LOCKHEED MARTIN ENERGY RESEARCH LIBRARIES

3 4456 0515871 7

CENTRAL RESEARCH LIBRARY
DOCUMENT COLLECTION

ORNL-4064
UC-32 – Mathematics and Computers

A SYSTEM FOR BIOMEDICAL DATA MANAGEMENT (DATASYS)

H. Katzan, Jr.

OAK RIDGE NATIONAL LABORATORY
operated by
UNION CARBIDE CORPORATION
for the
U.S. ATOMIC ENERGY COMMISSION

Printed in the United States of America. Available from Clearinghouse for Federal Scientific and Technical Information, National Bureau of Standards, U.S. Department of Commerce, Springfield, Virginia 22151
Price: Printed Copy $3.00; Microfiche $0.65

LEGAL NOTICE

This report was prepared as an account of Government sponsored work. Neither the United States, nor the Commission, nor any person acting on behalf of the Commission:

A. Makes any warranty or representation, expressed or implied, with respect to the accuracy, completeness, or usefulness of the information contained in this report, or that the use of any information, apparatus, method, or process disclosed in this report may not infringe privately owned rights; or

B. Assumes any liabilities with respect to the use of, or for damages resulting from the use of any information, apparatus, method, or process disclosed in this report.

As used in the above, "person acting on behalf of the Commission" includes any employee or contractor of the Commission, or employee of such contractor, to the extent that such employee or contractor of the Commission, or employee of such contractor prepares, disseminates, or provides access to, any information pursuant to his employment or contract with the Commission, or his employment with such contractor.

ORNL-4064

Contract No. W-7405-eng-26

MATHEMATICS DIVISION

A SYSTEM FOR BIOMEDICAL DATA MANAGEMENT (DATASYS)

H. Katzan, Jr.

JANUARY 1967

OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee
operated by
UNION CARBIDE CORPORATION
for the
U.S. ATOMIC ENERGY COMMISSION

LOCKHEED MARTIN ENERGY RESEARCH LIBRARIES

3 4456 0515871 7

CONTENTS

		Page
I.	Introduction	1
II.	System Organization	1
III.	System Description	2
IV.	System Operation	4
V.	General Scope of Applicability	6
VI.	Conclusions	8
VII.	DATASYS Publications	9

A SYSTEM FOR BIOMEDICAL DATA MANAGEMENT (DATASYS)

H. Katzan, Jr.

I. INTRODUCTION

The purpose of DATASYS is to provide the research scientist with a hierarchically organized set of easy-to-use programs for use with data management problems. More specifically, DATASYS is designed to facilitate the following types of operations:

(i) storing, updating, and manipulating data;
(ii) selecting, modifying, and sorting data for analysis;
(iii) summarizing data for verification or reports;
(iv) retrieving and collating lists of descriptive data;
(v) automating routine calculations; and
(vi) plotting data.

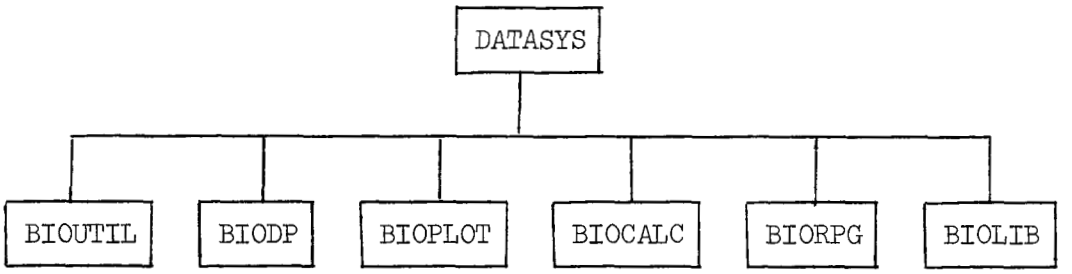
In addition, the system contains provisions for storing and executing, at will, frequently used statistical analysis programs.

Traditionally, data management problems are associated with manual procedures, duplication, multiple passes on the computer, and long lead times. Accordingly, the specific objectives in the design of DATASYS were to:

(i) minimize duplication;
(ii) decrease lead time;
(iii) minimize multiple passes on the computer;
(iv) minimize manual data handling; thereby
(v) lower costs.

II. SYSTEM ORGANIZATION

DATASYS is organized as a collection of programming modules under supervisory control of a control program which routes work through the system. Schematically, the system can be pictured as follows:



The DATASYS modules can be used in three ways:

(i) as subsystems under the DATASYS supervisor;
(ii) as stand-alone programs; and
(iii) as FORTRAN subroutines.

As subsystems under DATASYS, the modules can be used in succession during a single pass on the computer. As stand-alone programs, the modules can be used separately without going through the DATASYS supervisor. As Fortran subroutines, the modules are accessible to the programmer by using the CALL statement.

III. SYSTEM DESCRIPTION

The following paragraphs present a brief introduction to the DATASYS modules. Each module has its own general scope of applicability. Specific functions, such as the SELECT Command under BIODP, will not be considered.

> BIOUTIL is a general-purpose easy-to-use utility program which enables the user to create, copy, list, punch, convert, update, and otherwise manipulate large data files without having to use a detailed programming language. The user describes the work he wants performed with one or more simple control statements. The control statements are ingested by an interpretive program and the desired machine operations are performed by efficient subroutines. Sequence numbers may be included in the data itself or may be external to the data and maintained by the BIOUTIL system. The user may specify the size of his data records and the columns in which his sequence numbers are punched.

BIODP allows the user to execute standard data processing operations, such as sorting, collating, selecting, and the moving, listing, and retrieving of data by simply filling out a preprinted form, having the control information punched into card(s), and feeding these specifications into the program. Processing is performed by subroutines. These operations, which are ordinarily performed by EAM machines or special-purpose programs, can be combined and executed automatically so as to satisfy the data handling needs of an entire experiment.

BIOPLOT enables the user to plot graphs of discrete data without having to involve himself with the details of the electronic plotting equipment. Multiple curves may be generated on each graph. As with BIOUTIL, the user describes the work he wants performed with control statements and the plot tapes are generated by subroutines. Simple control statements may be used to specify titles, type (i.e., LOG, SEMILOG, etc.) symbols (for designating data points), source of data (i.e., bulk storage of system input data), and plotting mode (i.e., connected or disconnected points).

BIOCALC is a computer program designed to relieve the scientist or technician of performing tedious calculations either by hand or with a desk calculator. Generally, the system contains provisions for reading in columns of data, performing calculations on them in column-wise fashion, and printing the data and results in a tabular format. More specifically, the program contains facilities for specifying variable names, constants, and column heading; for generating data internally; and for executing a wide variety of arithmetic operations and mathematical functions.

BIORPG allows the user to generate reports directly from data files. Headings, footings, spacings, and the lines per page may be designated by control statements. BIORPG may be used with BIODP to present specific data in almost any prescribed format.

The BIOLIB facility of DATASYS allows the user to store frequently used programs on magnetic disk. This feature uses the capabilities of Operating System/360 to minimize card handling and to maximize system efficiency.

DATASYS, except for the BIOCALC module, is file oriented in that program data, other than "immediate" records, are maintained on auxiliary storage devices. This facilitates storage and provides a convenient way of storing the results of SELECT, MOVE, or RETRIEVE operations as well as others. In other words, the results of a SELECT operation on a data file is a new data file, just as the result of a COPY operation on a file is another file. Some operations, such as COLLATE and RETRIEVE, have more than data file as input and may require more than one file as output.

IV. SYSTEM OPERATION

All modules are stored as load modules in a partitioned dataset on magnetic disk. Under DATASYS, the user writes control statements as follows:

```
$EXECUTE        BIOUTIL
[input data for BIOUTIL]

$EXECUTE        BIODP
[input data for BIODP]

$EXECUTE        BIORPG
[input data for BIORPG]
.
.
.
$STOP
```

DATASYS interprets the '$EXECUTE' command and issues an Operating System/360 LINK macro to the requested module. Program control is returned to DATASYS so that modules can be executed successively.

As stand-alone programs, the programmer uses the following Operating System/360 Job Control Language statements to execute a module as a job step:

```
//BETA     JOB    ...
//JOBLIB   DD     DSNAME=DATASYS,UNIT=2311                    X
//          VOLUME=SER=BIOPAK,DISP=(OLD,KEEP)
//STEP1    EXEC   PGM=BIOUTIL
```

[additional Operating System/360 control cards for BIOUTIL]

[input data for BIOUTIL]

```
/*
//
```

Using the 'stand-alone' approach, the programmer is relieved of the need of submitting a program deck for each run on the computer and of maintaining a card file of programs. Frequently used programs may be placed on IBM System/360 magnetic disk as part of the DATASYS library (BIOLIB).

Under some circumstances, it is advantageous to be able to use data management programs with scientific Fortran programs. For this reasons DATASYS modules are available to the Fortran programmer by using the CALL statement and an appropriate subroutine as follows:

```
CALL  FORLINK(NAME)
```

where NAME is an argument with a length attribute of eight. This capability, intended for professional programmers, requires detailed coordination of Operating System/360 control information and DATASYS input data.

In all cases, the input and output of program data is coordinated to allow the maximum in system efficiency.

V. GENERAL SCOPE OF APPLICABILITY

DATASYS facilitates the storage and verification of experimental data. With the BIOUTIL module, data can be loaded onto an auxiliary storage device such as a magnetic tape or disk. As was mentioned previously, sequence numbers, if desired, can be a part of the data or be external to the data and maintained by the BIOUTIL system itself. Next, the data can be listed, evaluated, and updated to insure validity. In addition, the whole data file, or parts of it, can be listed, copied, or punched into cards — providing a versatile substitute for procedures that are usually performed manually. In an application where the data is not experimental, BIOUTIL is currently being used to maintain a dictionary for a pathology storage and retrieval system. Dictionary cards contain a medical term plus one or more pathology codes. Sequence numbers are external to the cards and are maintained by BIOUTIL which is used to insert and delete entries when necessary.

Data analysis very often involves the selection of appropriate data, the collating of two sets of related data, or the sorting of files in preparation for a special purpose statistical program. The BIODP module allows these functions to be performed conveniently during a single pass on the computer as well as providing facilities for moving fields within data records and for retrieving experimental data that correspond to generated lists of identifying information. In another recent example, the SELECT, COLLATE, and SORT operations of BIODP were used to develop the mutually exclusive subsets of eight disease states for approximately 1500 experimental animals. For each animal, one card

was punched for each disease and all of the cards were loaded onto disk. BIODP operations were executed repetitively until the desired classes were generated; they were subsequently used as input to statistical analysis programs.

By using the MOVE and SELECT command of the BIODP module along with BIORPG, selective data can be presented in almost any prescribed format. Fields which are not needed may be omitted. Important entries can be placed in a desired position on the printed page and appropriate titles and headings may be generated. Very often, the simple tabulating of experimental data requires a special computer program, which is unnecessary with DATASYS.

BIOPLOT allows data to be plotted on an electro-mechanical plotting device with a minimum of effort. This is convenient for "taking a look" at the data, which is a standard technique in preparation, perhaps, for a regression analysis. Moreover, with BIOPLOT, pictorial reports can be generated almost immediately from either cards or auxiliary data files.

It should be emphasized that a specialized data file generated by the SELECT, SORT, or COLLATE commands of BIODP can be used as an input file to any or all of the following modules:

(i) a statistical analysis program from BIOLIB;
(ii) BIOPLOT for graphic display of data;
(iii) BIORPG for reports containing experimental raw data;
(iv) BIODP for further delineation of data in preparation for analysis; or
(v) BIOUTIL for file maintenance.

In fact, the output of one module can be used as input to any other thereby providing a facility unavailable in most data handling systems.

Many calculations fall somewhere between the desk calculator level and the big computer category. Simple "data description" techniques may very well fall into this class. BIOCALC provides the necessary capability with a minimum of personal investment. It is most useful when long columns of numbers are involved, such as subtracting background noise from experimental data — or just for getting a feel for one's data. The module, which is in the same class as OMNITAB at the NBS, would be useful from a remote console.

VI. CONCLUSIONS

The system has been in operation since September, 1966, with very satisfactory results. The most significant aspect of the situation has been that fewer programmers could produce more results. Duplication has been almost eliminated and lead time has been helped considerably.

DATASYS has also aided in the assignment of computer personnel. Those with a flare for applications and systems design and little interest in detailed coding can still use the computer effectively thereby producing an intermediate link in the chain between programmers and research scientists. Furthermore, inexperienced programmers can now be placed into a productive status with a minimum of training. Experience with DATASYS provides an excellent background for later education in machine coding.

DATASYS was designed to be used by the occasional user or by the nonprogramming scientist. Little information is available on this type of activity but it remains as a potential source of savings in both time and money.

VII. DATASYS PUBLICATIONS

The DATASYS system is described in the following publications released as ORNL-TM-1693:

A SYSTEM FOR BIOMEDICAL DATA MANAGEMENT (DATASYS)

Part I: Concepts and Facilities

Part II: BIOUTIL

Part III: BIODP

Part IV: BIOPLOT

Part V: BIORPG

Part VI: BIOCALC

Part VII: Systems Information

ORNL-4064
UC-32 — Mathematics and Computers

INTERNAL DISTRIBUTION

1. Biology Library
2-4. Central Research Library
5-6. ORNL — Y-12 Technical Library
Document Reference Section
7-26. Laboratory Records Department
27. Laboratory Records, ORNL R.C.
28. N. B. Alexander
29. J. A. Auxier
30. R. E. Biggers
31. Barbara S. Bishop
32. A. A. Brooks
33. J. R. Buchanan
34. W. R. Busing
35. R. R. Coveyou
36. Arline Culkowski
37. W. Davis, Jr.
38. M. Feliciano
39. T. B. Fowler
40. D. A. Gardiner
41. D. G. Gosslee
42. W. L. Griffith
43. G. K. Haeuslein
44. T. L. Hebble
45. A. S. Householder
46. N. C. Hull
47. W. H. Jordan
48. K-25 Library
49. H. Katzan, Jr.
50. S. V. Kaye
51. B. Kuperstock
52. C. E. Larson
53. M. E. Laverne
54. Elmon Leach
55. W. W. Lee
56. H. G. MacPherson
57. R. P. Milford
58. S. E. Moore
59. Paul Nelson
60. C. W. Nestor, Jr.
61. J. S. Olson
62. P. Opstrup
63. B. C. Patten
64. Ruth Sanders Prince (K-25)
65. Ernest Schonfeld
66. M. J. Skinner
67. J. G. Sullivan
68. V. R. Rao Uppuluri
69. K. L. Vandersluis
70. A. M. Weinberg
71. C. S. Williams

EXTERNAL DISTRIBUTION

72. J. A. Swartout, 270 Park Avenue, New York 17, New York
73. J. H. Barrett, Mathematics Department, University of Tennessee
74. R. A. Charpie, Union Carbide Corporation, New York, New York
75. Research and Development Division, AEC, ORO
76-361. Given distribution as shown in TID-4500 under Mathematics and Computers category (25 copies — CFSTI)