MARTIN MARIETTA ENERGY SYSTEMS LIBRARIES

3 4456 0290230 3

ORNL/TM-10832

ornl

OAK RIDGE NATIONAL LABORATORY

MARTIN MARIETTA

The Marine Corps Enlisted Personnel Information Delivery System (MC-EPIDS): Prototype Design

L. F. Truett
C. S. Bowman
R. P. Rascoll
D. M. Flanagan

OAK RIDGE NATIONAL LABORATORY
CENTRAL RESEARCH LIBRARY
CIRCULATION SECTION
4500N ROOM 175
LIBRARY LOAN COPY
DO NOT TRANSFER TO ANOTHER PERSON
If you wish someone else to see this report, send in name with report and the library will arrange a loan.

OPERATED BY
MARTIN MARIETTA ENERGY SYSTEMS, INC.
FOR THE UNITED STATES
DEPARTMENT OF ENERGY

Printed in the United States of America. Available from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road, Springfield, Virginia 22161
NTIS price codes—Printed Copy: A04; Microfiche A01

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

ORNL/TM-10832

Energy Division

THE MARINE CORPS ENLISTED PERSONNEL INFORMATION DELIVERY SYSTEM (MC-EPIDS): PROTOTYPE DESIGN

L. F. Truett
C. S. Bowman
R. P. Rascoll*
D. M. Flanagan

*Evaluation Research Corporation

D. M. Flanagan
Project Manager

Date Published - March 1989

Prepared for the
NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER
San Diego, California
under Interagency Agreement DOE 40-1730-86

Prepared by the
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37831
operated by
MARTIN MARIETTA ENERGY SYSTEMS, INC.
for the
U.S. DEPARTMENT OF ENERGY
under Contract No. DE-AC05-84OR21400

MARTIN MARIETTA ENERGY SYSTEMS LIBRARIES

3 4456 0290230 3

CONTENTS

LIST OF FIGURES v

ACKNOWLEDGMENTS vii

ACRONYMS AND VARIABLE NAMES ix

ABSTRACT xi

1. INTRODUCTION 1

2. SYSTEM DESIGN 2

2.1 Background 2
2.1.1 The User Community and Current Operating Environment. 2
2.1.2 MC-EPIDS Requirements 3
2.1.3 Structured Analysis of MC-EPIDS 3
2.2 Data Base Design 8
2.2.1 NPRDC Requirements for Data Retrieval 8
2.2.2 Design Problem 9
2.2.3 Comparison of Design Alternatives 10
2.2.4 Creation and Modification of the Data Base 13
2.2.5 Overall System Configuration 15
2.3 User Interface Design 17
2.3.1 Tutorial 17
2.3.2 Screens Design 18
2.3.3 Help Files 20
2.3.4 Function Keys 21
2.3.5 Output Formats 22
2.4 Communications Options 22
2.4.1 Communication Lines 22
2.4.2 Hardware Requirements 23
2.4.3 Software for Terminal Emulation and Communication . 23

3. PROBLEMS AND CONSTRAINTS 26

4. SUMMARY AND CONCLUSIONS 29

APPENDIX A: Process Specifications and Data Dictionary for the MC-EPIDS Structured Analysis A-1
APPENDIX B: Alternative MC-EPIDS Data Base Designs B-1
APPENDIX C: Screens Viewed by the End User C-1
C.1 Tutorial
C.2 Menu Screens
C.3 Help Screens
C.4 Output Formats
C.5 Error Messages

LIST OF FIGURES

Fig.		Page
1	Context diagram	4
2	Level 0, diagram of the highest processes within MC-EPIDS..	5
3	Level 1.0, diagram of the breakdown of process 1	6
4	Level 2.0, diagram of the breakdown of process 2	7
5	MC-EPIDS data base tape end statistics	12
6	MC-EPIDS system configuration	16
7	MC-EPIDS program progression	19

ACKNOWLEDGMENTS

A significant contribution to this project was provided by Kevin A. Newman and Jonathan C. Goyert of Science Applications International Corporation (SAIC) of Oak Ridge, Tennessee.

ACRONYMS

CMS	Conversational Monitor System
CPU	Central Processing Unit
DBMS	Database Management System
DDN	Defense Data Network
MCDN	Marine Corps Data Network
MC-EPIDS	Marine Corps Enlisted Personnel Information Delivery System
ORNL	Oak Ridge National Laboratory
OS/VS1	Operating System, Version 1
NPRDC	Navy Personnel Research and Development Center
PC	Personal Computer
SAS	Statistical Analysis System
USERID	User Identifier
VM	Virtual Machine

VARIABLE NAMES

FY	Fiscal Year
OCC	Occupational Field Identifier
PAY	Paygrade Identifier
POP	Population Group Identifier
SEX	Sex Identifier
YOS	Year of Service Identifier

ABSTRACT

The Marine Corps Enlisted Personnel Information Delivery System (MC-EPIDS) provides Marine Corps manpower analysts with computer-based tools for retrieving, displaying, and analyzing historical enlisted manpower statistics. This report describes the prototype version of MC-EPIDS.

The prototype is implemented using the Statistical Analysis System (SAS) language. Data are stored in SAS datasets; procedural programming is in Base SAS; menu generation is supported by SAS/AF; and graphic displays are produced by SAS/GRAPH.

The development of MC-EPIDS required extensive research to overcome restraints caused by hardware, telecommunications, and SAS language limitations.

1. INTRODUCTION

Under an interagency agreement with the Navy Personnel Research and Development Center (NPRDC), the Oak Ridge National Laboratory (ORNL) developed a prototype for the Marine Corps Enlisted Personnel Information Delivery System (MC-EPIDS). On completion, system responsibility was transferred to NPRDC in San Diego, California. NPRDC will maintain the MC-EPIDS program and data base.

With MC-EPIDS, Marine Corps manpower analysts located in Arlington, Virginia, can rapidly retrieve and display a variety of historical enlisted manpower statistics in graphical and tabular formats. Typical operations include (1) examination of personnel statistics for single as well as multiple years, and (2) comparison of personnel statistics across a variety of data dimensions.

Based on an assessment of cost, feasibility, and maintainability, NPRDC requested a system based totally on the Statistical Analysis System (SAS). The current prototype is built on this concept.

MC-EPIDS was prototyped without life cycle management structure; NPRDC is responsible for supplying these principles to the final system development process. NPRDC is also responsible for writing all user manuals, for the training associated with the system, and for maintenance of the system.

This report has four sections. Section 2 presents the system design of the project MC-EPIDS. A structured analysis of MC-EPIDS was completed to help with design decisions. The data base design, user interface considerations, and communications options are discussed in Sect. 2. Section 3 lists problems and constraints to the design of this system. Section 4 summarizes the project and makes recommendations for improvements. Appendix A provides additional information on the structured analysis, and Appendix B discusses alternative data base designs. Appendix C contains copies of all screens viewed by the end users of MC-EPIDS.

2. SYSTEM DESIGN

This chapter presents the design of the MC-EPIDS prototype. To provide a background for discussion, the chapter first describes the current NPRDC operating environment and the goals which NPRDC wished to satisfy through the MC-EPIDS project.

2.1 Background

In implementing a system such as MC-EPIDS, ORNL system developers normally take the approach of interviewing the end-users of the system and studying first-hand the environment in which they work. However, for this phase of the project, NPRDC decided to act as an intermediary between the ORNL developers and the end-users. NPRDC made this decision because it felt that user needs and system requirements had already been determined and additional discussions with end-users were not vital to accomplish implementation objectives.

After learning about user requirements, ORNL analysts developed the functional definition of MC-EPIDS through use of structured analysis, which formed the basis of the prototype's functional definition.

2.1.1 The User Community and Current Operating Environment

The prototype is intended to serve about ten analysts at Marine Corps Headquarters in Arlington, Virginia. The analysts consist of primary users, who respond to queries for historical Marine Corps enlisted personnel data at all levels of detail, and secondary users, who depend on primary users for help in writing programs and manipulating data.

The historical Marine Corps manpower data is kept on 16 to 18 magnetic tapes and is accessed through an IBM 4341 mainframe computer system located on site at NPRDC in San Diego, California. This computer system operates under the VM/CMS operating system. The primary users may direct queries for historical data by submitting batch mode tape jobs as needed to the NPRDC computer system. The users submit these queries remotely from Arlington. The primary analysts have available any of the languages (such as FORTRAN, APL, and SAS) currently installed on the IBM 4341 computer system at NPRDC.

There are a number of problems with this arrangement. First, the primary users have to write a computer program for each new type of query they submit to the NPRDC system. This places a burden on the primary analysts and makes the secondary analysts dependent upon them for information. Second, these custom-written programs tend to generate different answers to similar queries. This is due to inconsistencies in the way programs are coded, in the way statistical techniques are employed, and in the practice

by analysts of mixing hand calculations with computer output in different ways. Finally, besides being error-prone, this arrangement is inefficient. Writing and compiling programs is a tedious process that requires manpower and training. Reading data from magnetic tapes is a relatively slow computer operation, because of the need for system operator intervention during program execution. Batch mode processing can be slow, especially during periods of peak CPU usage.

2.1.2 MC-EPIDS Requirements

As a way of reducing these problems, NPRDC sponsors wanted a software system installed on their IBM 4341 mainframe which users in Virginia could access interactively by means of TYMNET to produce statistical analyses quickly and efficiently. They wished to present a system to Marine Corps analysts that did not require coding and would provide reproducible answers to queries from data stored on disk. In addition, NPRDC wished to be able to provide the end-user with graphics output capability as well as textual output. The system had to be able to support a variety of terminal types; namely, Tektronix 4010 and 4105 terminals, VT100 terminals, and IBM PC or compatibles with asynchronous, ASCII access through TYMNET. NPRDC wanted the system to be easy to use, with the appropriate built-in help facilities, in order to give the secondary analysts freedom to work more directly with the data.

MC-EPIDS was developed using Base SAS, SAS/AF, and SAS/GRAPH. SAS/AF is used to design the screens for the menus. The SAS/GRAPH software is used to produce the graphics for MC-EPIDS. Many options normally available with SAS-GRAPH are not available in MC-EPIDS because of the limitations imposed by interactive applications over dial-up data lines. Collectively, the features of the Base SAS, SAS/AF, and SAS/GRAPH provide Marine Corps personnel planners and managers flexibility in the use of MC-EPIDS information. Because the end users only manipulate the data and never update the data base, the data is constant and reliable. (The data base is updated annually.) Thus, Marine Corps personnel planners can use the system without help from NPRDC.

2.1.3 Structured Analysis of MC-EPIDS

The logic flow of MC-EPIDS has been defined using structured analysis techniques. Structured analysis is a method that is best employed when the actual end-users of a system may be interviewed. For this project, NPRDC responded to the questions normally asked of the end users. The total logic flow is shown in the data flow diagrams presented in Figs. 1-4. A brief explanation of each of the processes that occur on these data flow diagrams is given below. Process descriptions and the data dictionary are provided in Appendix A.

Figure 1, the context diagram, shows all inputs and outputs of MC-EPIDS; that is, it defines the context of the system. As discussed in Sect. 3.2,

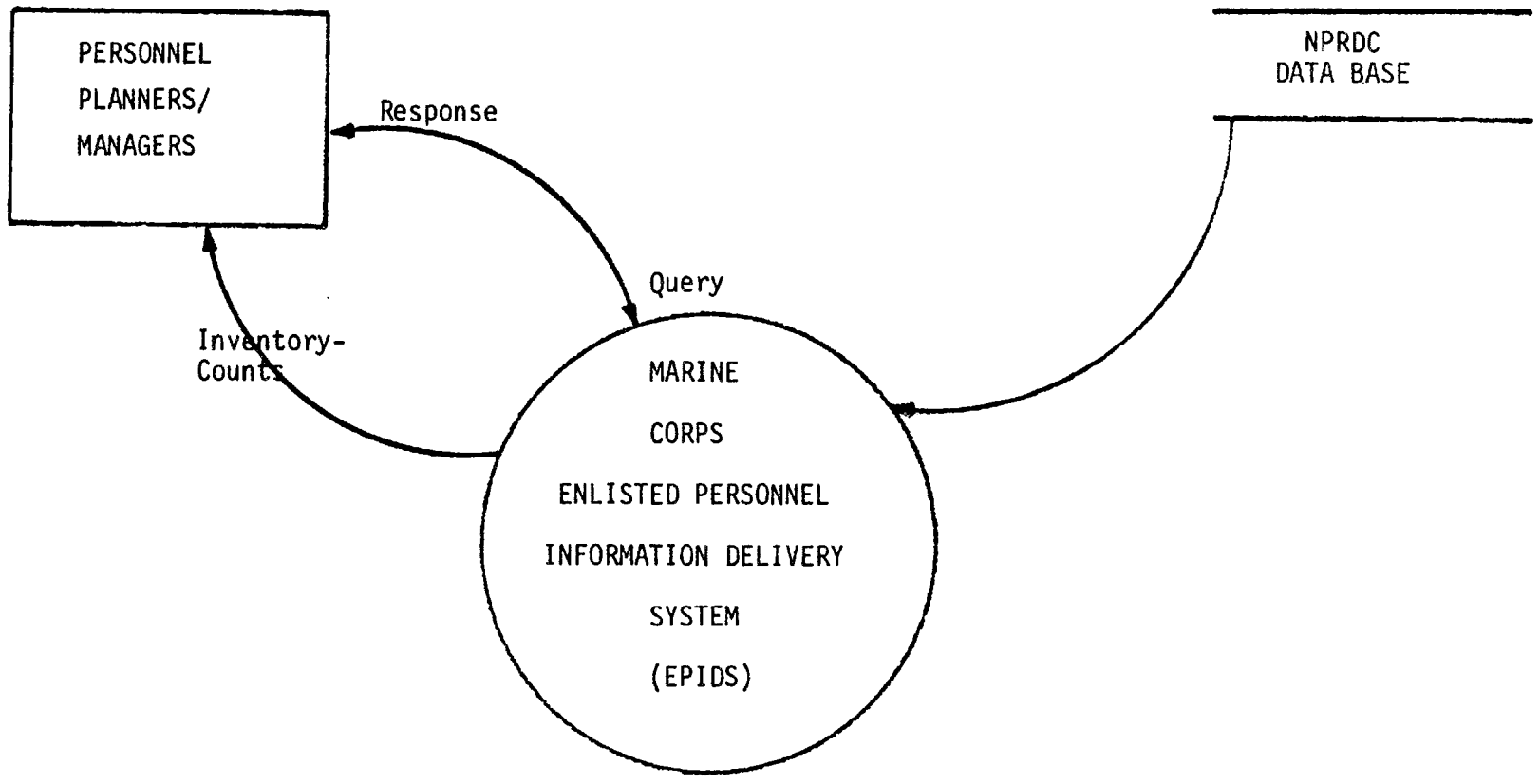


Fig. 1. Context diagram.

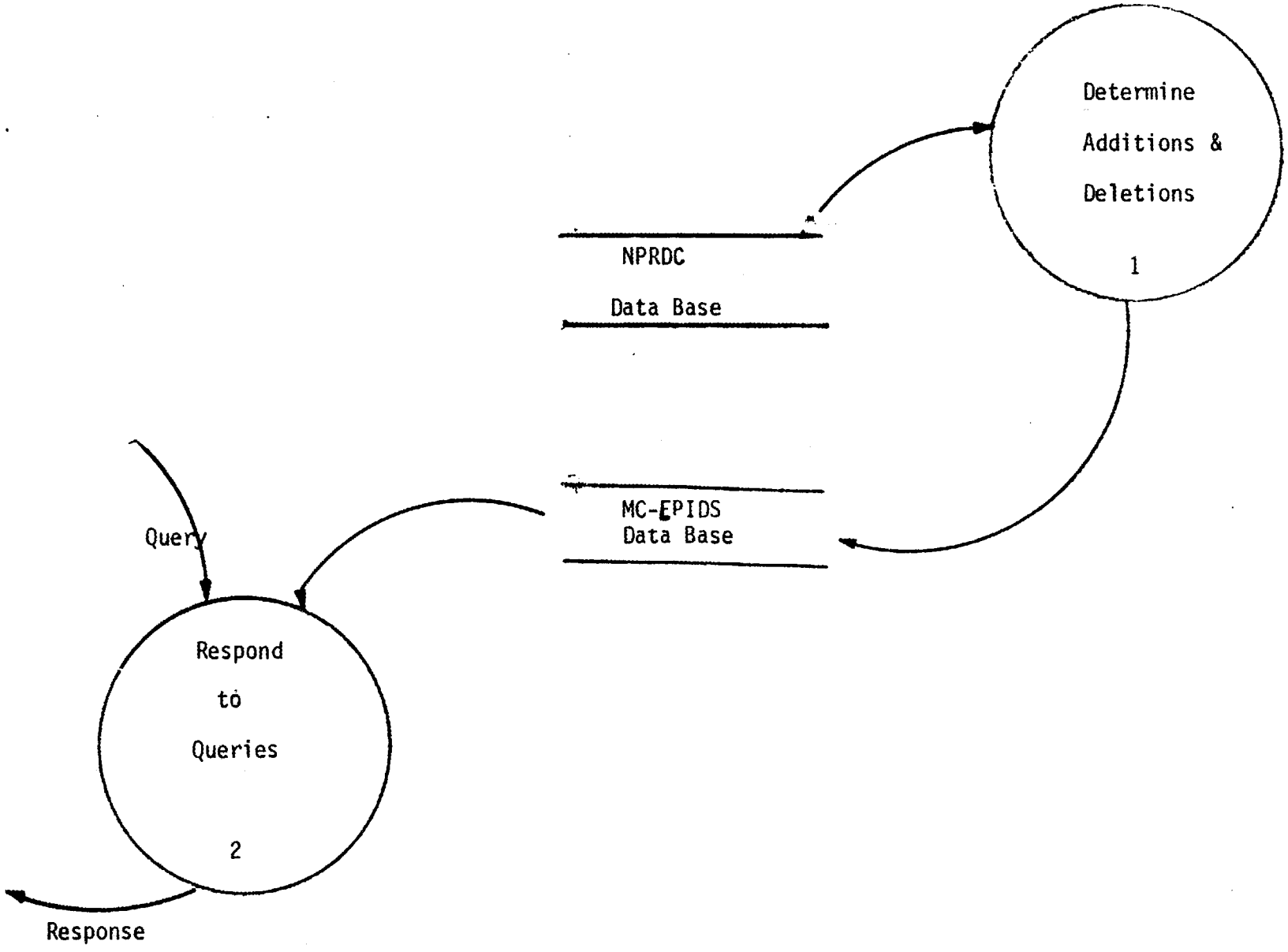


Fig. 2. Level 0, diagram of the highest processes within the MC-EPIDS.

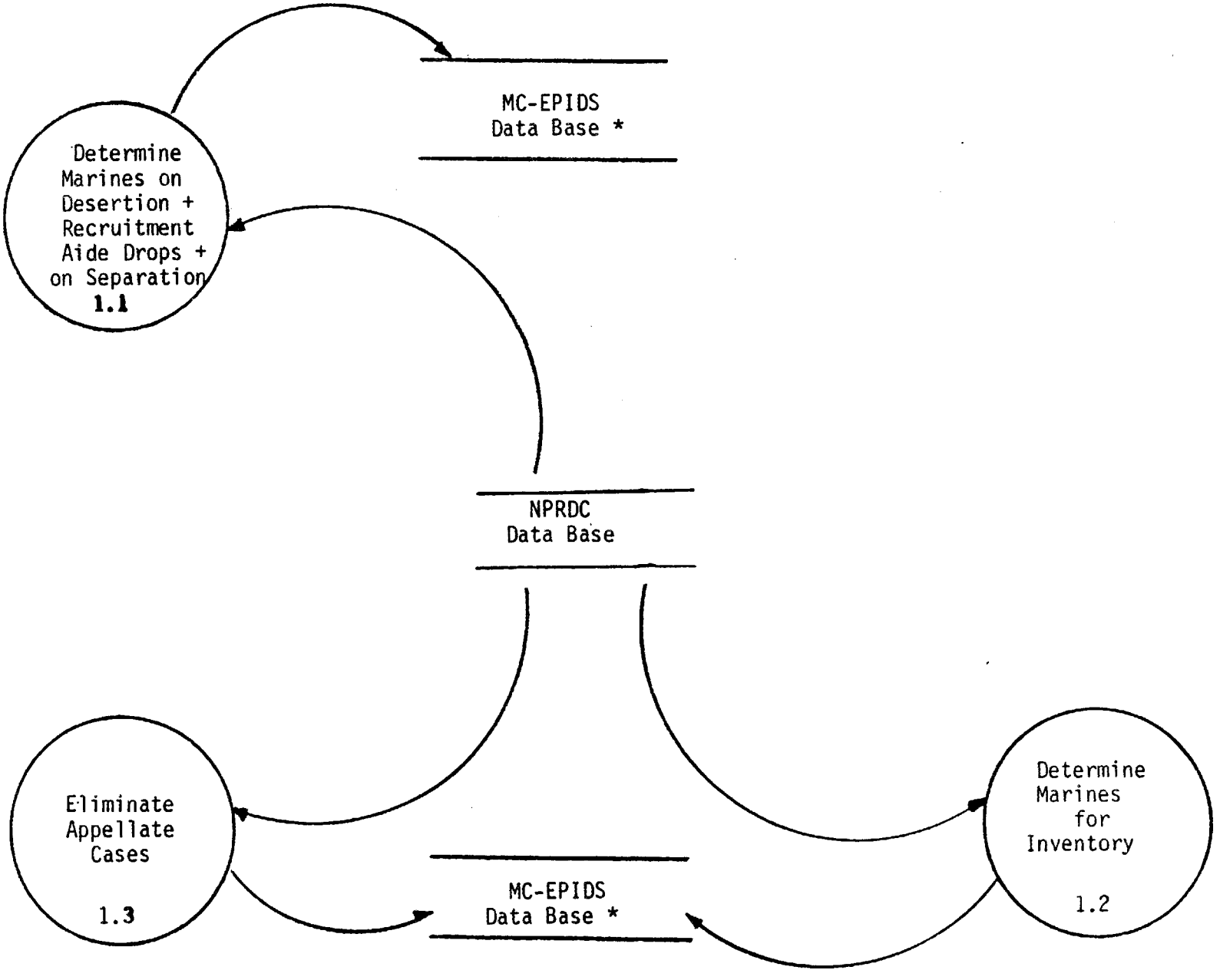


Fig. 3. Level 1.0, diagram of the breakdown of process 1.

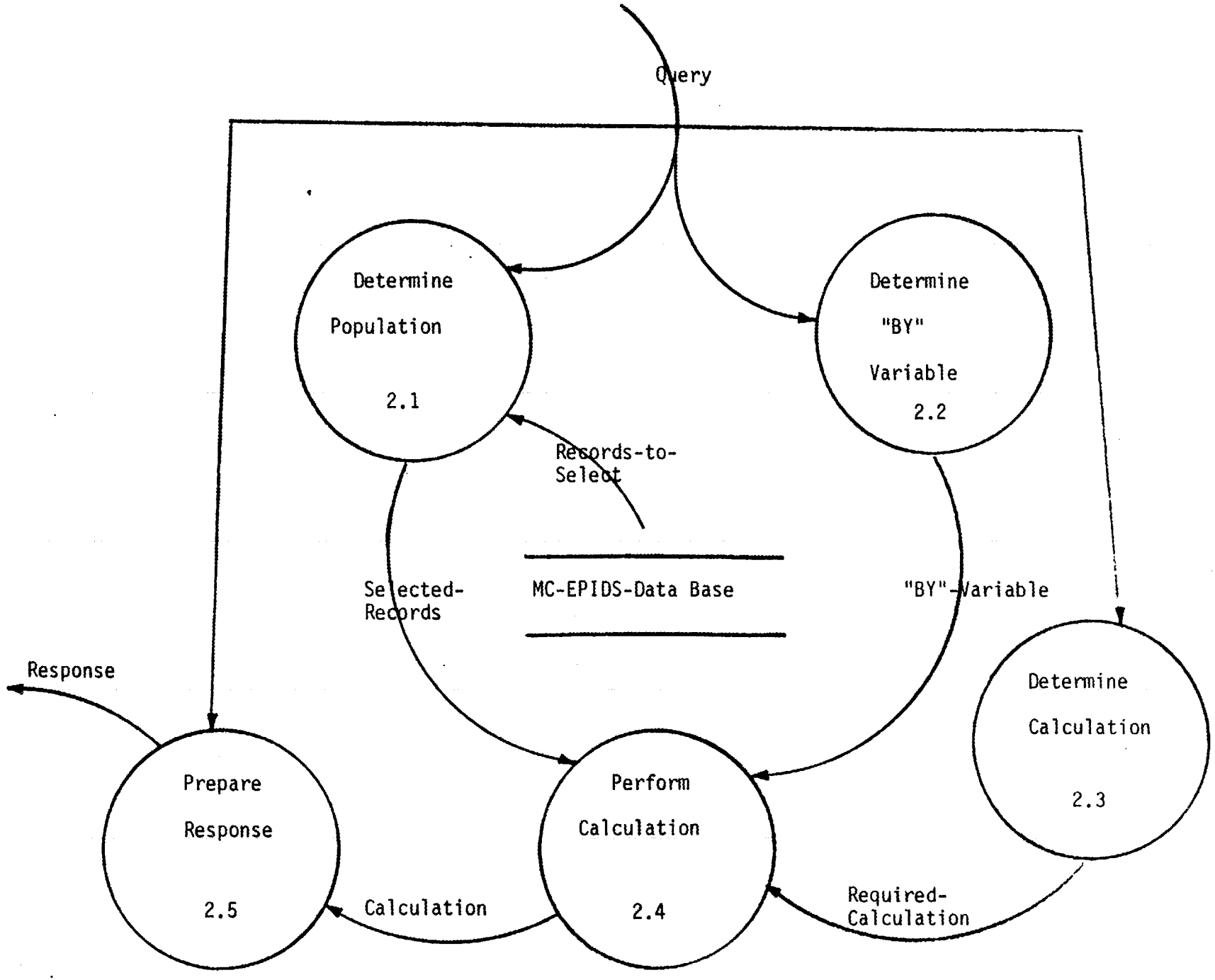


Fig. 4. Level 2.0, diagram of the breakdown of process 2.

the MC-EPIDS data base is derived from the NPRDC data base. The end users of MC-EPIDS, through queries of this data base, receive inventory counts and responses to the queries.

Figure 2 is a diagram of the two highest-level processes of MC-EPIDS. These processes are further explained in Figs. 3 and 4. Figure 3 describes how the MC-EPIDS data base is formed from the NPRDC data base. Figure 4 shows the data flows for the queries formulated by the MC-EPIDS users. Additional information on these figures is in Appendix A, which contains the process specifications and data dictionary for the structured analysis of MC-EPIDS.

2.2 Data Base Design

2.2.1 NPRDC Requirements for Data Retrieval

As originally proposed by NPRDC, the MC-EPIDS system exhibited a number of special challenges for the project design team. NPRDC expressed the need for a system which could direct queries of a particular type quickly and efficiently against a data base of moderate size which was composed of a fixed set of data elements. This set of data was not to be modified by the Marine Corps end user. In fact, NPRDC personnel only planned to update the data set, themselves, at one-year intervals. The use of a general-purpose data base management system, while still a viable design alternative under these conditions, was not an absolute necessity. An off-the-shelf data base management system usually provides many more features for manipulating data within a data base than is required by such an application. With this consideration in mind, and taking into account that NPRDC already had Base SAS and SAS/GRAPH installed on their mainframe computer, the project sponsor decided to acquire another SAS product, SAS/AF, to add interactive menu capability to their existing SAS software and to develop MC-EPIDS as a totally SAS-based system.

The NPRDC tape data, from which MC-EPIDS was to be built, contained on the order of 200,000 individual Marine records for each fiscal year. Data existed for seven fiscal years -- 1980 to 1986. One tape held one fiscal year's data. As each tape record contained approximately 150 bytes of information for each individual Marine, a data base storing tape data from seven years of data would have consumed about 200 megabytes of storage. However, it was unnecessary to store this much information for the type of queries which NPRDC wished to perform. For those queries, only the data stored within certain data fields of the individual tape records were needed. In addition, it was decided that the data base should not store individual records or even portions of those records at all. Instead, NPRDC required that the data be "aggregated" or stored in some pattern that would minimize disk space and speed up the retrieval of information by having some of the required computations for the most frequently occurring types of queries already performed and the results contained within the data base as part of the data structure. This would help optimize response times for those queries.

The following were the most important constraints placed upon the data base design:

- It should be SAS based;
- The data should be "aggregated" in such a way that records are not stored individually;
- Disk storage should be minimized;
- It should provide response times quick enough to support interactive queries.

The question of what constitutes a suitable response time is complicated by a number of human factors and was an open issue at the start of the project. For a large mainframe system, the response depends on how much user overhead is tasking the system at any given moment, as well as on the efficiency of the terminal network configuration in place on the system. Since this overhead must be included in the overall response characteristics of the software, it seemed important to try to keep the time for retrieval and calculations as short as possible, although it was understood from the beginning that response times as short as 15 sec would be impossible to guarantee. Nevertheless, the MC-EPIDS project team explored several data base designs in an effort to discover one that afforded a good approach to maximizing the speed of retrieval with a minimum number of calculations per query.

2.2.2 Design Problem

As a first step in developing a design concept for the MC-EPIDS data base, the NPRDC project team took the approach of trying to view in some formal way what these data represent. From discussions, the following description has emerged as a statement of the design problem.

The sample space for the MC-EPIDS data base is discrete and six-dimensional. We represent this sample space as a 6-tuple

(OCC, PAY, YOS, SEX, POP, FY)

where the six data dimensions are

OCC = occupational field
PAY = paygrade
YOS = years of service
SEX = sex
POP = population group
FY = fiscal year

If D represents any one of these data dimensions, then define N(D) to be the number of possible discrete values for dimension D; for example, if D is dimension PAY, then D has nine values

$$D = PAY = \{E1, E2, \ldots, E9\}$$

and

$$N(D) = 9.$$

Currently, the total number of data points in the MC-EPIDS sample space can be calculated to be

$$\begin{aligned} &N(OCC) * N(PAY) * N(YOS) * N(SEX) * N(POP) * N(FY) \\ &= 43 * 9 * 31 * 3 * 4 * 7 \\ &= 1{,}007{,}748 . \end{aligned}$$

This figure will increase as yearly data is added -- the incremental increase per year being uniform and independent of the FY data dimension. Thus for each fiscal year the number of sample points required is

$$\begin{aligned} &N(OCC) * N(PAY) * N(YOS) * N(SEX) * N(POP) \\ &= 143{,}964 \text{ points per year} . \end{aligned}$$

If an integer value equal to the count of all tape records whose corresponding data fields match those of the given point is assigned to each point of the sample space, then the six-dimensional vector of the sample space becomes a discrete, random variable with integer values. It represents a joint, unnormalized probability distribution. The MC-EPIDS project team believed that queries on this data base are all answerable through the construction of marginal distributions of one or more variables generated from the joint distribution of this sample space. The statistical functions required by the statement of work can all be determined from such distributions. Thus, the data base design problem appeared to be one of designing an efficient means of storage to allow for the rapid computation of these marginal distributions.

2.2.3 Comparison of Design Alternatives

One of the early difficulties in designing a data structure for the MC-EPIDS data base was that ORNL project team members had very little intuitive feeling for how the frequency counts in the sample space described above might be distributed. In particular, it was not known how many combinations of OCC, PAY, YOS, SEX, and POP would correspond to zero frequency counts. The response characteristics of any data structure for MC-EPIDS depends greatly upon how sparse the frequency distribution is. Therefore, in the early stages of design, the project team considered the worst-case possibility, in which most frequency counts were non-zero. It is generally recognized that SAS tends to slow down severely when the number of observations in a SAS data set approaches 100,000 or so. The possibility existed that data sets of this size might, indeed, arise in this application.

The team explored a number of different design approaches, all based in some way upon the idea of frequency distributions. They studied these designs in terms of response times, storage requirements, and how well the design fulfilled the sponsor's requirements. All of these approaches allowed aggregated data to be stored in pre-computed patterns that would serve to speed up the response for certain frequently asked queries. Unfortunately, some of these do not qualify as strictly SAS data sets and would have had to be interfaced with a driver routine written in another language such as PL/1. However, once the project team gained experience reading the tape data, it became apparent that these frequency distributions were actually extremely sparse. As much as 92% of the possible combinations of OCC, PAY, YOS, SEX, and POP corresponded to zero frequency counts. (Figure 5 shows the record counts and frequency counts obtained while reading the NPRDC tape data.)

The data set design finally used for this project is totally SAS-based. It represents a compromise between trying to optimize SAS computations by reducing the number of SAS observations in the data set and taking advantage of the high proportion of zero frequency counts. Below is a description of the final SAS data set design. Descriptions of the alternative designs are left in Appendix B.

The SAS data set design chosen for MC-EPIDS has the following schematic form:

	FY	OCC	PAY	POP	SEX	YOS1	YOS2	...	YOS31	TOT-COUNT
OBS1:										
OBS2:										
OBS3:										
.										
.										
.										

In this data set, "FY" is a two-byte tag that identifies to which fiscal year each observation belongs. The other variables have meanings as given in the following table:

CLASSIFICATION VARIABLES:		
FY	2 bytes	fiscal year identifier
OCC	2 bytes	occupational field identifier
PAY	1 byte	paygrade identifier
POP	1 byte	population group identifier
SEX	1 byte	sex identifier
FREQUENCY COUNTERS:		
YOS1 - YOS31	4 bytes each	year of service frequency counters
TOT-COUNT	4 bytes	total year of service counter
total	135 bytes	

FISCAL YEAR	TAPE NUMBER	RECORDS READ	RECORDS KEPT	NUMBER OF OBSERVATIONS IN SIMPLE FREQ. COUNT	NUMBER OF OBSERVATONS IN FREQ. COUNT PARTITIONED BY YOS
1980	90-1438	190,733	170,281	8196	1779
1981	90-1712	194,794	172,308	8385	1776
1982	90-2239	197,599	173,863	8815	1822
1983	90-2241	198,943	174,113	9104	1833
1984	90-2247	200,197	175,848	9200	1831
1985	90-2455	202,355	177,859	9400	1849
1986	90-2469	200,595	178,590	9601	1847
				TOTAL =	12,737

Fig. 5. MC-EPIDS data base tape end statistics.

The frequency counters YOS1 through YOS31 give the total number of observations for a given combination of classification variables FY, OCC, PAY, POP, and SEX with that particular value of year of service.

Experiments with the NPRDC tape data in which frequency counts on the individual records were performed to put that data into the above format showed that the total number of SAS observations required for each fiscal year was no more than 2000. After the data base was finally created, the number of observations for the entire seven years came to only 12,737. This data set size can be managed quite well by SAS. Altogether, the data base in this form occupies a total of 3.4 megabytes. Queries on this data base, depending on how much load is on the NPRDC computer system, usually take from one to three minutes.

2.2.4 Creation and Modification of the Data Base

The technique used to create the data base is relatively straightforward. A utility module was written in SAS to read the NPRDC tape data in yearly increments, process the data, and place it in the MC-EPIDS data base in the correct SAS format. As one fiscal year of data was being read, each individual record was examined to determine if it needed to be included in the data or rejected under criteria established by NPRDC. For those records kept, the relevant data fields were extracted and stored in a temporary SAS data set. Next, the utility module performed a simple frequency count of the observations in this data set and placed these counts in another temporary data set which gave the frequency distribution of all combinations of OCC, PAY, YOS, SEX, and POP that were read from the original tape data. From this frequency distribution, once more a new distribution was created which gave the frequencies counted by year of service as in the data set format above. This last data set was stored with the data for previous years that had already been added to MC-EPIDS. This procedure was continued for each year until all seven years were processed.

Limitations imposed by the NPRDC computer system require that users perform all tape processing interactively under the CMS operating system. This is because SAS has not been installed to run under OS at NPRDC. The steps involved in reading a data tape using the data base creation module may be given as follows:

1. Log on to the CMS USERID account where the data base and utility module reside and define temporary storage by entering the commands

```
DEFINE T3380 166 CYL 200
FORMAT 166 K
ACCESS 166 K
```

2. The utility module is located in a file called CREATE SAS. It is necessary for the individual who is performing the data base creation module to make one small modification to this program each time it is run. By going into XEDIT, the user must set the

fiscal identifier tag, "FISCAL," to the current year being added. The place in the program where this tag must be changed is near the first of the program and is well marked by comments. As an example, if the current year being added is 1987, then the tag, "FISCAL," should be changed to "87."

3. Issue a request to the system operator to mount the desired tape from which the new data are to be extracted. Use the command

   MOUNT ####### 181 6250 NO-WRITE

   where "########" is the tape number of the desired tape. Wait until the system operator responds with a message saying the tape is ready.

4. Type the command SAS CREATE to run the utility data base creation module. Wait until the program has completed. This can take from ten minutes to an hour, depending upon the user load on the system.

5. When the tape job has finished, type the command

   DET 166

   to let the operator know that the tape drive is no longer in use.

By following these steps, the user can add a complete year's worth of data to the MC-EPIDS data base. The above steps give the entire files maintenance procedures included in MC-EPIDS for the project sponsor at NPRDC. This was the process by which the original data set was built, starting with the year 1980 and adding one year at a time through 1986. Figure 5 gives the number of records read from each NPRDC data tape during this process along with the number of observations contained within the various frequency counts performed by the utility module. Data in future years may be added to the data set in exactly same way.

In addition to the utility data base creation module, two other small utility modules have been included with the software. One module, in the CMS file called ARCHIVE SAS, can be used to back up the current data set on tape. This module records the data in nonlabeled format and will erase any old data that happens to be on the tape. The other utility module, in the file called FETCH SAS, provides the facility to retrieve easily the data that were backed up using ARCHIVE SAS. It is only necessary to request the system operator to mount the desired tape as in STEP 3 above and issue the command SAS ARCHIVE or SAS FETCH to run the corresponding module.

2.2.5 Overall System Configuration

Having described how the data set was designed and created from the NPRDC data tapes, it remains to explain how the data are integrated into the overall MC-EPIDS operating software and how the data base is accessed by users of this system. Two independent primary program modules comprise most of the MC-EPIDS operating software. The first module was described above as the data base creation module. It is used only to generate the MC-EPIDS data base and to update it at yearly intervals. The second module is the user interface to the data base and provides the facility for performing the interactive queries within the system. This user interface module accesses the data in read-only mode. Figure 6 depicts how these two primary modules are related to each other and to the NPRDC tape data and MC-EPIDS data base.

The computer system at NPRDC is an IBM 4341 mainframe using the CMS operating system. Since SAS does not run under OS at this installation, it was necessary to build the data base within the file structures provided by CMS. It was decided to develop the software prototype within the environment of a normal CMS user minidisk. Alternatives to this approach may become feasible in the future.

NPRDC personnel established for this project one oversized CMS A-minidisk to contain all of the data base along with whatever portions of the operating software that could reside apart from the user. This A-disk is about 5.8 megabytes in size. The data base was created entirely on this CMS USERID area. The operating software portion of the project was developed primarily on another, smaller-sized USERID. Programmers using this account linked to the data base on the larger USERID in order to access the data in the same way the end user would have to do when the prototype was completed.

Users are linked automatically to the data base when they log on to the NPRDC system through the facility of a CMS EXEC procedure written by the project team and provided as part of the total MC-EPIDS package. This EXEC, called USER EXEC, also provides automatic allocation of temporary disk space required by the operating software during execution. The USER EXEC allocates temporary disk space on IBM 3380-type direct-access volumes in blocks of fifty cylinders per user. Project development included a thorough testing of these linking procedures, not only by the system programmers, but also by the project sponsor at NPRDC, who provided the means to experiment with simultaneously linking multiple users to the data. These experiments proved to be successful.

Each MC-EPIDS user is required to have his own CMS USERID account. Some of the operating software must reside on his CMS A-disk. The MC-EPIDS software modules that must be copied onto the user's area along with the amount of storage required for each module are as follows:

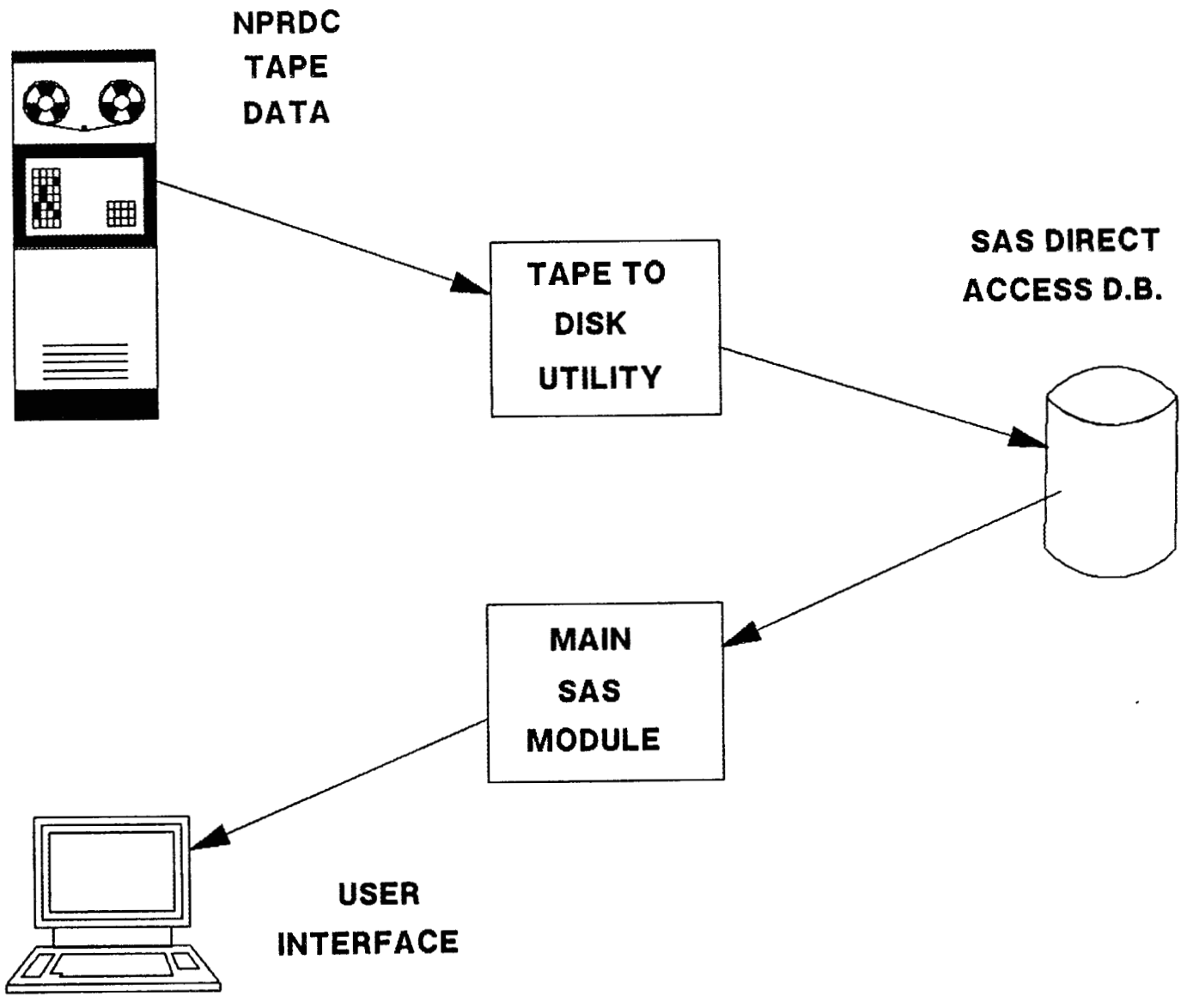


Fig. 6. MC-EPIDS system configuration.

PROFILE EXEC	1 BLK	
USER EXEC	1 BLK	
AF EXEC	3 BLK	
5PROFILE SASUSER	7 BLK	
NPRDC EXEC	1 BLK	
SAI SASEXEC	1 BLK	
5APPS NPRDC	57 BLK	
5APPS INVENT	149 BLK	
5APPS AVGYOS	151 BLK	
5APPS PCTPAY	151 BLK	
5APPS PCTYOS	152 BLK	
TOTAL =	674 BLK	= 690,000 BYTES

The remaining portions of the operating software reside with the data base on the oversized primary A-disk account. They are

FYDATA MCEPIDS.....	3325 BLK	
5APPS EPIDS	161 BLK	
NPRDCFM TXTLIB.....	13 BLK	
TOTAL =	3499 BLK	= 3.6 M-BYTES

2.3 User Interface Design

The MC-EPIDS is a menu-driven system. The data base is not maintained by MC-EPIDS users, and the users supply no data to the system. MC-EPIDS will provide the field users limited query capability. The queries must be entered via the menu format. The user interface for MC-EPIDS consists of a three-dimensional online approach: a brief tutorial, a sequence of five basic screens (the welcome screen and four menu screens), and help files. Attention was also placed on use of function keys. Finally, the output products are designed to contain all of the information needed by the MC-EPIDS users.

Using IBM PCs and PC emulators, the Marine Corps analysts access the NPRDC IBM 4341 via terminal emulation software, a 1200-Baud modem, and public-access phone lines (see also Sect. 2.5). After accessing the IBM 4341 mainframe computer, the analyst will type the command "NPRDC" and a carriage return to initiate the MC-EPIDS program.

2.3.1 Tutorial

The MC-EPIDS tutorial is a feature added by the ORNL design team and not specifically requested by the NPRDC Statement of Work. The tutorial consists of 20 screens and takes 10-20 minutes to complete, depending on

the speed with which the reader can absorb the material. A copy of the tutorial is included in Appendix C, Sect. C.1.

Although ORNL was not responsible for training or for training manuals, this tutorial serves to help train new users of MC-EPIDS. NPRDC and the Marine Corps users who previewed the tutorial thought it would be very helpful. The tutorial is a linear progression of text files, and the user can exit at any time. Thus, it can be used as a quick refresher course as well as an original training aid.

2.3.2 Screen Design

Although somewhat constrained by the SAS/AF program features, the MC-EPIDS screen has a consistent format. A chart of the progression of the MC-EPIDS program is given in Fig. 7. A complete set of menus is given in Appendix C, Sect. C.2.

When the user accesses the MC-EPIDS program, the terminal displays the "Welcome" message. The user is instructed to press function key 2 (F2) to continue. The second screen that the user views is the main menu. Options on this screen allow the user to enter the tutorial (Sect. 2.3.1), exit the program (i.e., return to the IBM 4341 system login prompt), or retrieve and display manpower statistics. The user enters the number of the desired menu choice and presses F2 to continue. If the manpower statistics option is selected, the user then views the type-of-calculation screen. The user selects one of four options by entering a single number and pressing F2.

The next screen (the selection screen) is the only one on which the user must actually enter data (rather than indicate a menu choice). Data entry fields are displayed on the screen. The tab key moves from one data field to the next. The up and down arrow keys and the Enter/Return key also move the cursor if the user does not wish to enter any data on a particular line. On the selection screen, the user also indicates a choice of a "By" variable. After making all the selections to properly subset the data, the user presses F2 to continue. The final screen is the output selection screen (see Sect. 2.3.5).

While on the selection screen, the arrow keys will move the cursor around the screen. If the user responds incorrectly to a question or wishes to change a response, the arrow keys are used to return to the field in question in order to revise the response. Once a user has indicated satisfaction with a particular menu screen and has pressed the function key to proceed, there is no mechanism for changing an entry on the previous screen. The only option is to exit and begin again. Although this procedure would be a considerable handicap in a large system with multiple menu selection screens, MC-EPIDS contains only five basic screens (only one of which requires data entry.) If the MC-EPIDS is expanded, this limitation should be readdressed.

There are 24 total horizontal lines on a screen. SAS-AF, however, requires the use of the top four lines. ORNL has tried to eliminate the

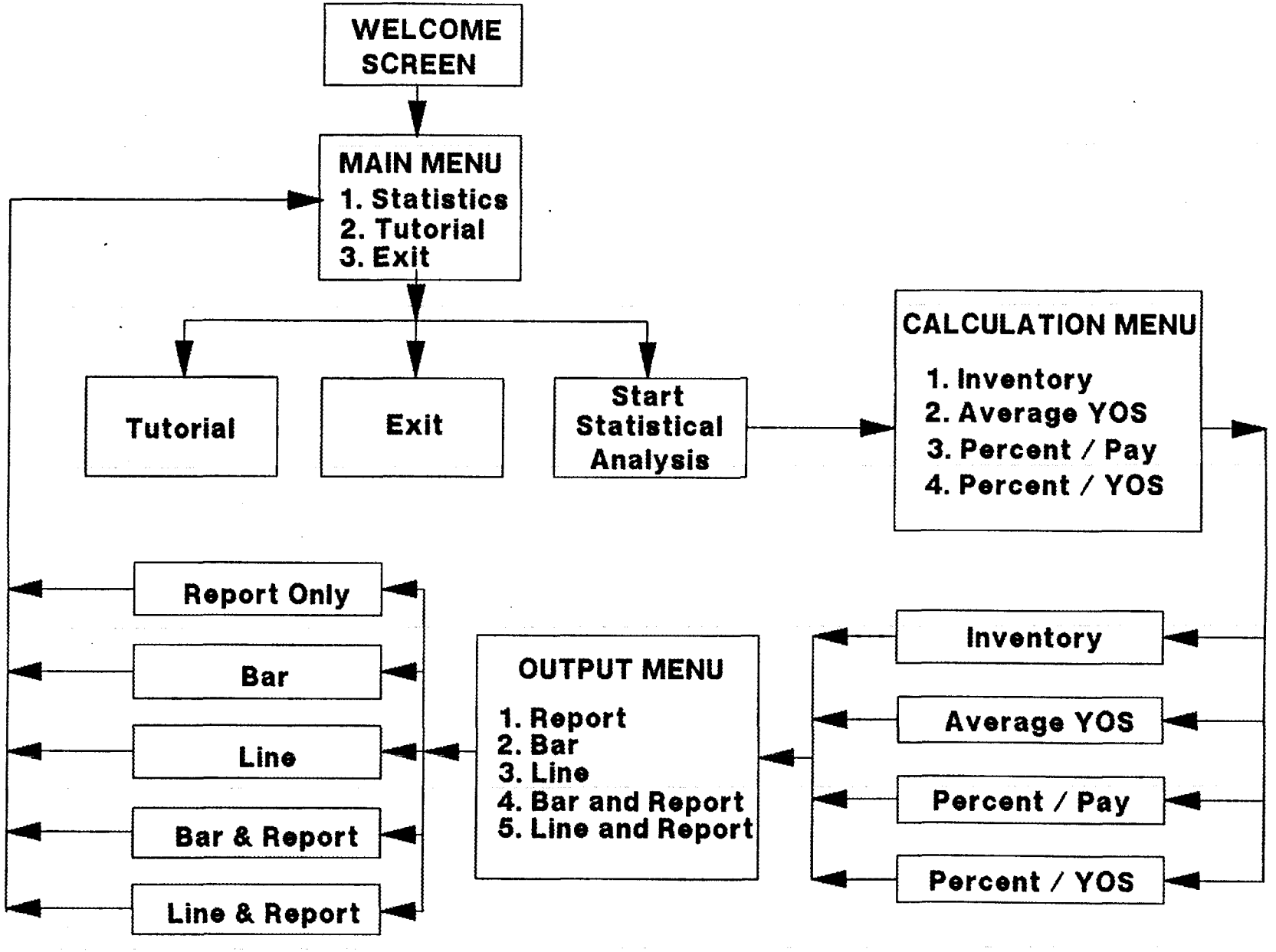


Fig. 7. MC-EPIDS program progression.

appearance on the screen of the default SAS messages, which are intended for the SAS programmers but will be meaningless, and perhaps confusing, to the actual users of MC-EPIDS. There appears to be no way, however, for deleting or masking these four lines. The MC-EPIDS user will be instructed to ignore these SAS messages. Thus, 20 lines are available for the menu option area and any necessary MC-EPIDS information.

Because most of the MC-EPIDS screens are menus, the user will usually be selecting an option by typing in a number. Therefore, the amount of text that must be entered is minimal. Designers of MC-EPIDS have considered this factor and have designed the option area to maximize the quantity of information included and yet keep the density of the screen as low as possible.

Words in the option area are usually in capitals and lower case (i.e., the initial word and all important words are capitalized). Certain important words and labels are in all capitals. No punctuation follows the options. When a menu is initially accessed, the cursor appears in the blank following the command line of the option area. The user enters the number of his/her choice and presses RETURN/ENTER. When the user accesses the selection screen, the cursor appears in the first data entry field.

The response time of the MC-EPIDS program is largely dependent on how many users are logged on to the NPRDC 4341 computer. ORNL designers felt that the users of MC-EPIDS should receive a "Working ..." message at the time when MC-EPIDS is processing data (i.e., on leaving the selection screen). Because of other problems that occurred during programming, this particular message was never incorporated. If the MC-EPIDS is expanded, this message should be added.

Error messages that convey appropriate information to the user of a computer system are an inherent part of a good user interface. The error checking of MC-EPIDS is as complete as possible. When the user indicates satisfaction with the choices of a particular screen and presses the appropriate function key to continue, every field that the user inserted data into is checked for correct format. If the entry is incorrect, the user receives a message such as "Invalid entry at ROW = 20 COLUMN = 22. Please reenter." The cursor is positioned at the row and column of the error. If the user has made more than one error, the system continues to identify the errors until the screen selections have been corrected. Appendix C.5 contains a sample screen showing an error message.

2.3.3 Help Files

The design of MC-EPIDS emphasizes conciseness. The screens aid the user in rapidly formulating a query. Initially, the users may need more information than is provided on the screen; therefore, a help file is available on request for every screen that requires user input. When the user determines a need for additional information, pressing F1 causes a new screen to appear. This screen contains information pertaining to valid

values, requirements, constraints, and miscellaneous notes. These help files will instruct the user in determining the proper response.

Copies of the help screens are given in Appendix C, Sect. C.3.

2.3.4 Function Keys

Because the user community will work with multiple types of keyboards, the user interface designer encouraged limiting the total number of function keys. The final design requires six keys. For keyboards with only four function keys, the users will need to determine the keyboard maps for F5 and F6 for their particular type of keyboard.

The following function keys are needed in MC-EPIDS.

F1 HELP
F2 CONTINUE TO NEXT MENU SCREEN
F3 ADVANCE TO NEXT PAGE (used while in "HELP" files, the TUTORIAL, or the summary of output selections)
F4 GO BACK ONE PAGE (used while in "HELP" files, the TUTORIAL, or the summary of output selections)
F5 RETURN TO MENU FROM WHICH "HELP" WAS ACCESSED (used to exit "HELP")
F6 EXIT TO OPENING MENU SCREEN

Two of these functions are used while in the main logic flow to go forward or to exit. It is not possible to "back up" in this program (i.e., to go back to a previous menu selection screen) without starting over. On any particular menu screen, the user may "back up" and revise as many entries as desired; however, after F2 ("continue to next menu screen"), has been pressed, the user is committed to the indicated set of choices. If, after continuing to the next menu screen, the user is dissatisfied with the indicated set of choices, the only option to change them is to exit to the opening menu screen (F6) and start over.

Four of the six function keys are used with the help function. To access the help function, the user presses F1; to go forward or back up while in help, the user presses either F3 or F4, respectively; and to exit help, the user presses F5. The function keys F3 and F4 are actually cursor movement keys to scroll the user throughout the "page" of help associated with a particular menu screen. The function keys F3 and F4 perform similar functions while viewing the tutorial and the output summary screens.

After receiving output to the screen, the user may print the screen by pressing the "PRINT SCREEN" key or its equivalent on his/her keyboard. The only other keyboard interface is that for saving output files. The specific keys to be used are unique to the software emulation package being used (see Chap. 2.5.3).

2.3.5 Output Formats

The user chooses one of five output options: report listing, bar chart, line graph, bar chart and report listing, or line graph and report listing. The user receives instructions on limitations to the choices. For example, the user can choose a line graph only if all fiscal years are included. This design was by direction of NPRDC, not because of a SAS constraint. If the user indicates an inappropriate choice, an error message appears. For example, if a line graph were chosen as output but all fiscal years had not been selected, the user would receive the following error message: "All fiscal years were not selected. No line graph will be produced. Press F6 to continue."

Before the report listing or graphic appears, the user views a summary of selections. The user may scroll through screens of this summary by pressing F3 and F4. When ready to see the final output screen, the user presses F6, which exits to the output screen.

The output is always to the screen, whether it be graphics or nongraphics. If the output is tabular, the user may scroll through screens by pressing F3 or F4. After viewing the output, the user may elect to obtain a hard-copy printout of the screen or to exit to the opening menu screen. If the user elects to exit, the query is deleted.

Sample output screens are provided in Appendix C (Sect. C.4).

2.4 COMMUNICATIONS OPTIONS

Currently, the Marine Corps analysts have two methods for running their computer models: via hard-wired lines and via a software communication package. First, they are using IBM PCs with an IRMA board installed. IRMA is a graphics board from DCA which enables the PC to emulate an IBM 3179 model G, IBM 3278, or IBM 3270 (all graphics terminals) and communicate with the Marine Corps computer via coaxial cable. Through this computer, they can access the computer system at Quantico. This method allows the analysts to obtain graphics output. Second, they use SmartCom II to allow the PC emulate a VT-100 or VT-52 terminal so that they can access a computer system in Washington, D.C., using a Hayes Smartmodem 1200. However, this method does not permit interactive graphics.

Users at NPRDC have terminals that are hard-wired to the NPRDC IBM 4341, or they have PCs with both terminal emulation and communication software packages (Grafpoint and Yaleterm). It is recommended that individuals at NPRDC who need access to MC-EPIDS continue using their current software and hardware arrangement.

2.4.1 Communication Lines

This project presented an interesting problem with respect to telecommunications. First, the users were located remotely from the computer on

which the SAS data set would reside. Second, the output would sometimes be in graphics mode and sometimes alphanumeric. Third, the query and response would be in interactive mode rather than batch mode.

Because of these factors, the telecommunications link is very important and a determining factor in communications/emulation software selection. Communication options considered by ORNL, in conjunction with the NPRDC, included the Marine Corps Data Network, the Defense Data Network, leased lines, and public packet-switching networks. Because of difficulties obtaining user access, neither the Marine Corps Data Network nor the Defense Data Network were considered feasible alternatives by NPRDC. The leased lines options seemed to be the alternative that presented the fewest technical problems; however, after considering the costs involved in additional hardware and leasing agreements, this option was also rejected.

Because NPRDC is a dialup node on the TYMNET public data network, use of TYMNET is preferred by NPRDC. The Marines access TYMNET through 1200-Baud asynchronous modems. Because ORNL believed that the slowness of the graphics response would significantly inhibit use of MC-EPIDS, ORNL recommended a trial period of the prototype. After the testing period, if the quality of the graphics were acceptable and the time response not acceptable, TYMNET could add an onsite packet assembler-disassembler (PAD) to which the users' PCs would be hardwired. A leased line would transmit data to the nearest node (Fairfax, Virginia) and then the transmittal would go via normal TYMNET lines. Initially, however, the procedure would be for the Marine Corps users to dial into the NPRDC computer via 1200-Baud modems over standard asynchronous TYMNET access.

2.4.2 Hardware Requirements

Communicating with the system at NPRDC involves going through an IBM 7171 protocol converter. The IBM 7171 provides the necessary translation between remote, asynchronous ASCII terminal devices and the standard IBM 327X type serial terminals using EBCDIC code. This converter turned out to be the source of a major problem in finding a workable communication link with the SAS/GRAPH software installed at NPRDC. This problem is discussed further in Sect. 3.

2.4.3 Software for Terminal Emulation and Communication

Project team members at ORNL examined a number of different terminal emulation software packages as possible candidates to be used by Marine Corps analysts. The criteria under which ORNL evaluated these packages were the following: the chosen product had to be able to display both alphanumeric and graphical data; it should be able to switch automatically between graphics and alphanumeric modes; it must be able to communicate reliably over 1200-baud phone lines with the NPRDC system; and it should be reasonably inexpensive. The cost was not the primary concern inasmuch as no more than ten copies of the product needed to be procured. The costs for the software packages examined by ORNL ranged from about $80 to $700.

One of the greatest problems encountered was to find a software package that could produce interactive graphics on the screen. Many popular software packages transmit data values, usually emulating a VT-100 or a terminal with similar attributes, but cannot transmit and produce graphics.

The software communication packages IRMA COM, PC-PLOT, SMARTERM 240, and TGRAF-07 were examined by ORNL.

The Marine Corps users are currently using DCA's IRMA software to communicate and produce graphics. IRMA and the IRMA family are products that are PC-based. The graphics packages in the IRMA family are IRMAX APA and IRMACOM APA. Users with IRMAX APA are locally attached to a cluster controller. For example an IBM 3274 or 3276 cluster controller allows up to 32 terminals to be hooked together via coaxial cable. After accessing the cluster controller, one leased line would be used to access the NPRDC computer in California. Users with IRMACOM APA attach to a synchronous modem which uses phone lines to attach to a synchronous modem at the other end. IRMACOM APA emulates exactly an IBM 3179 model G terminal. Although the Marine Corps currently own some IRMA software, additional hardware and software is needed for the IRMA line to function properly. Thus, if communication is over the public packet-switching networks, it is anticipated that the Marine Corps users will need to switch from the IRMA software to another software package.

PCPLOT is manufactured by Microplot Systems Company for use with IBM compatibles. Costing less than $100, it represents the least expensive of the graphics options tested. The product will emulate a VT100 terminal when it is in alphanumeric mode. In graphics mode it emulates fully a Tektronics 4010 terminal. Switching between the two modes is automatic.

PC-PLOT has a number of attractive features. Even though it is inexpensive, it was found to be the easiest to set up and use and provided a connection with Tymnet that was at least as reliable, if not better than, the links provided by the other products tested. PC-PLOT featured a simple scheme for automatic phone dialing and for entering local user commands that was easy to learn. Also furnished with the software was a set of drivers for capturing graphics data to disk and outputting to several different types of printers and plotters.

The main drawback to PC-PLOT is that it only emulates the Tektronix 4010, which is a monochrome device. Even though the software could partially emulate a Tektronix 4027 terminal, which is a color display device, it does not emulate well enough to interface with SAS/GRAPH. PC-PLOT simply does not recognize all the Tektronix 4027 cursor commands used by SAS/GRAPH. Nevertheless, for monochrome graphics, PC-PLOT appeared to be the best choice.

SmarTerm 240 provides VT100 and VT125 emulation in alphanumeric mode, and it provides either Tektronix 4010 and VT240 graphics terminal emulation in graphics mode. It is priced between $200 and $300. Although SmarTerm 240 furnishes a colorful, menu-driven user access and set up mode, it does not

support color displays in graphics emulation mode. In addition, in graphics mode only a partial screen display is presented. The user must scroll the graphics window in order to see the entire graphics image.

TGRAF-07 was the most promising software package examined for this application. Priced between $600 and $700, it was also the most expensive package seriously considered. TGRAF-07 emulates a VT100 in alphanumeric mode. In graphics mode it emulates a Tektronix 4107 graphics terminal, which is a more recent color graphics display device than the Tektronix 4010. It features a menu-driven set up and user access mode, and appears to be fairly easy to learn to use.

Use of TGRAF-07 on site at ORNL produced high-quality, full-screen color displays. The project sponsor reported the same results when used on site at the NPRDC. However, a major problem with TGRAF-07, and one that prevented its being highly recommended for project use, is that it did not communicate well with TYMNET. There seemed to be some inconsistency in the way TYMNET and TGRAF-07 set duplex modes with the result that double characters were generated on the screen. Making obvious changes to the set up parameters in TGRAF-07 did not correct this problem, and no other alternative solutions were presented in the software documentation.

3. PROBLEMS AND CONSTRAINTS

Although the MC-EPIDS was successfully demonstrated, it was developed under certain constraints that prevented a complete analysis and limited the final design. Problems that occurred during the system development precipitated conflicts that were not considered in advance. These problems and constraints have been alluded to within this report and are summarized below.

Because NPRDC had worked with the Marine Corps for a lengthy period of time, the NPRDC already had an extensive bank of knowledge about user requirements. Therefore, to prevent duplication of effort, ORNL information needs were supplied to the NPRDC, and NPRDC provided the responses. Both questions and responses were sometimes misinterpreted and/or less than fully answered because of the difficulty of working in a relay mode.

ORNL made some suggestions that would have made the MC-EPIDS a more generic system and, therfore, easier to modify. NPRDC, however, was constrained by both the amount of money that could be spent on this system and by the defined need to limit the Marine Corps to a certain set of queries.

Although NPRDC desired a 15-second time response for MC-EPIDS, the system was constrained to reside on the NPRDC IBM 4341, a notoriously slow computer, and be accessed via a 1200-Baud modem using public-access, packet-switching network. The 15-second response time could have possibly been met if a different computer were used or if a different method of telecommunication were used. These options were not available, however.

Not only were the users constrained to using the IBM 4341 and 1200-Baud lines, so were the developers. Development time also was constrained by the hours the IBM 4341 (in California) was available, which were sometimes shortened by computer downtime failures. The slow system response and turnaround times hindered programming and plagued the data base creation efforts. The IBM 4341 central processing unit at NPRDC is overloaded most of the day, which made testing and debugging difficult during normal working hours. Tape jobs that would take 10 to 15 min to run during off-hours sometimes took as much as one and a half to two hours to run during times of peak CPU usage. The constraint of having to run tape jobs interactively rather than in batch mode placed a burden on programmers by requiring them to maintain an open communication channel with the mainframe system for the duration of the tape jobs. Sometimes the loss of carrier signals over TYMNET required that the currently running tape job be halted and rerun. Such delays were common.

While the data base was being created and during the period when it was being tested, two problems developed with the NPRDC tape data itself. The first problem appeared as the tapes were being read the first time. The data base creation module stopped reading the FY 1985 tape when it found

some out-of-range data values in the year of service fields of certain Marine records. The year of service item in some records was set to a zero value, "00", and for some other records was set to value "99". MC-EPIDS requires that year of service be in the range of 1 to 31 years, "01" to "31". Since very few Marine records were involved, the project sponsor decided that this problem should be resolved by trapping those records with out of range values and setting all of them to year of service 1, "01". In this case the data were easy to repair.

The second problem appeared after the first prototype data base was set up and testing of statistical output begun. Some of the MC-EPIDS functions, particularly inventory values by paygrade for FY85, gave obviously incorrect numerical results. The project team tracked down the source of this problem and found it to be caused by missing occupational field data values in the NPRDC tape records of certain Marines. Normally the occupational field should contain a number in the range of 0 to 99. In these instances the occupational field contained blanks, and as a result the entire Marine record was dropped by the SAS data base creation module. Again, the fiscal 1985 tape was particularly badly affected, with 211 records involved. Also, tapes for some of the other years were found to have a few missing occupational field items. After consulting with the Marine Corps end users, the project sponsor at NPRDC decided that this problem should be resolved by trapping those records with missing occupational field data items and resetting those items to the value "00", which is the "unknown" category for OCC field. The data base creation module was modified to reflect this decision, and the MC-EPIDS data base was rebuilt. With these extra Marine personnel records now included in the frequency counts, the numerical output apparently is correct and corresponds exactly to the output expected by NPRDC.

As the MC-EPIDS became more clearly defined, the limitations of SAS as a programming tool became more obvious.

Although SAS/AF is a convenient mechanism for designing menus, the menus are limited in design capabilities. These limitations inhibit incorporation of a good user interface. For example, certain SAS character strings and automatic SAS messages appear inevitably on the screen. These messages and prompts, which are primarily needed by the developer, can be confusing to the user. The fact that they take up lines of otherwise usable space on the screen is also a handicap. Another problem with SAS/AF is that the cursor cannot be forced to return to a particular position on the screen. Working around these constraints was a challenge. The user interface for MC-EPIDS incorporates some of the SAS/AF automatic prompts. In addition, to prevent excessive screen density, the SAS help function was extensively used to provide background information.

SAS-GRAPH limits the output because the graphics must be designed to fit on the terminal screen. Because some analysts had only black and white terminals, no effort was made to code the output products in color. The default output is to a monochrome screen.

Communicating with the system at NPRDC involved going through TYMNET using terminal emulation software. These packages mostly follow an asynchronous-type of ASCII communicaton protocol. Interfacing with the IBM 4341 at NPRDC required that this asynchronous protocol be converted to the standard IBM 327X serial type of protocol using EBCDIC character code. The system at NPRDC made use of an IBM 7171 protocol converter to accomplish this. It turned out that this particular kind of protocol converter became the source of a major problem in finding a workable communication link with the SAS/GRAPH software installed at NPRDC.

It was found that SAS/GRAPH, used in combination with an IBM 7171 protocol converter, does not interface well with many of the common, readily available graphics terminal emulation software packages. The particular problem is that graphics displays will frequently not remain on the terminal screen after SAS/GRAPH has finished the drawing. Normally, the display should not be cleared until the user presses RETURN. Experimentation showed that the same problem occurred while working with SAS/GRAPH through an IBM 7171 protocol converter installed on sites at both NPRDC and ORNL.

Technical support by each of the software manufacturers failed to solve the screen clearing problem. The investigation served to apprise the manufacturers of the particular difficulties encountered by ORNL, but generated little attention among the vendors contacted. It should be noted that all the graphics terminal emulation software examined showed nearly identical characteristics. It was therefore speculated that the problem was in the communication link between the SAS/GRAPH device driver and the protocol converter rather than in the software packages. It was thought that, possibly, the protocol converter itself prematurely sends a carriage return to the interactive SAS/GRAPH software which in turns responds with a clear screen command sent to the emulation software at the user's end. During the investigation, other researchers at ORNL, who have used SAS/GRAPH and have encountered the same difficulties, expressed their interest in and hope that this problem will be successfully resolved at some future date inasmuch as SAS/GRAPH is generally well received. Its value as a research tool would be greatly enhanced.

4. SUMMARY AND CONCLUSIONS

ORNL provided NPRDC with a prototype of the MC-EPIDS. This system was successfully demonstrated to the Marine Corps users by NPRDC.

It is assumed that NPRDC will expand the MC-EPIDS to include the capability to analyze gain and loss data. It would have been preferable to add this capability in the original design; however, some thought has been given by the data base designer to the addition and it is possible.

The menu screens are easy to use, requiring very little input from the user. Although the category-selection screen is very dense, dividing the screen into two screens is not a logical solution.

Maintenance issues and costs for expansion of menus and other screens were not considered in this document. These issues should be addressed.

If NPRDC decides to expand this system to include additional options, most screens will need to be revised and the help files expanded. These changes, however, would probably present fewer problems than would be brought about by the changes in the data base design.

Although the tutorial will assist the users in learning to use MC-EPIDS, a set of training guides would be very helpful. It is anticipated that the greatest user interface problem will be in the area of data communications. The user must be familiar with the communications software package and its performance on the particular hardware being employed for the graphics output to be acceptable. Another major user interface problem will be in the area of response time. It is recommended that the public data network option be replaced with a different communications medium.

APPENDIX A

PROCESS SPECIFICATIONS AND DATA DICTIONARY FOR THE MC-EPIDS STRUCTURED ANALYSIS

PROCESS SPECIFICATION

Process Specification Number: 1.1

Process Specification Name: Determine Marines on Desertion, Recruitment Aide Drops and on Separation from the Service

Look at each record on the NPRDC-Data-Base for a given Fiscal-Year.

IF the record has a value of "D" OR "L" in Strength-Indicator

THEN, exclude this record in the created MC-EPIDS-Data-Base

ELSE, include this record in the created MC-EPIDS-Data-Base.

(*** Note - this removes all Marines from the MEDB data store that are:

- o on desertion - "D"
- o on separation from the Service - "L" ***)

PROCESS SPECIFICATION

Process Specification Number: 1.2

Process Specification Name: Determine Marines for Inventory

Look at each record on the NDB for a given Fiscal-Year.

If the record has a Strength-Indicator is Blank AND

The Component-Code	=	"11"	OR
		"12"	OR
		"13"	OR
		"AA"	OR
		"A1"	OR
		"A2"	OR
		"A3"	OR
		"A4"	OR
		"A5"	OR
		"A6"	OR
		"A7"	OR
		"A8"	OR
		"A9"	OR
		"CA"	OR
		"CB"	OR
		"CD"	OR
		"CH"	

THEN, include this record on the MC-EPIDS-Data-Base

ELSE, exclude this record on the MEDB.

(*** Note - this adds all Marines to the MEDB data store that are:

required, according to Figure 4, page 13, of the document entitled "An Assessment of Marine Corps Enlisted Personnel Data," NPRDC TR-87, dated March, 1987 ***)

PROCESS SPECIFICATION

Process Specification Number: 1.3

Process Specification Name: Eliminate Appellate Cases

Look at each record on the NDB for a given Fiscal-Year.

IF the record has a value of "A" in the Strength-Indicator

THEN, exclude this record from the MC-EPIDS-Data-Base

ELSE, include this record on the MEDB.

(*** Note - this removes all Marines from the MEDB data store that are:
- Confined Awaiting Trial
- IHCA or IHFA for more than 30 days
- Appellate Leave ***)

PROCESS SPECIFICATION

Process Specification Number: 2.1

Process Specification Name: Determine Population

Look at each Query received for a given Fiscal-Year.

IF the Query Population = Occupation-Field OR
Paygrade OR
Years-of-Service OR
Sex-Code OR
Population-Group OR
Fiscal-Year

THEN include each record on the Selected-Records that meets the Population

ELSE exclude the records on the Selected-Records.

(*** Note - The Query may ask with "Female Marines". The Population will view the MEDB data to segregate those marines that have a sex-code of female. Or the Query may address Marines with a Pay Grade of 09. All Pay Grade 09 Marines would then be segregated for the Population. ****).

PROCESS SPECIFICATION

Process Specification Number: 2.2

Process Specification Name: Determine "BY" Variable

Look at each Query received for a given Fiscal-Year.

IF the By-Variable of the Query is equal to any combination of

Occupation-Field	AND/OR
Paygrade	AND/OR
Years-of-Service	AND/OR
Sex-Code	AND/OR
Population-Group	AND/OR
Fiscal	

THEN prepare the By-Variable combinations

ELSE prepare a blank By-Variable

(*** Note - The "BY" Variable is a further, more detailed, segregation of the MEDB data to respond to a Query. The query requesting female data may be further segregated by a request of female data by paygrade and years of service. Or, the query may have asked for all Marines with a paygrade of 09, the "BY" Variable may further define this request by years of service and assignment location. The table shown below represents the possible combinations of the "BY" Variable. One or more of the "BY" Variables may be present in any Marine Corps Personnel Planner query ***)

Calculation Requested Analysis	OCCUPAT. FIELD	PAYGRADE	YEAR OF SERVICE	SEX	POPULAT GROUP
Inventory Values	X	X	X	X	X
Average Yrs of Service	X	X		X	X
Percent by Grade	X		X	X	X
% Years of Service	X	X		X	X

"BY" VARIABLES

PROCESS SPECIFICATION

Process Specification Number: 2.3

Process Specification Name: Determine Calculation

Look at the Query received.

IF the Query contains a Calculation containing one or more of:

Inventory-Values for:	Occupational-Field or Paygrade or Years-of-Service or Sex-Code or Population-Group or Fiscal-Year	OR
Average-Years-of-Service for:	Occupational-Field or Paygrade or Sex-Code or Population-Group or Fiscal-Year	OR
Percent-by-Paygrade for:	Occupational-Field or Years-of-Service or Sex-Code or Population-Group or Fiscal-Year	OR
Percent-Years-of-Service for:	Occupational-Field or Paygrade or Sex-Code or Population-Group or Fiscal-Year	

THEN prepare a Required-Calculation containing these

ELSE prepare a Required-Calculation-Null.

PROCESS SPECIFICATION

Process Specification Number: 2.4

Process Specification Name: Perform Calculation

Look at Selected-Records + the By-Variable + the Required-Calculation.

IF the Required-Calculation is a Required-Calculation-Null

THEN use only the Selected-Records + the By-Variable.

IF the By-Variable is Null + there is a Required-Calculation-Null

THEN prepare Calculation using all Selected-Records.

IF the By-Variable is NOT Null and the Required-Calculation is NOT Null

THEN prepare Calculation using:

Selected-Records +
By-Variable +
Calculation-Required.

PROCESS SPECIFICATION

Process Specification Number: 2.5

Process Specification Name: Prepare Response

Look at Calculation received.

THEN format the Calculation for response to Marine Corps Personnel Planners according to Screen-Layouts.

EPIDS DATA DICTIONARY

Act-In-Out-Yr-Exten-Code	=	* Values: 1 = In 2 = 2 In 3 = Out 4 = 1 In and 1 Out 5 = 2 Out *
Admin-Reduction-Code	=	* A one position code to indicate that an admin. reduction has occurred *
Armed-Forces-Active-Duty-Date	=	Year + Month + Day
By-Variable	=	* The portion of each query where the selected records for review are further defined for output *
Calculation	=	* The specified part of a query that is used to specify the calculations to be performed for the output requested *
Calendar-Year-Month	=	* A two position numeric code. Value: 01 through 12 *
Component-Code	=	* A two position Code. Valid codes are found in USMC Codes Manual *
Continuity-Of-Service	=	* One Numeric Code. Values: 1 or 2 or 3 or 0 (Zero) *
Curr-On-Extension-Code	=	* One position numeric code. Values: 1 or 0 (Zero) *
Date-Current-Tour-Began	=	Year + Month + Day
Date-Detached-Last-Comm	=	Year + Month + Day
Date-Joined-Present-RU	=	Year + Month + Day
Date-Of-Birth	=	Year + Month + Day
Date-Of-Enlistment	=	Year + Month + Day

Day	=	* Two position numeric code. Valid values 01 through 31 *
Delayed-Entry-Program-Cd	=	* One position numeric code. Values: 1 or 0 (zero) *
Distribution	=	
Early-Reenlistment-Code	=	* One position numeric code. Values: 1 or 0 (Zero) *
EAS	=	Year + Month + Day
ECC	=	Year + Month + Day
Education-Code	=	* A one position numeric code. Values: 1 = Non-High School Graduate 2 = High School Graduate or other credential 3 = High School Diploma Grad 4 = Unknown *
First-Additional-MOS	=	* Four position identifier with values found in the MOS Table *
Fiscal-Year	=	* A two numeric position year code. Values: 78 through 87 *
Gain-Indicator	=	* One position numeric code. Values: 1 or 0 (Zero) *
Identification-Number	=	* The Military Id Number. This is the Social-Security Number of the Marine *
Inter-OCC-Field-Move-Cd	=	* One position numeric code. Values: 1 or 0 (Zer0) *
Inventory-Counts	=	
Length-Of-Enlistment	=	* One position code. Values: 0 (Zero) or 1 or 2 or 3 or 4 or 5 or 6 or o (Alpha) *
Marital-Status	=	* Values: M = Married S = Single D = Divorced O = Ohter

U = Unknown *

MC-EPIDS-Data-Base = * The data base created from the NPRDC-Data-Base. The MEDB is used to respond to all Querries. *

MEDB = MC-EPIDS-Data-Base

Mental-Category-Code = * A one position numeric code. Values:
1 = Category I
2 = Category II
3 = Category IIIA
4 = Category OOOB
5 = Category IVA
6 = Category IVB
7 = Category V
8 = Unknown *

Meritorious-Promotion-Cd = * One position numeric code. Values: 0 (zero) or 1 or 2 or 3 *

MID = Non-Marine-Code + Identification-Number

Month = * Two position numeric code. Values 01 through 12 *

NDB = NPRDC-Data-Base

No-Extension-Cur-Qtr = * One position numeric code. Values: 0 (Zero) or 1 or 2 or 3 *

Non-Marine-Code = * One position Code. Values:
0 (Zero) = Count for Strength
D = Desertion end of quarter
L = Strenght loss in quarter*

NPRDC-Data-Base =

Act-In-Out-Yr-Exten-Code +
Admin-Reduction-Code +
Armed-Forces-Active-Duty-Date
Calendar-Year-Month +
Component-Code +
Continuity-Of-Service +
Curr-On-Extension-Code +
Date-Current-Tour-Began +
Date-Detached-Last-Comm +
Date-Joined-Present-RU +

```
                                  Date-Of-Birth              +
                                  Date-Of-Enlistment         +
                                  Delayed-Entry-Program-Cd   +
                                  Early-Reenlistment-Code    +
                                  EAS                        +
                                  ECC                        +
                                  Education-Code             +
                                  First-Additional-MOS       +
                                  Fiscal- Year               +
                                  Gain-Indicator             +
                                  Inter-OCC-Field-Move-Cd    +
                                  Length-Of-Enlistment       +
                                  Marital-Status             +
                                  Mental-Category-Code       +
                                  Meritorious-Promotion-Cd   +
                                  MID                        +
                                  No-Extension-Cur-Qtr       +
                                  Number-Of-Dependents       +
                                  Occupation-Code-Indicat    +
                                  Occupation-Field           +
                                  Paygrade                   +
                                  Population-Group           +
                                  Pot-In-Out-Yr-Exten-Code   +
                                  Present-Date-Of-Rank       +
                                  Present-MCC                +
                                  Present-RUC                +
                                  Previous-Primary-MOS       +
                                  Primary-MOS                +
                                  Punitive-Reduction-Code    +
                                  Reclassification-Indicat   +
                                  Recruit-MOS-To-Basic-MOS   +
                                  Regular-Promotion-Code     +
                                  Second-Additional-MOS      +
                                  Sex-Code                   +
                                  Strength-Indicator         +
                                  Years-Of-Service           +

Number-Of-Dependents          =   * Values:
                                    1 = One Dependent
                                    2 = Two Dependents
                                    .
                                    .
                                   10 = Ten or More Dependents *

Occupation-Code-Indicator     =   * One position numeric code.
                                    Values:  0 (Zero) or 1.

Occupation-Field              =   * Two position numeric field.
                                    Valid values found on Encl.
                                    #2, Mar. 16, 1987 EPIDS
                                    Statement of Work *
```

Output-Format	=	* A part of the query whereby the requestor specifies the output format (graphic, tables, report, etc.) *
Paygrade	=	* A one position numeric code. Values: 1 = E1 Pvt 2 = E2 PFC 3 = E3 LCPL 4 = E4 CPL 5 = E5 SGT 6 = E6 SSGT 7 = E7 GNYSGT 8 = MSGT 9 = MGYSGT
Population	=	* The primary part of the query whereby the requestor identifies the records to be reviewed in the NDB for the output response *
Population-Group	=	* Is the Race Code. Values: 1 = White 2 = Black 3 = Hispanic 4 = Other 5 = Unknown *
Pot-In-Out-Yr-Exten-Code	=	* One position numeric code. Values: 0 = No potential extension of service indicated 1 = In first out-year 2 = In second out-year 3 = In third out-year *
Present-Date-Of-Rank	=	Year + Month + Day
Present-MCC	=	* Three position field. Values found in USMC Codes Manual *
Present-RUC	=	* Five position field. Values found in USMC Codes Manual *
Previous-Primary-MOS	=	* Four position field - Shows previous primary occupation speciality. Values found in USMC Codes Manual *

Primary-MOS	=	* Four position field - Shows primary military occupation speciality. Values found in USMC Codes Manual *
Punitive-Reduction-Code	=	* One position numeric code. Values: 0 (Zero) or 1 or 2 or 3 or 4 *
Query	=	Population + By-Variable + Calculation + Output-Format
Reclassification-Indicat	=	* One position numeric code. Values: 0 (Zero) or 1 or 2 or 3 *
Recruit-MOS-To-Basic-MOS	=	* One position numeric code. Values: 0 (Zero) or 1 *
Regular-Promotion-Code	=	* One position numeric code. Values: 0 (Zero) or 1 or 2*
Required-Calculation	=	* The calculation required by the requestor on the Query. *
Required-Calculation-Null	=	* A requestor Query that does NOT contain any Required-Calculation. *
Response	=	* Output to requestor that contains the Calculation of the Selected-Records + the By-Variable + Required-Calculation. Response is in accordance with Screen-Layouts. *
Second-Additional-MOS	=	* Four position second additional military occupational speciality. Values found in USMC Codes Manual *
Selected-Records	=	* Those records that are selected for a response to a query *

Sex-Code	=	* Values: 1 = Male 2 = Female 3 = Unknown *
Strength-Indicator	=	* Values: ' ' Count for strength at end of quarter 'D' On Desertion at end of quarter 'L' Strength loss this quarter 'B' Black hole this quarter HMF1 w/no HMF2
Year	=	* Two position numeric. Values: 01 through 87 *
Years-Of-Service	=	* Two position numeric field. Values: 01 to 31 *

APPENDIX B

ALTERNATIVE MC-EPIDS DATA BASE DESIGNS

APPENDIX B

ALTERNATIVE MC-EPIDS DATA BASE DESIGNS

The project team discussed three alternative approaches to the problem of aggregating the data in the MC-EPIDS data base. These are presented below for completeness.

SAS Data Set Approach

Aggregating the MC-EPIDS data in the form of a five-dimensional frequency distribution of record counts requires a maximum of 143,964 data elements per FY. SAS has the useful facility of being able to pack data automatically in its data set storage. In the SAS data set approach, individual records are stored in the SAS analogy to the frequency distribution. The data set would be a matrix of the following form:

```
        OCC   PAY  YOS  SEX  POP  FY81  ..... FY86
OBS1
OBS2
 .
 .
 .
```

Each variable of the type FY81, etc., would hold the frequency count of that particular occurrence of the classification variables OCC, PAY, YOS, SEX, and POP for the corresponding FY. An upper bound for storage can be established by assuming a requirement of 60 bytes per observation. This will translate into a requirement of about 8.6 megabytes. This is likely to be overestimated at least 25%. The final data base design selected for this project is basically a modification of this simple frequency distribution.

Augmented PDF Approach:

In the interest of increasing efficiency and performance, a couple of relational-type data base designs specialized for this particular application were considered. The most promising of these is a combinatorial approach, which is a generalization of the simple frequency distribution of the previous design. In this approach not only are frequency counts of records stored, but also frequency counts for all possible combinations of classification variables. The data can be stored as an array of counts; e.g.

A (OCC, PAY, YOS, SEX, POP)

The classification variables become indices into the array. A special index value, an "ALL" value, is assigned to each classification variable in order to allow for counts of combinations. For example, the value stored in location

A (10, 3, *, 2, 1)

will be the count of all white, female personnel in occupation 10 with a paygrade of E3. The special index value * in the YOS index indicates that the count is taken over all years of service.

In array position

A (*, 3, *, *, *)

will be the count of all personnel with paygrade E3.

Many of the most frequently occurring sums, as these examples demonstrate, will already be stored in such a data base, thus minimizing the amount of processing required for each query. However, an array of this type is not, strictly speaking, a SAS data set. Such a data set would have to be interfaced with SAS/AF and BASE SAS by writing a customized SAS procedure in PLI. A design of this type would be extremely tedious and time consuming to program.

Memory requirements for this type of data base can be computed in a straightforward manner. It is proposed that the data be stored as an external, partitioned data set in blocks of one fiscal year each. Each fiscal year would require a total of 281,600 data elements. Using a four-byte numeric field for the data elements gives a total storage requirement of 1,126,400 bytes per year. By comparison, storage for one fiscal year's worth of individual records requires approximately 1,200,000 bytes, assuming 200,000 individual records at six bytes per record. It is also likely that the array data can be packed to save space.

For the augmented PDF implementation, response time will be chiefly constrained by the time for disk I/O operations. The number of required I/O operations will be determined by the amount of paging that takes place within the system. Again, the safest time estimates can probably be derived through benchmark testing, but assuming the number of I/O operations to be around 20 and a bytes per second disk data rate of 10 , then approximately 20 seconds of I/O time would be required to access data over all seven fiscal years. Much faster response times would occur for queries over just one or two fiscal years. CPU time would be negligible compared to these figures.

The CDF Data Base Design:

The CDF design is another variation of the basic frequency distribution. It offers the advantage of a data base of reduced size at the cost of decreased execution speed. However, the total number of I/O operations involved in execution should be either the same or slightly less than that of the previous augmented PDF method. The total response time for the CDF method should not differ significantly from the previous design.

As in the augmented PDF design, this approach would store the data in an array. However, instead of an array of simple frequencies, the array would consist of a five-dimensional cumulative frequency distribution. Frequency counts of data in this array may be computed by means of a 32-term difference equation in which the terms themselves are found by simple accesses of data elements in the CDF.

Overall, storage requirements for the CDF method are about half that of the augmented PDF method. It would be 143,964 data elements or 575,856 bytes per fiscal year. Implementation of the CDF approach would require development of custom written SAS procedures using PLI.

APPENDIX C

SCREENS VIEWED BY THE END USER

C.1 TUTORIAL
C.2 MENU SCREENS
C.3 HELP SCREENS
C.4 OUTPUT FORMATS
C.5 ERROR MESSAGES

Appendix C.1

TUTORIAL

TUTORIAL

Command ===>

MARINE CORPS

ENLISTED PERSONNEL INFORMATION DELIVERY SYSTEM
(MC-EPIDS)

This tutorial is a brief introduction to MC-EPIDS, a menu-driven, SAS-based information delivery system. In this tutorial, the menu screens will be explained. You will be given the opportunity to view some typical queries, and you will see tabular output.

At any time during the tutorial, if you wish to exit, simply press F6, and you will be returned to the MC-EPIDS opening menu screen. To continue, press F3. To see previous screen(s), press F4.

To continue, press F3. To exit, press F6.

TUTORIAL

Command ===>

In MC-EPIDS, you will view the following menu screens:

* Welcome Screen
* Opening Menu Screen
* Type-of-Calculation Screen
* Selection Screen (Population of Interest and "By Variable")
* Output Screen

To continue, press F3. To exit, press F6.

TUTORIAL

Command ===>

In MC-EPIDS, you will use the following function keys.

F1 HELP
F2 CONTINUE TO NEXT MENU SCREEN
F3 ADVANCE TO NEXT PAGE (used while in "HELP" files or the TUTORIAL)
F4 GO BACK ONE SCREEN (used while in "HELP" files or the TUTORIAL)
F5 RETURN TO MENU FROM WHICH "HELP" WAS ACCESSED (used to exit "HELP")
F6 EXIT TO OPENING MENU SCREEN

To continue, press F3. To exit, press F6.

TUTORIAL

Command ===>

NOTE: Two unsupported messages appear in the upper right-hand corner of some MC-EPIDS screens. These messages are "Press END to return" and "Press FORWARD for more." Because these keys are not supported in MC-EPIDS, please ignore them.

To continue, press F3. To exit, press F6.

TUTORIAL

Command ===>

After you leave the MC-EPIDS Welcome Screen, you will view the Opening Menu Screen. On this screen, you must decide which of the following options you wish to pursue:

1. Retrieve and display manpower statistics
2. View the MC-EPIDS tutorial
3. Exit MC-EPIDS

Option 2 allows you to view this tutorial. Option 3 returns you to your login area on the NPRDC computer.

If you choose option 1, you will proceed through the regular MC-EPIDS program. We shall assume that you have selected option 1.

To continue, press F3. To exit, press F6.

TUTORIAL

Command ===>

If you select option 1 (Retrieve and display manpower statistics), you will then see the Type-of-Calculation Screen. On this screen, you may choose one of the following calculations.

1. Inventory values
2. Average year of service
3. Percent by paygrade
4. Percent by year of service

The most general type of calculation is choice 1 (Inventory values), which yields the total number of Marines on active duty at the end of a selected fiscal year. We shall assume that you have selected choice 1.

To continue, press F3. To exit, press F6.

TUTORIAL

Command ===>

On the Selection Screen, you will choose the population of interest and "by variable." Options are Fiscal Year, Occupational Field, Paygrade, Years of Service, Population Group, and Sex. You can use the tab key or the arrow keys to move the cursor within this screen.

A list of valid values is available by pressing F1 (HELP) after first moving the cursor to the field in question.

The default value for Fiscal Year is the most recent year of complete data. The default value for each of the other categories is all values. For example, if you choose no specific paygrade, output values will be summed across all paygrades.

To continue, press F3. To exit, press F6.

TUTORIAL

Command ===>

The following example illustrates how to mark selections for a sample query. This query does not contain a "by variable." (The screens you will see below are abbreviated versions of the actual screens in MC-EPIDS.)

"At the end of 1986, how many female Hispanic Marines with six years of service were in occupational fields 61, 63, 64, 65?

```
                                                        "BY"
          POPULATION OF INTEREST                      VARIABLE
FISCAL YEAR: 86_ ___ ___ ___ ___ ___ ___
 OCC. FIELD: 61_ 63_ 64_ 65_ ___ ___ ___ ___ ___ ___      _
   PAYGRADE: ___ ___ ___ ___ ___ ___ ___ ___ ___          _
        YOS: 6__ ___ ___ ___ ___ ___ ___ ___ ___          _
 POP. GROUP: 3__ ___ ___                                  _
        SEX: 2__ ___ ___                                  _
```

To continue, press F3. To exit, press F6.

TUTORIAL

Command ===>

To complete the selection screen, you indicate your desired "by variable(s)". In the previous example, if you mark choices of occupational fields 61, 63, 64, and 65 and wish to see values for each field separately, place an "X" in the "by variable" column to the right of OCC. FIELD. (If you want to see the values for the four occupational fields summed together, leave the "by variable" column blank.)

If you choose more than one fiscal year, fiscal year will automatically become your first "by variable". You do not need to enter an "X" in the "by variable" column.

To continue, press F3. To exit, press F6.

TUTORIAL

Command ===>

The following example illustrates how to mark selections for a sample query with a single "by variable".

"For each fiscal year, how many black females are in occupational field 72?"

```
                                                        "BY"
          POPULATION OF INTEREST                      VARIABLE
FISCAL YEAR: 80_ 81_ 82_ 83_ 84_ 85_ 86_
 OCC. FIELD: 72_ ___ ___ ___ ___ ___ ___ ___ ___ ___      _
   PAYGRADE: ___ ___ ___ ___ ___ ___ ___ ___ ___          _
        YOS: ___ ___ ___ ___ ___ ___ ___ ___ ___ ___      _
 POP. GROUP: 2__ ___ ___ ___                              _
        SEX: 2__ ___ ___                                  _
```

Because paygrade and year of service (YOS) were not mentioned, assume that the population covers all paygrades and YOS. "All" is the default for these categories.

To continue, press F3. To exit, press F6.

TUTORIAL

Command ===>

The following example illustrates how to mark selections for a sample query with two "by variables".

"What is the distribution of the enlisted force by population group by sex?"

	POPULATION OF INTEREST	"BY" VARIABLE
FISCAL YEAR:	___ ___ ___ ___ ___ ___ ___	
OCC. FIELD:	___ ___ ___ ___ ___ ___ ___ ___ ___ ___	_
PAYGRADE:	___ ___ ___ ___ ___ ___ ___ ___ ___	_
YOS:	___ ___ ___ ___ ___ ___ ___ ___ ___ ___	_
POP. GROUP:	___ ___ ___ ___	X
SEX:	___ ___ ___	X

The output values will be summed across all occupational fields, paygrades and YOS. Only values for the most recent fiscal year (the default value) will be displayed. A separate value will be calculated for each combination of sex and population group.

To continue, press F3. To exit, press F6.

TUTORIAL

Command ===>

After you have completed the selection screen by choosing (1) the population of interest and (2) the "by variable(s)", you should review your selections carefully. If you want to make changes, use the arrow keys to move the cursor to the field you wish to change, and type over the incorrect entries.

When you are satisfied with your selections, press F2 to proceed. Note that once you press F2 and leave the selection screen, you may not return to it to make additional changes.

MC-EPIDS will then check to ensure that you have entered legitimate values. Erroneous entries will be identified and must be corrected before you can proceed.

To continue, press F3. To exit, press F6.

TUTORIAL

Command ===>

The Output Screen lists the following output choices:

1. Report Listing
2. Bar Chart (data from selected fiscal years)
3. Line Graph (data from all fiscal years)
4. Bar Chart and Report Listing
5. Line Graph and Report Listing

We shall assume that you wish to view a report listing (option 1). MC-EPIDS will supply titles and column headings for the report lists.

To continue, press F3. To exit, press F6.

TUTORIAL

Command ===>

Consider the second query given above, "For each fiscal year, how many black females are in occupational field 72?"

If you selected output option 1, you would view the following tabular listing.

INVENTORY DISTRIBUTION
FOR
FISCAL YEARS 80 81 82 83 84 85 86

FY	TOTAL
1980	2
1981	1
1982	1
1983	1
1984	3
1985	4
1986	4

To continue, press F3. To exit, press F6.

TUTORIAL

Command ===>

You have now seen an example of MC-EPIDS tabular output of inventory-type values.

Although it is beyond the capability of this tutorial to portray graphics, we will briefly discuss options 2-5 of the Output Screen. To view an actual sample of MC-EPIDS graphical output, please see the MC-EPIDS Users' Manual.

For all graphs, the titles and labels will be provided by MC-EPIDS.

It should be noted that legibility of the graphs deteriorates as the number of data points increases.

To continue, press F3. To exit, press F6.

TUTORIAL

Command ===>

Consider again options 2-5 of the Output Screen, all of which produce a graph.

2. Bar Chart (data from selected fiscal years)
3. Line Graph (data from all fiscal years)
4. Bar Chart and Report Listing
5. Line Graph and Report Listing

Note that a line graph (options 3 and 5) can be selected only if all fiscal years are selected. If you have not chosen all fiscal year values, you must select option 2 or 4 (bar graph). If you make an invalid selection, your program cannot be completed.

To continue, press F3. To exit, press F6.

TUTORIAL

Command ===>

The "by variable" column is hierarchically ordered from top to bottom. Thus, on the graphs produced by MC-EPIDS:

1. If more than one fiscal year of data are requested, the

horizontal axis is the fiscal year variable. If a second "by variable" is selected, the bar segments (options 2 and 4) or lines (options 3 and 5) will represent the different levels of the second "by variable". If you choose more than one fiscal year, fiscal year is AUTOMATICALLY the first "by variable".

2. If only one fiscal year of data is requested, then the horizontal axis is the first "by variable" marked with an "X". If a second "by variable" is selected, the bar segments will represent the different levels of the second "by variable".

To continue, press F3. To exit, press F6.

TUTORIAL

Command ===>

Although a maximum of two "by variables" may be shown on the graphic display, more than two may be selected on the Selection Screen.

When more than two "by variables" are selected, the first two marked are used in the graph. The tabular output, however, will provide output of the multiple "by variables", as requested.

To continue, press F3. To exit, press F6.

TUTORIAL

Command ===>

If you request graphic output (options 2-5), you will be queried for device type. If you are using TGRAF by Grafpoint, type TEK4105 and then press enter/return. Consult the Users' Guide if you are using another terminal emulation package. (You will be asked for this information only during your first request for a graph.)

After you supply the device name, your graph will be painted on the screen. If you wish, you may screen-print the graph. If you choose option 4 or 5, you must press ENTER/RETURN in order to view the tabular output.

After you have viewed the graphical and/or tabular display of MC-EPIDS, press F6 to exit to the Opening Menu.

To continue, press F3. To exit, press F6.

TUTORIAL

Command ===>

At this point in the tutorial, we return to the Type-of-Calculation menu.

1. Inventory values
2. Average year of service
3. Percent by paygrade
4. Percent by year of service

You have examined how to mark screens and make decisions for performing inventory calculations (option 1). The procedures for marking screens and making decisions for each of the other options are very similar. For more information, please see the MC-EPIDS Users' Manual.

When ready to exit the tutorial, press F6.

Appendix C.2

MENU SCREENS

```
Command ===>

* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
*                                                                             *
*                                  MC-EPIDS                                   *
*                                                                             *
*                                 WELCOME TO THE                              *
*                                                                             *
*                        MARINE CORPS ENLISTED PERSONNEL                      *
*                                                                             *
*                          INFORMATION DELIVERY SYSTEM                        *
*                                  Version 1.0                                *
*                                                                             *
*                                                                             *
*                                 Provided by the                             *
*                 Navy Personnel Research and Development Center              *
*                              San Diego, California                          *
*                                                                             *
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

                                                  Press F2 to continue.
```

Select Option ===> 1 Press END to return.

Select one of the following options (1-3); then press ENTER/RETURN.

1. Retrieve and display manpower statistics

2. View the MC-EPIDS tutorial

3. Exit MC-EPIDS

Select Option ===> Press END to return.

Enter type of calculation desired (1-4); then press ENTER/RETURN.

1. Inventory values

2. Average year of service

3. Percent by paygrade

4. Percent by year of service

**Notes:
For help, press F1.
To exit, press F6.

Command ===>

Select the population of interest and "by" variable(s). To see a listing of valid values for any category, place the cursor in any blank of the category and press F1.

	POPULATION OF INTEREST	"BY" VARIABLE
FISCAL YEAR:	___ ___ ___ ___ ___ ___ ___	
OCC. FIELD:	___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___	_
	___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___	
	___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___	
	___ ___ ___ ___ ___ ___ ___ ___	
PAYGRADE:	___ ___ ___ ___ ___ ___ ___ ___ ___	_
YEARS OF SERVICE:	___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___	_
	___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___	
	___ ___ ___ ___ ___ ___ ___ ___ ___	
POP. GROUP:	___ ___ ___ ___	_
SEX:	___ ___ ___	_

**Notes:
When satisfied with your selections, press F2 to continue.
To exit without performing calculation, press F6.

Select Option ===> Press END to return.

Select the desired output (1-5); then press ENTER/RETURN.

1. Report Listing
2. Bar Chart (data from selected Fiscal Years)
3. Line Graph (data from all Fiscal Years)
4. Bar Chart and Report Listing
5. Line Graph and Report Listing

Appendix C.3

HELP SCREENS

Command ===> Press FORWARD for more.

HELP FOR TYPE-OF-CALCULATION SCREEN

1. INVENTORY VALUES: For a given population, this function yields a count of the total number of Marines on active duty. The precise population is defined by limiting (or not limiting) the variables fiscal year, occupation field, paygrade, years of service, population group, and sex. Each of these variables may also be used as a "by" variable.

2. AVERAGE YEAR OF SERVICE (YOS): For a given population, this function yields the average YOS of the members of the population. A Marine in YOS cell n is assumed to have >(n - 1) and <= n years of service. Thus, calculation for a Marine in YOS cell n is based on n - 0.5 actual years of service. Example: suppose a given population contains 10 Marines in YOS cell 10, 20 Marines in YOS cell 11, and 20 Marines in YOS cell 12. The average YOS for this population is 10.7, ((10 * 9.5) + (20 * 10.5) + (20 * 11.5)) /50.

**Notes:
To view page 2, press F3.
To return to Type-of-Calculation screen, press F5.

Command ===> Press END to return.

3. PERCENT BY PAYGRADE: For a given population, this function yields the percentage of the population in each paygrade (E1 - E9). Example: suppose a population contains 15 E1's, 20 E2's, 35 E3's, and 30 E4's. The percentages are: E1 = 15% (15/100), E2 = 20% (20/100), E3 = 35% (35/100), and E4 = 30% (30/100).

4. PERCENT BY YEAR OF SERVICE: For a given population, this function yields the percentage of the population in each cell. Example: suppose a population is defined for Marines in YOS 11 - 13. The The populations of those cells are 50, 40, and 30, respectively. The YOS percentages appear as YOS 11: 41.7% (50/120), YOS 12: 33.3% (40/120), and YOS 13: 25% (30/120).

**Notes:
To view page 1, press F4.
To return to Type-of-Calculation screen, press F5.

Command ===> Press FORWARD for more.

VALID VALUES FOR THE VARIABLE "FISCAL YEAR"

```
80  (1980)
81  (1981)
82  (1982)
83  (1983)
84  (1984)
85  (1985)
86  (1986)
```

**Notes:
If fiscal year is a "by" variable, data for each fiscal year will be grouped separately.
If fiscal year is NOT a "by" variable, data for only the most recent fiscal year will be used.

To return to Selection screen, press F5.

Command ====> Press FORWARD for more.

VALID VALUES FOR THE VARIABLE "OCC. FIELD"

00	UNKNOWN	28	DATA/COMM MAINT
01	PERSONNEL & ADMIN	30	SUPPLY ADMINISTRATION
02	INTELLIGENCE	31	TRAFFIC MANAGEMENT
03	INFANTRY	33	FOOD SERVICE
04	LOGISTICS	34	AUDITING FINAN/ACCTNG
08	FIELD ARTILLERY	35	MOTOR TRANSPORT
11	UTILITIES	40	DATA SYSTEMS
13	ENGINEER CONSTR/EQUIP	41	MARINE CORPS EXCHANGE
14	DRAFTING SURVEY/MAPPING	43	PUBLIC AFFAIRS
15	PRINTING/REPRODUCTION	44	LEGAL SERVICES
18	TANK & ASSAULT AMPHIB	46	TRAINING & AV SUPPORT
21	ORDNANCE	55	BAND
23	AMMUNITION & EOD	57	NUCLEAR, BIO & CHEM
25	OPERATIONS COMM	58	MILITARY POLICE
26	SIGNALS INTELLIGENCE	59	ELECTRONICS MAINTENANCE

**Notes:
To view page 2, press F3.
To return to Selection screen, press F5.

Command ===> Press END to return.

60 AIRCRAFT MAINTENANCE
61 AVIATION MAIN (HELO)
63 AVIONICS
64 AVIONICS
65 AVIATION ORDNANCE
66 DISCONTINUED
68 WEATHER SERVICE
70 AIRFIELD SERVICES
72 AIR CONTROL
73 AIR TRAFFIC CONTROL
89 1STSGT/SGTMAJ
98 MARINE CORPS BAND
99 RECRUIT

**Notes:
To view page 1, press F4.
To return to Selection screen, press F5.

Command ===> Press END to return.

VALID VALUES FOR THE VARIABLE "PAYGRADE"

1 (E1)
2 (E2)
3 (E3)
4 (E4)
5 (E5)
6 (E6)
7 (E7)
8 (E8)
9 (E9)

**Notes:
To return to Selection screen, press F5.

Command ===> Press END to return.

VALID VALUES FOR THE VARIABLE "YEARS OF SERVICE"

1	(> 0 YOS <= 1)	17	(> 16 YOS <= 17)
2	(> 1 YOS <= 2)	18	(> 17 YOS <= 18)
3	(> 2 YOS <= 3)	19	(> 18 YOS <= 19)
4	(> 3 YOS <= 4)	20	(> 19 YOS <= 20)
5	(> 4 YOS <= 5)	21	(> 20 YOS <= 21)
6	(> 5 YOS <= 6)	22	(> 21 YOS <= 22)
7	(> 6 YOS <= 7)	23	(> 22 YOS <= 23)
8	(> 7 YOS <= 8)	24	(> 23 YOS <= 24)
9	(> 8 YOS <= 9)	25	(> 24 YOS <= 25)
10	(> 9 YOS <= 10)	26	(> 25 YOS <= 26)
11	(> 10 YOS <= 11)	27	(> 26 YOS <= 27)
12	(> 11 YOS <= 12)	28	(> 27 YOS <= 28)
13	(> 12 YOS <= 13)	29	(> 28 YOS <= 29)
14	(> 13 YOS <= 14)	30	(> 29 YOS <= 30)
15	(> 14 YOS <= 15)	31	(all values > 30)
16	(> 15 yos <= 16)		

**Notes:
To return to Selection screen, press F5.

Command ===> Press END to return.

VALID VALUES FOR THE VARIABLE "POP. GROUP"

1 (WHITE)
2 (BLACK)
3 (HISPANIC)
4 (OTHER)

**Notes:
To return to Selection screen, press F5.

Command ===> Press END to return.

VALID VALUES FOR THE VARIABLE "SEX"

1 (MALE)
2 (FEMALE)
3 (UNKNOWN)

**Notes:
To return to Selection screen, press F5.

Appendix C.4

OUTPUT FORMATS

After the user completes selections of both data and output format, the MC-EPIDS provides a summary of the selection criteria. Three sample summaries are provided below.

AS A RESULT OF THE SELECTIONS MADE AGAINST THE MC-EPIDS DATABASE
THE DATABASE CONTAINS INFORMATION FOR THE FOLLOWING CATEGORIES

FISCAL YEAR	OCC. FIELD	PAYGRADE	YOS	POP. GROUP	SEX
86	00	E1	1	WHITE	MALE
	01	E2	2	BLACK	FEMALE
	02	E3	3	HISPANIC	
	03	E4	4	OTHER	
	04	E5	5		
	08	E6	6		
	11	E7	7		
	13	E8	8		
	14	E9	9		
	15		10		
	18		11		
	21		12		
	23		13		
	25		14		

Press F3 to advance one page, F4 to go back one page.
After the bell, press F6 to exit.

AS A RESULT OF THE SELECTIONS MADE AGAINST THE MC-EPIDS DATABASE
THE DATABASE CONTAINS INFORMATION FOR THE FOLLOWING CATEGORIES

FISCAL YEAR	OCC. FIELD	PAYGRADE	YOS	POP. GROUP	SEX
	26		15		
	28		16		
	30		17		
	31		18		
	33		19		
	34		20		
	35		21		
	40		22		
	41		23		
	43		24		
	44		25		
	46		26		
	55		27		
	57		28		

Press F3 to advance one page, F4 to go back one page.
After the bell, press F6 to exit.

AS A RESULT OF THE SELECTIONS MADE AGAINST THE MC-EPIDS DATABASE
THE DATABASE CONTAINS INFORMATION FOR THE FOLLOWING CATEGORIES

FISCAL YEAR	OCC. FIELD	PAYGRADE	YOS	POP. GROUP	SEX
	58		29		
	59		30		
	60		31		
	61		.		
	63		.		
	64		.		
	65		.		
	68		.		
	70		.		
	72		.		
	73		.		
	89		.		
	98		.		
	99		.		

Press F3 to advance one page, F4 to go back one page.
After the bell, press F6 to exit.

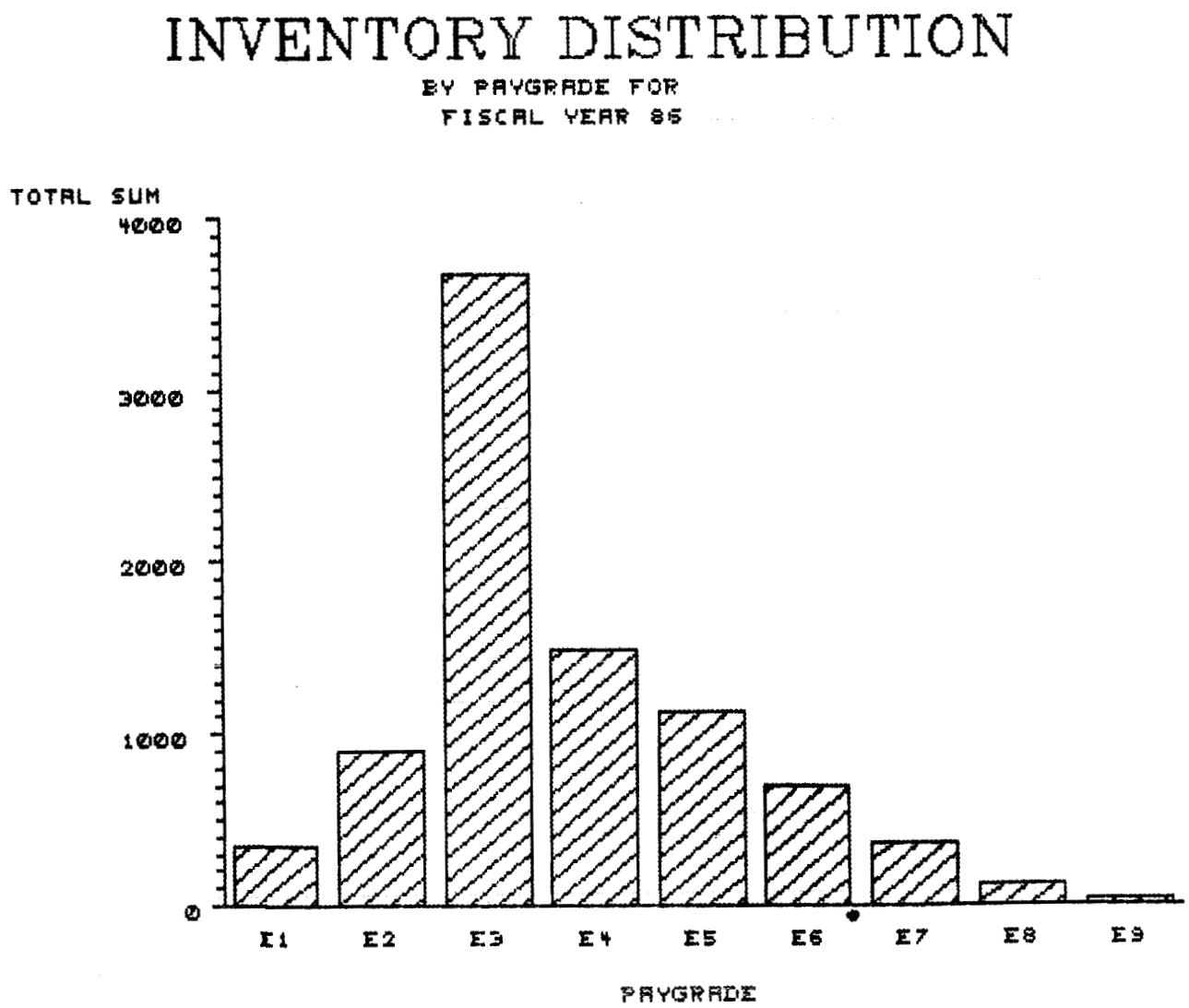
A sample of tabular output for the MC-EPIDS is provided below.

INVENTORY DISTRIBUTION
BY SEX FOR
FISCAL YEAR 86

SEX	TOTAL
MALE	169354
FEMALE	9236

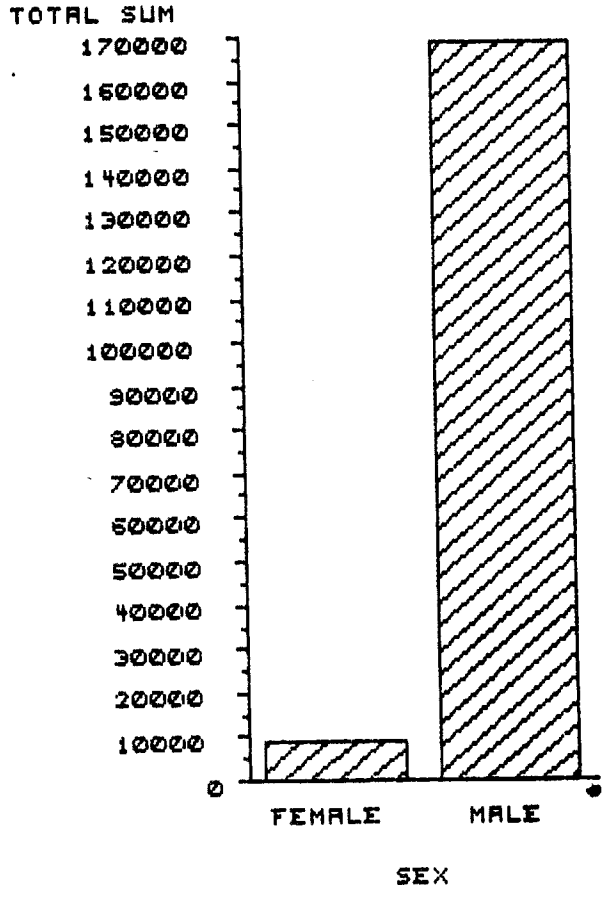
Press F3 to advance one page, F4 to go back one page.
After the bell, press F6 to exit.

Samples of graphical output are provided below and on the next page.



INVENTORY DISTRIBUTION

BY SEX FOR
FISCAL YEAR 85



Appendix C.5

ERROR MESSAGES

Sample error messages supplied by the MC-EPIDS are supplied below.

```
Command ===>
ERROR: Invalid entry at ROW = 20 COLUMN = 22. Please reenter
Select the population of interest and "by" variable(s).  To see a listing
of valid values for any category, place the cursor in any blank of the
category and press F1.
                                                                    "BY"
                    POPULATION OF INTEREST                        VARIABLE

       FISCAL YEAR: __ __ __ __ __ __ __
        OCC. FIELD: __ __ __ __ __ __ __ __ __ __ __                 _
                    __ __ __ __ __ __ __ __ __ __ __
                    __ __ __ __ __ __ __ __ __ __ __
                    __ __ __ __ __ __ __ __
          PAYGRADE: __ __ __ __ __ __ __ __ __
  YEARS OF SERVICE: __ __ __ __ __ __ __ __ __ __ __                 _
                    __ __ __ __ __ __ __ __ __ __ __
                    __ __ __ __ __ __ __ __ __
        POP. GROUP: __ __ __ __                                      X
               SEX: F  __ __                                         _

 **Notes:
   When satisfied with your selections, press F2 to continue.
   To exit without performing calculation, press F6.
```

AS A RESULT OF THE SELECTIONS MADE AGAINST THE MC-EPIDS DATA BASE
THE DATA BASE CONTAINS INFORMATION FOR THE FOLLOWING CATEGORIES

MESSAGE

All fiscal years were not selected.

No line graph will be produced.

Press F6 to continue.

Press F3 to advance one page, F4 to go back one page.
After the bell, press F6 to exit.

INTERNAL DISTRIBUTION LIST

1.	C. B. Grillot
2.	R. G. Edwards
3.	W. Fulkerson
4.	R. B. Honea
5.	H. L. Hwang
6.	J. Kolb
7.	George T. Privon
8-12.	Tykey Truett
13.	D. P. Vogt
14.	T. J. Wilbanks
15.	D. E. Reichle
16.	R. B. Shelton
17.	Eleanor T. Rogers
18.	Central Research Library
19.	Document Reference Section
20.-23.	Laboratory Records
24.	Laboratory Records - RC
25.	Patent Office
26-27.	C. S. Bowman

EXTERNAL DISTRIBUTION LIST

28. Office of Assistant Manager for Energy Research & Development, Department of Energy, Oak Ridge Operations Office, Oak Ridge, TN 37831.

29-38. Office of Scientific and Technical Information, P.O. Box 62, Oak Ridge, TN 37831.

39. J. J. Cuttica, Vice President of Research and Development, Gas Research Institute, 8600 W. Bryn Mawr Avenue, Chicago, IL 60631.

40. J. P. Kalt, Professor of Economics, Kennedy School of Government, Harvard University, 79 John F. Kennedy Street, Cambridge, MA 02138.

41. D. E. Morrison, Professor of Sociology, Michigan State University, 201 Berkey Hall, East Lansing, MI 48824-1111.

42. R. L. Perrine, Professor, Engineering and Applied Sciences, Civil Engineering Department, Engineering I, Room 2066, University of California, Los Angeles, CA 90024.

43. Debbie Flanagan, 1216 Live Oak Circle, Knoxville, TN 37932.

44. R. P. Rascoll, Structured Systems, 20 Argonne, Suite 350, Oak Ridge, TN 37830.

45-49. Carol Mullins, Navy Personnel Research & Development Center, Code 61, San Diego, CA 92152-6800.