DOCUMENT RESUME

ED 050 763 LI 002 823

AUTHOR	Ludwig, Herbert; Vallee, Jacques
TITLE	Medical Data Management in Time-Sharing: Findings of the DIRAC Project.
INSTITUTION	Stanford Univ., Calif. Computation Center.
REPORT NO	RR-3
PUB DATE	Nov 70
NOTE	110p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$6.58
DESCRIPTORS	*Computers, Data Bases, *Information Networks, *Information Processing, Information Retrieval, *Information Systems, Medicine, *On Line Systems, *Programing Languages
IDENTIFIERS	DIRAC, Direct Access Project

ABSTRACT

In terms of examples drawn from clinical and research data files, one of the objectives of this study is to illustrate several factors that have combined to delay the implementation of medical data bases. A primary factor has been inherent in the design cf computer software. The languages currently on the market are procedural in nature: they demand the complete definition of a problem before a program can be run on the machine. This is a serious difficulty because medical information -- as opposed to mathematical quantities -- cannot be cast in a definite format, and questions of interest to a potential investigator cannot always be enumerated before the data is acquired. The constraints placed on data retrieval in the environment of a Blood Bank, for example, are such that on-line interaction with short response time is the basic design objective. In order to clarify the problems of interactive data handling in time-sharing under such constraints, a series of five experiments using the Direct Access Project (DIRAC) language prototype is explained. (Author/MM)

MEDICAL DATA MANAGEMENT IN TIME-SHARING: FINDINGS OF THE DIRAC PROJECT

RESEARCH REPORT NUMBER III

NOVEMBER 1970

HERBERT LUDWIG AND JACQUES VALLEE

STANFORD UNIVERSITY
COMPUTATION CENTER
INFORMATION SYSTEMS

MEDICAL DATA MANAGEMENT IN TIME-SHARING:
FINDINGS OF THE DIRAC PROJECT

This is the third and final report by the Information Systems Group at Stanford University Computation Center. The previous reports were entitled:

I) THE DIRAC LANGUAGE: CONCEPTS & FACILITIES (MAY 1970).

II) SCIENTIFIC INFORMATION NETWORKS: A CASE STUDY (SEPTEMBER 1970).

They may be obtained from the authors.

TABLE OF CONTENTS

EXPERIMENTS WITH A CANCER RESEARCH FILE

INTRODUCTION

Two previous reports in this series have described a data-base oriented system named DIRAC and some of its applications, notably in the implementation of a documentation network for use by astronomers. In the present report we have gathered several documents pertaining to the application of the DIRAC prototype to the medical environment.

In his classic book, "Use of Computers in Biology and Medicine" (McGraw Hill, 1965), Ledley pointed out:

> Perhaps the greatest utilization of computers will be in biomedical applications. The problems that arise here characteristically involve large numbers of data and many complicated interrelating factors, and it is just these types of problems for which computers are primarily suited.

Several factors, however, have combined to delay the implementation of medical data-bases, and one of the objectives of our study was to illustrate these obstacles in terms of examples drawn from clinical and research data files. A primary factor has been inherent in the design of computer software. The languages currently on the market are procedural in nature: they demand the complete definition of a problem before a program can be run on the machine. This is a serious difficulty because medical information -- as opposed to mathematical quantities -- cannot be cast in a definite format, and questions of interest to a potential investigator cannot always be enumerated before the data is acquired. The constraints placed on data retrieval in the environment of a Blood Bank, for example, are such that on-line interaction with short response time is the basic design objective. In order to clarify the problems of interactive data handling in time-sharing under such constraints, we conducted a series of experiments with the DIRAC language prototype. These experiments will be described in five reports.

"Applications of a Generalized Data-Management Language in Medicine: Some Recent Experience at Stanford University", presents a survey of the features of DIRAC that make it suitable in the medical environment. It then reviews various applications with an emphasis on language design problems.

"Progress Towards A Direct-Access Hematology Data Base: Stanford's Experience with the DIRAC Language", describes in detail the design and implementation of an interactive information system for bone marrow analysis.

"Interactive Computer Management of Clinical Data: A New Approach in the Blood Bank", is a step-by-step illustration of the same approach in the design of a model for a transfusion center.

"Forms Design in Medical Information Retrieval", reviews the problems of information acquisition in the cancer research environment.

"Experiments with a Cancer Research File", completes the preceding report with a description of a computational interface providing interactive statistical information to a terminal user.

ACKNOWLEDGMENTS

The experiments reported here were made possible by the support extended to us by members of the Stanford medical community; our special thanks go to Dr. Paul Wolf, Director of the Clinical Laboratory and of the Stanford Blood Bank, for many valuable discussions during which he offered comments and advice crucial to the success of the DIRAC prototype. Dr. Gordon Ray spent many hours with us studying the data acqusition and interface problems that arose in connection with the implementation of the cancer research file. Dr. Bagshaw, Dr. Zatz, Dr. Corn, Dr. Sussman, and Dr. Downey, have made their time and experience available to us in a series of most valuable meetings.

We are also grateful to Mrs. Shirley Bardo, Mrs. Caroline Newell, and the entire Staff of the Stanford Clinical Laboratory for their guidance through many technical problems. Finally we are indebted to the Campus Facility of the Computation Center first under Mr. Roderick Fredrickson, and later under Mrs. Susan Kolasa, for supporting us during the life of the DIRAC project.

APPLICATIONS OF A GENERALIZED DATA-MANAGEMENT LANGUAGE IN MEDICINE: SOME RECENT EXPERIENCE AT STANFORD UNIVERSITY

by

Herbert R. Ludwig, Ph.D.

1. INTRODUCTION

An interactive file-oriented language that allows the user to interface with a text-editor and with his own FORTRAN or assembly language code has been implemented for the IBM 360/67 computer of the Campus Facility at Stanford University by Dr. Jacques Vallee, Manager of the Information Systems Department. This paper is concerned with one set of applications, medical, for which this file-oriented language has been used.

The first section of the paper is a brief description of a file-oriented language called DIRAC. The four basic modes of operation -- CREATE, UPDATE, QUERY, and STATUS -- are discussed. Also, a brief discussion on file structures which will be useful for the prospective user of file-oriented systems.

The second section concerns itself with the different applications that DIRAC has had at the Stanford Medical Center, and the impact the introduction of such a system has had upon the cost and time spent on projects of similar nature.

Conclusions and some recommendations for the future are discussed in the third section. The impact of such a system upon the medical community, the patient and the doctor, with respect to its ability to accelerate the research process as well as treatment and diagnostic procedures are also discussed. Future recommendations are given in light of the experience gained from these studies.

2. THE DIRAC SYSTEM

2.1 Background

DIRAC (Date, Integer, Real, Alphanumeric, and Coded) is an information retrieval language which provides the user the ability to operate under four modes: CREATE, UPDATE, QUERY, and STATUS. (1)

(1) The CREATE mode allows the user to completely define the terminology and structure of his own file. The structure of the empty file is constructed under this mode.

(2) The UPDATE mode allows such operations as adding, deleting modifying, or replacing records.

(3) The QUERY mode of DIRAC allows the user to obtain information about selected subsets of his file at any level of the record structure so designated by the user during the CREATE mode. The different commands through which a file may be queried are described in a later section of this article.

(4) The STATUS mode provides the user with an up-to-date status report for his particular file. Field identification, description of the fields, statistics and validation information are displayed in a standard report form.

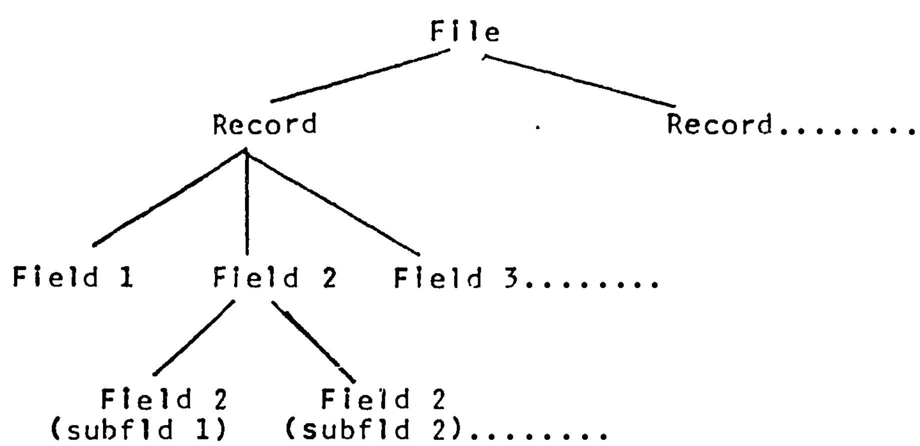
2.2 File Structures for DIRAC

A file is defined here as a collection of related records containing data needed for subsequent processing. The medical field lends itself nicely to strictly defined file structures in that structures beyond the subfield structure are usually not required. This has been the case at Stanford Hospital for the three applications discussed in this paper.

2.2.1 Fields and Subfields:

Within a DIRAC record every property is identified as an individual field: a patient's name in a hospital record, an address, an invoice number. Once record structure is determined by the user, the fields are declared to DIRAC and named and identified during file creation. They are then available for any type of retrieval response from the file. Fields of a record can be numeric, integer such as 'Age', numeric real such as 'purchases' on a charge account (XX.XX), alphabetic such as 'Name' or 'Address'; they can also be dates or codes.

A record consists of fields which may themselves be formed from two or more subfields. This process of subdivision (tree structure) can theoretically be continued.



However, in the first version of DIRAC file structure representations were not supported beyond the subfield level.

2.2.2 File Construction Under DIRAC:

DIRAC provides the user with the opportunity to completely specify his own file organization or structure. Thus, the user does not have to be concerned about using a fixed field or fixed word length or format. He is not bound by a set of rigid rules pertaining to record size, length, etc., and these parameters are not even apparent to him.

The user should first compile a working list of all fields to be included in a record, specifying whether or not a field is singular or multiple (subfields). Example: Suppose that we want to create a DIRAC file of heart-transplant patients; we have determined that we want to include the following information (fields) in a patient's record:

FIELD DESCRIPTION	SINGLE/MULTIPLE
Name of Patient	[illegible]
Hospital Record Number	S
Home Address	S
Date of Birth	S
Home Phone	S
Cardiac Diagnosis	M
Vascular Diagnosis	M
Admission Dates	M
Dates of Operations	M
Status at Time of Discharge	M
Date of Death	S

A typical Patient Record would have the structure:

Name	Record #	Address	Birthdate	Home Phone
John L. Smith	237863	33 Fifth Ave. Glendale, Ca.	19080325	459-7414

Cardiac Diagnosis	Vascular Diagnosis	Admission Dates	Operation Dates	Status	Date of Death
(text)	(text)	19680325	----	(text)	
(text)	(text)	19681103	19681120	(text)	
. .	. .	. .	. .	. .	

Note that the fields Cardiac Diagnosis, Vascular Diagnosis, Admission Dates, Operation Dates, and Status at Time of Discharge are multiple. In other words, a patient might have been admitted several times for some type of diagnostic checkup or operation over the past year(s); each time the patient was admitted pertinent data was recorded in the appropriate field.

The user must also determine the "type" of each field which he includes as part of a record. For example, patient's name would be alphanumeric (ALPHA), whereas Hospital Record Number would be INTEGER; Date of Birth would be DATE type, as well as Admission Dates and Dates of Operations. However, in the latter two cases a distinction is made to depict the MULTIPLE status of these two fields.

After determining the type of each field and whether or not that field is singular or multiple, the fields may be numbered and given a name as follows:

FIELD	NAME	DESCRIPTION
1	Name	Name of Patient
2	Number	Hospital Record Number
3	Address	Home Address
4	BirthDate	Date of Birth
5	Phone	Home Phone
6	CDiagnosis	Cardiac Diagnosis
7	VDiagnosis	Vascular Diagnosis
8	AdmDate	Admission Dates
9	OpDate	Operation Dates
10	Status	Status at Time of Discharge
11	DDate	Date of Death

A delimeter will be picked from a set of special characters (such as @,$,#) to denote a field in DIRAC. (The user can pick any delimeter out of the list which is convenient to him, thus avoiding the need for a rigid standard notation imposed by most existing systems. If the user does not select a notation for record and field, the standard notation $ and @ are used respectively.)

DIRAC will prompt the user for Type and Multiplicity of the fields within a record. In our example the following information would then be typed at the terminal: (where every line beginning with a colon is typed by the user)

```
7.  SUPPLY DATA TYPE AND MULTIPLICITY
:      ALPHA SINGLE @3 @1 @5
:      INTEGER SINGLE @2
:      DATE SINGLE @4 @11
:      DATE MULTIPLE @8 @9
:      ALPHA MULTIPLE @6 @10 @7
```

Field specifications can be input in any order. Also note that the delimeter "@" was used to specify fields. "Integer Single" means that the value to be stored in field 2 will be a single integer number. "Alpha Multiple" means that there exists a multiple field in which alphanumeric information is stored. (The same holds true for a DATE MULTIPLE specification.) From the example we note that fields @6 - @10 are multiple. Thus, when reference is made to @8(1) -- admission date -- the cardiac diagnosis, vascular diagnosis, date of operation (if there was an operation) for that visit (or duration of stay), and a statement concerning the condition of the patient at time of discharge are contained in @6(1), @7(1), @9(1), and @10(1), respectively.

2.2.3 Medical Files:

Most files encountered in the medical field have the following properties in common:

(1) They are patient-oriented files, i.e., they are set up by patient name.

(2) Each patient has fields which are similar to other patients, even though these may be in a completely different file, e.g., address, phone, hospital I.D. number, blood type, etc.

(3) Most files do not go beyond the subfield structure, e.g., 'treatment' might be the name of a field which is designated ALPHA MULTIPLE. This means that alphanumeric information is stored in this field and that successive information within each subfield (such as in the case of an array item in FORTRAN) refers to a treatment given the patient on a given date, whereas 'RxDate' might be the name of a multiple field which stores the dates in each successive subfield for each successive treatment.

3. DIRAC APPLICATIONS IN THE HOSPITAL ENVIRONMENT

3.1 Hematology Clinic

3.1.1 Statement of Problem:

The rapidly expanding volume of bone marrow examinations performed at Stanford Medical Center has presented many problems for the physician with respect to his daily work patterns. These problems, for which no classical solution is available, have only recently been recognized. The ability to store and retrieve pertinent patient record information is a case in point. Since it is essential to be able to analyze these data in order to perfect new methods of diagnosis and treatment of hematologic disease, the use of a file-oriented language seemed inevitable. By the introduction of DIRAC the following objectives were met:

(1) The production of a simple bone marrow report format showing all of the essential information necessary for meaningful identification and analysis.

(2) Interfacing this report generator with an efficient file organization within the DIRAC environment.

A brief description of the problem of data-base implementation under these constraints is now presented along with an example of 'conversational' interrogation.

3.1.2 Data-Base Implementation:

A description of the bone marrow report system is presented in Figure 1. Note that the bone marrow reports were typed by a secretary and then filed manually either by a clerk or by the secretary. This method proved to be quite cumbersome because when specific bone marrow reports were desired for study a manual sort had to be made to retrieve the document(s). This was often time consuming and a very burdensome task. Note also that there existed no feedback to the patient from the bone marrow reports after they were filed.

Figure 2 describes the first phase in automating the bone marrow report system described in Figure 1. The bone marrow diagnosis was typed into a working data set in computer storage. The reports could now be checked for accuracy and verified. Additions and corrections could be made with ease utilizing WYLBUR, the Stanford text editor (2). A report generator was written which would generate a formatted general report for each patient record. Copies of these reports are now distributed to different physicians and to the medical records section of the hospital.

This automatic report generating system was a vast improvement over the simplified file system described by Figure 1. The problem of distributing bone marrow reports to various departments was now solved. Since the bone marrow data was now being input remotely into the computer memory, there now existed a data-base within the

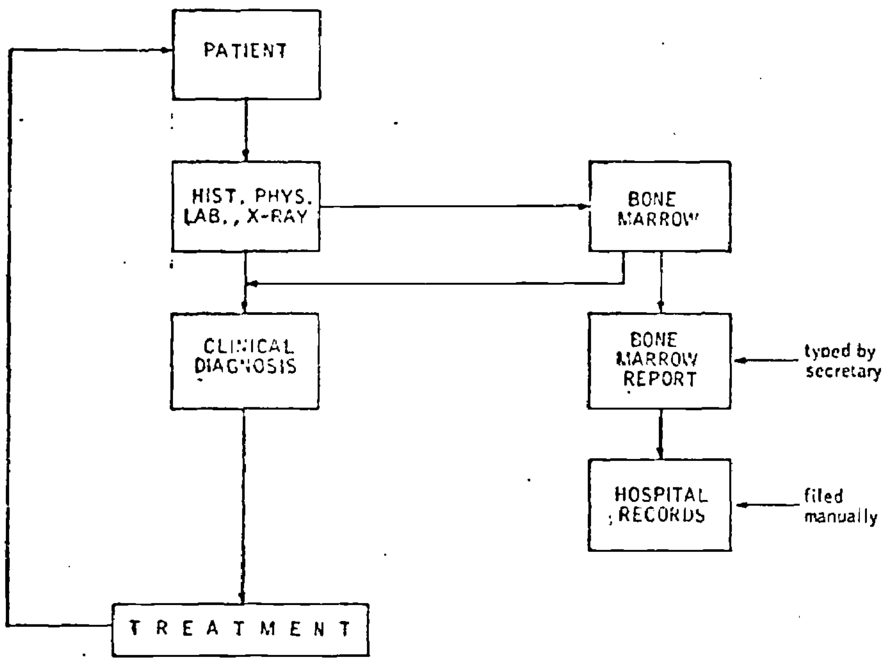


Figure 1.

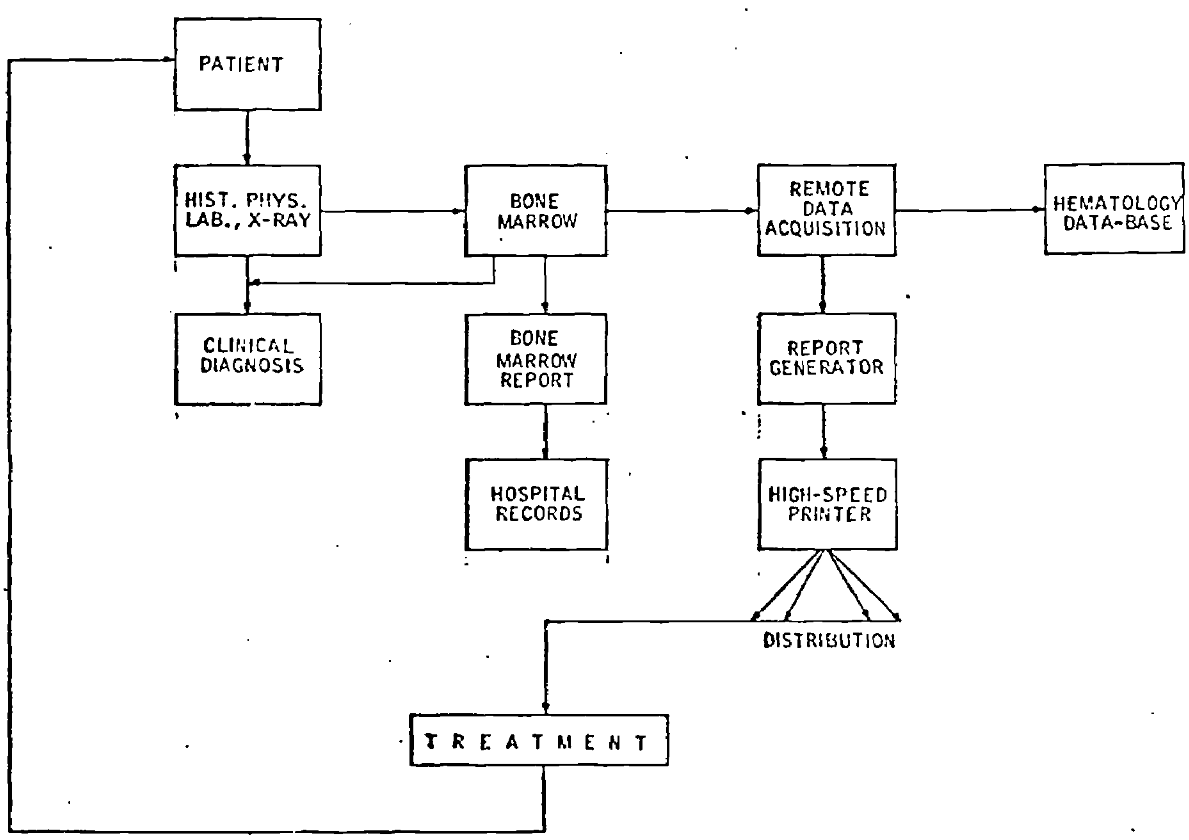


Figure 2.

computer. Thus, the problem of storage & retrieval of specific patient information when needed from the bone marrow file had been partially solved -- a data base existed in machine readable form -- but there existed no facility to interrogate the file. If a physician wished to have information regarding a specific patient or obtain statistical data concerning certain parameters within the records he still had to search the file manually. The clerical task of retrieving records for review and research analysis was solved by the introduction of DIRAC, thus, transforming Figure 2 into Figure 3. Note that there now is feedback to medical personnel on the diagnosis. This feedback is practically instantaneous, and thus, alleviates the clerical burden of searching manually for records within the file.

3.1.3 Conversational Interrogation:

Another advantage gained by introducing a data-management system is that it allows research and administrative procedures to be simplified. A whole file of bone marrow reports is now available for interrogation by medical researchers. By use of its interface capability DIRAC can perform statistical analyses on pertinent records of the file when directed to do so. Interrogation by users of the file can also be displayed on a video unit or typed out at the user terminal, or can be printed on a high speed printer, whichever is more convenient based on the query and need of the user. Figure 4 is a status report for the Bone Marrow File. Note field identifications, statistics, and validations given to the user.

The following example shows how the user can carry on a simple 'conversation' with DIRAC. Browsing the file, retaining subsets of selected queries, and displaying pertinent information are all

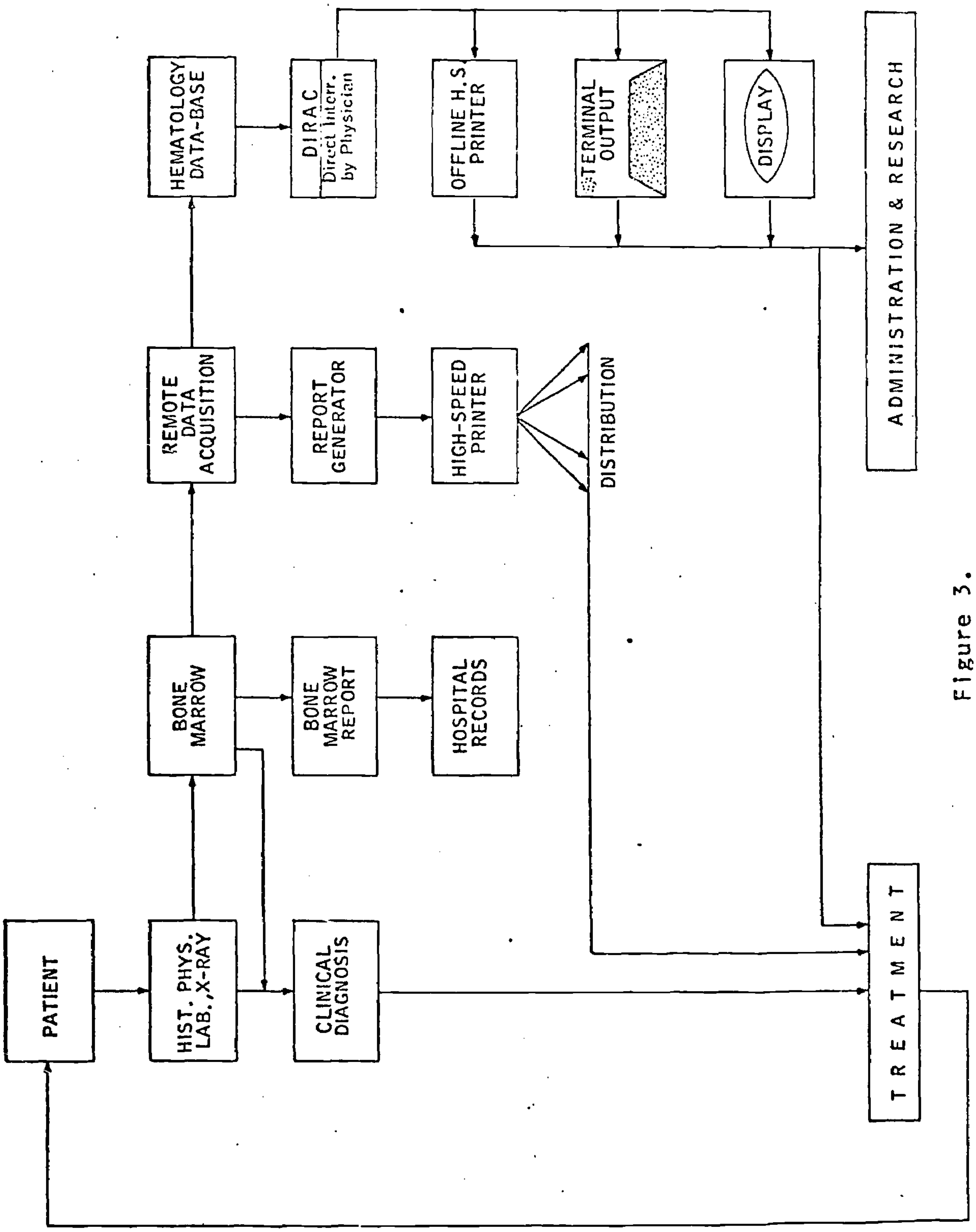

PATIENT
HIST. PHYS. LAB., X-RAY
CLINICAL DIAGNOSIS
BONE MARROW
BONE MARROW REPORT
HOSPITAL RECORDS
REMOTE DATA ACQUISITION
REPORT GENERATOR
HIGH-SPEED PRINTER
DISTRIBUTION
HEMATOLOGY DATA-BASE
DIRAC Direct Interr. by Physician
OFFLINE H.S. PRINTER
TERMINAL OUTPUT
DISPLAY
ADMINISTRATION & RESEARCH
TREATMENT

Figure 3.

STANFORD UNIVERSITY COMPUTATION CENTER	STATUS REPORT FOR FILE M010 1/JUL/1970	(CLASS 1) LANGUAGE: DIRAC-1A

DESCRIPTION : BONE MARROW REPORT

CREATION DATE : 15/MAY/1970	1 = TYPE	FILE NAME : M010
RECORD NOTATION : $VS	2 = MULTIPLICITY	FILE CREATED BY : Wolf
FIELD NOTATION : DF	3 = INDEXING	RECORD LENGTH : 512.
DISPOSITION : PRIVATE	4 = CODE RESIDENCE	NO.OF FIELDS : 29
NO. OF RECORDS : 235	5 = CODE TYPE	LATEST UPDATE ON : 15/JUN/1970

F I E L D   I D E N T I F I C A T I O N ,   S T A T I S T I C S   A N D   V A L I D A T I O N   I N F O R M A T I O N

FLD	NAME	DESCRIPTION	STORAGE 1	2	3	4	5	VALIDATIONS REC.	SIZE	SUB.	STATISTICS SIZE	DEC.	SUB.	EXISTENCE REC.	PCT
1	Name	Name of Patient	A	S	0			YES	0	0	24	0	0	235	100.00%
2	Sex		A	S	0				0	0	6	0	0	235	100.00%
3	Birth	Birthdate	A	S	0				0	0	9	0	0	234	99.57%
4	Ward	Ward No.	A	S	0				0	0	10	0	0	235	100.00%
5	Phys	Ref. Physician	A	S	0				0	0	13	0	0	218	92.77%
6	Record	Medical Record No.	A	S	0			YES	0	0	9	0	0	235	100.00%
7	MarrowNo	Bone Marrow No.	A	S	0			YES	0	0	8	0	0	235	100.00%
8	Date	Examination Date	A	S	0				0	0	8	0	0	235	100.00%
9	Previous	Previous Marrow	A	S	0				0	0	3	0	0	157	66.81%
10	Clinical	Clinical Information	A	S	0				0	0	144	0	0	219	93.19%
11	RBC	Peripheral Smear Section	A	S	0				0	0	363	0	0	208	88.51%
12	Platelets	Peripheral Smear Section	A	S	0				0	0	57	0	0	202	85.96%
13	WBC	Peripheral Smear Section	A	S	0				0	0	139	0	0	200	85.11%
14	C1	Peripheral Smear Section Comment	A	S	0				0	0	67	0	0	43	18.30%
15	Quality	Bone Marrow Aspirate Section	A	S	0				0	0	31	0	0	230	97.87%
16	Ratio	Myeloid/Erythroid --Bone Mar. Se	A	S	0				0	0	39	0	0	194	82.55%
17	Meg	Megakaryocytes -- Bone Marrow Se	A	S	0				0	0	73	0	0	217	92.34%
18	Myeloid	Myeloid Elements -- B.M.Sect.	A	S	0				0	0	144	0	0	217	92.34%
19	Erythroid	Erythroid Elements -- B.M.Sect.	A	S	0				0	0	96	0	0	217	92.34%
20	C2	Bone Marrow Section Comments	A	S	0				0	0	188	0	0	180	76.60%
21	Amount	Iron Stores Section	A	S	0				0	0	32	0	0	195	82.98%
22	Location	Iron Stores Section	A	S	0				0	0	37	0	0	116	49.36%
23	C3	Iron Stores Section Comments	A	S	0				0	0	56	0	0	39	16.60%
24	Impression	General Comments	A	S	0				0	0	268	0	0	228	97.02%
25	Qual	Spicule Section	A	S	0				0	0	159	0	0	82	34.89%
26	Confirms	Confirms Other Data -- Spicule S	A	S	0				0	0	3	0	0	115	48.94%
27	C4	Spicule Section Comments	A	S	0				0	0	222	0	0	130	55.32%
28	Hematologist		A	S	0			YES	0	0	19	0	0	235	100.00%
29	Director		A	S	0			YES	0	0	25	0	0	235	100.00%

Figure 4.

displayed. (Each line beginning with a colon is the user's response to a DIRAC prompt)

EXAMPLE OF CONVERSATIONAL QUERY

```
:    SELECT ALL
     ______________________
     401 RECORDS SELECTED
:    Sex CONTAINS FEMALE END
     ______________________
     168 RECORDS SELECTED
:    RETAIN

:    Date CONTAINS 67 END
     ______________________
     27 RECORDS SELECTED
:    Director CONTAINS WOLF END
     ______________________
     3 RECORDS SELECTED
:    TYPE Name MarrowNo END
_________________________
     55
Name          Mary Q. Smith
MarrowNo      55-77-25
_________________________
     96
Name          Joan R. Jones
MarrowNo      77-22-28
_________________________
:    RELEASE
:    Impression CONTAINS "HODGKINS DISEASE" END
_________________________
     50 RECORDS SELECTED
:    Date CONTAINS "3-11-70" AND Quality CONTAINS GOOD END
_________________________
     2 RECORDS SELECTED
:    TYPE Name Record Date Impression Quality END
_________________________
     106
Name          John R. Hopkins
Record        453-789
Date          3-16-70
Quality       GOOD
Impression    EOSINOPHILIA. PLASMACYTOSIS. AND INCREASED
              R.E. CELL IRON.
_________________________
     368
Name          Kenneth M. Lawrence
Record        583-91
Date          3-11-70
Quality       GOOD
Impression    NON-DIAGNOSTIC MARROW. PLASMA CYTOSIS AND
              MILD EOSINOPHILIA.
_________________________
:    RELEASE
       .
       .
```

Figure 5

In the above example the exclamation mark stands for a wild character. Thus, searches may be made over a specific set of digits in this case. Also the use of logical expressions is valid.

3.2 Stanford Blood Bank

3.2.1 Statement of Problem:

The Stanford Blood Bank, headed by Dr. Paul Wolf, has initiated a research effort leading to the implementation of an automated transfusion center prototype. This research effort has combined the advanced hardware and software technology available at Stanford with the Hospital's experience in Blood Bank design. It centers on one aspect of medical information systems design, and leads to results that will be generally applicable throughout the Clinical Laboratory.

Even with the advent of better transfusion techniques and devices, the Hospital still envisions an ever increasing use in whole blood. Increased patient traffic, due to the advent of new technology and greater capacity of the hospital, will add to this expansion. Thus, a control system that will govern the use and allocation of blood units throughout the hospital is a definite need. Along with this increase in whole blood use will come an increase in the paper load both for administrative and research purposes.

With the introduction of a non-procedural, file oriented, data-base management system, namely DIRAC, the above problems can be directly confronted and solved. The "non-procedural" technique is characterized here by the ability for the blood bank clerk, the research assistant, medical technologist, or a doctor, etc., to directly interrogate or access a specified file without programmer intervention or programming knowledge.

3.2.2 Systems Design & Analysis:

The objectives of the project were threefold:

(1) To decrease the paper-volume and paper-traffic through the Blood Bank.

(2) To design an information system which will allow medical personnel to access the files directly.

(3) To set up inventory and control systems for all units passing through the Blood Bank.

The system as it existed prior to the design phase is described in Figure 6. Note that many of the services performed by the clerks could easily be performed by the computer, e.g., updating donor cards, the entering of recipient information in the S.U.H. logbook. etc. The main files in this system are the Stanford Donor File, the Recipient File, and the Two Logbooks -- Stanford University Hospital Logbook and the Blood-from-Other-Banks Logbook.

Figure 7 shows how the existing system was redesigned to make use of an interactive data-base management language. Note the three sequences into which the system has been partitioned: the matching, inventory, and update sequence. The flow of paper is now administered internally by the blood bank clerk through the use of this data-base management system.

3.2.3 Emergency Interrogation of Blood Donor File:

Assume that an emergency has arisen at a participating hospital. The Central Blood Bank receives a call to supply or locate donors with O NEG Blood type. The problem for the blood bank

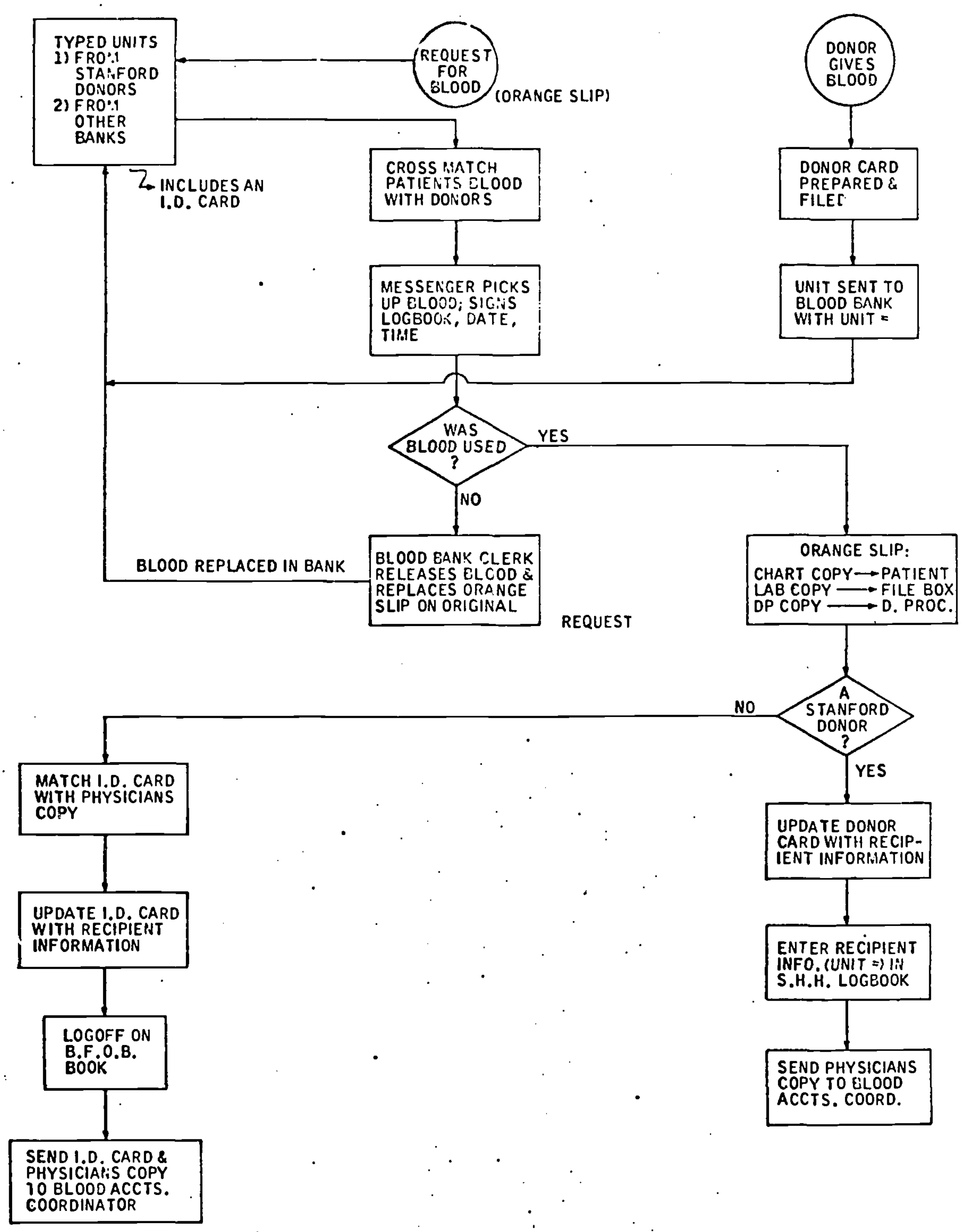

TYPED UNITS
1) FROM STANFORD DONORS
2) FROM OTHER BANKS
REQUEST FOR BLOOD
(ORANGE SLIP)
DONOR GIVES BLOOD
INCLUDES AN I.D. CARD
CROSS MATCH PATIENTS BLOOD WITH DONORS
DONOR CARD PREPARED & FILED
MESSENGER PICKS UP BLOOD; SIGNS LOGBOOK, DATE, TIME
UNIT SENT TO BLOOD BANK WITH UNIT #
WAS BLOOD USED ?
YES
NO
BLOOD REPLACED IN BANK
BLOOD BANK CLERK RELEASES BLOOD & REPLACES ORANGE SLIP ON ORIGINAL
REQUEST
ORANGE SLIP:
CHART COPY → PATIENT
LAB COPY → FILE BOX
DP COPY → D. PROC.
A STANFORD DONOR ?
NO
YES
MATCH I.D. CARD WITH PHYSICIANS COPY
UPDATE I.D. CARD WITH RECIPIENT INFORMATION
LOGOFF ON B.F.O.B. BOOK
SEND I.D. CARD & PHYSICIANS COPY TO BLOOD ACCTS. COORDINATOR
UPDATE DONOR CARD WITH RECIPIENT INFORMATION
ENTER RECIPIENT INFO. (UNIT #) IN S.H.H. LOGBOOK
SEND PHYSICIANS COPY TO BLOOD ACCTS. COORD.

Figure 6.

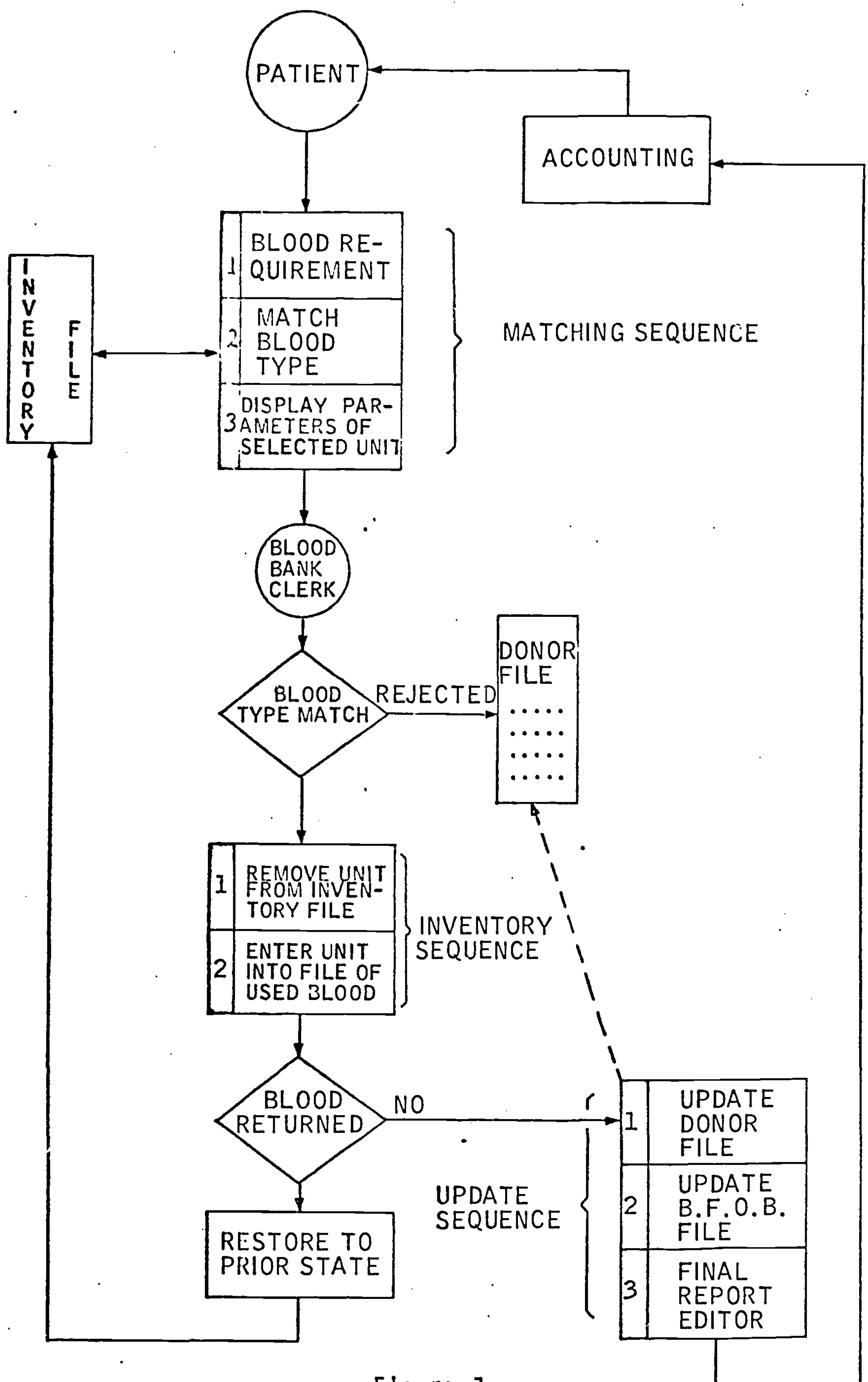

PATIENT
ACCOUNTING
INVENTORY FILE
1 BLOOD RE-QUIREMENT
2 MATCH BLOOD TYPE
3 DISPLAY PAR-AMETERS OF SELECTED UNIT
MATCHING SEQUENCE
BLOOD BANK CLERK
BLOOD TYPE MATCH
REJECTED
DONOR FILE
1 REMOVE UNIT FROM INVEN-TORY FILE
2 ENTER UNIT INTO FILE OF USED BLOOD
INVENTORY SEQUENCE
BLOOD RETURNED
NO
RESTORE TO PRIOR STATE
UPDATE SEQUENCE
1 UPDATE DONOR FILE
2 UPDATE B.F.O.B. FILE
3 FINAL REPORT EDITOR

Figure 7.

is to check their inventory and determine whether or not they can afford to delete their supply of O NEG blood or locate donors and send them as fast as possible to the participating hospital. Assume the doctor decides on the latter alternative. A clerk will immediately establish communication with the computer (at night, a terminal located in the director's home can be used for this purpose) and the interrogation of the donor file under the DIRAC system takes place as follows: (where every line beginning with a colon is typed by the user)

```
:  Serum CONTAINS O END
_______________________________________________
512 RECORDS SELECTED
:  RETAIN
:  RHType CONTAINS NEG END
_______________________________________________
28 RECORDS SELECTED
:  City CONTAINS "Belmont" END
_______________________________________________
20 RECORDS SELECTED
:  DIFFERENCE (PresentDate - LastDate) > 60 END
_______________________________________________
18 RECORDS SELECTED
:  HomePH CONTAINS "593-" END
_______________________________________________
11 RECORDS SELECTED
:  TYPE Name Address City HomePH Bus.PH Serum RHType END
_______________________________________________
     23
Name              Best, John S.
Address           1721 Hiawatha Dr.
City              Belmont
HomePH            593-1789
Bus.PH            488-9161
Serum             O
RHType            NEG
_______________________________________________
     55
Name              Smith, Roger D.
Address           275 Fifth Ave.
   .                 .
   .                 .
   .                 .
```

Figure 8

The field 'Serum' is tested to determine whether or not the blood-type is 0. The resulting 512 records are retained for further interrogation. 'RHType' is now tested and 28 records are found to exist with an RH factor NEG. A successive filtering on each set of resulting records obtained from each query results in the selection of 11 records. The purpose of this query was to find all those persons whose blood type was 0 NEG, who lived in Belmont, who hadn't given blood in the last 60 days, and whose home phone started with digits '593-'. The blood bank clerk can now print out the pertinent parameters for these individuals (phone no.), call them, and send them to the participating hospital to donate blood. This problem could be solved even further if the participating hospital had at its disposal a terminal which communicated with the local Blood Bank. Thus, the participating hospital could directly interrogate the Local Blood Donor Files and locate the proper individuals. This network concept of information retrieval has been under study and results are forthcoming. (3)

Querying a file in this manner seems highly advantageous as opposed to searching by hand through a card file. The front and backside of the donor information card are shown in Figure 9. Attempting to search through a few thousand of these donor cards for specific blood types could become a tedious and highly burdensome task. Having the ability to interrogate this information internally stored in the computer gives the clerk a freedom that can be applied to more meaningful tasks.

3.2.4 Management Reports:

Figure 10 is a description of the donor file as it now exists in prototype form. This is a standard report form which is printed out by DIRAC at the user's discretion. Note that the report gives

Date	STANFORD UNIVERSITY HOSPITAL BLOOD BANK STANFORD, CALIFORNIA 94305	Donor Number

Name ______ (LAST) ______ (FIRST) ______ (MIDDLE) Birth Date ______ Minor? ______ Sex ______

Address ______ City ______ Home Ph. ______

Date last donation ______ No. of donations in last 12 mos. ______ Bus. Ph. ______

Credit to ______

CBR ______

Successful bleeding ______ Unsuccessful ______ Why ______

Reaction ______

REMARKS: ______

M.D. ______ Phlebotomist: ______

Hemantigen: ______

Serum Grouping ______ Cell Grouping ______ RH Typing ______ Serology ______

Miscellaneous: ______

______ Technologist: ______

Expiration Date ______ Recipient ______ Hospital ______

HISTORY Donor: Do you have or have you ever been subject to the following diseases and conditions?

	NO	YES		NO	YES		NO	YES	PHYSICAL EXAM
Malaria			Rheumatic Fever			Dental Ext. (Last 72 Hrs.)			TEMP.
Malaria Area or Drugs			Kidney Disease			Under Physician Care			PULSE:
Jaundice, Liver Disease, Hepatitis			Infectious Mononucleosis			Medications			B P:
Jaundice or Hepatitis Contact			Swelling of Feet			Immunization or Injection *			WEIGHT:
Allergies			Convulsions, Fainting Spells			Surgery			ANT CUB FOSSA
Skin Eczema, Boils, Dermatitis			Unexplained weight loss			Hospitalization			HGB:
Ulcers			Shortness of Breath			Other illness within 1 Yr.			$CuSO_4$
Tuberculosis			Pain in Chest			Pregnancy			CYANMETH
Diabetes			Heart Condition			Transfusions			
Hazardous Occupation			Recent Cold or Sore Throat			Venereal Diseases			
Undulant or Prolonged Fevers			Persistent Cough			Food or Drink within 4 hrs.			
Bleeding Tendency			Any Malignancy			Tattoos			

* INCLUDES IMMUNIZATION TO HUMAN BLOOD CELL ANTIGENS

REMARKS:

STATEMENT AND RELEASE

I hereby certify that I have been questioned concerning the diseases, illnesses, symptoms and other physical conditions appearing on this card; that I have answered truthfully concerning each of the same and all questions asked me and that I have never suffered any of said diseases, illnesses or symptoms, except as indicated

I voluntarily donate my blood to the STANFORD UNIVERSITY HOSPITAL BLOOD BANK, STANFORD, CALIFORNIA 94035, to be used as decided by the said Blood Bank. I hereby release the Stanford University Hospital, its members, agents, representatives and employees from all claims and demands whatsoever which I have or may have by reason of the taking of blood to which I have submitted or am about to submit, and consequences resulting therefrom.

______
Donor's Signature

Accepted ______

Refused ______ Why ______ M.D. ______

Figure 9.

STANFORD UNIVERSITY
COMPUTATION CENTER

STATUS REPORT FOR FILE M200 (CLASS 1)
28/SEP/1970

LANGUAGE:
DIRAC-1A

DESCRIPTION : Blood Donors at Stanford Hospital

CREATION DATE : 18/SEP/1970
RECORD NOTATION : N
FIELD NOTATION : @F
DISPOSITION : PRIVATE
NO. OF RECORDS : 0

1 = TYPE
2 = MULTIPLICITY
3 = INDEXING
4 = CODE RESIDENCE
5 = CODE TYPE

FILE NAME : M200
FILE CREATED BY : WOLF
RECORD LENGTH : 512
NO.OF FIELDS : 33
LATEST UPDATE ON : 18/SEP/1970

F I E L D   I D E N T I F I C A T I O N ,   S T A T I S T I C S   A N D   V A L I D A T I O N   I N F O R M A T I O N

FLD	NAME	DESCRIPTION	STORAGE 1	2	3	4	5	VALIDATIONS NEC.	SIZE	SUB.	STATISTICS SIZE	DEC.	SUB.	EXISTENCE REC.	PCT
1	Name	Name of Donor	A	S	0			YES	0	0	0	0	0	0	0.0 %
2	BirthDate	Date of Birth	D	S	0				0	0	11	0	0	0	0.0 %
3	Minor	Under 21 Years of Age	A	S	0				0	0	0	0	0	0	0.0 %
4	Sex		A	S	0				0	0	0	0	0	0	0.0 %
5	Address	Address of Donor	A	S	0			YES	0	0	0	0	0	0	0.0 %
6	City		A	S	0			YES	0	0	0	0	0	0	0.0 %
7	HomePH		A	S	0			YES	0	0	0	0	0	0	0.0 %
8	BusinessPH	Business Phone	A	S	0				0	0	0	0	0	0	0.0 %
9	Date	Date of Donation	D	S	0			YES	0	0	11	0	0	0	0.0 %
10	Number	Blood Bank Number	I	S	0				0	0	0	0	0	0	0.0 %
11	LastDate	Date of Last Donation	D	S	0				0	0	11	0	0	0	0.0 %
12	Donations	No. of Donations in Last 12 Mos.	I	S	0				0	0	0	0	0	0	0.0 %
13	Credit	Credit to...	A	S	0				0	0	0	0	0	0	0.0 %
14	CBR		A	S	0				0	0	0	0	0	0	0.0 %
15	Bleed	Succ. & Unsucc. & Remarks	A	M	0				0	0	0	0	0	0	0.0 %
16	Reaction	Type of Reaction	A	S	0				0	0	0	0	0	0	0.0 %
17	Remarks	General Comments	A	S	0				0	0	0	0	0	0	0.0 %
18	M.D.	Doctor Present	A	S	0				0	0	0	0	0	0	0.0 %
19	Phlebotomist	Phlebotomist Present	A	S	0				0	0	0	0	0	0	0.0 %
20	Hemantigen		A	S	0				0	0	0	0	0	0	0.0 %
21	Serum	Serum Grouping	A	S	0				0	0	0	0	0	0	0.0 %
22	Cell	Cell Grouping	A	S	0				0	0	0	0	0	0	0.0 %
23	RhType	Rh Typing	A	S	0				0	0	0	0	0	0	0.0 %
24	Serology	Serology	A	S	0				0	0	0	0	0	0	0.0 %
25	Misc	Miscellaneous	A	S	0				0	0	0	0	0	0	0.0 %
26	ExpireDate	Expiration Date	A	S	0				0	0	0	0	0	0	0.0 %
27	Rechnologist		A	S	0				0	0	0	0	0	0	0.0 %
28	Recipient		A	S	0				0	0	0	0	0	0	0.0 %
29	Hospital		A	S	0				0	0	0	0	0	0	0.0 %
30	History	History of Donor	A	M	0				0	0	0	0	0	0	0.0 %
31	Physical	Physical Examination	A	M	0				0	0	0	0	0	0	0.0 %
32	Immunization	Inc. Imm. to Human Bld. Cell Ant	A	S	0				0	0	0	0	0	0	0.0 %
33	Comments	General Remarks	A	S	0				0	0	0	0	0	0	0.0 %

Figure 10.

all vital statistics of the file such as creation date, disposition of the file, who created the file, and information on its contents, as well as internal records information necessary for updating and querying the file. The user also has the option of writing other formats of reports which would be meaningful to him in his work and then querying the file and selecting the necessary records and displaying them through this medium. In other words, the user is limited only by his own imagination as to what data he can display through the various types of DIRAC interfaces available to him (Terminal, Scope, and WYLBUR).

3.2.5 Interactive Updating of Files:

The environment under which the user can update files or records within files is unique to DIRAC. The user has the option of either updating his file on a real time basis by terminal interaction or by working through an interface provided with the text editor, WYLBUR. Thus, if the user wishes he can catalogue his updates and drive them (update the file) through WYLBUR. The following example will demonstrate this procedure thoroughly: Each line beginning with a colon refers to the user's response. (Figure 11)

EXAMPLE OF UPDATE MODE IN DIRAC

```
NAME OF USER
   :    Smith
PLEASE TYPE EXECUTION MODE
   :  UPDATE                      WYLBUR data set containing
FILE IDENTIFICATION               original file records in
   :  X990                        working storage area.  See
FIXED FORMAT ?                    Figure 12.
   :  NONE
INTERFACE ?
   :  WYLBUR
MAXIMUM RECORD LENGTH :   326 WORDS.
THE FILE CONTAINS   4067 RECORDS.
UPDATE EXECUTION TERMINATED.
            .
            .
```

Figure 11

The user identifies himself and specifies the execution mode which he will be using. Identification for the file is M010 (Bone Marrow) and the format of the file is not fixed. DIRAC now prompts the user to designate whether updating will take place through WYLBUR or terminal interface. At this point if the user's data set (where the records for updating should be-- in WYLBUR) is empty, DIRAC will prompt the user and tell him that his data set is in fact empty and that he now is given the opportunity to bring those records into his data set either from the disk where they are stored or typing them now. (The latter alternative is, of course, an option, however, the reader should realize that this procedure would then be the same as terminal interaction on the part of the user.) The real advantage in using WYLBUR for updating procedures is that large amounts of data can be input into the file without the user having to sit at the terminal and type in each record. Of course, the record information has to be keypunched somewhere, but by placing it in a working storage area which WYLBUR provides, the user can edit his information prior to update making sure that most typing errors, etc. are corrected. Also, terminal time would be expensive, as well as CPU time if the user were to sit at the terminal and input the record information without going through the WYLBUR interface.

An example of the data and how it might appear in WYLBUR ready for input into DIRAC is shown in Figure 12.

EXAMPLE OF WYLBUR UPDATE FILE

```
NEW
@1 "MARKSON, JAMES"
@2 MALE
@3 "12-10-30"
@4 "E1A"
   HOLMAN
```

EXAMPLE OF WYLBUR UPDATE FILE (con't)

```
@6 "45-10-26"
@7 "A70-845"
@8 "7-24-70"
@9 YES
@10 "ACUTE MYELOGENOUS LEUKEMIA."
@14 "NONE SUBMITTED."
@15 "SCANTY SPECIMEN"
@17 "NONE SEEN."
@18 SPARSE
@19 SPARSE
@24 "MARROW INFILTRATED WITH MYELOMONOBLASTS; ? SLIGHT
INCREASE IN MATURITY OF BLASTS SINCE LAST MARROW"
@27 "NONE SUBMITTED."
@28 "ABRAHAM POTOLSKY, M.D."
@29 "GEORGE WALTUCH, M.D."
NEW
@1 "SIMM, ROBERTA A."
@2 FEMALE
@3 "4-21-36"
@4 "E1B"
@5 "J.WARTMAN"
@6 "34-60-87"
@7 "C70-765"
@8 "8-21-70"
@9 NO
    .
    .
    .
```

Figure 12

The user may wish to delete, revise, or replace a record in his file. Usually these revisions or changes are of a lesser magnitude than those of initially inputting -- updating -- the data into the file. i.e. using the WYLBUR interface. However, there may be times when the user might desire to use WYLBUR for this type of update. In the following example the terminal interaction is displayed only. Each line beginning with a colon refers to the user's response. (Figure 13)

INTERACTIVE UPDATING OF DIRAC FILE

```
   .
   .
FIXED FORMAT ?
   :    NONE
NEW OR OLD  ?
   :   OLD
____________
ACTION
   :   REPLACE $1$
____________
RECORD NO.     1 DEL.
____________
     :  @1 "John R. Smith"
     :  @2 "585 El Camino, Palo Alto, California"
         .
         .
         .
     :  NEW
____________
RECORD No.     4068
____________
     :  @1 "Larry S. Jones"
     :  @2 "45 Fifth Ave., New York"
         .
         .
         .
   :   END
MAXIMUM RECORD LENGTH :    326 WORDS.
THE FILE CONTAINS 4068 RECORDS.
UPDATE EXECUTION TERMINATED.
   .
   .
```

Figure 13

In the above example the user first designates that he is going to update 'OLD' records. Thus, he is replacing record number 1 with another record or maybe a similar record with part of the field information altered. He then designates the mode 'NEW' which means that he is going to input a new record into the file. At the present time there are 4067 records existing in the file. Therefore, DIRAC prompts the user for the 4068th record. The user types out the information and terminates his update with 'END'.

3.3 The Radio-Therapy Clinic

3.3.1 Statement of Problem:

The radio-therapy clinic at Stanford Hospital has undertaken a study on 400 patients who have incurred prostate cancer. These patients have been followed-up for approximately 3-8 years and the data which has been amassed in each patient file is considerable. It is not the object here to draw conclusion on the study per se (this is being done in the medical journals), but to present the reason for which the file was created and implemented under DIRAC. Specifically, to obtain results in this study, various subsets of record information are needed to be retrieved and then correlated either against time or against each other. To do this by hand would be useless. To write a dedicated computer program to retrieve these subsets and correlate them would solve the problem partially. However, this method limits the investigators to before-the-fact analysis, and provides no procedure by which statistics and correlations other than those that were predetermined and programmed can be obtained. Thus, the investigator is faced with a reprogramming task to answer those questions that were not anticipated. DIRAC provides a solution to this problem by allowing the user to browse his file and interrogate his file in a non-procedural mode not having to anticipate questions that might arise.

3.3.2 The Prostate Cancer File:

Figure 14 is a status report of the Prostate Cancer File. Note that this file contains a large number of fields due to the large amount of diagnostic information obtained from the patients. Note also that there are a large amount of 'DATE' fields, thus allowing

3.3 The Radio-Therapy Clinic

3.3.1 Statement of Problem:

The radio-therapy clinic at Stanford Hospital has undertaken a study on 400 patients who have incurred prostate cancer. These patients have been followed-up for approximately 3-8 years and the data which has been amassed in each patient file is considerable. It is not the object here to draw conclusion on the study per se (this is being done in the medical journals), but to present the reason for which the file was created and implemented under DIRAC. Specifically, to obtain results in this study, various subsets of record information are needed to be retrieved and then correlated either against time or against each other. To do this by hand would be useless. To write a dedicated computer program to retrieve these subsets and correlate them would solve the problem partially. However, this method limits the investigators to before-the-fact analysis, and provides no procedure by which statistics and correlations other than those that were predetermined and programmed can be obtained. Thus, the investigator is faced with a reprogramming task to answer those questions that were not anticipated. DIRAC provides a solution to this problem by allowing the user to browse his file and interrogate his file in a non-procedural mode not having to anticipate questions that might arise.

3.3.2 The Prostate Cancer File:

Figure 14 is a status report of the Prostate Cancer file. Note that this file contains a large number of fields due to the large amount of diagnostic information obtained from the patients. Note also that there are a large amount of 'DATE' fields. thus allowing

STANFORD UNIVERSITY COMPUTATION CENTER — STATUS REPORT FOR FILE M300 (CLASS 1) — 28/SEP/1970 — LANGUAGE: DIRAC-1

DESCRIPTION : File of Prostate Cancer Patients

CREATION DATE : 16/SEP/1970	1 = TYPE	FILE NAME : Prostate
RECORD NOTATION : SNS	2 = MULTIPLICITY	FILE CREATED BY : RAY
FIELD NOTATION : OF	3 = INDEXING	RECORD LENGTH : 1024
DISPOSITION : PUBLIC	4 = CODE RESIDENCE	NO. OF FIELDS : 68
NO. OF RECORDS : 139	5 = CODE TYPE	LATEST UPDATE ON : 16/SEP/1970

F I E L D   I D E N T I F I C A T I O N ,   S T A T I S T I C S   A N D   V A L I D A T I O N   I N F O R M A T I O N

FLD	NAME	DESCRIPTION	STORAGE 1	2	3	4	5	VALIDATIONS NEC.	SIZE	SUB.	STATISTICS SIZE	DEC.	SUB.	EXISTENCE REC.	PCT
1	Name	Patient's Name	A	S	0			YES	0	0	22	0	0	139	100.00%
2	SUH#	Stanford University Hosp. No.	A	S	0			YES	0	0	8	0	0	139	100.00%
3	Birthdate		D	S	0				0	0	11	0	0	139	100.00%
4	Race		A	S	0				0	0	9	0	0	138	99.28%
5	Physician	Referral Physician	A	S	0				0	0	29	0	0	139	100.00%
6	Address		A	S	0				0	0	56	0	0	137	98.56%
7	Phone		A	S	0				0	0	12	0	0	65	46.75%
8	DT1	First Visit DT	D	S	0			YES	0	0	11	0	0	139	100.00%
9	Symptoms	Symptoms & Signs	A	M	0				0	0	24	0	8	139	100.00%
10	Dur1	Dur of Symptoms & Signs	I	M	0				0	0	3	0	16	112	80.58%
11	Status	Hormone Status	A	M	0				0	0	20	0	7	139	100.00%
12	DT2	Hormone DTs	D	M	0				0	0	11	0	3	65	46.75%
13	Estrogen	Estrogen Type	A	S	0				0	0	2	0	0	135	97.12%
14	Family	Family History	A	M	0				0	0	20	0	5	102	73.38%
15	Past	Past Medical History	A	M	0				0	0	21	0	9	135	97 12%
16	DT3	IVP DT (Pre Rx)	D	M	0				0	0	11	0	2	121	87.05%
17	IVP	IVP - Roentgen Studies	A	M	0				0	0	26	0	2	139	100.00%
18	Chest	Chest X-Ray	A	M	0				0	0	32	0	13	139	100.00%
19	DT4	DTs - Chest, Liver, Brain	D	M	0				0	0	11	0	4	127	91.37%
20	Scans	Liver & Brain Scans	A	M	0				0	0	1	0	5	139	100.00%
21	Laboratory	Laboratory - Pre Rx	A	M	0				0	0	1	0	9	114	82.01%
22	Recent	Most Rec. Dx Evaluation	A	M	0				0	0	10	0	2	139	100.00%
23	Long	Long Bones	A	M	0				0	0	6	0	2	12	8.63%
24	Cervical	Cervical Spine	A	M	0				0	0	4	0	2	13	9.35%
25	Thoracic	Thoracic Spine	A	M	0				0	0	4	0	2	17	12.23%
26	Lumbar	Lumbar Spine	A	M	0				0	0	16	0	2	27	19.42%
27	Sacrum	Sacrum	A	M	0				0	0	4	0	2	23	16.55%
28	RInnominate	R. Innominate Bone	A	M	0				0	0	4	0	2	18	12.95%
29	LInnominate	L. Innominate Bone	A	M	0				0	0	4	0	2	19	13.67%
30	Pubis	Pubis	A	M	0				0	0	4	0	2	19	13.67%
31	Skull	Skull	A	M	0				0	0	4	0	2	19	13.67%
32	RIlium	R. Ilium	A	M	0				0	0	14	0	2	22	15.83%
33	LIlium	L. Ilium	A	M	0				0	0	10	0	2	14	10.07%
34	Ribs	Ribs	A	M	0				0	0	4	0	2	9	6.47%
35	Pelvis	Bony Pelvis-other	A	M	0				0	0	4	0	2	11	7.91%
36	Bones	Bones - other	A	M	0				0	0	16	0	2	8	5.76%
37	DT5	Diagnosis DT	D	S	0				0	0	11	0	0	134	96.40%
38	Method	Dx Method	A	M	0				0	0	10	0	8	139	100.00%
39	Cardiovascular	Physical Examination	A	M	0				0	0	11	0	3	139	100.00%
40	GU	GU-Prostate, Phys. Exam.	A	M	0				0	0	14	0	8	139	100.00%
41	Lymph	Lymph Nodes	A	M	0				0	0	8	0	4	139	100.00%
42	Leg	Leg Edema	A	M	0				0	0	2	0	3	139	100.00%
43	Staging	Staging Summary	A	M	0				0	0	8	0	5	139	100.00%
44	PreviousRx	Previous Rx	A	S	0				0	0	11	0	0	138	99.28%
45	Therapeutic	Therapeutic Plan	A	M	0				0	0	1	0	2	137	98.56%
46	Bladder	Bladder Infection	A	M	0				0	0	3	0	2	139	100.00%
47	Histopathology	Histopathology	A	M	0				0	0	32	0	11	139	100.00%
48	Radiation	Radiation	A	M	0				0	0	26	0	13	138	99.28%
49	Area	Area in CM	I	M	0				0	0	4	0	6	136	97.84%
50	DT6	DT of 1st Rx	D	S	0			YES	0	0	11	0	0	139	100.00%
51	DT7	DT of Last Rx	D	S	0				0	0	11	0	0	136	97.84%
52	Size	Size of Gland	A	M	0				0	0	1	0	2	136	97.84%
53	BonyMets	Bony Mets	A	M	0				0	0	2	0	4	35	25.18%
54	Dur2	Leg Edema Dur	A	M	0				0	0	2	0	3	6	4.32%
55	Complicatoin	Complications	A	M	0				0	0	65	0	13	135	97.12%
56	Dur3	Dur of Complications	I	M	0				0	0	3	0	12	80	57.55%
57	Impotency	Impotency	A	S	0				0	0	2	0	0	49	35.25%
58	DT8	DT of Last Follow Up	D	S	0				0	0	11	0	0	137	98.56%
59	Follow	Follow-Up Status	A	M	0				0	0	28	0	4	137	98.56%
60	Failure	First Failure	A	S	0				0	0	1	0	0	134	96.40%
61	DT9	DT of First Failure	D	S	0				0	0	11	0	0	45	32.37%
62	Site	Site	A	M	0				0	0	14	0	9	47	33.81%
63	DT10	DT of Liver Mets	D	S	0				0	0	11	0	0	1	0.72%
64	FailureRx	Rx of Failures	A	S	0				0	0	4	0	0	112	80.58%
65	Autops	Autopsy	A	M	0				0	0	35	0	7	137	98.56%
66	DT11	DT of Brain Mets	D	S	0				0	0	11	0	0	10	7.19%
67	Remarks	General Comments	A	S	0				0	0	130	0	0	46	33.09%
68	Cause	Cause of Death	A	S	0				0	0	63	0	0	36	25 90%

Figure 14.

the user (the doctors in this case) to interrogate the file with respect to date or time limitation. (this aspect is further explained in the next section). This file differs from the other files discussed in this paper for the following reasons:

(1) A large number of fields exists.

(2) A complex mixture of types is represented.

(3) The file is static, i.e., it does not change over time.

3.3.3 Interactive Creation & Revision of Files:

The environment under which the user can create or revise files or records within files is unique to DIRAC. A typical example of the error recovery and revision procedures that DIRAC allows the user during the creation phase is as follows: (Figs. 15 & 16)

EXAMPLE OF CREATE MODE REVISION PROCESS

```
1.  FILE NAME
    :    M300
2.  FILE "DESCRIPTION"
    :    "File of Prostate Cancer Patients"
3.  DISPOSITION (PUBLIC/PRIVATE)
    :    PUBLIC
         ENTER "AUTHORIZED UPDATE USERS"
    :    "Ray"
         ___________
         REVISIONS ?
    :    REVISE 3
3.  DISPOSITION (PUBLIC/PRIVATE)
    :    PRIVATE
         ENTER "AUTHORIZED QUERY USERS"
         "Ray Bagshaw"
    :    ENTER "AUTHORIZED UPDATE USERS"
         "Ray Bagshaw"
         ___________
         REVISIONS ?
    :    NONE
           .
           .
           .
```

Figure 15

DIRAC first asks the user to type a 'file name' and then a 'verbal description' of the file which he is creating -- (1) & (2). It then asks the user to designate the disposition of the file, whether it will be public or private -- (3). When this section has been completed the user may discover that he forgot to specify that the file be kept private and to place a certain name under "AUTHORIZED UPDATE USERS". Thus, he now has the option to revise a specific paragraph as shown in the illustration and to enter the name(s) he wishes.

EXAMPLE OF CREATE MODE ERROR RECOVERY PROCESS
(Con't from previous example)

```
4.  SPECIAL NOTATION FOR RECORD AND FIELD?
    :     NONE
5.  RECORD LENGTH ?
    :     1024
          REVISIONS ?
          ____________
    :     REVISE 5
5.  RECORD LENGTH ?
    :     511
          ***!
                              ______________ 53
                              LENGTH MUST BE A POWER OF 2, FROM
                              16 TO 2048 WORDS.

5.  RECORD LENGTH ?
    :     256
          ____________
          REVISIONS ?
    :     NONE
6.  SUPPLY NAME AND "DESCRIPTION" FOR ALL THE FIELDS

@1 ?
    :     Name "Name of Donor"
                .
                .
                .
@33?
    :     Comments "General Remarks"
@34?
    :     END
          ____________
          REVISIONS ?
    :     NONE
```

```
7.  SUPPLY DATA TYPE AND MULTIPLICITY
    :     ALPHA SINGLE @1 @2 @4 @5 @6 @1> . . . .
    :     DATE SINGLE @3 @8 @37 @50 @51 . . . . .
    :     INTEGER MULTIPLE @10 @49
    :     ALPHA MULTIPLE @15 @30 @33
    :     DATE MULTIPLE @12 @16 . . .

8.  VALIDATION
    :     @1 NECESSARY
    :     @2 NECESSARY
    :     @18 SUBFIELDS 13
    :     @50 SIZE 1
    :     @9 NECESSARY
    :     END
          __________
          REVISIONS ?
    :     NOPE
          ****!                  __________________51
                                 PROPER RESPONSE IS EITHER "NONE"
                                 OR "REVISE"
          __________
          REVISIONS ?
    :     NONE
_________
THE FILE HAS BEEN CREATED.
```

Figure 16

Each request by DIRAC has associated with it a specific error recovery message. If an error is made either in typing or responding with the wrong type of 'VALUES' then the recovery process is triggered such as is displayed in Figure 16.

It must be remembered that during the create phase of DIRAC field information for each record is not input into the file. This is done during the UPDATE phase. The CREATE phase is solely concerned with the structure of the file and the parameters which allow the user to define this structure.

4. CONCLUSIONS

The three applications discussed in this paper are exemplary of the manner by which a file-oriented language might be applied in Medicine. Even though the applications are very different in nature, they have the same basic structure. In one application -- Radio Therapy -- the files are used primarily for research purposes and afford no administrative or accounting function. The other two applications discussed -- Hematology and Blood Bank -- utilize interface techniques which provide inventory control as well as file management reports. These applications clearly point to the significance of a flexible and inexpensible file management tool in the medical environment.

REFERENCES

(1) Vallee, J. and Ludwig, H.R., "DIRAC: An Overview of an Interactive Retrieval Language", JOURNAL OF DATA MANAGEMENT, (Feb. 1971).

(2) Fredrickson, R., WYLBUR ON THE IBM 360/67, Stanford University Computation Center, Stanford, California (1969).

(3) Stengle, J.M., 'The National Blood Resource Program and Data-Processing Systems', TRANSFUSION 10,3 (May/June 1970).

(4) Hogman, C.F., and Ramgren, O. 'Computer System for Blood Transfusion Service' TRANSFUSION 10,3 (May/June 1970).

(5) Stewart, R.A., and Stewart, W.B., 'Computer Program for a Hospital Blood Bank', TRANSFUSION 9,2 (Mar/April 1969).

(6) Vallee, J., "Scientific Information Networks: A Case Study." Research Report No. 2. Information Systems Group, Stanford University Computation Center, (Oct. 1970).

(7) Vallee, J. and Ludwig, H.R., "The DIRAC Language: Concepts and Facilities." Research Report No. 1. Information Systems Group, Stanford University Computation Center, (May 1970).

PROGRESS TOWARDS A DIRECT-ACCESS HEMATOLOGY DATA-BASE: STANFORD'S EXPERIENCE WITH THE DIRAC LANGUAGE.

by

Paul L. Wolf, M.D.
(Stanford University Medical Hospital)

Herbert R. Ludwig, Ph.D.
(Stanford Computation Center)

Jacques F. Vallee, Ph.D.
(Stanford Computation Center)

SYNOPSIS-ABSTRACT:

A computer system has been implemented to permit the efficient storage, retrieval, dissemination and analysis of bone marrow examination reports at the Stanford University Hospital Hematology clinic. The data-base structure described in this article relies on an interactive data management language named DIRAC that allows physicians to create, update and interrogate private files without any programmer intervention. An example of actual interrogation of a sample file is reviewed in detail. The impact this system has had on the administration and work patterns of the Hematology clinic is described.

KEYWORDS: Bone-Marrow examinations, medical data retrieval, DIRAC Language, interactive information management, direct-access techniques.

I. INTRODUCTION

The adaptation of the physician's work patterns in an increasingly complex information environment presents many problems that have been recognized only recently, and for which no classical solution is available. The rapidly expanding volume of bone marrow examinations performed at Stanford Medical Center is a case in point. At the Stanford University Hospital Clinical Laboratory the number of examinations performed every year is such that it has been extremely difficult to store and retrieve pertinent information. Since it is essential to be able to analyze these data in order to perfect new methods of diagnosis and treatment of hematologic disease, we have long been interested in obtaining a simple (and, if possible, inexpensive) computerized storage and retrieval method. Early in 1970, this led us to request the assistance of the Information Systems group at the Computation Center, then in the process of testing a language prototype called DIRAC. (1)

A primary objective of our joint project was to produce a simple bone marrow report format showing all of the essential information necessary for meaningful identification and analysis, and to interface this report generator with an efficient file organization within the DIRAC environment.

Section II describes the problem of data-base implementation under these constraints. Section III is a brief overview of the DIRAC language, in particular, the modes under which it operates: CREATE, UPDATE, QUERY, and STATUS. Section V contains the conclusions and recommendations for future systems of this type.

II. DATA-BASE IMPLEMENTATION

Figure A is a description of the bone marrow report system at its inception. Note that there was no initial feedback to the ward. The reports were typed and given to the file clerk for storing. When information was to be obtained from the records a person had to search the file until the desired record(s) was obtained. This method proved to be unsuccessful since recovery of records and their distribution took too long.

Figure B describes the first phase in automating the bone marrow report system described in Figure A. The bone marrow diagnosis was typed into a working data set in computer storage. The reports could now be checked for accuracy and verified. Additions and corrections could be made with ease utilizing WYLBUR, the Stanford text editor (2). A report generator was written which would generate a formatted general report for each patient record (Fig.C). Copies of these reports are now distributed to different physicians and to the medical records section of the hospital.

This automatic report generating system is a significant improvement over the simplified file system described by Figure A. However, the problem of retrieving specific patient information when needed from the bone marrow file had not yet been solved . If a physician wished to have information regarding a specific patient or obtain statistical data concerning certain parameters within the records he had to search the file manually. This, of course, was time consuming and especially inefficient when the information was crucial to a

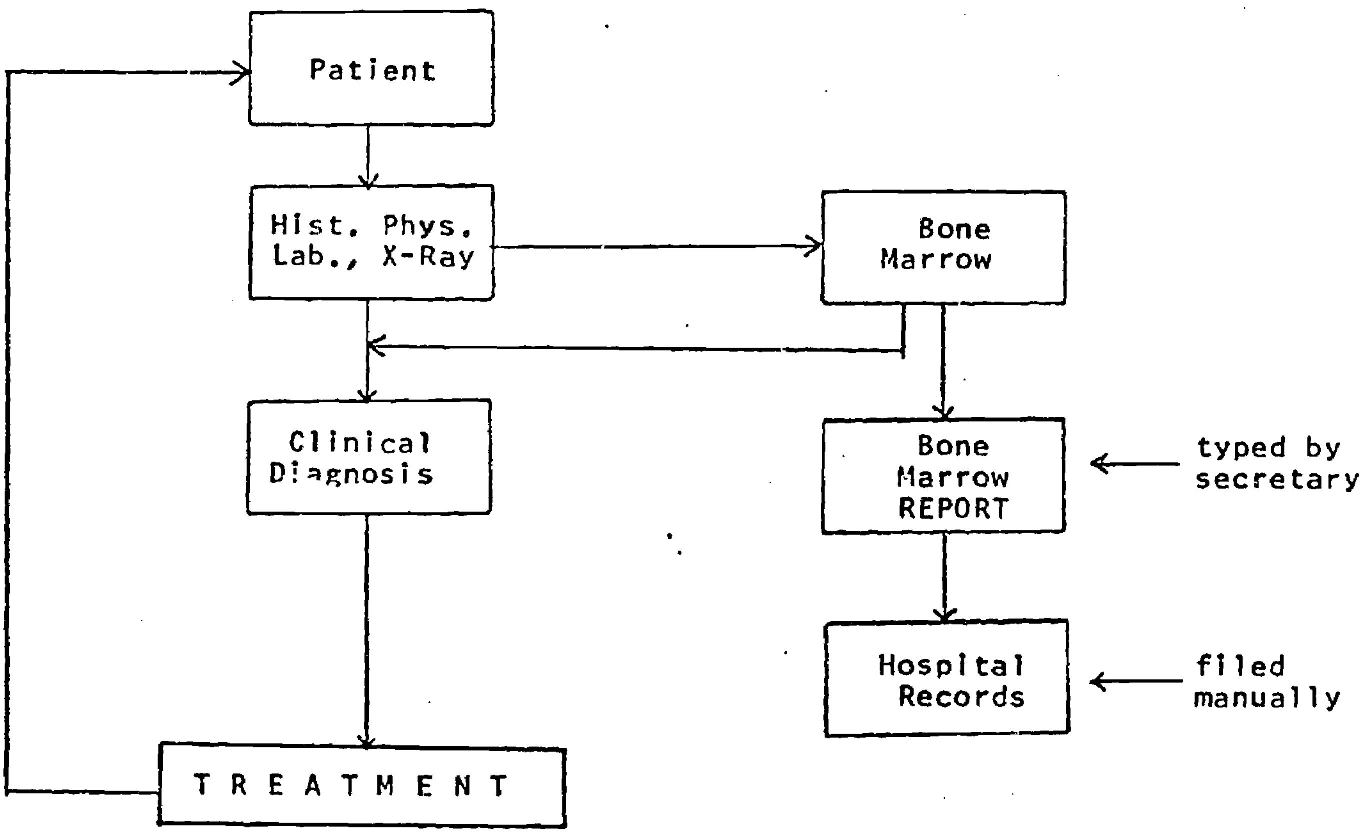
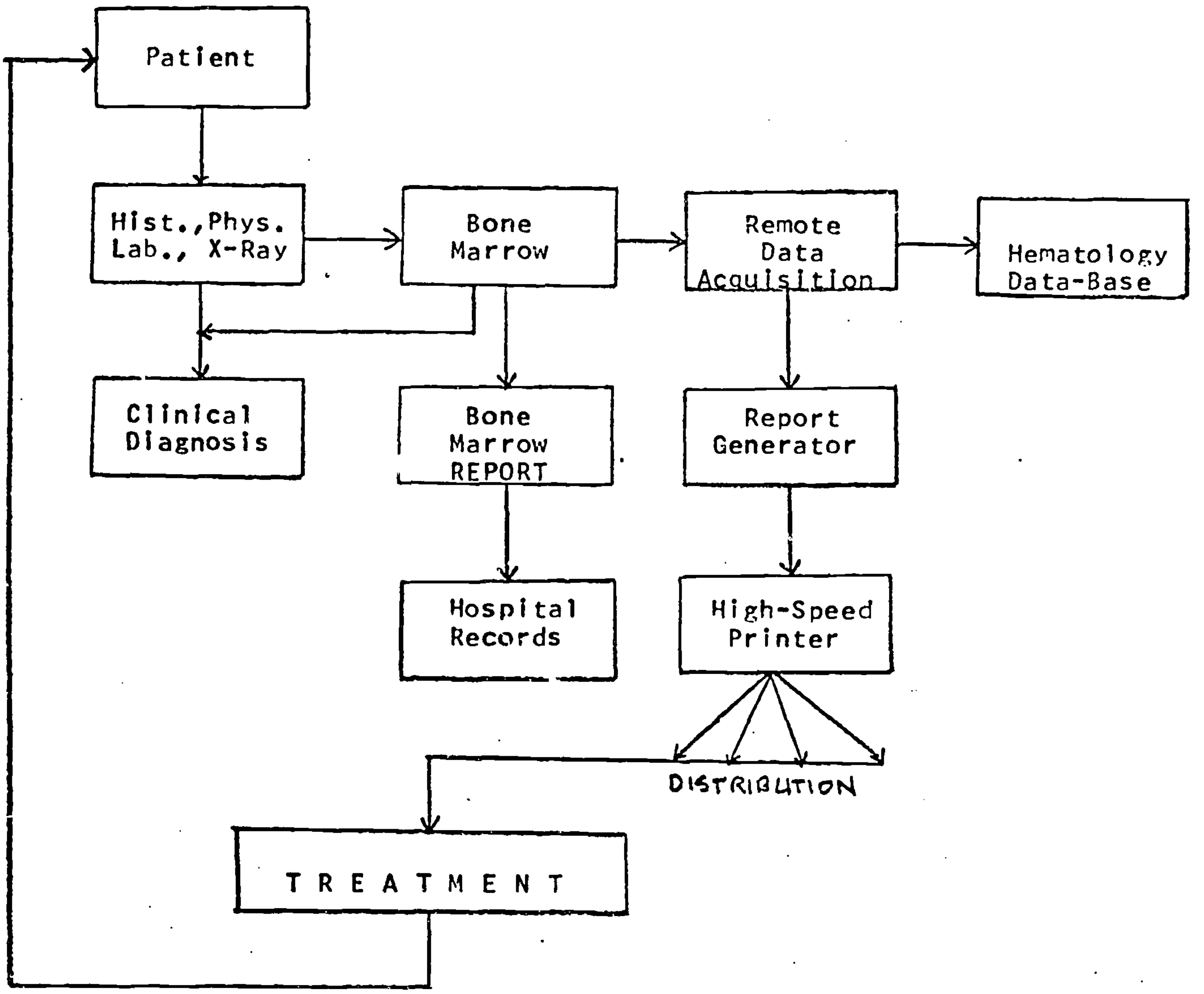

Patient
Hist. Phys. Lab., X-Ray
Bone Marrow
Clinical Diagnosis
Bone Marrow REPORT
typed by secretary
Hospital Records
filed manually
TREATMENT

Figure A.



Patient
Hist.,Phys. Lab., X-Ray
Bone Marrow
Remote Data Acquisition
Hematology Data-Base
Clinical Diagnosis
Bone Marrow REPORT
Report Generator
Hospital Records
High-Speed Printer
DISTRIBUTION
TREATMENT

Figure 8.

1. NAME: WATERS, SUSAN 2. SEX: FEMALE 3. BIRTHDATE: 12-31-15 4. WARD NO: CRC 5. REF. PHYS: MERIGAN	STANFORD UNIVERSITY MEDICAL CENTER **** BONE MARROW REPORT	6. MED. REC. NO: 32-03-40 7. BONE MARROW NO: 870-456 8. EXAM DATE: 4-23-70 9. PREVIOUS MARROW?

10. CLINICAL INFORMATION: MULTIPLE MYELOMA, ON TREATMENT.

PERIPHERAL SMEAR
****

11. RBC: UNREMARKABLE EXCEPT OCCASIONAL TARGET CELL.

12. PLATELETS: DECREASED SIGNIFICANTLY.

13. WBC: UNREMARKABLE

14. COMMENTS:

BONE MARROW ASPIRATE
****

15. QUALITY: ADEQUATE 16. MYELOID/ERYTHROID RATIO: 8:1
17. MEGAKARYOCYTES: DECREASED. MORPHOLOGY DISTURED. POOR GRANULARITY.

18. MYELOID ELEMENTS: ACTIVE. LEFT-SHIFTED.

19. ERYTHROID ELEMENTS: DECREASED. NORMOBLASTIC.

20. COMMENTS: PLASMACYTOSIS, MARKED. PLASMABLASTS PRESENT.

IRON STORES
****

21. AMOUNT: PRESENT/ADEQUATE 22. LOCATION: SOME IN RE CELLS
23. COMMENTS:

24. IMPRESSION
****

MULTIPLE MYELOMA. THROMBOCYTOPENIA WITH DECREASED MEGAKARYOCYTES IN BONE MARROW.

SPICULE SECTION
****

25. QUALITY: 26. CONFIRMS OTHER DATA: YES
27. COMMENTS: MULTIPLE MYELOMA

28. RALPH GEORGE, M.D.
(RESIDENT HEMATOLOGIST)

29. PAUL L. WOLF, M.D.
(DIRECTOR, CLINICAL LABORATORY)

task at hand. The clerical task of retrieving records for review and research analysis has been solved by the introduction of DIRAC, the first prototype in a family of information oriented languages designed primarily for application areas that demand flexible interaction with large files. It is non-procedural and demands no previous computer experience on the part of the user. (A more detailed analysis of DIRAC is given in the next section) By the introduction of DIRAC figure B is now transformed to Figure D. Note that there now is feedback to medical personnel on the diagnosis. This feedback is practically instantaneous and thus, alleviates the clerical burden of searching manually for records within the file.

Another advantage gained by introducing this data-management system is found in the that research and administrative procedures are greatly simplified. A whole file of bone marrow reports is now available for interrogation by medical researchers. DIRAC automatically does statistical analyses on pertinent records of the file when directed to do so. Interrogation by users of the file can also be displayed on a video unit or typed out at the user terminal, or can be printed on a high speed printer, whichever is more convenient based on the query and need of the user.

III. DIRAC - AN OVERVIEW

DIRAC (The name stands for 'DIRect ACcess', and indicates the five types of data one may wish to use -- Date, Integer, Real, Alphanumeric, and Coded) is an information retrieval language which gives the user the ability to operate on his

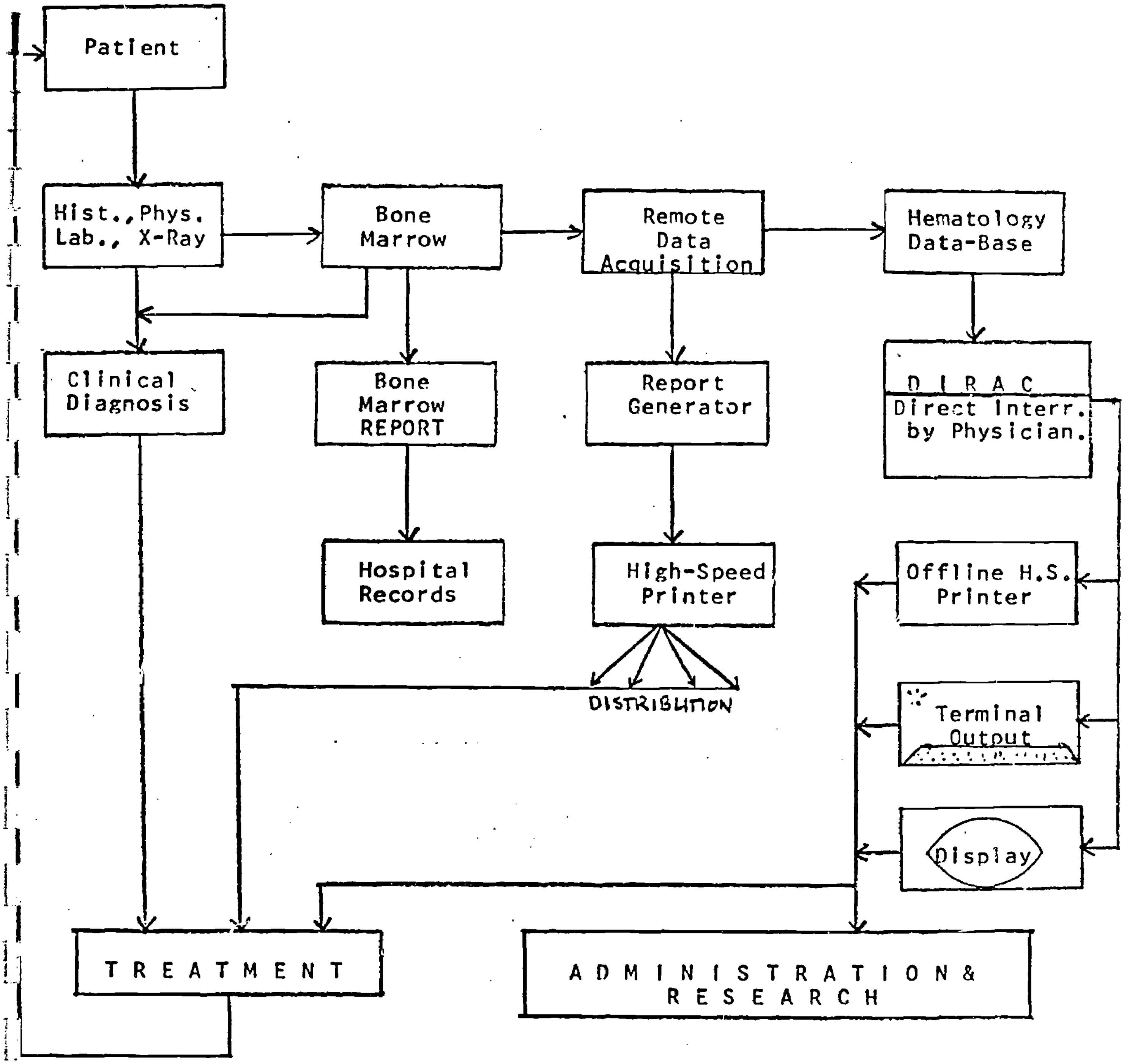

Patient
Hist., Phys.
Lab., X-Ray
Bone
Marrow
Remote
Data
Acquisition
Hematology
Data-Base
Clinical
Diagnosis
Bone
Marrow
REPORT
Report
Generator
D I R A C
Direct Interr.
by Physician.
Hospital
Records
High-Speed
Printer
Offline H.S.
Printer
DISTRIBUTION
Terminal
Output
Display
T R E A T M E N T
A D M I N I S T R A T I O N &
R E S E A R C H

Figure D.

private data-base under four modes: CREATE, UPDATE, QUERY, and STATUS.

(1) The CREATE mode allows the user to completely define the terminology and structure of his own file.

(2) The UPDATE mode allows such operations as adding, deleting or replacing records.

(3) The QUERY mode allows the user to obtain information about selected subsets of his file at any level of the records structure. The different commands through which a file may be queried are described in this article.

(4) The STATUS mode provides the user with an up-to-date status report for his particular file. Field identification, description of the fields, statistics and validation information are displayed in a standard report form.

CREATE Mode:

During the create phase of DIRAC the user describes and structures his file completely. DIRAC will lead the user through a series of steps, prompting him for all information necessary for the file creation. The user will specify the general organization of his data, whether the fields are singular or multiple, what limitations he would like to impose such as field size, maximum number of subfields, etc. DIRAC also lets the user specify those persons allowed to access the file and those persons who are allowed to alter the file during UPDATE, thus providing the widest scope of file security.

A description of the hematology file created at Stanford University Medical Center is given in Figure E. Note the field numbers, names, and the various validations used. Figure E is the standard report form generated by DIRAC for the user describing the status of his file. (The 'status' of a particular file is empty until that file is updated).

UPDATE Mode:

During UPDATE mode the user can add, delete, or replace records in his file. In other words he need not input all information at once into the file. The user may have a delay problem in obtaining some of the necessary field information, but may want to use the other information which he already has acquired. He can input this information into the file and at a later date complete the record. Figure E is a description of a subset of the Hematology file containing 235 records used here for the purpose of illustration. Note the number of records, the percentage of fields which have actual data in them, and the various validations.

QUERY Mode:

There are five fundamental commands utilized in QUERY mode:

(1) The SELECT command defines a subset of the primary data file by means of selection rules.

(2) The DISPLAY or TYPE command is used to either display information on a scope or type out information about the particular subset currently selec-

ted. When the volume of information is large, however, the display or type action can be triggered through the text editor and printing can now be done off-line on a high-speed printer.

(3) The RETAIN command is used to save the current subset. The resulting records are usually processed again by further selection until the search has been narrowed to the desired information.

(4) The RELEASE command completes the browsing facility by allowing re-initialization of the search to the entire file. In later versions of DIRAC this command will be combined with a subset designation to allow a hierarchy of embedded subsets rather than the simpler concept of a single filter, as currently implemented.

(5) The EXTRACT command, similar in form to the DISPLAY or TYPE command, transmits specified field information through a computational interface with FORTRAN. The user's own code can then operate along with DIRAC modules to achieve complex computations that are not possible within the basic file-oriented commands. As a default, the current implementation generates cross-tabulation of extracted fields and can be expanded to include standard post-processing for any particular application.

IV. DIRAC - AN EXAMPLE

If a physician wishes to interrogate the file that contains a number of bone marrow reports. He might want to obtain the following information:

(1) Determine all those patients whose bone marrow examinations were supervised by 'Wolf'.

(2) Retain this subset of the total set for further investigation.

(3) Determine from this subset all those persons whose gender is female.

(4) Determine from this subset all those persons who did not have a previous bone marrow taken.

(5) Determine from this subset all those persons whose 'Impression field' contains the word Plasmacytosis.

(6) Determine from this subset all those persons whose bone marrows were taken during the period March 10-19.

(7) Type out on the terminal the fields, 'Name', MarrowNo', 'Date', and 'Impression', for those records selected.

This procedure is described as it actually would be done using DIRAC in Figure F. Note that three records were finally selected which satisfied all those conditions specified by the interrogator.

V. CONCLUSIONS & RECOMMENDATIONS

An interactive storage, retrieval and dissemination system has been developed jointly by the Hematology Clinic and the Information Systems group at Stanford University. This system is based on a generalized data management language, named DIRAC, that allows file creation, updating and interrogation by users without previous computer experience.

Placing the system into operation at our facility required only the re-training of one secretary so that bone marrow reports could be entered directly from a terminal into main memory. Interrogation is done remotely by the physicians themselves as needed, without ANY programmer intervention whatsoever.

As a result of the use of DIRAC in this work, the physician is now provided with essential information at three levels: 1) The standard clinical diagnosis 2) An improved bone marrow report, permanently available for checking and editing, 3) A retrieval and display language capable of answering unpredictable queries in conversational mode.

This system is now well beyond the prototype stage and is applied routinely to the administration and analysis of our clinical information as new bone marrow examinations are entered into the data-base as well as records from previous years. The system is also of considerable value to the physician in:

(1) developing new methods of treatment and diagnosis;
(2) collecting and comparing data on a great variety of Hematologic diseases;
(3) evaluating effects of new treatment methods on Hematologic diseases and bone marrow and peripheral blood morphology;
(4) correlating bone marrow smears with bone marrow sections.

Placing him now in a position to make judgements based on up-to-date statistical data whose acquisition was too time-consuming, too expensive or too complex under previous methods.

```
Director CONTAINS WOLF END

227 RECORDS SELECTED
:    RETAIN
:    Sex CONTAINS FEMALE END

120 RECORDS SELECTED
:    Previous CONTAINS N END

 59 RECORDS SELECTED
:    Impression CONTAINS PLASMACYTOSIS END

  6 RECORDS SELECTED
:    Date CONTAINS "3-11-70" END

  3 RECORDS SELECTED
:    TYPE Name MarrowNo Date Impression END

 29
Name          SMITH,JUDY
MarrowNo      B70-345
Date          3-11-70
Impression    ROULEAUX FORMATION, TOXIC GRANULATION OF PMN,
              EOSINOPHILIA, AND PLASMACYTOSIS.

 32
Name          JONES, JOAN
MarrowNo      B70-453
Date          3-13-70
Impression    PERIPHERAL THROMBOCYTOPENIA AND ADEQUATE MEGAKARYO-
              CYTES SUGGEST PERIPHERAL REMOVAL MECHANISM.  MYELOID
              HYPERPLASIA.  EOSINOPHILIA, PLASMACYTOSIS, AND IN-
              CREASED R.E. CELL IRON.  ERYTHROGENESIS APPEARS
              DECREASED.

 45
Name          MARK, JANET
MarrowNo      B70-256
Date          3-16-70
Impression    NONDIAGNOSTIC MARROW.  PLASMACYTOSIS AND MILD
              EOSINOPHILIA.
```

Fig. F

STANFORD UNIVERSITY COMPUTATION CENTER — STATUS REPORT FOR FILE M010 (CLASS 1) — 1/JUL/1970 — LANGUAGE: DIRAC-1A

DESCRIPTION : BONE MARROW REPORT

CREATION DATE : 15/MAY/1970	1 = TYPE	FILE NAME : M010
RECORD NOTATION : V	2 = MULTIPLICITY	FILE CREATED BY : Wolf
FIELD NOTATION : @F	3 = INDEXING	RECORD LENGTH : 512
DISPOSITION : PRIVATE	4 = CODE RESIDENCE	NO.OF FIELDS : 29
NO. OF RECORDS : 235	5 = CODE TYPE	LATEST UPDATE ON : 15/JUN/1970

F I E L D   I D E N T I F I C A T I O N ,   S T A T I S T I C S   A N D   V A L I D A T I O N   I N F O R M A T I O N

			STORAGE					VALIDATIONS			STATISTICS			EXISTENCE	
FLD	NAME	DESCRIPTION	1	2	3	4	5	NEC.	SIZE	SUB.	SIZE	DEC.	SUB.	REC.	PCT
1	Name	Name of Patient	A	S	0			YES	0	0	24	0	0	235	100.00%
2	Sex		A	S	0				0	0	6	0	0	235	100.00%
3	Birth	Birthdate	A	S	0				0	0	9	0	0	234	99.57%
4	Ward	Ward No.	A	S	0				0	0	10	0	0	235	100.00%
5	Phys	Ref. Physician	A	S	0				0	0	13	0	0	218	92.77%
6	Record	Medical Record No.	A	S	0			YES	0	0	9	0	0	235	100.00%
7	MarrowNo	Bone Marrow No.	A	S	0			YES	0	0	9	0	0	235	100.00%
8	Date	Examination Date	A	S	0				0	0	8	0	0	235	100.00%
9	Previous	Previous Marrow	A	S	0				0	0	3	0	0	157	66.81%
10	Clinical	Clinical Information	A	S	0				0	0	144	0	0	219	93.19%
11	RBC	Peripheral Smear Section	A	S	0				0	0	363	0	0	208	88.51%
12	Platelets	Peripheral Smear Section	A	S	0				0	0	57	0	0	202	85.96%
13	WBC	Peripheral Smear Section	A	S	0				0	0	139	0	0	200	85.11%
14	C1	Peripheral Smear Section Comment	A	S	0				0	0	67	0	0	43	18.30%
15	Quality	Bone Marrow Aspirate Section	A	S	0				0	0	31	0	0	230	97.97%
16	Ratio	Myeloid/Erythroid --Bone Mar. Se	A	S	0				0	0	39	0	0	194	82.55%
17	Meg	Megakaryocytes -- Bone Marrow Se	A	S	0				0	0	73	0	0	217	92.34%
18	Myeloid	Myeloid Elements -- B.M.Sect.	A	S	0				0	0	144	0	0	217	92.34%
19	Erythroid	Erythroid Elements -- B.M.Sect.	A	S	0				0	0	96	0	0	217	92.34%
20	C2	Bone Marrow Section Comments	A	S	0				0	0	180	0	0	180	76.60%
21	Amount	Iron Stores Section	A	S	0				0	0	32	0	0	195	82.98%
22	Location	Iron Stores Section	A	S	0				0	0	17	0	0	116	49.36%
23	C3	Iron Stores Section Comments	A	S	0				0	0	56	0	0	39	16.60%
24	Impression	General Comments	A	S	0				0	0	268	0	0	228	97.02%
25	Qual	Spicule Section	A	S	0				0	0	159	0	0	82	34.89%
26	Confirms	Confirms Other Data -- Spicule S	A	S	0				0	0	3	0	0	115	48.94%
27	C4	Spicule Section Comments	A	S	0				0	0	222	0	0	130	55.32%
28	Hematologist		A	S	0			YES	0	0	19	0	0	235	100.00%
29	Director		A	S	0			YES	0	0	25	0	0	235	100.00%

REFERENCES:

1. THE DIRAC LANGUAGE: Concepts and Facilities, by J.F.Vallee and H.R.Ludwig. Research Report no.1, Information Systems Group, Stanford University Computation Center, May 1970.
2. WYLBUR ON THE IBM 360/67, by Roderick Fredrickson, Stanford University Computation Center.

INTERACTIVE COMPUTER MANAGEMENT OF CLINICAL DATA: A NEW APPROACH IN THE BLOOD BANK

By

Paul Wolf, M.D.

Herbert R. Ludwig, Ph.D.

Jacques Vallee, Ph.D.

1. INTRODUCTION

Recent experience at Stanford University confirms observations made at the regional and federal level concerning the urgent need for the application of computer techniques to the automation of Blood Banks. In view of the rapidly increasing volume of work handled by such a facility, and in view of the expanding list of services it must provide in a modern hospital, the design of computer-based systems for the blood banks is imperative; recent experiments in the United States (1,3) and in other countries (2) point to the feasibility of computerized transfusion services that provide both rapid identification of donors and blood inventory control, as well as other administrative services.

The current status of the work in this field was summarized in a recent TRANSFUSION editorial (1) that noted:

> So far donor characteristics and motivations have been surveyed and partly analyzed; the organization and management of donor groups have been studied; a forecasting procedure for short-term blood usage has been proposed; a heuristic allocation model has been developed; several hospital inventory control models have been studied; and a time-sharing inventory control tracking system with a single remote terminal is being field tested. Clearly, a great deal has been accomplished, but even when these projects are completed, much remains to be done.

The purpose of this paper is to present a computer approach to the problem that makes it possible to identify general functions and to support economically the activities of the Transfusion Center in a time-sharing environment.

The Stanford Blood Bank, headed by Dr. Paul Wolf, has initiated a research effort leading to a practical model of an automated tranfusion center prototype. This research effort has combined the advanced hardware and software technology available at Stanford with the Hospital's experience in Blood Bank design. It centers on a specific aspect of medical information systems design, but leads to results that will be generally applicable throughout the Clinical Laboratory.

The system under discussion here is unique in the sense that it makes use of a generalized language in the creation, updating and interrogation of donor, patient, and blood component files on the Stanford University Computation Center Facilities, and makes these files available to medical personnel through a simple yet powerful set of commands. Use of these techniques developed by the computation center's Information Systems Group have made it possible to conceive a system of considerably lower cost than currently available in other implementations across the nation.

Figure 1 shows the total units of whole blood used at Stanford Hospital through the years 1962 - present. Note the tendency of increased use of units even though there are periods of low peaks. Even with the advent of better transfusion techniques and devices, the Hospital envisions an ever increasing use in whole blood. Increased patient flow, due to the advent of new technology and greater capacity of the hospital, will add to this expansion. Thus, a control system that will govern the use and allocation of blood units throughout the hospital is a definite need. Along with this increase in whole blood use will come an increase in the paper load both for administrative and research purposes.

With the introduction of a non-procedural, file oriented system the above problems can be directly confronted and solved. (4) The "non-procedural" technique is characterized here by the ability for the blood bank clerk, the research assistant, medical technologist, or Doctor, to directly interrogate or access a specific file without programmer intervention.

2. SYSTEMS DESIGN

The objectives of the project are threefold:

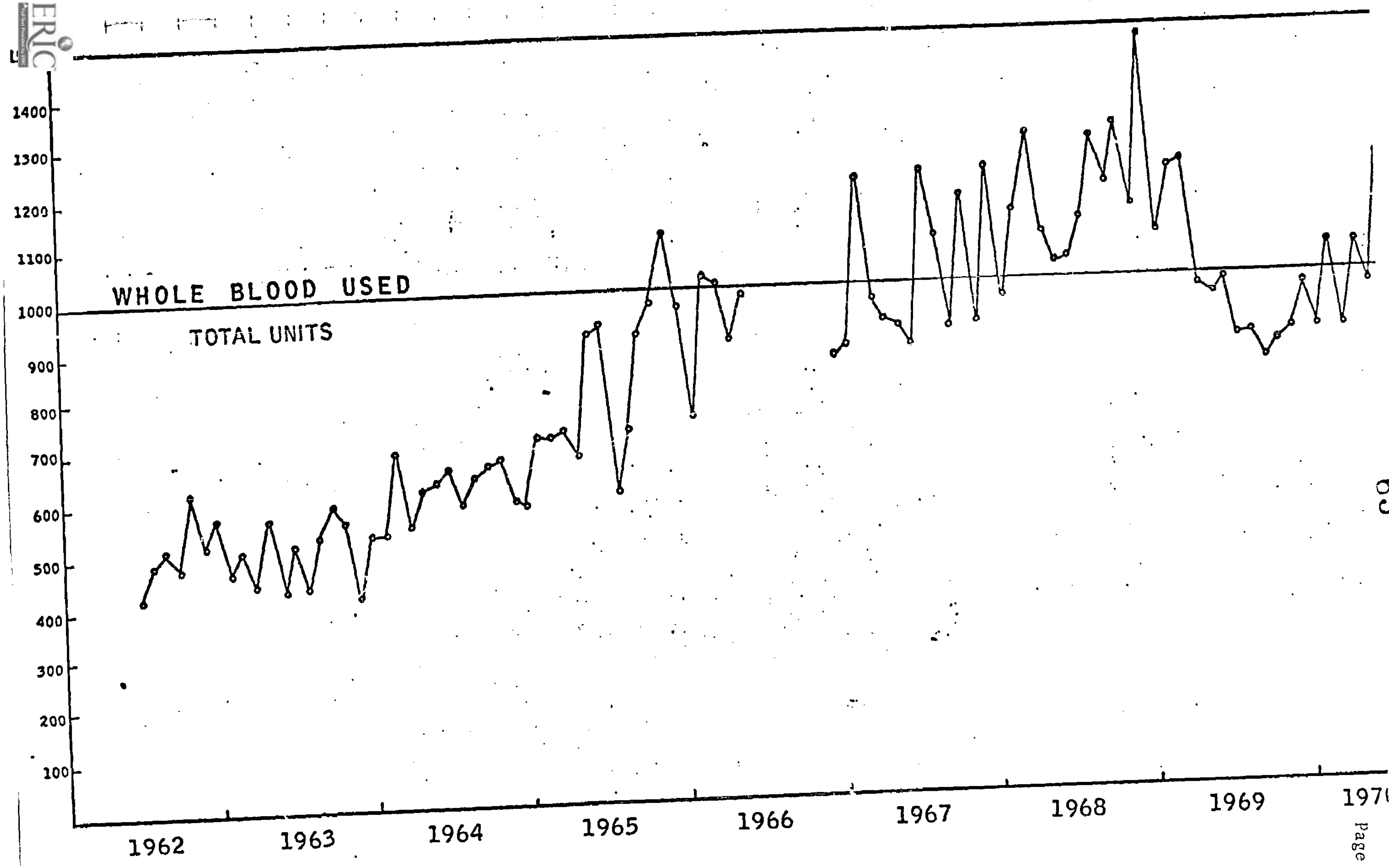
(1) To decrease the paper volume and paper flow through the Blood Bank.
(2) To Design an information system which will allow medical personnel to access the files directly.
(3) To Set up Inventory and Control systems for all units passing through the Blood Bank.

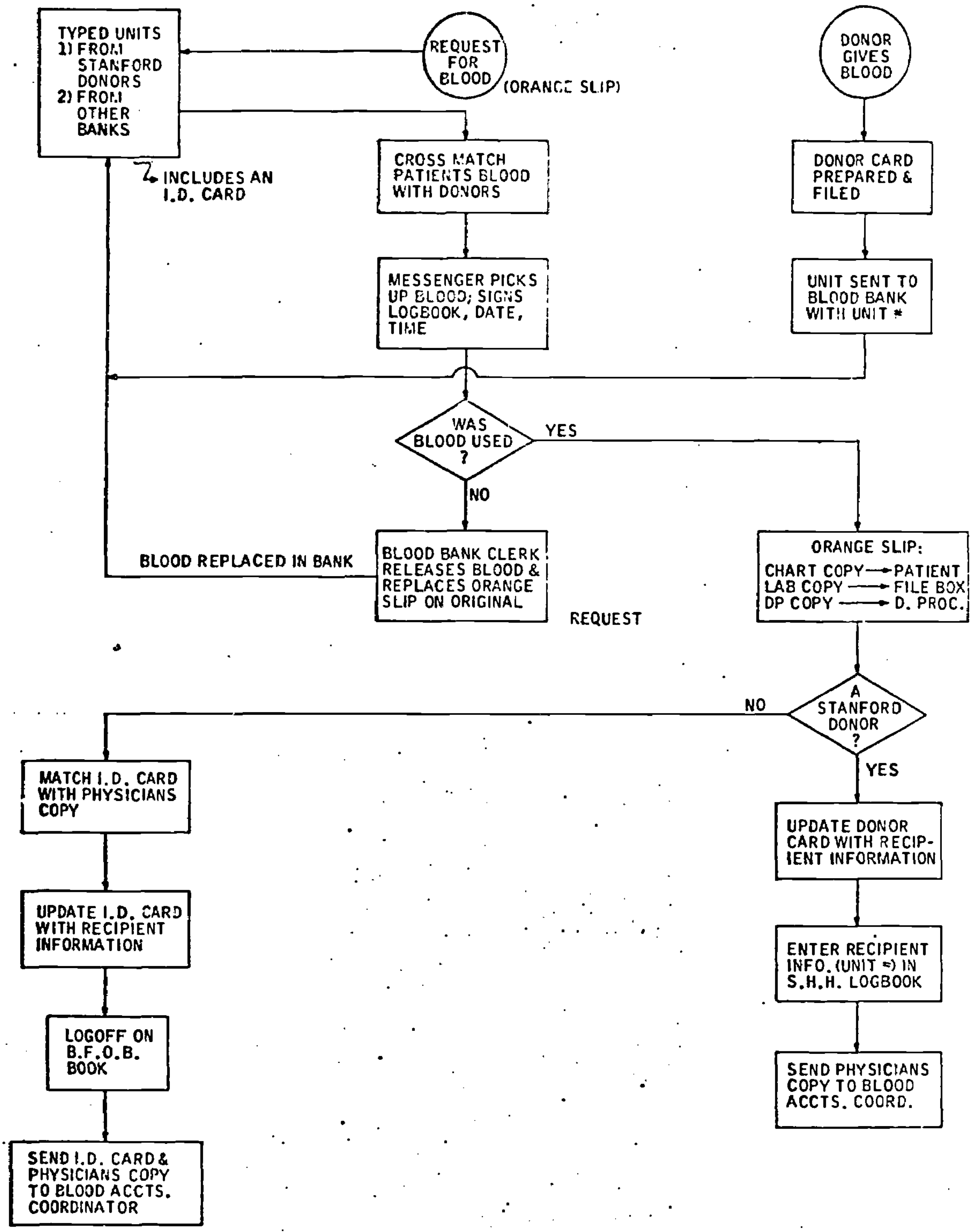
The system as it existed prior to the design phase is described in Figure 2. Note that many of the services performed by the clerks can easily be performed by the computer: The updating of the donor cards, the entering of recipient information in the S.U.H. logbook, etc. Duplication of files also exists between the Blood Accounts Coordinator and the Blood Bank Clerk. The main files in this system are the Stanford donor file, the recipient file, and the two logbooks, Stanford University Hospital logbook and the Blood-from- Other-Banks Logbook.

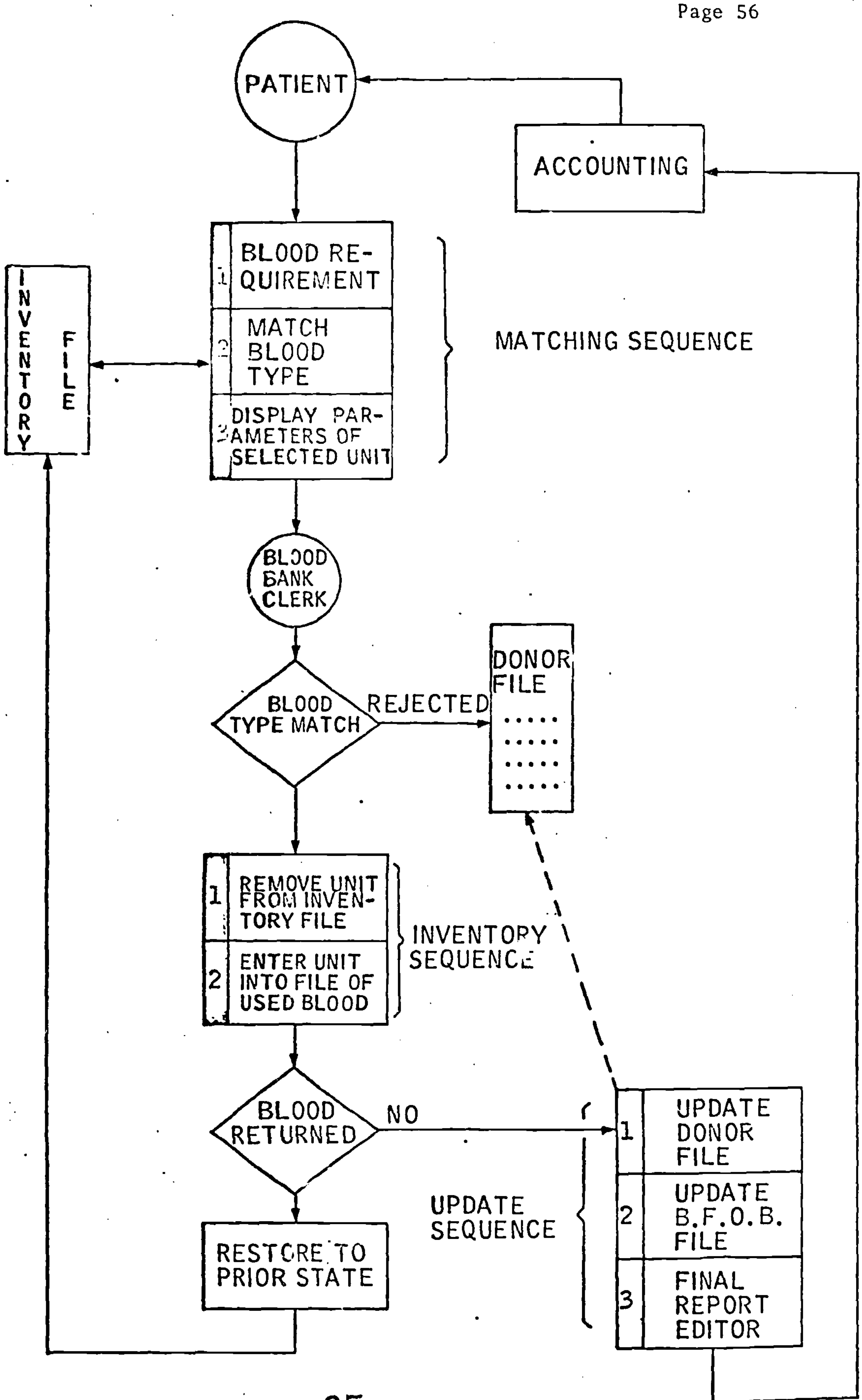
Figure 3 shows how the existing system can be redesigned to make use of an interactive data-base management language. Note the three sequences into which the system has been partitioned: the matching, inventory, and update sequence. The flow of paper is now administered internally by the blood bank clerk through the use of this management system.

Matching Sequence:

As blood units are requested a specific match of particular components required must be made by the blood bank clerk. The particular components needed are checked off or written on the "request-for-blood" form received by the blood bank clerk. Using


WHOLE BLOOD USED
TOTAL UNITS
1400
1300
1200
1100
1000
900
800
700
600
500
400
300
200
100
1962
1963
1964
1965
1966
1967
1968
1969
197


TYPED UNITS
1) FROM STANFORD DONORS
2) FROM OTHER BANKS
INCLUDES AN I.D. CARD
REQUEST FOR BLOOD
(ORANGE SLIP)
DONOR GIVES BLOOD
CROSS MATCH PATIENTS BLOOD WITH DONORS
DONOR CARD PREPARED & FILED
MESSENGER PICKS UP BLOOD; SIGNS LOGBOOK, DATE, TIME
UNIT SENT TO BLOOD BANK WITH UNIT #
WAS BLOOD USED ?
YES
NO
BLOOD REPLACED IN BANK
BLOOD BANK CLERK RELEASES BLOOD & REPLACES ORANGE SLIP ON ORIGINAL
REQUEST
ORANGE SLIP:
CHART COPY → PATIENT
LAB COPY → FILE BOX
DP COPY → D. PROC.
NO
A STANFORD DONOR ?
YES
MATCH I.D. CARD WITH PHYSICIANS COPY
UPDATE DONOR CARD WITH RECIPIENT INFORMATION
UPDATE I.D. CARD WITH RECIPIENT INFORMATION
ENTER RECIPIENT INFO. (UNIT #) IN S.H.H. LOGBOOK
LOGOFF ON B.F.O.B. BOOK
SEND PHYSICIANS COPY TO BLOOD ACCTS. COORD.
SEND I.D. CARD & PHYSICIANS COPY TO BLOOD ACCTS. COORDINATOR


PATIENT
ACCOUNTING
INVENTORY FILE
1 BLOOD RE-QUIREMENT
2 MATCH BLOOD TYPE
3 DISPLAY PAR-AMETERS OF SELECTED UNIT
MATCHING SEQUENCE
BLOOD BANK CLERK
BLOOD TYPE MATCH
REJECTED
DONOR FILE
1 REMOVE UNIT FROM INVEN-TORY FILE
2 ENTER UNIT INTO FILE OF USED BLOOD
INVENTORY SEQUENCE
BLOOD RETURNED
NO
RESTORE TO PRIOR STATE
UPDATE SEQUENCE
1 UPDATE DONOR FILE
2 UPDATE B.F.O.B. FILE
3 FINAL REPORT EDITOR

the file management system called DIRAC the blood bank clerk interrogates the inventory file to determine if a unit exists which satisfies the "request-for-blood" form. If a unit does exist the particular parameters in question are displayed either at the terminal or on a cathode-ray tube scope.

Another very important aspect of the blood bank is the Plasma program using the plasmapheresis procedure. Plasmapheresis is a technical procedure in which whole blood is obtained from a donor and is fractionated into multiple bag units. The red blood cells are given back to the donor and the plasma and its components are kept in the blood bank and fractionalized into different components. This program insures a greater supply of plasma and its components since donors can be bled more frequently and are not depleted of their red cells. A tissue typing program has been established in the Blood Bank to specifically match donors and recipients with regard to platelets and Cryoprecipitates. The computer can identify specifically matched donors for recipients who need repeated transfusions of platelets and Cryoprecipitates. Thus, chances for survival in the recipient are greatly increased by this type of program.

Inventory Sequence:

If the parameters displayed match the "request-for-blood" form the blood bank clerk will instruct a medical technician to remove unit number XXXX from inventory (refrigerator). This action will be reflected in the removal of the unit from the inventory file maintained by DIRAC. The unit number in The unit number is now entered into the file of used blood and remains there until processed by the UPDATE sequence.

Those components of blood which appear in the inventory of the Stanford Blood Bank are as follows:

Whole Blood	Packed Red Blood Cells
Leukocyte-Poor Blood	Platelets
Single Donor Fresh Froz. Plasma	Single Donor Fresh Plasma
Single Donor Factor VIII-Rich Cryoprecipitate & AHG Conc.	Factor IX Complex
	Fibrinogen
Immune Serum Globulin	Rh (D) Immune Globulin (Rhogan)
Plasma Substitutes	

Update Sequence:

If the blood is used (transfused) the blood bank clerk will update the donor file with information concerning the recipient of the blood. If the blood which was transfused was received from another blood bank, then this file will be updated with the recipient information also. A report will be generated which will be sent to the participating blood bank and also internal reports will be generated for administrative purposes. If the blood was not outdated the blood bank clerk restores the inventory file to prior state.

3. FILE MANAGEMENT UNDER DIRAC

Background:

The Language described here is the first prototype in a family of information oriented languages studies at the Stanford Computation Center. It is non-procedural and demands no previous computer experience on the part of the user. It allows creation, updating, bookkeeping operations, and the querying of data files in conversational mode under a time-sharing monitor on the IBM 360/67. It interfaces with the Stanford Text Editor, WYLBUR (4), and with the user's own FORTRAN code when complex computations or management reports on the contents or status of the files are required. (For a more detailed description of the DIRAC language see reference (5)).

Conversational Interrogation:

Assume that an emergency has presented itself. We are called upon to locate donors with O NEG blood type. The problem for the blood bank is to find these donors and get them into the hospital as fast as possible. A clerk will immediately establish communication with the computer (at night, a terminal located in a doctor's home can be used for this purpose) and the interrogation of the donor file under the DIRAC system takes place as follows: (where every line beginning by a colon is typed by the user)

EXAMPLE OF CONVERSATIONAL INTERROGATION USING DIRAC:

```
: Serum CONTAINS O END
________________________________________
512 RECORDS SELECTED

: RETAIN

: RHType CONTAINS NEG END
________________________________________
28 RECORDS SELECTED

: City CONTAINS "Palo Alto" END
________________________________________
20 RECORDS SELECTED

: DIFFERENCE (PresentDate - LastDate) > 60 END
________________________________________
18 RECORDS SELECTED

: HomePH CONTAINS "326-" END
________________________________________
11 RECORDS SELECTED

: TYPE Name Address City HomePH Bus.PH Serum RHType END
```

23	
Name	Best, John S.
Address	1721 Hiawatha Dr.
City	Palo Alto
HomePH	326-1789
Bus.PH	488-9161
Serum	O
RHType	NEG
55	
Name	Smith, Roger D.
Address	275 Fifth Ave.
.	.
.	.
.	.

Figure 5

The field 'Serum' is tested to determine whether or not the blood type is O. The resulting 512 records are retained for further interrogation. 'RHType' is now tested and 28 records are found to exist with an RH factor NEG. A successive filtering on each set of resulting records obtained from each query results in the selection of 11 records. The purpose of this query was to find all those persons whose blood type was O NEG, who lived in Palo Alto, who had not given blood in the last 60 days, and whose home phone number started with digits '326-'. We can now print out the pertinent parameters for these individuals (phone number) and call them to the hospital to donate blood.

Querying a file in this manner seems highly advantageous as opposed to searching by hand through a card file. The front and backside of the donor information card are shown in Figure 6. Attempting to search through a few thousand of these donor cards for specific blood types could become a tedious and highly burdensome task. Having the ability to interrogate this information internally stored in the computer gives the clerk a freedom that can be applied to more meaningful tasks.

Management Reports:

Figure 7 is a description of the donor file as it now exists in prototype form. This is a standard report form which is printed out by DIRAC at the user's discretion. Note that the report gives all vital statistics of the file such as creation date, disposition of the file, who created the file, and information on its contents, as well as internal records information necessary for updating and querying the file. The user also has the option of writing other formats of reports which would be meaningful to him in his work and then querying the file and selecting the necessary records and displaying them through this medium. In other words, the user is limited in the data he can display through the various types of C interfaces available to him. (Terminal, Scope, and WYLBUR)

Date	STANFORD UNIVERSITY HOSPITAL BLOOD BANK STANFORD, CALIFORNIA 94305	Donor Number

Name ____________ (LAST) (FIRST) (MIDDLE) Birth Date ________ Minor? ________ Sex ________

Address ____________ City ____________ Home Ph. ____________

Date last donation ____________ No. of donations in last 12 mos. ________ Bus. Ph. ________

Credit to ____________

CBR ____________

Successful bleeding ________ Unsuccessful ________ Why ____________

Reaction ____________

REMARKS: ____________

M.D. ____________ Phlebotomist: ____________

Hemantigen: ____________

Serum Grouping ________ Cell Grouping ________ RH Typing ________ Serology ________

Miscellaneous: ____________

Technologist: ____________

Expiration Date ________ Recipient ____________ Hospital ________

HISTORY Donor: Do you have or have you ever been subject to the following diseases and conditions?

	NO	YES		NO	YES		NO	YES	PHYSICAL EXAM
Malaria			Rheumatic Fever			Dental Ext. (Last 72 Hrs)			TEMP.
Malaria Area or Drugs			Kidney Disease			Under Physician Care			PULSE:
Jaundice, Liver Disease, Hepatitis			Infectious Mononucleosis			Medications			B.P.:
Jaundice or Hepatitis Contact			Swelling of Feet			Immunization or Injection *			WEIGHT:
Allergies			Convulsions, Fainting Spells			Surgery			ANT. CUB. FOSSA
Skin Eczema, Boils, Dermatitis			Unexplained weight loss			Hospitalization			HGB:
Ulcers			Shortness of Breath			Other illness within 1 Yr.			$CuSO_4$
Tuberculosis			Pain in Chest			Pregnancy			CYANMETH
Diabetes			Heart Condition			Transfusions			
Hazardous Occupation			Recent Cold or Sore Throat			Venereal Diseases			
Undulant or Prolonged Fevers			Persistent Cough			Food or Drink within 4 hrs.			
Bleeding Tendency			Any Malignancy			Tattoos			

* INCLUDES IMMUNIZATION TO HUMAN BLOOD CELL ANTIGENS

REMARKS:

STATEMENT AND RELEASE

I hereby certify that I have been questioned concerning the diseases, illnesses, symptoms and other physical conditions appearing on this card; that I have answered truthfully concerning each of the same and all questions asked me and that I have never suffered any of said diseases, illnesses or symptoms, except as indicated.

I voluntarily donate my blood to the STANFORD UNIVERSITY HOSPITAL BLOOD BANK, STANFORD, CALIFORNIA 94035, to be used as decided by the said Blood Bank. I hereby release the Stanford University Hospital, its members, agents, representatives and employees from all claims and demands whatsoever which I have or may have by reason of the taking of blood to which I have submitted or am about to submit, and consequences resulting therefrom.

____________ Donor's Signature

Accepted ____________

Refused ________ Why ____________ M.D. ____________

Figure 9.

ORD UNIVERSITY
TATION CENTER

STATUS REPORT FOR FILE M200 (CLASS 1)
18/SEP/1970

LANGUAGE:
DIRAC-1A

DESCRIPTION : Blood Donors at Stanford Hospital

CREATION DATE : 18/SEP/1970
RECORD NOTATION : SNT
FIELD NOTATION : ar
DISPOSITION : PRIVATE
NO. OF RECORDS : 0

1 = TYPE
2 = MULTIPLICITY
3 = INDEXING
4 = CODE RESIDENCE
5 = CODE TYPE

FILE NAME : M200
FILE CREATED BY : WOLF
RECORD LENGTH : 512
NO.OF FIELDS : 33
LATEST UPDATE ON : 18/SEP/1970

F I E L D I D E N T I F I C A T I O N , S T A T I S T I C S A N D V A L I D A T I O N I N F O R M A T I O N

FLD	NAME	DESCRIPTION	STORAGE 1	2	3	4	5	VALIDATIONS NEC.	SIZE	SUB.	STATISTICS SIZE	DEC.	SUB.	EXISTENCE REC.	PCT
1	Name	Name of Donor	A	S	0			YES	0	0	0	0	0	0	0.0 %
2	BirthDate	Date of Birth	D	S	0				0	0	11	0	0	0	0.0 %
3	Minor	Under 21 Years of Age	A	S	0				0	0	0	0	0	0	0.0 %
4	Sex		A	S	0				0	0	0	0	0	0	0.0 %
5	Address	Address of Donor	A	S	0			YES	0	0	0	0	0	0	0.0 %
6	City		A	S	0			YES	0	0	0	0	0	0	0.0 %
7	HomePH		A	S	0			YES	0	0	0	0	0	0	0.0 %
8	BusinessPH	Business Phone	A	S	0				0	0	0	0	0	0	0.0 %
9	Date	Date of Donation	D	S	0			YES	0	0	11	0	0	0	0.0 %
10	Number	Blood Bank Number	I	S	0				0	0	0	0	0	0	0.0 %
11	LastDate	Date of Last Donation	D	S	0				0	0	11	0	0	0	0.0 %
12	Donations	No. of Donations in Last 12 Mos.	I	S	0				0	0	0	0	0	0	0.0 %
13	Credit	Credit to...	A	S	0				0	0	0	0	0	0	0.0 %
14	CBR		A	S	0				0	0	0	0	0	0	0.0 %
15	Bleed	Succ. & Unsucc. & Remarks	A	M	0				0	0	0	0	0	0	0.0 %
16	Reaction	Type of Reaction	A	S	0				0	0	0	0	0	0	0.0 %
17	Remarks	General Comments	A	S	0				0	0	0	0	0	0	0.0 %
18	M.D.	Doctor Present	A	S	0				0	0	0	0	0	0	0.0 %
19	Phlebotomist	Phlebotomist Present	A	S	0				0	0	0	0	0	0	0.0 %
20	Hemantigen		A	S	0				0	0	0	0	0	0	0.0 %
21	Serum	Serum Grouping	A	S	0				0	0	0	0	0	0	0.0 %
22	Cell	Cell Grouping	A	S	0				0	0	0	0	0	0	0.0 %
23	RhType	Rh Typing	A	S	0				0	0	0	0	0	0	0.0 %
24	Serology	Serology	A	S	0				0	0	0	0	0	0	0.0 %
25	Misc	Miscellaneous	A	S	0				0	0	0	0	0	0	0.0 %
26	ExpireDate	Expiration Date	A	S	0				0	0	0	0	0	0	0.0 %
27	Technologist		A	S	0				0	0	0	0	0	0	0.0 %
28	Recipient		A	S	0				0	0	0	0	0	0	0.0 %
29	Hospital		A	S	0				0	0	0	0	0	0	0.0 %
30	History	History of Donor	A	M	0				0	0	0	0	0	0	0.0 %
31	Physical	Physical Examination	A	M	0				0	0	0	0	0	0	0.0 %
32	Immunization	Inc. Imm. to Human Bld. Cell Ant	A	S	0				0	0	0	0	0	0	0.0 %
33	Comments	General Remarks	A	S	0				0	0	0	0	0	0	0.0 %

Interactive Creation & Revision of Files:

The environment under which the user can create or revise files or records within files is unique to DIRAC. A typical example of the error recovery and revision procedures that DIRAC allowss the during the creation phase is as follows: (Figs. 8 & 9)

EXAMPLE OF CREATE MODE REVISION PROCESS

```
1.  FILE NAME
    :     X990
2.  FILE "DESCRIPTION"
    :     "Blood Donors at Stanford Hospital"
3.  DISPOSITION (PUBLIC/PRIVATE)
    :     PUBLIC
          ENTER "AUTHORIZED UPDATE USERS"
    :     "Wolf"

          REVISIONS ?
    :     REVISE 3
3.  DISPOSITION (PUBLIC/PRIVATE)
    :     PRIVATE
          ENTER "AUTHORIZED QUERY USERS"
          "Wolf, Downey"
    :     ENTER "AUTHORIZED UPDATE USERS"
          "Wolf Downey"

          REVISIONS ?
    :     NONE
            .
            .
```

Figure 8

DIRAC first asks the user to designate the name of the file he is creating and to describe it. It then asks the user to designate the disposition of the file, whether it will be public or private. When this section has been completed the user may discover that he forgot to specify that the file be kept private and to place a certain name under "AUTHORIZED UPDATE USERS". Thus, he now has the option to revise a specific paragraph as shown in the illustration and to enter the name(s) he wishes.

The DIRAC processor contains as a subset a powerful debugging and recovery system that are not normally apparent to, but remain at the service of, the on-line user. If an error is made either in typing or responding with an inappropriate command or an erroneous value, then the recovery process is triggered as is displayed in Figure 9.

It must be remembered that during the create phase of DIRAC, field information for each record is not input into the file. This is done during the UPDATE phase. The CREATE phase is solely concerned with the structure of the file and the parameters which allow the user to define this structure.

EXAMPLE OF CREATE MODE ERROR RECOVERY PROCESS IN DIRAC-1

(Con't from previous example)

```
4.  SPECIAL NOTATION FOR RECORD AND FIELD?
    :     NONE
5.  RECORD LENGTH ?
    :     512
          REVISIONS ?
          ____________
    :     REVISE 5
5.  RECORD LENGTH ?
    :     257
          ***!
                                        ______________53
                                        LENGTH MUST BE A POWER OF 2, FROM
                                        16 TO 2048 WORDS.
5.  RECORD LENGTH ?
    :     256
          ____________
          REVISIONS ?
    :     NONE
6.  SUPPLY NAME AND "DESCRIPTION" FOR ALL THE FIELDS

@1 ?
    :     Name "Name of Donor"
                   .
                   .
@33?
    :     Comments "General Remarks"
@34?
    :    END
          ____________
          REVISIONS ?
          NONE
7.  SUPPLY DATA TYPE AND MULTIPLICITY
    :     ALPHA SINGLE @1 @3 @4 @5 @6 @7 ..........
    :     DATE SINGLE @2 @11 @9
    :     INTEGER SINGLE @10 @12
    :     ALPHA MULTIPLE @15 @30 @33
8.  VALIDATION
    :     @1 NECESSARY
    :     @9 NECESSARY
    :     @30 SUBFIELDS 30
    :     @10 SIZE 5
    :     END
          ____________
          REVISIONS ?
    :     NOPE
          ***!
                                        PROPER RESPONSE IS EITHER "NONE"
                                        OR "REVISE"
          ____________
          REVISIONS ?
    :     NONE
__________
THE FILE HAS BEEN CREATED.
```

Figure 9

CONCLUSION

The problem of transfusion center automation affords medical personnel an opportunity to apply and evaluate the new software techniques that are now coming into existence. It should be recalled that computers have been designed with numerical calculation, rather than information processing in mind. The field of software design has only begun to emerge from this mathematical orientation by making available data-management languages that are truly user-oriented. When combined with the flexibility of the time-sharing technique, these languages are seen to have special significance in the long-neglected field of medical information processing.

REFERENCES

(1) Stengle, J.M., 'The National Blood Resource Program and Data-Processing Systems', TRANSFUSION 10,3 (May/June 1970).

(2) Hogman, C.F., and Ramgren, O. 'Computer System for Blood Transfusion Service' TRANSFUSION 10,3 (May/June 1970).

(3) Stewart, R.A., and Stewart, W.B., 'Computer Program for a Hospital Blood Bank', TRANSFUSION 9,2 (Mar/April 1969).

(4) Vallee, J. and Ludwig, H.R., "Dirac: An Overview of an Interactive Retrieval Language", EDUCATIONAL NEWSLETTER, (Sept. 1970).

(5) Fredrickson, R., WYLBUR ON THE IBM 360/67, Stanford University Computation Center, Stanford, California (1969).

(6) Vallee, J. and Ludwig, H.R., "The DIRAC Language: Concepts and Facilities." Research Report No. 1. Information Systems Group, Stanford University Computation Center, (May 1970).

FORMS DESIGN IN MEDICAL INFORMATION RETRIEVAL

by

Herbert R. Ludwig, Ph.D.

1. INTRODUCTION

The ability to interrogate files of information in a non-procedural mode has one distinct advantage over a procedural mode inquiry: the user need not anticipate all the questions he might want to ask of a specific file when the file is created. The user will, of course. have some idea of the questions he wishes to ask, but those questions which crop up unexpectedly can send the cost skyrocketing since they usually involve some redesign or reprogramming.

The design of the form used by the physician to collect his data must serve two purposes: first, it must be an easy form for the physician to fill out, and second, it must be set up in such a fashion to make keypunching a reliable task. Thus, the design of the form utilized for the task at hand is crucial and if the design is erroneous, then the data fed into the computer will be of no use, and retrieval will be futile.

Rather than discussing forms design in an abstract manner, it is of interest to consider a specific problem. Such a problem presented itself recently at Stanford Hospital when the Radio-Therapy Clinic began to use a computer in its analysis of 400 patients who have incurred prostate cancer. patients have been followed-up for approximately 3-8 years and the data which has been amassed in each patient file is considerable. It is not the object here to draw conclusion on the study per se (this is being done in the medical journals), but to present (1) the method by which the forms for extraction of data from patient record charts were re-designed, (2) how the file has been structured, and (3) the implementation of this prostate cancer file under a non-procedural information retrieval language called DIRAC. (1)

Specifically, to obtain results in this study, various subsets of record information must be retrieved and then correlated either against time or against each other. To do this by hand is out of the question. To write a dedicated computer program to retrieve these subsets and correlate them would solve the problem partially. However, as stated above, this method limits the investigators to before-the-fact analysis, and provides no procedure by which statistics and correlations other than those that were predetermined and programmed can be obtained. Thus, the investigator is faced with a reprogramming task to answer those questions that were not anticipated. A language prototype that provides a solution to this problem by allowing the user to browse through his file interactively has been tested in this environment.

2. The DIRAC Concept

2.1 Introduction

DIRAC (Date, Integer, Real, Alphanumeric, and Coded), is a file-oriented retrieval language developed by J. Vallee. During 1970, this language was used for a series of tests on the IBM 360/67 at the Stanford Computation Center. DIRAC allows the user to interact with his file in a time-share mode. The user can "converse" with his file in his own terminology after having "declared" it during creation. He also has the ability to perform catalogued interrogations, presenting a set of questions to the file which were predetermined. The former -- conversational mode -- [illegible]ows the user to ask questions of his file which had not been anticipated before. The latter -- catalogued mode -- allows him to rapidly process predetermined queries.

It is not the purpose in this article to describe the DIRAC language in its entirety. This has been done elsewhere. (2) However, a brief overview of the language and its basic concepts is necessary for those readers who are unfamiliar with file-oriented languages. The introductory comments on DIRAC will, hopefully, set the framework by which forms design is discussed in the later sections of this paper. It must be remembered that with the advent of these new file-oriented languages, programmer intervention is minimized. The user must design his forms -- which leads to his file structure and design -- so that the data can be easily entered onto the form by the physician, easily keypunched by an operator, and acceptable to the retrieval system.

2.2 Modes of Operation

There are four modes of operation in DIRAC -- CREATE, UPDATE, QUERY, and STATUS. Each mode encompasses a certain set of commands which enable to user to apply a particular type of strategy on his file.

2.2.1 The CREATE Mode:

The CREATE mode allows the user to specify the structure of his file. In this mode the user does not input information into his file, he merely sets up the structure for the file. Among the declarations which the user is instructed to specify are the following:

File Name

File Description

File Status (PUBLIC or PRIVATE?)

Authorized QUERY Users
Authorized UPDATE Users

Maximum Record Length

Record & Field Delimiters

Field Name & Description

Type designation (Alpha, Integer, etc.)

Field specification (Single or Multiple?)

Validations:

Field Size
Number of Subfields
Necessary condition

The user is prompted for each one of the above. Once he has answered correctly to all prompts and is satisfied that no revisions should be made the file will have been created.

2.2.2 The UPDATE Mode:

The UPDATE mode allows the user to add, delete, or replace record or field information within his file. Thus, it is in this mode that the user "fills" his file up with information. The user has the option to either update his file from a terminal or through an interface with a text-editor. In the particular tests we describe, DIRAC was interfaced with an excellent text-editing language named WYLBUR, developed by J. Borgelt. (3) If the latter is used whole files can be updated by one single command. (This is known as a catalogued update procedure where all the data necessary for update is stored in some data set by the computer and is accessed by DIRAC upon the user's command.)

2.2.3 <u>The QUERY Mode:</u>

This mode allows the user to interrogate his file and retrieve selected records. There are six basic selection commands:

(1) The CONTAINS Command:
EXAMPLE: @1 CONTAINS "H R DAVIS" END
(Field 1 contains the text "H R Davis")

(2) >,<,=, (or combination)
EXAMPLE: @3> 'VALUE' END
(Field 3 is greater than some 'VALUE')

(3) RETAIN: Retains a selected set of records for further processing by selection commands. Each successive set of records selected is retained once this commands has been executed. Allows for a filtering down process.

(4) RELEASE: Releases selection process to whole file.

(5) TYPE or DISPLAY Command: Allows user to either print selected information at terminal, cathode ray tube scope, or on high speed printer.

(6) EXTRACT Command: Allows the user to extract field information from the whole set or subset of records in the file. The information is transferred to a user area where i can be processed by user-supplied statistical packages. programs through the use of the DIRAC interface capability.

2.2.4 The STATUS Mode:

This mode prints out the status of the user's file when needed. It describes in detail all file parameters, including file name, description, file creator, date created, etc. It also describes the file in detail giving field names and descriptions, whether each field is singular or multiple, alpha, real, integer, etc, and the various validations imposed by the user on his fields. (An example of a STATUS report is shown in section 4.1, Figure 6)

3. FORMS DESIGN FOR PROSTATE CANCER FILE

3.1 Initial Design

When the prostate cancer study was in its infancy the forms design aspect of the study did not appear to be critical to the physicians, who were going to have a dedicated procedural program (written in PL/1) to retrieve the answers and compile statistics. An initial form was designed which was thought suitable for what they had in mind. This form was quite lengthy and, therefore, only parts of it will be displayed in this paper in order to conserve space. However, the physicians also realized that if a procedure was programmed to obtain answers to specific questions, then unanticipated questions would go unanswered. As they could not possibly anticipate all the questions which they might want to ask of the file the initial project was abandoned and the Information Systems group, working on a DIRAC prototype, was contacted. The physicians realized that the data they were collecting for the study might not be in the correct form required by DIRAC. Therefore, we analyzed and redesigned the form so that it could be easily handled by the Physician collecting the data, the terminal operator who would input the data into a WYLBUR data set and DIRAC.

The following objectives were defined:

(1) Design the form so that the physician can easily understand and fill out the form.

(2) Design the form so that the operator can easily type the entries with the least amount of frustration and with a minimum amount of special instructions.

(3) Design the form so as not to mix the types of field data.

The first objective meant that the physician should be relieved of the time consuming task of skipping from page to page in a random fashion entering data. A logical sequence should be followed. The physician should only be required to mark off multiple choice entries. In only a few cases, where it was absolutely necessary, should he have to write text. Also an indication should be given to the physician that alerts him to specific field information that is 'NECESSARY' i.e., those fields which have to be included in the file, otherwise DIRAC would reject the record upon input.

The second objective was decided upon because a study of this importance could not have erroneous data. Therefore, the ease by which data could be transferred from the design form to a WYLBUR Data Set was a major consideration. Since most of the entries were marking off some type of code letter it would be very difficult to verify each entry. Some flagrant keypunch errors could probably be caught, however, for the most part, each separate form could not be re-checked against the keypunched data for accuracy. This task would just be too time-consuming. Aligning the entries of the form answers in such a manner so that the keypuncher's eyes would not have to skip around the page, either up or down, or in some zig-zag fashion was the ultimate goal.

It was felt that if a suitable combination of the above two objectives were obtained then the routine of extracting the data from patient records to a WYLBUR data set would be a relatively simple task. The process by which this was accomplished is shown in Figure 1, which is a systems flow-chart of the task. Note the feedback loop for field entries which were either punched wrong or written wrong by the physican filling out the form. This type of error recovery was accomplished by a processor (written in FORTRAN) that would edit the keypunched data and manipulate the data into the correct form and sequence for input into DIRAC. If an error occurred (a recognizer was written into this processor to do this) which was not an obvious keypunching error the original design form would be re-routed back to the physician who filled it out for rectification. Of course, the whole form would not be keypunched again. Only the portion (usually just a field entry) would be corrected by using the WYLBUR text-editing commands. Thus, this simple cycle would lead to reliable data for input into DIRAC.

SYSTEMS FLOW CHART FOR PROSTATE CANCER FORM

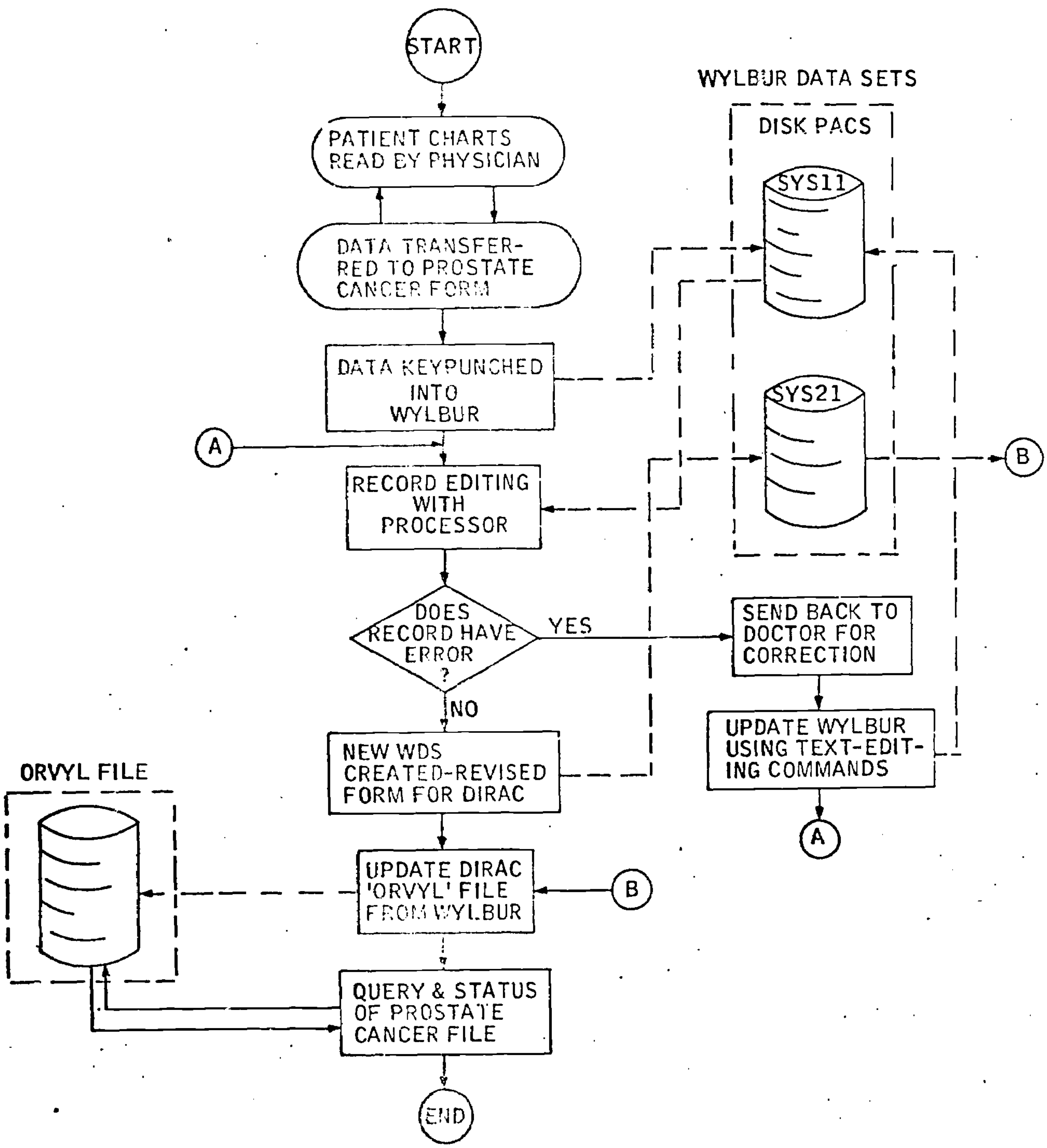


Figure 1.

3.2 Examples of Redesign

3.2.1 Separating Mixed Mode Entries:

Figure 2 (A,B, & C) is an example of three sections of the original form as designed by the medical personnel. One record consists of 170 fields. It was unsatisfactory for the following reasons:

(1) The DIRAC prototype could not handle more than 80 field specifications. Although this constraint could be changed, it was felt that this limit was a reasonable one that should be sufficient even for complex files. The initially large number of fields was seen as an indication that the problem had been insufficiently analyzed, and this assumption proved correct.

(2) Certain field entries contained mixed mode data. This can be tolerated by DIRAC, but it makes retrieval very cumbersome. DIRAC provides specific sets of commands by which the different types of data (Date, Integer, Real, Alpha, or Coded) can be retrieved or selected. (See Objective 3)

Note @9 in Figure 2A (where '@' is the field delimiter). This field contains codes as well as an integer number describing 'DURATION'. Thus, if the symptom 'PROSTATISM' occurred with the signs of 'DRIBBLING' and 'NOCTURIA', and the duration for these were 5 and 8 months respectively, the code representation would be 'ACF12'. Now, if the physican wished to retrieve all those patients which showed the symptom of 'PROSTATISM' and the sign of 'DRIBBLING', with a duration of 6-11 months, an error would occur since the above-described entries would be retrieved: it contains the codes A,C, and 2 . It must be pointed out here that the '2'

SYMPTOMS & SIGNS

Codes for Duration:
1 0 - 5 Months
2 6- 11 "
3 12- 17 "
4 > 18 "
5 No Information

88 ASYMPTOMATIC(AS)
A Yes
B No
C No Information

39 PROSTATISM(AS)(DURATION FROM 1ST RX)
A Yes
B No
C Dribbling
D Obstruction
E Hesitancy
F Nocturia
G Hematospermia
H Hematuria
I No Information

DURATION ___________

810 LOCAL ANAL PAIN(AS)
A Yes
B No
C No Information

DURATION ___________

812 SENSATION OF MASS(AS)
A Yes
B No
C No Information

DURATION ___________

813 BONE PAIN < 1 YEAR(AS)
A Yes
B No
C Cervical Spine
D Thoracic Spine
E Lumbar Spine
F Pelvis
G Other ___________
H No Information

DURATION ___________

814 1ST SYMPTOM & INTERVAL(AS)
A Prostatism
B Local Anal Pain
C Perineal Pain
D Sensation of Mass
E Bone Pain

DURATION ___________

Figure 2A

FAMILY HISTORY

024 HISTORY OF CARCINOMA(AS)
- A Yes
- B No
- C No Information.

025 HORMONE DEPENDENT TUMORS(AS)
- A Yes
- B No
- C Prostate
- D Breast
- E Endometrial
- F Kidney
- G No Information

026 ASHD(AS)
- A Yes
- B No
- C No Information

027 ASCVD(AS)
- A Yes
- B No
- C No Information

PAST MEDICAL HISTORY

028 HISTORY OF CARCINOMA(AS)
- A Yes
- B No
- C No Information

029 HORMONE DEPENDENT TUMORS(AS)
- A Yes
- B No
- C Breast
- D Kidney
- E No Information

030 ASHD(AS)
- A Yes
- B No
- C No Information

031 ASCVD(AS)
- A Yes
- B No
- C No Information

032 RECTAL SURG.(AS)
- A Yes
- B No
- C No Information

033 PROSTATITIS(AS)
- A Yes
- B No
- C No Information

034 COLITIS
- A Yes
- B No
- C No Information

035 IMPOTENCY(AS)
- A Yes
- B No
- C No Information

Figure 2B

BONES:
@50 MOST REC. PRE-RX DX EVAL.(AS)
A PERFORMED
B NOT PERFORMED
C IVP + - ? (EQUIV.)
D REGIONAL X-RAY + - ?
E CHEST X-RAY + - ?
F BONE SURVEY + - ?
G BONE SCAN + - ?
H NO INFORMATION

CODES FOR BONE STUDIES:
A IVP + ? (EQUIV.)
B REGIONAL X-RAY + ?
C CHEST X-RAY + ?
D BONE SURVEY + ?
E BONE SCAN + ?
F NO INFORMATION

@51 LONG BONES(AS) ______

@52 CERVICAL SPINE(AS) ______

@53 THORACIC SPINE(AS) ______

@54 LUMBAR SPINE(AS) ______

@55 SACRUM(AS) ______

@56 R INNOMINATE BONE(AS) ______

@57 L INNOMINATE BONE(AS) ______

@58 PUBIS(AS) ______

@59 SKULL(AS) ______

@60 R ILIUM(AS) ______

@61 L ILIUM(AS) ______

@62 RIBS(AS) ______

@63 BONY PELVIS-OTHER(AS) ______

@64 BONES - OTHER(AS) ______

@65 POST-RX DX EVAL.(AS)
(AT TIME OF DEMON. OF 1ST METS)
A PERFORMED
B NOT PERFORMED
C IVP + - ? (EQUIV.)
D REGIONAL X-RAY + - ?
E CHEST X-RAY + - ?
F BONE SURVEY + - ?
G BONE SCAN + - ?
H NO INFORMATION

REGION - 1ST BONY METS
@66 LONG BONES(AS) ______

@67 CERVICAL SPINE(AS) ______

@68 THORACIC SPINE(AS) ______

@69 LUMBAR SPINE(AS) ______

@70 SACRUM(AS) ______

@71 R INNOMINATE BONE(AS) ______

@72 L INNOMINATE BONE(AS) ______

@73 PUBIS(AS) ______

@74 SKULL(AS) ______

@75 R ILIUM(AS) ______

@76 L ILIUM(AS) ______

@77 RIBS(AS) ______

@78 BONY PELVIS-OTHER(AS) ______

@79 BONES - OTHER(AS) ______

REGION - PREV EQUIV - SUBQ +
@80 LONG BONES(AS) ______

@81 CERVICAL SPINE(AS) ______

@82 THORACIC SPINE(AS) ______

@83 LUMBAR SPINE(AS) ______

@84 SACRUM(AS) ______

@85 R INNOMINATE BONE(AS) ______

@86 L INNOMINATE BONE(AS) ______

@87 PUBIS(AS) ______

@88 SKULL(AS) ______

@89 R ILIUM(AS) ______

@90 L ILIUM(AS) ______

@91 RIBS(AS)

@92 BONY PELVIS-OTHER(AS) ______

@93 BONES - OTHER(AS) ______

Figure 2C.

code belongs with the duration of the NOCTURIA sign. Figure 3 is a redesign of this whole section, 'SYMPTOMS & SIGNS". Note that field 10, which is called duration, is now a multiple field, i.e., a field with subfields. Note also how the various field entries describing symptoms and signs were categorized into one multiple field, @9. Thus, @9(1) is connected to @10(1): 'DRIBBLING' is related to duration of 'DRIBBLING'. The difference in mode is now retained, and retrieval can be accomplished without complication. If the physician wishes to retrieve all those patients who had 'DRIBBLING' for 8 months he can specifically test @9(2) for an 'A' and @10(2) for an '8' . (Example of retrieval commands and procedure are given in the next section) Note the alignment of entries in columns so that the keypuncher can easily select the correct entry to be punched. Also, the old form utilized seven fields for this section and mixed modes (@8 - @13). The redesigned section utilizes only two fields (@9,@10), and it clearly separates coded from integer values. (DRIBBLING from DURATION)

3.2.2 Combining Entries:

Another example of redesign is shown by the transition of Figure 2B to Figure 4. This particular example demonstrates how related fields can be combined in such a way as to retain their individual significance and complete meaning. Sections headed 'FAMILY HISTORY' and 'PAST MEDICAL HISTORY' are the cases in point. Note that @24 - @35 are used for these entries (Figure 2B). Note also the duplication of certain fields under each of these sections, e.g., @24-@27 are similar to @28-@31 respectively. Instead of duplicating these fields the design should attempt to place them in such a way that one heading suffices for both entries. Thus, space will be saved on the form and number of fields will be decreased,

*** PRE RX ***

9 -- SYMPTOMS & SIGNS --			Duration 10
Omit 9(2) - 9(7) if 9(1) is Y			
(1) Asymptomatic	Y	N	
(2) Prostatism:			(1)____
Dribbling		A	(2)____
Obstruction		B	(3)____
Hesistancy		C	(4)____
Nocturia		D	(5)____
Hematospermia		E	(6)____
Hematuria		F	(7)____
Dysuria		G	(8)____
Frequency		H	(9)____
Decreased Stream		I	(10)____
Urgency		J	(11)____
No		N	
(3) Local Anal Pain:	Y	N	(12)
(4) Perineal Pain:	Y	N	(13)
(5) Sensation of Mass:	Y	N	(14)
(6) Bone Pain (< 1 Yr)			(15)____
Cervical Spine		A	
Thoracic "		B	
Lumbar "		C	
Pelvis		D	
Other________		E	
None		N	
(7) Other:			(16)
(8) 1st Symptom:			
Prostatism		A	
Local Anal Pain		B	
Perineal Pain		C	
Sensation of Mass		D	
Bone Pain		E	
None		N	

-- HORMONE STATUS --

11 Status - Pre Rx			
(1) Castration	Y	N	
(2) Estrogen (start)	Y	N	
(3) " (Stop)	Y	N	

12 Dates:

(1) Castration	________
(2) Estrogen (start)	________
(3) " (stop)	

13 Estrogen (Type):		
Stilbesterol	A	
Tace	B	
Cytoproteroneacetate	C	
No	N	

Figure 3.

14 FAMILY HISTORY			15 PAST MED. HISTORY			
(1) Carcinoma	Y	N		(1)	Y	N
(2) ASHD	Y	N		(2)	Y	N
(3) ASCVD	Y	N		(3)	Y	N
(4) CA Type: ______			(4) ______			
			Rect. Surgery	(6)	Y	N
			Prostatitis	(7)	Y	N
			Colitis	(8)	Y	N
			Impotency	(9)	Y	N

Hormone Dep. Tumors:

(5)	Breast	A	(5)	A
	Kidney	B		B
	Prostate	C		C
	None	N		N

-- ROENGTEN STUDIES --

16(1) IVP Date: (Pre Rx)

17(1) IVP: (Pre Rx)

Normal	A
Displaced R Ureter	B
Displaced L "	C
Obst. Uropathy	D
Prostatic Enlargement	E
Bladder Invasion	F
Bony Mets	G
Other: ______	H
Not Performed	N

18 Pre Rx: | 19 Date

(1) Chest X-Ray | 19(1) ______

Normal	A
Parenchymal Mets	B
Mediastinal Mets	C
Bony Mets	D
Abn. Other: ______	E
Not Performed	N

Liver Scan 19(2) ______

Brain Scan 19(3) ______

20 Scans:	(1) Liver	(2) Brain
Normal	A	A
Abn. with Mets	B	B
Abn. other	C	C
Not Performed	N	N

(3) Other (remarks):

21 -- LABORATORY -- (Pre Rx)

	Normal	Abnormal
(1) Acid Phosphat.: No Estro.	A	B
(2) " " On "	A	B
(5) Alk. Phosphatase - Pre Rx	A	B

Figure 4.

with corresponding savings in storage. 'FAMILY HISTORY' is a field of five subfields whereas 'PAST MEDICAL HISTORY' is a field of nine subfields. Note how the columns are aligned for each entry. Thus, the physician as well as the operator should have no trouble in filling out and keypunching the data respectively. The physician's pencil need not wander around the page seeking a place to enter the data. The operator's eyes need not wander around the form searching for the next entry, but just follows a downward path to the bottom of the page.

3.2.3 Reorganization & Condensing:

The study had within its scope a 'PRE RX' section and a 'POST RX' section, i.e., Pre treatment vs. Post treatment. Since the record of these treatments would appear in different sections of the patients record chart they would have to appear in different parts of the form to keep in line with objective (2) -- keeping chronological order throughout the form if possible even though such a structure placed serious constraints on the complexity of the DIRAC input processor. The old form divided all sections into single fields (no subfields). In the redesigned form, sections which were exactly similar except that one was 'PRE' and the other 'POST' were designated as multiple fields. thus, continuity of field designation was accomplished throughout the form. Figure 5A is a description of the 'BONES' section of the new form, @22(1) - @36(1). ON the old form, Figure 2C, the 'BONES' section was described by field entries 50-64 for PRE and 65-93 for POST. Note now the POST section in Figure 5B. It is exactly similar except for the subfield number notation, '2'. In other words, subfield '1' designates 'PRE', while subfield '2' designates 'POST'. The ability to designate similar fields as a subfield structure in different

-- BONES --				
22(1) Most Rec. Dx Evaluation (Pre Rx): (? = Equiv.)				
IVP	A	+	-	?
Regional x-ray	B	+	-	?
Chest "	C	+	-	?
Bone Survey	D	+	-	?
Bone Scan	E	+	-	?
Not Performed	N			

Codes for Bone Studies: ↑ ↑ ↑

Bone Analysis:	Pre Rx
23(1) Long Bones :	
24(1) Cervical Spine :	
25(1) Thoracic Spine :	
26(1) Lumbar Spine :	
27(1) Sacrum :	
28(1) R Innominate Bone:	
29(1) L Innominate Bone:	
30(1) Pubis :	
31(1) R Ilium :	
32(1) L Ilium :	
33(1) Bony Pelvis-other:	
34(1) Skull :	
35(1) Ribs :	
36(1) Bones - other :	

-- DIAGNOSIS --

37 Date:

38(1) Dx Method:

Transrectal Needle Bx	A
Transperineal " "	B
" Open	C
Needle Bx Unk. Type	D
Retropubic Prostatect	E
RAD " "	F
RAD Perineal "	G
Aspiration Bx	H
TUR	I
No	N

38(2) Tur. - Pre Rx Dx Method:

< 25%	A
25 - 50%	B
> 50%	C

Reason for Bx:

38(3)	Obstructive Symptoms	Y	N
38(4)	Abn. Prostate Gland	Y	N

Figure 5A.

*** POST RX *** (Con't)

-- BONES --

22(2) Most Rc. Dx Evaluation (Post Rx):
(? = Equiv.)

IVP	A	+	-	?
Regional X-Ray	B	+	-	?
Chest	C	+	-	?
Bone Survey	D	+	-	?
Bone Scan	E	+	-	?
None	N			

Codes for Bone Studies: ↑ ↑ ↑

23(2)	Long Bones	:
24(2)	Cervical Spine	:
25(2)	Thoracic Spine	:
26(2)	Lumbar Spine	:
27(2)	Sacrum	:
28(2)	R Innominate Bone	:
29(2)	L Innominate Bone	:
30(2)	Pubis	:
31(2)	R Ilium	:
32(2)	L Ilium	:
33(2)	Bony Pelvis-other	:
34(2)	Skull	:
35(2)	Ribs	:
36(2)	Bones - other	:

53 Bony Mets: Omit (1) - (4) if 22(2) is Negative or No Information

(1)	Lytic	A
	Blastic	B
	A & B	C
	None	N
(2)	Painful	A
	Non-painful	B
	No Bony Mets	N

Pain Remission:

(3)	Estrogen Rx	Y	N
(4)	X-Ray Rx	Y	N

41 Lymph Nodes - Post Rx:

(Equiv. = ?)	Clin. (3)	Bx (4)
R. S/C	A + ?	A + ?
L. S/C	B + ?	B + ?
Inguinal	C + ?	C + ?
R. Iliac	D + ?	D + ?
L. Iliac	E + ?	E + ?
R. Pan	F + ?	F + ?
L. Pan	G + ?	G + ?
None	N	N

Figure 5B.

parts of the form enables the user to keep the continuity of his file throughout the entire form, instead of giving a field number to every entry of the form. This cuts down the number of separate fields and helps in the retrieval aspects as shown in the next section.

3.3 General Rules for Forms Design

3.3.1 Determine type of answers you want, i.e., codes, integers, real numbers. text, etc.

3.3.2 Align fields for easy reading, marking, and key-punching.

3.3.3 Place "similar" pieces of information together whenever possible, making that field multiple.

3.3.4 Keep field designations to minimum.

3.3.5 Do not mix entries, e.g., Do not set up a field which can be both integer and DATE.

3.3.6 Keeping the number of pages to minimum is, of course, important; however, make the form pleasing to the eye so that it separates itself into obvious sections, and the user does not have difficulty deciding where boundaries occur.

4. USING THE PROSTATE CANCER FILE

4.1 File Structure & STATUS Report

Figure 6 is a STATUS report for the prostate cancer file generated by DIRAC. Note that the total file structure is described by this report. Each field is identified by name and description, as well as type (ALPHA, INTEGER, etc.) and multiplicity (SINGLE or MULTIPLE). Note also the various validation features which are presented such as the NECESSARY, SIZE, and SUBFIELD validations.

STANFORD UNIVERSITY
COMPUTATION CENTER

STATUS REPORT FOR FILE M400
4/NOV/1970

(CLASS 1)

LANGUAGE:
DIRAC-1A

DESCRIPTION : File of Prostate Cancer Patients

CREATION DATE : 25/OCT/1970	1 = TYPE	FILE NAME : Prostate	
RECORD NOTATION : $MS	2 = MULTIPLICITY	FILE CREATED BY : RAY	
FIELD NOTATION : GF	3 = INDEXING	RECORD LENGTH : 1024	
DISPOSITION : PUBLIC	4 = CODE RESIDENCE	NO.OF FIELDS : 68	
NO. OF RECORDS : 343	5 = CODE TYPE	LATEST UPDATE ON : 25/OCT/1970	

F I E L D   I D E N T I F I C A T I O N ,   S T A T I S T I C S   A N D   V A L I D A T I O N   I N F O R M A T I O N

FLD	NAME	DESCRIPTION	STORAGE 1	2	3	4	5	VALIDATIONS NEC.	SIZE	SUB.	STATISTICS SIZE	DEC.	SUB.	EXISTENCE REC.	PCT
1	Name	Patient's Name	A	S	0			YES	0	0	22	0	0	343	100.00%
2	SUH#	Stanford University Hosp. No.	A	S	0			YES	0	0	8	0	0	343	100.00%
3	Birthdate		D	S	0				0	0	11	0	0	343	100.00%
4	Race		A	S	0				0	0	9	0	0	341	99.42%
5	Physician	Referral Physician	A	S	0				0	0	29	0	0	338	98.54%
6	Address		A	S	0				0	0	56	0	0	335	97.67%
7	Phone		A	S	0				0	0	12	0	0	118	34.40%
8	DT1	First Visit DT	D	S	0			YES	0	0	11	0	0	343	100.00%
9	Symptoms	Symptoms & Signs	A	M	0				0	0	24	0	8	342	99.71%
10	Dur1	Dur of Symptoms & Signs	I	M	0				0	0	3	0	16	275	80.17%
11	Status	Hormone Status	A	M	0				0	0	36	0	7	322	93.88%
12	DT2	Hormone DTs	D	M	0				0	0	11	0	3	147	42.86%
13	Estrogen	Estrogen Type	A	S	0				0	0	2	0	0	297	86.59%
14	Family	Family History	A	M	0				0	0	31	0	5	260	75.80%
15	Past	Past Medical History	A	M	0				0	0	39	0	9	331	96.50%
16	DT3	IVP DT (Pre Rx)	D	M	0				0	0	11	0	2	266	77.55%
17	IVP	IVP - Roengten Studies	A	M	0				0	0	51	0	2	339	98.83%
18	Chest	Chest X-Ray	A	M	0				0	0	35	0	13	340	99.13%
19	DT4	DTs - Chest, Liver, Brain	D	M	0				0	0	11	0	4	280	81.63%
20	Scans	Liver & Brain Scans	A	M	0				0	0	1	0	5	340	99.13%
21	Labnratory	Laboratory - Pre Rx	A	M	0				0	0	25	0	9	287	83.67%
22	Recent	Most Rec. Dx Evaluation	A	M	0				0	0	10	0	2	342	99.71%
23	Long	Long Bones	A	M	0				0	0	8	0	2	20	5.83%
24	Cervical	Cervical Spine	A	M	0				0	0	8	0	2	19	5.54%
25	Thoracic	Thoracic Spine	A	M	0				0	0	8	0	2	31	9.04%
26	Lumbar	Lumbar Spine	A	M	0				0	0	16	0	2	59	17.20%
27	Sacrum	Sacrum	A	M	0				0	0	8	0	2	41	11.95%
28	RInnominate	R. Innominate Bone	A	M	0				0	0	8	0	2	35	10.20%
29	LInnominate	L. Innominate Bone	A	M	0				0	0	8	0	2	36	10.50%
30	Pubis	Pubis	A	M	0				0	0	8	0	2	34	9.91%
31	Skull	Skull	A	M	0				0	0	8	0	2	38	11.08%
32	RIlium	R. Ilium	A	M	0				0	0	14	0	2	37	10.79%
33	LIlium	L. Ilium	A	M	0				0	0	24	0	2	28	8.16%
34	Ribs	Ribs	A	M	0				0	0	8	0	2	15	4.37%
35	Pelvis	Bony Pelvis-other	A	M	0				0	0	8	0	2	24	7.00%
36	Bones	Bones - other	A	M	0				0	0	16	0	2	17	4.96%
37	DT5	Diagnosis DT	D	S	0				0	0	11	0	0	330	96.21%
38	Method	Dx Method	A	M	0				0	0	19	0	8	341	99.42%
39	Cardiovascular	Physical Examination	A	M	0				0	0	11	0	3	340	99.13%
40	GU	GU-Prostate, Phys. Exam.	A	M	0				0	0	24	0	11	340	99.13%
41	Lymph	Lymph Nodes	A	M	0				0	0	11	0	4	331	96.50%
42	Leg	Leg Edema	A	M	0				0	0	3	0	3	331	96.50%
43	Staging	Staging Summary	A	M	0				0	0	8	0	5	342	99.71%
44	PreviousRx	Previous Rx	A	S	0				0	0	30	0	0	330	96.21%
45	Therapeutic	Therapeutic Plan	A	M	0				0	0	39	0	3	338	98.54%
46	Bladder	Bladder Infection	A	M	0				0	0	3	0	2	329	95.92%
47	Histopathology	Histopathology	A	M	0				0	0	32	0	11	339	98.83%
48	Radiation	Radiation	A	M	0				0	0	26	0	13	333	97.08%
49	Area	Area in CM	I	M	0				0	0	4	0	6	323	94.17%
50	DT6	DT of 1st Rx	D	S	0				0	0	11	0	0	321	93.59%
51	DT7	DT of Last Rx	D	S	0				0	0	11	0	0	320	93.29%
52	Size	Size of Gland	A	M	0				0	0	1	0	2	315	91.84%
53	BonyMets	Bony Mets	A	M	0				0	0	2	0	4	84	24.49%
54	Dur2	Leg Edema Dur	A	M	0				0	0	2	0	3	15	4.37%
55	Complicatoin	Complications	A	M	0				0	0	15	0	13	315	91.84%
56	Dur3	Dur of Complications	I	M	0				0	0	3	0	12	191	55.69%
57	Impotency	Impotency	A	S	0				0	0	24	0	0	113	32.94%
58	DT8	DT of Last Follow Up	D	S	0				0	0	11	0	0	319	93.00%
59	Follow	Follow-Up Status	A	M	0				0	0	38	0	4	329	95.92%
60	Failure	First Failure	A	S	0				0	0	1	0	0	300	87.46%
61	DT9	DT of First Failure	D	S	0				0	0	11	0	0	104	30.32%
62	Site	Site	A	M	0				0	0	18	0	9	108	31.49%
63	DT10	DT of Liver Mets	D	S	0				0	0	11	0	0	3	0.87%
64	FailureRx	Rx of Failures	A	S	0				0	0	4	0	0	232	67.64%
65	Autopsy	Autopsy	A	M	0				0	0	35	0	7	298	86.88%
66	DT11	DT of Brain Mets	D	S	0				0	0	11	0	0	17	4.96%
67	Remarks	General Comments	A	S	0				0	0	144	0	0	126	36.73%
68	Cause	Cause of Death	A	S	0				0	0	96	0	0	73	21.28%

Figure 6.

Should any validation criterion be violated by the input data, a warning would be issued and a recovery process triggered. Also, the status of all field entries is shown in the last few columns. It is this status report which is used by the physician as a guide when he wishes to query the file.

4.2 Interrogating the Prostate Cancer File

4.2.1 Catalogued Interrogations:

If the physician knows exactly what questions he wants to ask of his file on a routine basis (for example, to generate a weekly report on the list of patients with certain characteristics) then there is no need for him to sit at the terminal and spend time typing them. He might as well input the questions into a WYLBUR DATA SET, and then let the DATA SET drive DIRAC, i.e., have DIRAC receive its commands from the text-editor. Figure 7 is a description of this process by which catalogued interrogations stored under a text-editor can be used to drive a data-base management language in a way which is, to the best of our knowledge, unique to the DIRAC language.

CATALOGUED INTERROGATION
(lines beginning with a colon contain the user's response)

```
NAME OF USER
    :     RAY
EXECUTION MODE
    :     QUERY
NAME OF FILE
    :     M300
INTERFACE
    :     WYLBUR
            .
            .
            .
YOU CAN NOW DESIGNATE A NEW EXECUTION MODE OR
TYPE AN EXCLAMATION POINT FOR EXIT FROM DIRAC
     :     !
EXIT FROM DIRAC
```

Figure 7.

The name of the user, execution mode, and file name are all requested by DIRAC. The next prompt asks the user to designate whether or not the file will be queried from WYLBUR or from the terminal. Figure 8 is a partial description of the types of commands which might be stored in WYLBUR. As these questions or commands are executed by DIRAC the answers will be printed out at the terminal. This process is described in the next section.

WYLBUR DATA SET OF CATALOGUED QUERIES

```
SELECT ALL
@9(2) CONTAINS AGI END
@10(2) < 5 END
@10(7) < 3 AND @10(9) < 6 END
@16(1) < 1968 END
@18(1) CONTAINS A END
TYPE Name Birthdate Physican END
```

Figure 8.

4.2.2 Conversational Interrogations:

For those questions which cannot be anticipated or catalogued such as the previous section described then the user will want to set at the terminal himself and ask questions as they come to his mind. Figure 9 is an example of this type of query mode.

CONVERSATIONAL INTERROGATION USING DIRAC

```
:    SELECT ALL
785 RECORDS SELECTED
:    RETAIN
:    @9(2) CONTAINS AGI END
43 RECORDS SELECTED
:    @10(2) < 5 END
26 RECORDS SELECTED
:    @10(7) < 3 AND @10(9) < 4 END
11 RECORDS SELECTED
:    @16(1) < 1965 END
4  RECORDS SELECTED
:    @18(1) CONTAINS A END
2  RECORDS SELECTED
:    TYPE Name Birthdate Physician END
     203
Name         John Smith
Birthdate    10-9-39
Physician    Watson
     368
Name         Joan Marshall
Birthdate    1-14-34
Physician    Farmer
:    RELEASE
        .
        .
        .
```

Figure 9.

5. CONCLUSIONS AND RECOMMENDATIONS

The problem of information retrieval has been with us for a long time. The classical method of solving the problem for a given application has always been to write a procedural program. However, with the new data-management languages the user need not be concerned with anticipating every single question he wishes to ask of a specific file. He need only collect his data and input this data correctly into the computer; then he can retrieve information regarding those questions which he already had anticipated, but he can also ask an unpredicted one without re-design or re-programming. Thus, the design and use of forms by which information is collected, organized, and input into the computer becomes critical. In this paper we have tried to show how this problem was approached and solved in a specific application.

REFERENCES

(1) Vallee, J. and Ludwig, H.R., "DIRAC: An Overview of an Interactive Retrieval Language". JOURNAL OF DATA MANAGEMENT. (Feb. 1971).

(2) Vallee, J. and Ludwig, H.R., "The DIRAC Language: Concepts and Facilities." Research Report No. 1. Information Systems Group, Stanford University Computation Center, (May 1970).

(3) Fredrickson, R., WYLBUR ON THE IBM 360/67, Stanford University Computation Center, Stanford, California (1969).

(4) Vallee, J. "DIRAC: An Interactive Retrieval Language with Computational Interface." Information Storage & Retrieval Journal, Vol 6, No. 5 (Dec. 1970).

EXPERIMENTS WITH A CANCER RESEARCH FILE

Problem Description

The previous article entitled "Forms Design in Medical Information Retrieval", has reviewed the problem of data acquisition and the creation of a cancer research file suitable for interactive inquiry. The present report will describe in detail the techniques that have been used to convert the raw data into a form acceptable to the DIRAC update processor, and to extract specialized statistical information through an interface between DIRAC and FORTRAN.

Figure 1 is a general diagram of the data flow through the entire system.

Initial Input Pre-Processor

The form utilized to collect the data on the prostate cancer patients was designed for the convenience of the physician and the keypunch operator -- The former reading the patient record charts and transcribing information onto the form; the latter transcribing the information on the form into a WYLBUR DATA SET. (1) Thus once the data was entered into WYLBUR, it was not in correct order for input into the DIRAC file system. A pre-processor was written in FORTRAN which re-arranged the data in proper order and checked for obvious errors. A description of the pre-processor is given in Figure 2.

Interface Specifications

This section describes the statistical subroutines that generate survival rates and a Berkson-Gage table for selected subsets of the prostate cancer file. These subroutines are interfaced with modules of the DIRAC data-base management system and operate on information retrieved by the DIRAC 'selection' commands

GENERAL DATA FLOW

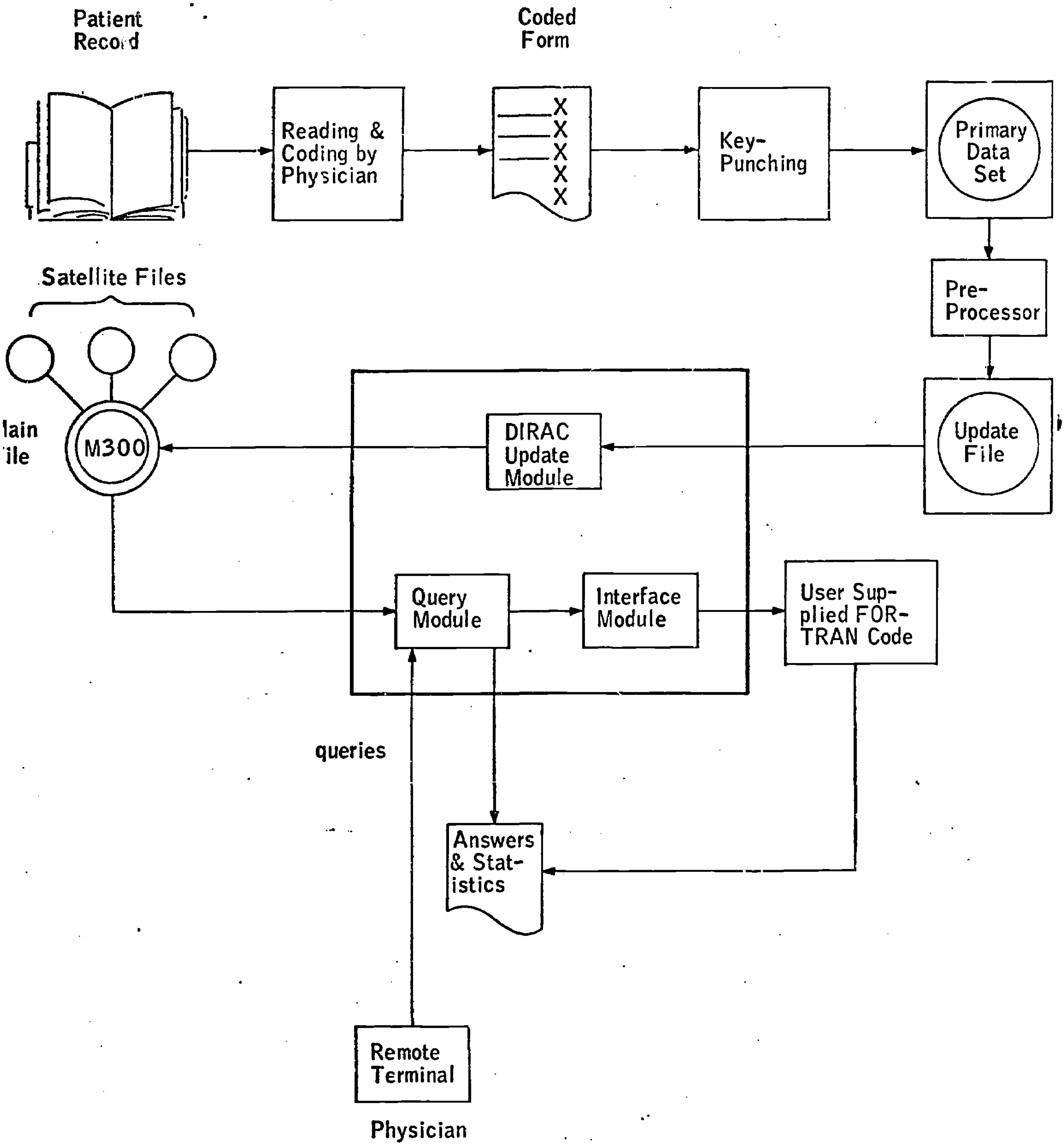


Figure 1.

GENERAL FLOWCHART OF PRE-PROCESSOR

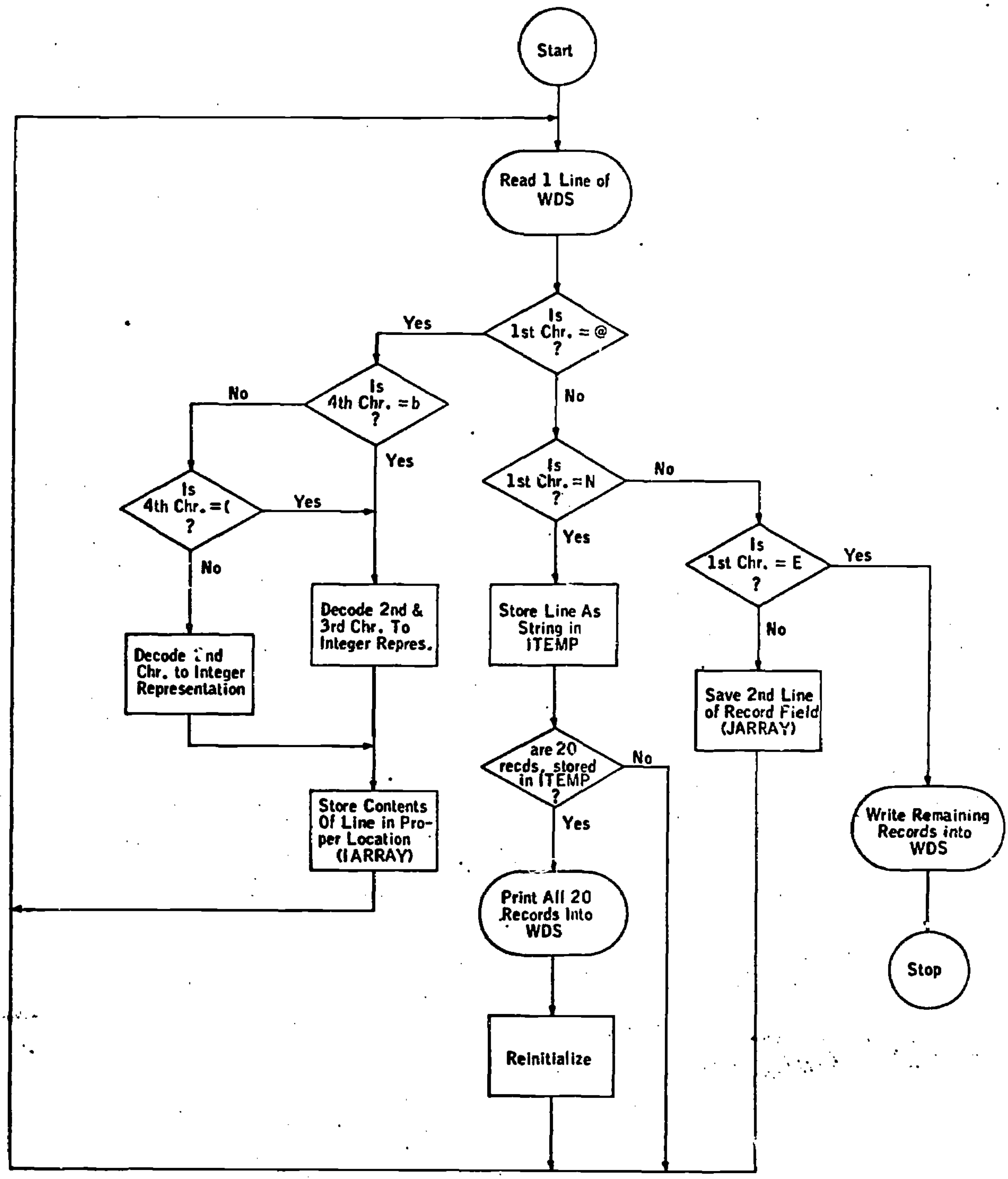


Figure 2.

from file M300.

Two application-oriented subroutines (S1 and S2) are involved in this interaction. Their general organization is shown on Figures 3 and 4. The main purpose of these FORTRAN-coded routines was to perform survival rate calculations in accordance with the standard statistical presentation of Cancer Staging and End Results.

Such a standard was developed by a special American Cancer Committee formed in January 1959, whose conclusions are contained in a booklet issued in July, 1963. It involved certain calulations that cannot be expected to be included as a subset of a language like DIRAC. Rather, it is of interest to observe how DIRAC can be interfaced with the modules we have mentioned, whose flow-charts are given as Figures 3 and 4. Calls to these modules are triggered by an EXTRACT command, as shown in the following examples.

Conversational Inquiry

Figure 5 shows an actual interrogation of M300 from the terminal. The researcher was interested in the following subsets: The search has been restricted to patients with cancer limited to the prostate gland. For how many of these does the field "Failure" contain a Yes ? This question yields fourteen records. For this subset we generate survival rates and a statistical table.

We then continue with the DIRAC interrogation and we ask how many patients have disease limited to the gland and increased acid phosphatase: four patients are found. Going back to the whole file, we continue the search by selecting patients born between 1910 and 1920, etc.

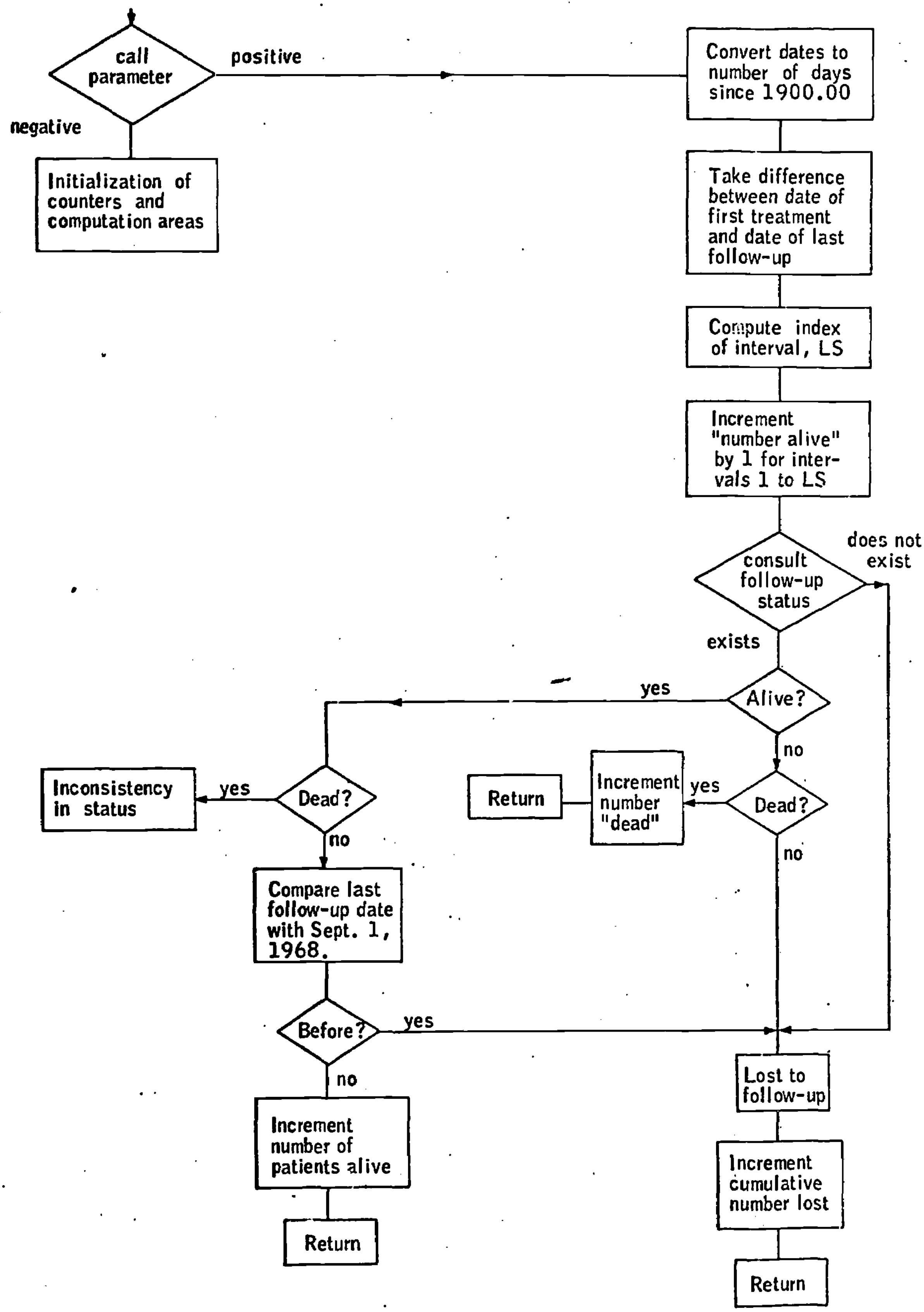


Figure 3. General Flow-Chart, Module S-1

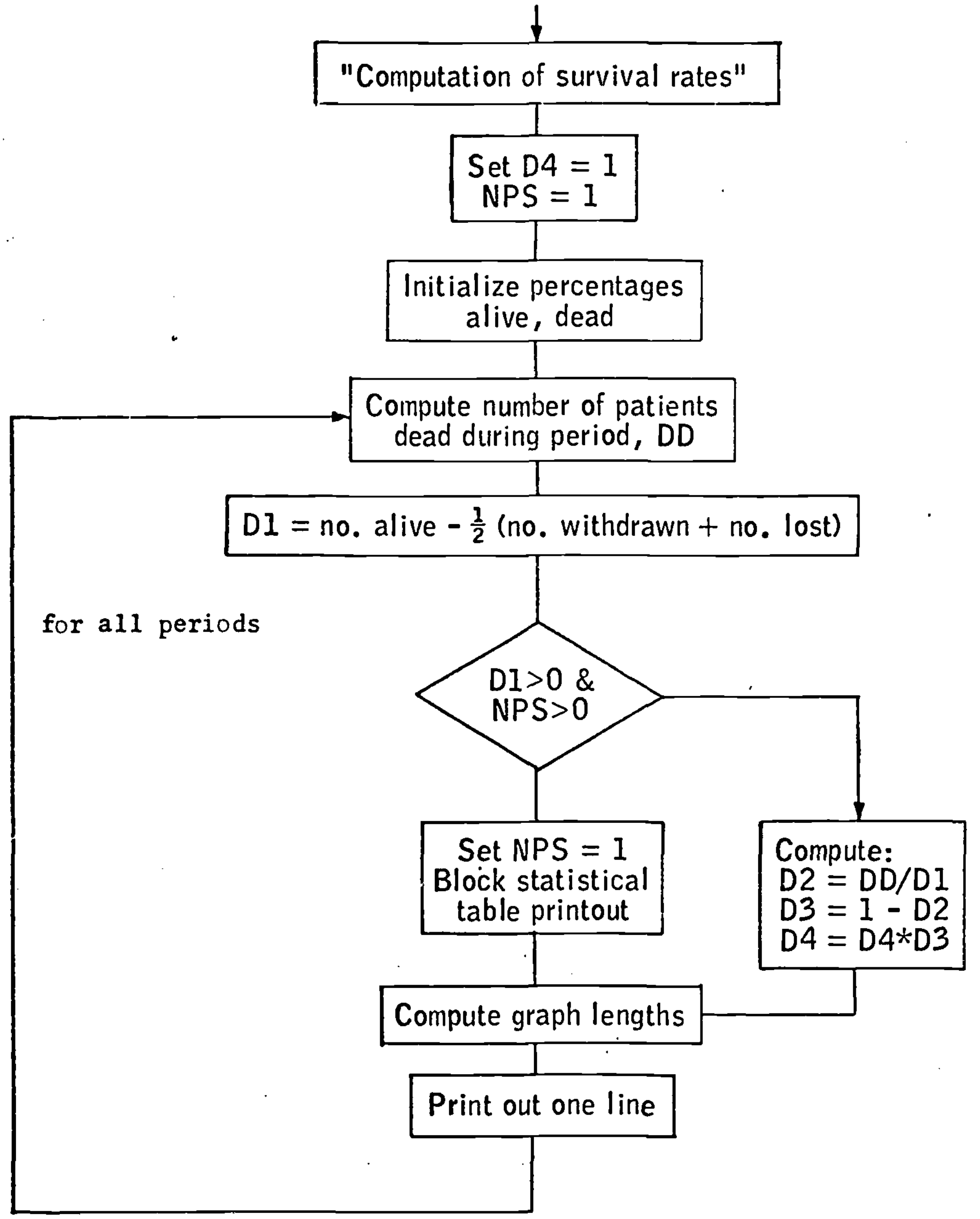


Figure 4. General Flow-Chart, Module S-2

```
:   Q60 CONTAINS Y END
______________________
  14 RECORDS SELECTED

ACTION
  :    EXTRACT 1 DT6 END
  19620212  19681119  ALIVE  B     6.8 YEAR
  19581103  19640301   DEAD  D     5.3 YEAR
  19651018  19680313   DEAD  D     2.8 YEAR
  19650501  19680618   DEAD  D     3.0 YEAR
  19611013  19671127   DEAD  C     6.1 YEAR
  19640922  19690508  ALIVE  D     4.6 YEAR
  19571002  19690306  ALIVE  D    11.4 YEAR
  19680329  19700709  ALIVE  C     2.3 YEAR
  19661006  19691125   DEAD  D     3.1 YEAR
  19640730  19650222   DEAD  D     0.6 YEAR
  19651122  19670110   DEAD  D     1.1 YEAR
  19650930  19700526  ALIVE  D     4.7 YEAR
  19570730  19661209   DEAD  D     9.4 YEAR
  19620426  19680601                      6.1  LOST TO FOLLOW-UP

COMPUTATION OF SURVIVAL RATES
STATUS OF PATIENTS AT THE END OF EVERY 6-MONTH PERIOD SINCE FIRST TREATMENT

     YR    A    D   T.D   W  LST          % ALIVE                 % DEAD          l'x    qx    px    Px         SURVIVAL CURVE

 1  0.0  14.   0.   0.   0.   0.  XXXXXXXXXXXXXXXXXXXX |                        14.00  0.0   1.00  1.00  ------------------------------*
 2  0.5  14.   1.   0.   0.   0.  XXXXXXXXXXXXXXXXXXXX |                        14.00  0.07  0.93  0.93  ---------------------------*
 3  1.0  13.   1.   1.   0.   0.    XXXXXXXXXXXXXXXXXX | X                      13.00  0.08  0.92  0.86  -------------------------*
 4  1.5  12.   0.   2.   0.   0.     XXXXXXXXXXXXXXXXX | XX                     12.00  0.0   1.00  0.86  -------------------------*
 5  2.0  12.   0.   2.   1.   0.     XXXXXXXXXXXXXXXXX | XX                     11.50  0.0   1.00  0.86  -------------------------*
 6  2.5  11.   1.   2.   0.   0.       XXXXXXXXXXXXXXX | XX                     11.00  0.09  0.91  0.78  -----------------------*
 7  3.0  10.   2.   3.   0.   0.        XXXXXXXXXXXXXX | XXXX                   10.00  0.20  0.80  0.62  ------------------*
 8  3.5   8.   0.   5.   0.   0.           XXXXXXXXXXX | XXXXXXX                 8.00  0.0   1.00  0.62  ------------------*
 9  4.0   8.   0.   5.   0.   0.           XXXXXXXXXXX | XXXXXXX                 8.00  0.0   1.00  0.62  ------------------*
10  4.5   8.   0.   5.   2.   0.           XXXXXXXXXXX | XXXXXXX                 7.00  0.0   1.00  0.62  ------------------*
11  5.0   6.   1.   5.   0.   0.              XXXXXXXX | XXXXXXX                 6.00  0.17  0.83  0.52  ---------------*
12  5.5   5.   0.   6.   0.   0.               XXXXXXX | XXXXXXXX                5.00  0.0   1.00  0.52  ---------------*
13  6.0   5.   1.   6.   0.   0.               XXXXXXX | XXXXXXXX                4.50  0.22  0.78  0.40  ------------*
14  6.5   3.   0.   7.   1.   1.                  XXXX | XXXXXXXXXX              2.50  0.0   1.00  0.40  ------------*
15  7.0   2.   0.   7.   0.   1.                    XX | XXXXXXXXXX              2.00  0.0   1.00  0.40  ------------*
16  7.5   2.   0.   7.   0.   1.                    XX | XXXXXXXXXX              2.00  0.0   1.00  0.40  ------------*
17  8.0   2.   0.   7.   0.   1.                    XX | XXXXXXXXXX              2.00  0.0   1.00  0.40  ------------*
18  8.5   2.   0.   7.   0.   1.                    XX | XXXXXXXXXX              2.00  0.0   1.00  0.40  ------------*
19  9.0   2.   1.   7.   0.   1.                    XX | XXXXXXXXXX              2.00  0.50  0.50  0.20  ------*
20  9.5   1.   0.   8.   0.   1.                     X | XXXXXXXXXXX             1.00  0.0   1.00  0.20  ------*
21 10.0   1.   0.   8.   0.   1.                     X | XXXXXXXXXXX             1.00  0.0   1.00  0.20  ------*
22 10.5   1.   0.   8.   0.   1.                     X | XXXXXXXXXXX             1.00  0.0   1.00  0.20  ------*
23 11.0   1.   0.   8.   1.   1.                     X | XXXXXXXXXXX             0.50  0.0   1.00  0.20  ------*
24 11.5   0.   0.   8.   0.   1.                       | XXXXXXXXXXX             0.0 ******************  -
25 12.0   0.   0.   8.   0.   1.                       | XXXXXXXXXXX             0.0 ******************  -
26 12.5   0.   0.   8.   0.   1.                       | XXXXXXXXXXX             0.0 ******************  -
27 13.0   0.   0.   8.   0.   1.                       | XXXXXXXXXXX             0.0 ******************  -
28 13.5   0.   0.   8.   0.   1.                       | XXXXXXXXXXX             0.0 ******************  -
29 14.0   0.   0.   8.   0.   1.                       | XXXXXXXXXXX             0.0 ******************  -
INSTANTANEOUS DEATH RATE =    0.06


ACTION
  :    RELEASE
______________________
 139 RECORDS SELECTED

ACTION
  :    Q43(1) CONTAINS Y AND Q43(5) CONTAINS Y END
______________________
   4 RECORDS SELECTED

ACTION
  :    Birthdate < 1920 AND Birthdate > 1910 END
______________________
  21 RECORDS SELECTED

ACTION
  :    EXTRACT 1 DT6 END
```

Figure 5.

Conclusion

This experiment has demonstrated that it was possible to create a flexible interaction between medical researchers and a significantly large data base. The medical results of this study will be published elsewhere, but it is apparent from the examples given here that the efficiency of the man/maching interace can be greatly improved when a generalized file-oriented language is used.

This observation points to a series of measures that can be taken by language designers in the medical environment. These measures include:

- Greater emphasis on the use of non-procedural techniques for medical data-management.
- Increased reliance on medical personnel for the creation, updating and interrogation of their files.
- Better and clearer definition of interfaces between data management modules and classical software (FORTRAN, PL/1, Statistical Packages) required in specific applications.