Wolf, Ludwig and Vallee:
Progress towards a Direct-Access
Hematology Data-Base.

(in: Archives of Pathology, June 71)

Progress Towards a Direct-Access

Stanford's Experience With the DIRAC Language

Paul L. Wolf, MD; Herbert R. Ludwig, PhD; and Jacques F. Vallee, PhD.

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 Hematopathology Section. 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. The user communicates directly with the computer in natural language context. An example of actual interrogation of a sample file is reviewed in detail. The impact that this system has had on the administration and work patterns of the Hematopathology Section is described.

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 University Medical Center is a case in point. At the Stanford 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 Stanford 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.

This report first describes the problem of data-base implementation under these constraints. Following that is a brief overview of the DIRAC language, in particular, the following modes under which it operates: CREATE, UPDATE, QUERY, and STATUS. Lastly the report contains the conclusions and recommendations for future systems of this type.

Data-Base Implementation

Figure 1 represents a composite of the sequential steps of development of the direct-access hematology data-base system. Step one of Fig 1 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

Accepted for publication Jan 11, 1971.

From the Clinical Laboratory, Stanford University Medical Center (Dr. Wolf), and the Stanford Computation Center (Drs. Ludwig and Vallee), Stanford, Calif.

Reprint requests to the Clinical Laboratory, Stanford University Medical Center, Stanford, Calif 94305 (Dr. Wolf).

Hematology Data-Base

Stanford, Calif

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 was obtained. This method proved to be unsuccessful, since recovery of records and their distribution took too long.

Step two of Fig 1 describes the first phase in automating the bone marrow report system described in step one. 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 general report in format for each patient record (Fig 2). 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 step one. 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 factors 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 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 nonprocedural 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, step two is now transformed to Fig 1 in total. 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 that the 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, 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.

DIRAC: An Overview

The name, DIRAC, stands for "DIRect ACcess" and indicates the five types of data one may wish to use—Date (Jan 1, 1970), Integer number, Real number, Alphanumeric (text information such as a comment or

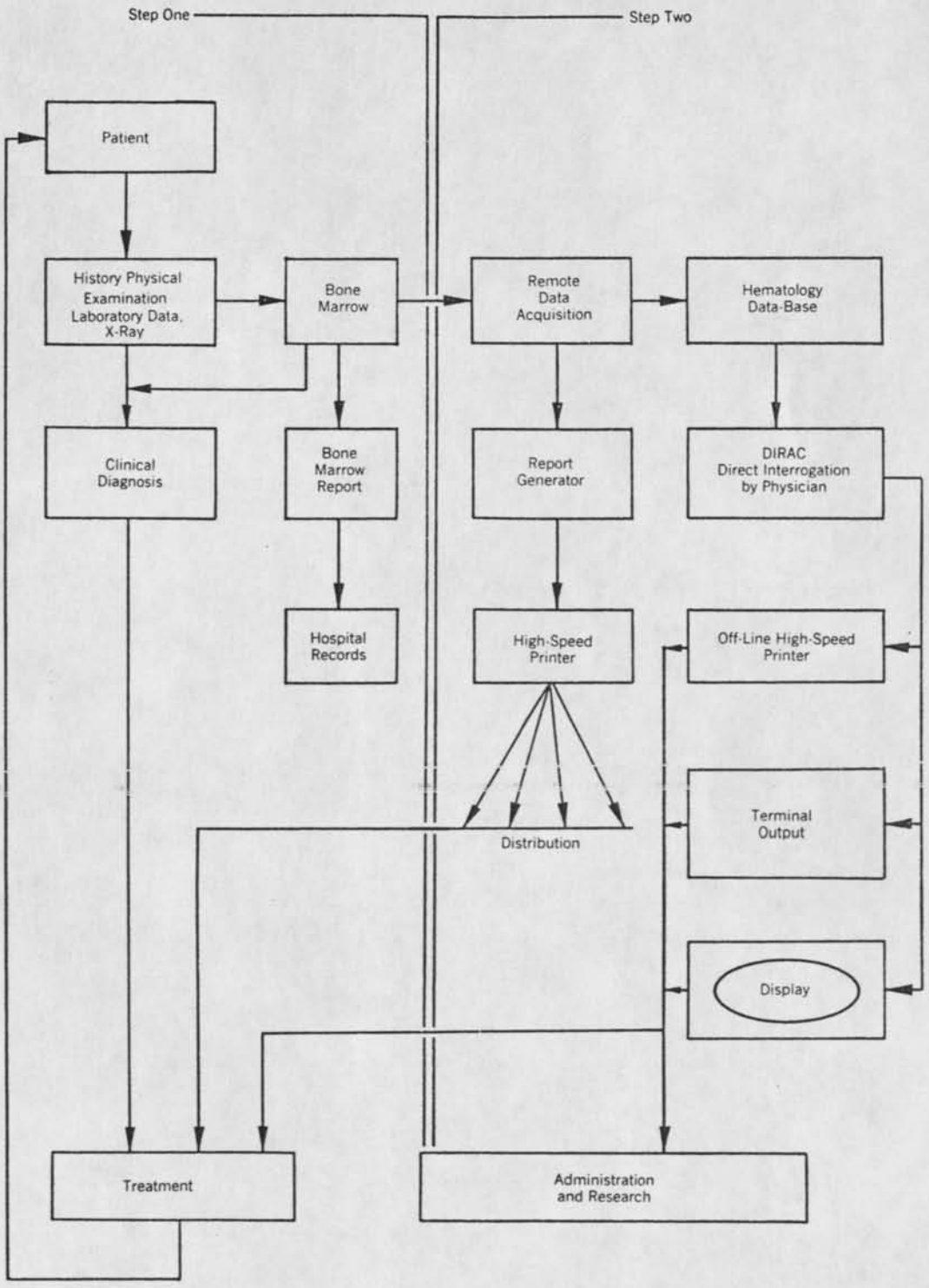

Fig 1.—Composite of the sequential steps of development of the direct-access hematology data-base system.

impression), and Coded (which would vary dependent on the patient application). DIRAC is an information-retrieval language which gives the user the ability to operate on his private data-base under the following 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, and what limitations he would like to impose, such as field size, maximum number of subfields, etc. Only those who have access to the security key are allowed access to the file in the UPDATE or QUERY mode, thus providing file security.

A description of the hematology file created at Stanford University Medical Center is given in Fig 3. Note the field numbers, names, and the various validations used. Figure 3 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 put 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 put this information into the file and at a later date complete the record. Figure 3 is a description of a subset of the hematology file containing approximately 670 records used here for the purpose of illustration. The file is increasing at the rate of four bone marrow reports per day. Note the number of records, the percentage of fields which have actual data in them, and the various validations.

QUERY Mode.—The following five fundamental commands are 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 selected. 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 the search to begin again with 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. The current implementation generates cross-tabulation of extracted fields and can be expanded to include standard post-

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

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.  MEGARARYOCYTES:  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. THROMBOPENIA WITH DECREASED MEGAKARYOCYTES IN BONE MARROW.

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

28.  _____                          29.  _____
     RALPH GEORGE, M.D.                  PAUL L. WOLF, M.D.
     (RESIDENT HEMATOLOGIST)             (DIRECTOR, CLINICAL LABORATORY)
```

Fig 2.—Typical bone marrow report generated for each bone marrow examination at Stanford University Medical Center utilizing DIRAC system. Reports are sent to the patient's chart, referring physician, hematologist, and clinical laboratory file, and are stored in the computer.

processing for any particular application.

DIRAC: An Example

Assume that 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 how many female patients are in this subset.
4. Determine from this subset all those

STANFORD UNIVERSITY
COMPUTATION CENTER

STATUS REPORT FOR FILE M010 (CLASS 1)
8/JUL/1970

LANGUAGE: DIRAC-1A

DESCRIPTION : BONE MARROW REPORT

CREATION DATE : 15/MAY/1970
RECORD NOTATION : N
FIELD NOTATION : QF
DISPOSITION : PRIVATE
NO. OF RECORDS : 235

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

FILE NAME : M010
FILE CREATED BY : Wolf
RECORD LENGTH : 512
NO.OF FIELDS : 29
LATEST UPDATE ON : 15/JUN/1970

FIELD IDENTIFICATION, STATISTICS AND VALIDATION INFORMATION

FLD	NAME	DESCRIPTION	STORAGE 1	2	3	4	5	VALIDATIONS NEC.	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	Megararyocytes -- 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%

Fig 3.—Typical standard report form generated by DIRAC for the user describing the status of his file.

```
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 "3-1!-70" END

-------------------------------------------------------

3 RECORDS SELECTED
:   TYPE Name MarrowNo Date Impression END

-------------------------------------------------------

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

-------------------------------------------------------

32
Name
MarrowNo  B70-453
Date      3-13-70
Impression PERIPHERAL THROMBOCYTOPENIA AND ADEQUATE MEGAKARYOCYTES
           SUGGEST PERIPHERAL REMOVAL MECHANISM. MYELOID HYPERPLASIA,
           EOSINOPHILIA, PLASMACYTOSIS, AND INCREASED R.E. CELL IRON.
           ERYTHROGENESIS APPEARS DECREASED.

-------------------------------------------------------

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

Fig 4.—Typical DIRAC procedure which a physician utilizes to interrogate the file to obtain specific information.

persons who did not have a previous bone marrow report taken.

5. Determine from this subset all persons whose "Impression" field contains the word, *plasmacytosis*.
6. Determine from this subset all those persons whose bone marrow reports were taken during the period March 10 to 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 Fig 4. Note that three records were finally selected which satisfied all those conditions specified by the interrogator.

The DIRAC Environment

DIRAC has proven itself to be useful in two other specialized applications at Stanford University Medical Center. One has been a research effort leading to a practical model of a computerized blood bank transfusion center prototype.[3] The other has been the use of DIRAC in the classical retrieval sense, providing the Department of Radio Therapy personnel the ability to scan a file of 400 prostate-cancer patients and interrogate this file in a nonprocedural mode.

The cost effectiveness of such a system is immediately apparent. If the development cost of the system can be spread over many applications, the system cost per application is minimized.

From the experience gained in developing this system and other nonprocedural systems, the DIRAC system has proven to be the most economic among several alternatives. The major advantage is the cost savings because of the nonprocedural mode, ie, the elimination of the system's programmer and the ability of the user to directly communicate with the computer in natural-language context. The cost per bone marrow report has been estimated to be approximately 25 cents. The DIRAC program is proprietary and is not available; however, the WYLBUR text is available upon inquiry from the Stanford Computation Center.

Conclusions and Recommendations

An interactive storage, retrieval, and dissemination system has been developed jointly by the Hematopathology Section, Stanford University Medical Center, 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 retraining 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 the following 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 the following areas: (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, and (4) correlating bone marrow smears with bone marrow sections. Information in these areas places the physician 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.

References

1. Vallee JF, Ludwig HR: *The DIRAC Language: Concepts and Facilities*, Stanford University Computation Center Research Report 1. Stanford, Calif, Stanford University Press, 1970.

2. Frederickson R: *WYLBUR on the IBM 360/67*, Stanford University Computation Center Report. Stanford, Calif, Stanford University Press, 1969.

3. Wolf PL, Ludwig HR, Vallee JR: Interactive computer management of clinical data: A new unique approach in the blood bank. Read before the American Association of Blood Banks, San Francisco, 1970.