Program Specifications, Report No. 64-31 (AD 434 572), New Developments Research Branch, Personnel Research Division, Bureau of Naval Personnel, Operations Research Incorporated, Silver Spring, Maryland, pp vi + 161.

Channell, Ralph C. and Tolcott, Martin A., 1960. An Introduction to Human Engineering, Special Devices Center Report 151-1-18, U. S. Naval Special Devices Center, Groton, Connecticut, pp vii + 87.

Coombs, Clyde, 1964. A Theory of Data, John Wiley and Sons, Inc., New York, pp xviii + 585.

Demaree, R. C., 1961. Development of Training Equipment Planning Information, ASD-TR-61-533 (AD 267 236), Aeronautical Systems Division, Wright-Patterson Air Force Base, Ohio.

Devoe, Donald B. and Saul, Ezra V., 1959. The Tufts Index to Human Engineering Literature Human Factors, Vol. IV, Medford, Massachusetts, pp 47-54.

Dixon, W. J. (Editor), 1965. BMD: Biomedical Computer Programs, Health Sciences Computing Facility, Department of Preventive Medicine and Public Health, School of Medicine, University of California, Los Angeles, California, pp viii + 620.

Doyle, Lauren B., 1966. Recent Developments in the Plan for Statistical Classification of Large Numbers of Documents, System Development Corporation, Santa Monica, California.

Drew, D. L.; Summit, R. K; Tanaka, R. I.; and Whitney, R. B., 1966. An On-Line Technical Library Reference Retrieval System, American Documentation, Vol. 1, pp 3-7.

Evans, Carol S., 1959. "A Marginal Punched Card System for Human Factors Literature." Human Factors, Vol. IV, pp 32-46.

FFS Generation, Modification, and Updating, 1966. RADC-TR-66-119, Rome Air Development Center, Research and Technology Division, Air Force Systems Command, Griffiss Air Force Base, New York, pp i + A-5.

FFS Retrieval Techniques, 1966. RADC-TR-66-121, Rome Air Development Center, Research and Technology Division, Air Force Systems Command, Griffiss Air Force Base, New York, pp i + C-7.

Folley, J. D., 1964. Development of an Improved Method of Task Analysis and Beginnings of a Theory of Training, T. R. NAVRADEVCEN 1218-1, U. S. Navy Training Device Center, Port Washington, New York.

Gafarian, A. W. and Walsh, John E., 1966. Statistical Approach for Validating Simulation Models by Comparison with Operational Systems - Illustrated for Traffic Flow, SP-2367, System Development Corporation, Santa Monica, California, pp 17.

Grant, E. E., 1965. The LUCID User's Manual, TM-2354/001/00, System Development Corporation, Santa Monica, California, pp 1 + 69.

Hammid, Ibrahim B.; Dresher, Lillian; and Wade, Carol S., 1965. Survey of Life Sciences Computer Programs, AMRL-TR-65-113 (AD 627 152), Aerospace Medical Research Laboratories, Aerospace Medical Division, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio, pp iv + 173.

Hannah, L. Duncan, and Reed, Lawrence E., 1965. Basic Human Factors Task Data Relationships in Aerospace Systems Design and Development, AMRL-TR-65-231 (AD 630 638), Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio, pp iii + 55.

Hannah, L. Duncan; Boldovici, John A.; Altman, James W.; and Manion, Raymond C., 1965. The Role of Human Factors Task Data in Aerospace System Design and Development, AMRL-TR-65-131 (AD 621 379), Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio, pp vi + 87.

Hilsenrath, Joseph; Ziegler, Guy G.; Messina, Carlin G.; Walsh, Philip J.; and Herbold, Robert J., 1965. OMNITAB: A General Purpose Interpretative Program for the Calculation of Tables of Functions and Statistical and Numerical Analysis, National Bureau of Standards, pp x + 256.

Hogan, M. A., 1966. A General Simulation Routine (AD 633 425), U. S. Naval Radiological Defense Laboratory, San Francisco, California, pp i + 37.

Hopkins, J. S., 1966. The Basic Language Specifications for TDMS, TM-2745/800/00, System Development Corporation, Santa Monica, California, pp 1 + 21.

Howard, W. J. and Inaba, K., 1963. The TITAN II Inertial Guidance System Category II PSTE/MIRR Program: Vol. II, A Guide for Implementing the Program, SERENDIPITY and Associates, Sherman Oaks, California, pp IV + 301.

Janning, Edward A., 1963. Establishment of a Coordinate Indexing Retrieval System for the Air Force Materials Laboratory, RTD-TDR-63-4263 (AD 428 423), Research and Technology Division, Wright-Patterson Air Force Base, Ohio, pp iii + 71.

Janning, E. A., 1965. The Modification of an Information Retrieval System by Improving Vocabulary Control, Indexing Consistency, and Search Capabilities, AFML-TR-65-20 (AD 613 301), Air Force Materials Laboratory, Wright-Patterson Air Force Base, Ohio, pp iv + 81.

Janning, E. A., 1966. Operations of a Document Retrieval System Using a Controlled Vocabulary, AFML-TR-66-36 (AD 633 614), Air Force Materials Laboratory, Wright-Patterson Air Force Base, Ohio, pp iv + 9.

Keenan, James J., 1965. Handbook for the Assessment of Human Performance in Air Force Systems, Vol. 1, Technology for Assessing Human Performance, Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio. (unpublished)

Kribs, C. A., 1966. Central Tables and File Design for TDMS, TM-2745/025/01, System Development Corporation, Santa Monica, California, pp 1 + 58.

Kellogg, Charles H., 1966. An Approach to the On-Line Interrogation of Structured Files of Facts Using Natural Language, SP-243, System Development Corporation, Santa Monica, California, pp 1 + 85.

Kirby, D. B., 1966. Optimization Using a Remote Console for a Digital Computer, Simulation, Simulation Councils, Incorporated, San Diego, California.

Kurke, Martin I., 1961. Operational Sequence Diagrams, Human Factors, Vol. 3, pp 66 + 73.

Lackner, Michael R., 1964. Digital Simulation and System Theory, SP-1612, System Development Corporation, Santa Monica, California, pp 41.

LaJeunesse, D. J.; Weis, E. G. Jr.; and Hogan, T. J. Jr., 1965. SYSTRAN: A Digital Computer Program, AMRL-TR-65-133 (AD 624 468), Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio, pp 315.

Losee, John E.; Payfer, G. E.; Frahm, W. F.; and Eisenberg, Bernie, 1961. Methods for Computing Manpower Requirements for Weapon Systems Under Development, ASD Technical Report 61-361 (AD 264 435), Aeronautical Systems Division, Wright-Patterson Air Force Base, Ohio, pp v + 163.

McLaughlin, J.; Carbee, R.; Reynolds, L; Whitaker, A.; Wiesinger, G.; and Coursen, J., 1963. Design Control Specification for Simulator, LEM Mission, No. LSP-440-43100A, Grumman Aircraft Engineering Corporation, Bethpage, Long Island, New York, pp ii + 82.

McMillan, W., 1964. The Initial Scientific and Technical Vocabulary, System Development Corporation, Santa Monica, California.

Meister, David and Farr, Donald E., 1965. Development of Tests to Measure the Utilization of Human Factors Information by Designers (AD 626 638), Performed for the Engineering Psychology Branch, Office of Naval Research, Systems Effectiveness Department, Bunker-Ramo Corporation, Canoga Park, California, pp ii + 131.

Miller, Robert B., 1953. A Method for Man-Machine Task Analysis, WADC-TR-59-784 (AD 15 921), Wright Air Development Center, Wright-Patterson Air Force Base, Ohio, p 37.

Moore, William Jr.; Meeker, Robert J.; and Shure, Gerald H., 1965. TRACE-Model 1 Time-Shared Routines for Analysis, Classification, and Evaluation, TM-2621, System Development Corporation, Santa Monica, California, p 57.

Murray, Hubert Jr., 1966. Methods for Satisfying the Needs of the Scientist and the Engineer for Scientific and Technical Information, RSIC-510 (AD 627 845), Redstone Scientific Information Center, Redstone Arsenal, Alabama, pp iii + 12.

Olson, Dewey E., 1966. " A Shortcut to Disseminating Information," Electronic Industries, Vol. III, pp 139-140.

Parker, Edwin B., 1966. "The User's Place in an Information System," American Documentation, Vol. 1, pp 26-27.

Pulscak, M. W.; Cleveland, R. H.; Hardy, R. W.; Raphaelson, Elliot; Beeker, George; and Morse, A. R. Jr., 1965. The CAPRI System for Naval Personnel Program Management, Volume I, Description and Operation, TR 322 (AD 468 096), Operations Research Incorporated, Silver Spring, Maryland, pp i + 203.

Purvis, R. E.; McLaughlin, R. L.; and Mallory, W. K., 1964. Queuing Tables for Determining System Manning and Related Support Requirements, AMRL-TR-64-125 (AD 458 206), Behavioral Sciences Laboratory, Aerospace Medical Division, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio, pp iv + 811.

Purvis, R. E.; Mallory, W. K.; and McLaughlin, R. L., 1965. Validation of Queuing Techniques for Determining System Manning and Related Support Requirements, AMRL-65-32 (AD 615 436), Behavioral Sciences Laboratory, Aerospace Medical Research Laboratories, Aerospace Medical Division, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio, pp vii + 126.

Reed, Lawrence E.; Foley, John P. Jr.; Graham, Ralph S.; and Hilgeman, J. B., 1963. A Methodological Approach to the Analysis and Automatic Handling of Task Information for Systems in the Conceptual Phase, AMRL-TR-63-78 (AD 419 018), 6570th Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio, pp V + 120.

Ranco, Paul G., 1960. Human Engineering Information and Analysis Service, Tufts University.

Rubinoff, Morris, 1965. A Rapid Procedure for Launching a Microthesaurus, Paper circulated to IEEE Source Indexing Program.

Sable, J., 1964. Design of Reliability Central Data Management Subsystem, RADC-TR-65-189 (AD 469 269), Rome Air Development Center, New York, pp iii + 5 - 11.

Schaeffer, K. H., and Shapero, Albert, 1964. The Structuring and Analysis of Complex Systems Problems, Paper presented at the Operations Research Society of America, Chicago, Illinois.

Shapero, Albert, and Bates, Charles Jr., 1959. A Method for Performing Human Engineering Analysis of Weapon Systems, WADC-TR-59-784 (AD 235 920), Aerospace Medical Research Laboratory, Wright Air Development Center, Wright-Patterson Air Force Base, Ohio, pp v + 68.

Smith, Bertram J., 1965. Task Analysis Methods Compared for Application to Training Equipment Development, NAVTRADEV CEN-TR-1218-5, U. S. Naval Training Device Center, Port Washington, New York, pp vi + 140.

Tomlinson, Helen, 1965. Classification of Information Topics by Clustering Interest Profiles, PLR-TR-65-19 (AD 628 597), Personnel Research Laboratory, Lackland Air Force Base, Texas, pp iii + 13.

Topmiller, Donald A., 1964. A Factor Analytic Approach to Human Engineering Analysis and Predicting of System Maintainability, AMRL-TR-64-115 (AD 610 210), Behavioral Sciences Laboratory, Aerospace Medical Research Laboratories, Aerospace Medical Division, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio, pp vi + 78.

Ullmann, H. D., 1966. Indexing Techniques, SP-2317, System Development Corporation, Santa Monica, California, pp i + 17.

Van Buskirk, Roger C. and Huebner, Walter J., 1962. Human Initiated Malfunctions and System Performance Evaluation, AMRL-TDR-62-105 (AD 289 036), Behavioral Sciences Laboratory, Aerospace Medical Research Laboratories, Aerospace Medical Division, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio, pp vii + 47.

Whiteman, Irvin R., 1965. The Role of Computers in Handling Aerospace Systems Human Factors Task Data, AMRL-TR-65-206 (AD 631 182), Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio, pp ix + 173.

Williams, William F., 1965. Principles of Automated Information Retrieval, Fundamentals of Indexing, Elmhurst, Illinois, The Business Press, pp 155-144.

Willis, M. P., 1961. Deriving Training Device Implications From Learning Theory Principles, T.R. NAVTRADEVCEN 784-1, U. S. Navy Training Device Center, Port Washington, New York.

Woodson, Wesley E., and Conover, Donald W., 1964. Human Engineering Guide for Equipment Designers, University of California Press, Berkeley, California, pp 1-1 + 7-8.

Young, Gifford A., 1966. NASA Selective Dissemination of Information Program, NASA TM X-57001, National Aeronautics and Space Administration, Washington, D. C.

REGULATIONS MANUALS, MILITARY SPECIFICATIONS, STANDARDS, AND EXHIBITS

The following Air Force and NASA Regulations and Manuals, and Military Specifications and Standards and Exhibits were also used:

AFR 30-8. Development of a personnel subsystem for aerospace systems. May, 1964.

AFR 80-16. Training equipment characteristics (Fifth edition).

AFR 80-29. Scientific and technical information. May, 1964.

AFSCM 80-3-13. Handbook of instructions for aerospace personnel subsystems design. May, 1966.

AFSCR 80-16. Personnel subsystem program for aerospace, support, and command and contract systems. May, 1963.

AFSCM/AFLCM 310-1-C. Management of contractor data and reports. February, 1966.

AFSCM 375-1. Configuration Management. June, 1964.

AFSCM 375-2. System program management and industrial management assistance surveys. June, 1963.

AFSCM 375-3. System Program Office Manual. June, 1964.

AFSCM 375-4. Systems program management manual. June, 1964 (Final Coordination Draft).

AFSCM 375-5. Systems engineering management procedures. March, 1966.

AFSCM 375-6. Development Engineering management. June, 1964.

BSD Exhibit 65-14. Personnel subsystem definition and development. May, 1965.

BUWEPS Exhibit XWR-30. Maintenance Engineering Analysis Record (MEAR).

MIL-STD-803. Human Engineering Design Criteria for Aerospace Systems and Equipment.

MSFC-STD-391 (NASA). Standard, human factors engineering program. July, 1965.

NPC 500-1. Apollo configuration management manual. May, 1964.

ONR-25. Manual for Building a Technical Thesaurus. April, 1966.

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) *System Development Corporation 2500 Colorado Ave. Santa Monica, California 90406	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP N/A

3. REPORT TITLE

THE DEVELOPMENT AND APPLICATION OF COMPUTER SOFTWARE TECHNIQUES TO HUMAN FACTORS TASK DATA HANDLING PROBLEMS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Final report 21 June 1965 - 21 June 1966

5. AUTHOR(S) (Last name, first name, initial)

Potter, K.W.
Tulley, A.T.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
December 1966	165	66

8a. CONTRACT OR GRANT NO. AF 19(628)-3418	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO. 1710	
c. Task no. 171006	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	AMRL-TR-66-200

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
Funded in part by NASA, No. PR-115	Aerospace Medical Research Laboratories Aerospace Medical Division, Air Force Systems Command, Wright-Patterson AFB, Ohio

13. ABSTRACT

Research leading to the application of computer software techniques for handling human factors task data generated in support of aerospace system development programs is discussed. It is recognized that data handling techniques must be developed in context with their total operative environment. A concept of an operational data management system for storing, processing, and retrieving human factors task data in a government/contractor environment is discussed and illustrated. This concept is predicated on the assumption that a user-oriented computerized data system will help draw human factors specialists closer to their data. Five problem areas, considered to be fundamental to the development of data handling techniques, were researched. These areas are: (1) analysis of human factors task data, data relationships, and classification schemes, (2) application of vocabulary and thesaurus techniques to increase the effectiveness of communication among man/machine/software functions, (3) application of computer storage and retrieval techniques to human factors task data, and (5) application of current awareness techniques to provide notifications of data availability.

DD FORM 1 JAN 64 1473

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Personnel subsystem data						
Task analysis data						
Information storage						
Information retrieval						
Data processing						
Data classification						
Current awareness						
Automated systems						
Data management						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through
_______________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through
_______________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through
_______________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.