Annals of Economic and Social Measurement, 1/2, 1972

SOCIAL SCIENCE COMPUTING AT THE UNIVERSITY OF WISCONSIN: SIMS AND SEOSYS

BY MAX E. ELLIS

INTRODUCTION

For the past three years, the Data and Computation Center for the Social Sciences (DACC) at the University of Wisconsin has been engaged in developing software for social science applications. The main effort has been research and development of systems for describing and processing hierarchical data files. Emphasis has been placed on the design of user languages for describing data already in machine readable form and on the development of efficient algorithms and systems for retrieval and editing of large data files. Two such systems are described in this paper. SIMS, a Social Science Information Management System, is now under development and is our ultimate goal in providing the social scientist with a complete modular and transportable system for processing complex structured files. SEOSYS, the Survey of Economic Opportunity System, is a system developed specifically for retrieval of information from the Survey of Economic Opportunity data files and has been used as a model for the design and implementation of SIMS.

The University of Wisconsin has a Univac 1108 system with batch terminals at remote sites throughout the University. DACC has a Univac 9200 computer serving as an Input/Output terminal to the 1108. The 9200 communicates with the 1108 via coaxial cable and provides card I/O and printing at the social science building site. Magnetic tape files are stored at the central 1108 site and are accessible to all remote terminals. The 1108 hardware configuration consists of a central processing unit, 4 memory units of 65K 36 bit words each, 2 Fastrand II drum storage devices consisting of 22 million words each, 4 flying head drums consisting of 262K words each, 10 tape drives, a printer, card reader, punch and the communication devices to handle the more than 10 remote batch terminals.

The minimum computer system configuration in which SIMS can operate must have the following attributes:

—A multi-processing capability with facility for creation and execution of a job control stream from a user program.

—An ANSI Fortran IV or a comparable Fortran compiler which through Fortran system routines or special routines called from a Fortran program, allows I/O to a random access device such as drum or disk. Also needed are I/O functions comparable to the UNIVAC or CDC Fortran BUFFERIN, BUFFEROUT, DECODE and ENCODE [3, 5].

—Provides users with an equivalent of 50K 36 bit words or greater core for the program and common block and at least an equivalent of one million 36 bit words of random storage.

—Allows collection or mapping of precompiled relocatable routines, routines compiled at execution and labelled common blocks.

—A compiler for ANSI Cobol.

—At least 3 tape units are required for certain processing functions.

SEOSYS, described in the last section, is written in Fortran and only requires the hardware normally made available to standard Fortran programs. Since all SEOSYS I/O is tape, no use is made of the random storage devices. The size of SEOSYS is well within the limitation of 65K words set by the Fortran compilers.

SIMS: A Social Science Information Management System

SIMS incorporates a number of integrated processing functions for the complete processing of simple and complex data files consisting of fixed length data items. Facilities exist for describing hierarchical structured files which are already in machine readable form and for the complete editing of such files [2]. These two basic functions are complemented by a series of analytical functions such as cross-tabulation, correlation, etc. The modular construction of the system enables additional analytical routines to be added, including user supplied Fortran subroutines. The user oriented command language of SIMS provides the social science researcher with an interface to the system which is familiar to him. The syntax *and* semantics of this language may be easily altered by a programmer to handle any idiosyncrasies in the terminology used by a particular class of users, or to change the user interface entirely to conform to users other than the social scientist.

Figure 1 is a sample SIMS request with explanations of the input statements. It provides a general feel for the system and some properties of the language. This example combines a number of different processing functions in one request or job. The user has survey data on cards and is using SIMS to "familiarize" himself with his data. Assume this is the first time the data has been processed by the computer. In a single SIMS run, the researcher can describe the data (*DESCRIPTION), validate and perform consistency checks on data items (*EDIT) and produce some preliminary cross-tabulations (*CROSSTABS).

An input request may be catalogued and retrieved at a later date for updating or execution. The file description may be entered into the SIMS library and stored in machine readable form. When the file described is referenced in subsequent runs (using the *INPUT statement) the file's description is automatically retrieved and made available to the SIMS retrieval and analytical routines.

Initially SIMS will be limited in its statistical analysis capability since this type of processing is readily available via other systems or generalized routines and the file handling features of SIMS provide for complete editing, reformatting and extracting of data for such statistical programs. The main objective of SIMS is to provide a researcher with a file processing tool that he can use without the aid of a programmer. Figure 2 is a list of the commands for the first SIMS system. Details on the parameters of each statement are not given but the brief descriptions of each should serve to summarize the features of SIMS.

The first version of SIMS is scheduled for release by the end of 1972. This version will be batch operational and will run under the EXEC 8 operating system of the Univac 1108. Most routines have been written in ANSI Fortran IV or Cobol with additional DACC Fortran coding standards applied [1].

A generalized system for implementing applications software systems has been developed for the implementation of SIMS. LENS (Language intErface with Natural Semantics) [4] is a system which writes or generates programs from input

SAMPLE SIMS REQUEST

```
*BEGIN, USER=SMITH,ACCOUNT=2908,MODE=PROD,RUN-ID=1971-SURVEY-TABLES

*TITLE ANALYSIS OF SURVEY DATA

*INPUT, 1971-SURVEY

*EDIT, TYPE=OBSERVATIONS,MAX-ERRORS=100

 VALIDATE VARIABLES SEX,INCOME,AGE,OCCUPATION(HEAD)
 CHECK, IF(SEX(HEAD)•IS•MALE•AND•AGE(HEAD)•GT•21•AND•INCOME•GT•2000)
 CHECK, IF(SEX(HEAD)•IS•MALE•AND•V6•EQ•1•AND•SEX(SPOUSE)•IS•FEMALE)

*CROSSTABS:  CELLS ARE FREQUENCIES, ROW-PERCENT, COLUMN-PERCENT

 TABLE, ROW=OCCUPATION, COLUMN=SEX(HEAD)
 TABLE, ROW=OCCUPATION, COLUMN=WORK-CODE, PAGE=SEX(HEAD), CELLS ARE FREQUENCIES

*DESCRIPTION, FILE-NAME=1971-SURVEY

 ABSTRACT:  1971 SURVEY OF HEADS OF HOUSEHOLDS IN S.E. WISCONSIN
            THIS DATA OBTAINED FROM DEPT. OF WELFARE.
 STORAGE-DESCRIPTION:  STORAGE-DEVICE=CARDS
                       RECORD-IDENTIFICATION=CARD-NO
 OBSERVATION-IDENTIFICATION:  NAME = HEAD
                              ID = HEAD-NUMBER
 RECORD-DESCRIPTION:  NAME = HEAD, CARD-NO = 1
  VARIABLE 1:  NAME = HEAD-NUMBER, FORMAT = 1/I4
  VARIABLE 2:  NAME = CARD-NO, FORMAT = 5/I1
               BOUND 1:  NAME = HEAD-CD, VALUE = 1
               BOUND 2:  NAME = SPOUSE-CD, VALUE = 2
               BOUND 3:  NAME = HEAD-CASH, VALUE = 3
  VARIABLE 3:  NAME = SEX, FORMAT = 6/I1
               BOUND 1:  NAME = MALE, VALUE = 1
               BOUND 2:  NAME = FEMALE, VALUE = 2
  VARIABLE 4:  NAME = AGE, FORMAT = 7/I2  DETAIL = 00 IMPLIES NO AGE GIVEN
  VARIABLE 5:  NAME = RACE, FORMAT = 9/A5
  VARIABLE 6:  NAME = MARITAL-STAT, FORMAT = 14/I1
               BOUND 1:  NAME = MARRIED, VALUE = 1
               BOUND 2:  NAME = SINGLE, VALUE = 2
  VARIABLE 7:  NAME = WORK-CODE, FORMAT = 15/I1
               BOUND 1:  NAME = NOT-WORKING, VALUE = 0
               BOUND 2:  NAME = WORKING,     VALUE = 1
  VARIABLE 8:  NAME = OCCUPATION, FORMAT = 16/I2  DETAIL = NOT ALL OCCUPATIONS ARE GIVEN
               BOUND 1:  NAME = BRICKLAYER, VALUE = 1
               BOUND 2:  NAME = CARPENTER, VALUE = 2
               BOUND 3:  NAME = OTHER, VALUE = 3-98
               BOUND 4:  NAME = MISSING, VALUE = 99
```

EXPLANATION OF STATEMENTS

This is a production run for SMITH, the input stream catalogued under account 2908 and the given run identification 1971-SURVEY-TABLES. This title appears on all pages of printer output.

Input is the 1971-SURVEY file described under *DESCRIPTION.

Validate the codes for the variables listed and perform the consistency checks stated. Continue until MAX-ERRORS=100. Check each entry or observation and print error message if expression is false.

Produce the following two contingency tables giving frequencies of occurrence (or counts) and percentages. The second table is 3-dimensional. For table 2 the global parameters of the *CROSSTABS statement are overridden by CELLS ARE FREQUENCIES.

The survey file is on cards with 1 to 3 cards per observation or entry depending whether a spouse is present and if head worked. Cards are identified by CARD-NO, and observations by HEAD-NUMBER. Card 1 is HEAD infor., Card 2 is SPOUSE and 3 income infor. of HEAD. The statements between *DESCRIPTION and *DATA are substatements of the Data Description Language.

The FORMAT is the "starting column"/"Fortran Format". The BOUND is the code or value of a variable or item. The HEAD-NUMBER appears in Cols. 1-4 on every card or record. The CARD-NO. is in Col. 5 of every card. VALUES may be referenced by their name, e.g. SEX IS MALE. VARIABLES may be referenced by their 12 char. name or unique number.

A detailed description of a variable may be given and continued on additional cards if necessary (e.g. AGE on the left).

MARITAL-STAT indicates if SPOUSE card should be present. If SPOUSE present and this VALUE = 2 then a validation error will be indicated.

WORK-CODE indicates if HEAD-CASH card present. Only one SPOUSE card and one HEAD-CASH card may appear for a HEAD. This is stated in the STRUCUTURE-DESCRIPTION.

If OCCUPATION was not given a MISSING VALUE of 99 was assigned.

Figure 1

SAMPLE SIMS REQUEST

```
RECORD-DESCRIPTION:  SPOUSE, 2 (SAME AS HEAD RECORD VARIABLE 1-5)

 VARIABLE 9:  HEAD-NUMBER 1/I4  WHO SPOUSE BELONGS TO

 VARIABLE 10: CARD-NO     5/I1  CARD/RECORD IDENTIFICATION
              BOUND 1:  HEAD-CD     1     HEAD DEMOGRAPHIC INFO.
              BOUND 2:  SPOUSE-CD   2     SPOUSES DEMOGRAPHIC INFO.
              BOUND 3:  HEAD-CASH   3     MONEY VALUES IF HEAD WORKED

 VARIABLE 11: SEX        6/I1
              BOUND 1:  MALE       1
              BOUND 2:  FEMALE     2

 VARIABLE 12: AGE        7/I2      AGE IN YEARS

 VARIABLE 13: RACE       9/A5      THE VALUE IS ALPHANUMERIC
              BOUND 1:  1          WHITE
              BOUND 2:  2          BLACK
              BOUND 3:  3          OTHER

RECORD-DESCRIPTION:  HEAD-CASH,   3

 VARIABLE 14: INCOME          40/F10.2      GROSS INCOME/YEAR

 VARIABLE 15: ASSETS          50/F10.2      TOTAL ASSETS

 VARIABLE 16: LIABILITIES     60/F10.2      TOTAL LIABILITIES

STRUCTURE-DESCRIPTION:  HEAD RECORD IS FOLLOWED BY HEAD-CASH RECORD
                        IF WORK-CODE IS WORKING ELSE IS FOLLOWED BY SPOUSE RECORD
                        IF MARITAL-STAT EQUALS 1.
                            SPOUSE RECORD IS FOLLOWED BY HEAD RECORD.
                            HEAD-CASH RECORD IS FOLLOWED BY SPOUSE RECORD.
                        IF MARITAL-STAT EQUALS 1 ELSE IS FOLLOWED BY HEAD RECORD.
*DATA, FILE-NAME = 1971-SURVEY
    (Data Cards for 1971 Survey File)

*END
```

EXPLANATION OF STATEMENTS

The SPOUSE record description appears with parameter names missing and with more detail description for each variable. SPOUSE cards are coded "2" in the record ID code or CARD-NO. This value, 2, is listed on the RECORD-DESCRIPTION statement. Delimiters are optional as only "blanks" are required.

Note that RACE value names are numerals and the actual coded values are names.

Bounds need not be specified as can be seen from these continuous money values.

The STRUCTURE-DESCRIPTION describes the logical relation among the 3 cards or record types. Note that HEAD Card can be followed by any one of the 3 card types.

SPOUSE Card can be followed only by a HEAD Card, and HEAD-CASH Card by a HEAD or a SPOUSE Card depending on marital status.

A data card file must be preceded by an *DATA statement. The file name must agree with that appearing on the *INPUT and *DESCRIPTION statements.

End of SIMS input request.

Figure 1 (Continued)

SIMS Statements

Control Statements	
*BEGIN	This statement precedes each SIMS request and identifies the user and job.
*END	A SIMS request is terminated by *END. More than one request may be submitted and is identified by a beginning *BEGIN and an ending *END.
*LIST and *NOLIST	These statements if embedded in the input request either turn on or turn off a listing of the input request cards.
*REMOVE	All input requests are catalogued under the RUN-ID of the *BEGIN statement (if present) and are removed or uncatalogued with the *REMOVE statement.
*USER	Users Fortran subroutines must be preceded by this statement.
*DATA	Data on cards submitted as part of the SIMS input request are preceded by this statement.
Input Statements	
*INPUT or INPUT	This statement identifies the file that is to be the input to the processing function or functions specified. The major statement, *INPUT, is global to all processing functions unless a major processing function statement (e.g. *CROSSTAB, *EDIT etc.) is followed by an INPUT statement (no asterisk). Then the file listed on the INPUT statement will be used as input to the function requested.
*SELECT or SELECT *OMIT or OMIT	These statements are used to select or omit observations and/or variables for processing. The same global relation as explained for *INPUT and INPUT applies to these statements.
*DESCRIPTION	This statement and its 12 substatements (Not listed here) represent the Data Description Language (DDL) of SIMS.
*DEFINE	This statement and its 4 substatements (Not listed here) represent the SIMS variable redefinition capability. The statements are used to recode variables or compute new variables, and define and assign values and value names to variables.
*RESTRUCTURE	This statement is analogous to the STRUCTURE statement of the DDL (See the sample request). It enables the logical structure of the file to be respecified at execution time thereby increasing the retrieval efficiency.

Figure 2

Output Statements	
*OUTPUT or OUTPUT	This statement identifies an output file. The same global relation as explained for *INPUT and INPUT applies to this statement.
*TITLE	The title specified appears on every page of output.
*EDIT	This statement and its 2 substatements (Not listed here but appearing on the sample) represent the SIMS file validation and consistency checking capabilities. Edit operations on the input file specified include validation of 1) observation structure, 2) variable formats, and 3) variable codes or values and consistency checking among variables.
*UPDATE	This statement accompanied by update transaction cards provides a means for deleting, adding or correcting observations or variables of the input file specified.
*DUMP	Records of the input file specified are dumped or printed in readable form in a format dependent on the recording mode of the file and options specified by the user.
*COPY	The input file specified is copied in the same format.
*EXTRACT	This statement specifies conditional extraction of observations or variables producing subpopulations of the input file specified.
*SORT	Version 1 of SIMS assumes serial or sequential processing of data, Sorting of a specified input file is specified using the *SORT statement.
*MERGE	Two files of the same logical structure may be merged. The merge criteria and "hit-miss" options are specified on this statement.
*SAMPLE	A random sample or a sample which includes rare occurring values for variables is produced from the input file specified. The variables used as the sampling criteria are listed as part of this statement.
*MARGINALS	One-dimensional or marginal frequency distributions on variables are produced from the options and variable list of this statement.
*CROSSTABS	N-dimensional tables of frequencies, means, sums, standard deviations, row percentages or column percentages are specified using this statement as well as associated statistics such as chi-square, variance, standard deviation, etc.
*MOMENTS	Raw moment matrices or matrices of selected variables sums and sums of cross-products are produced from the options and variables listed in this statement.
*CORRELATIONS	Correlation matrices or selected variables are produced from the variables in this statement.
*TEACH	When in batch mode sections of the SIMS machine readable documentation will be printed according to the problem areas the user has indicated as part of this statement. In interactive or on-line mode this statement initiates an interactive teaching function in which the user answers questions relevant to this problem. The interactive TEACH function will not be available in version 1 of SIMS.

Figure 2 (Continued)

describing the *source* language (SIMS statements) and the *target* language (the generated or precompiled SIMS job stream which is to be executed). Rules are given to LENS for the mapping of the source language to the target language. In the case of SIMS, the rules are the complete description of the SIMS command language. For some other general applications program the rules would be the description of the resultant program's control cards and control card processing. During the mapping process detailed error messages are printed as statements of the source language are checked for syntax errors, completeness and order. Statements of the target language are stored on a random access device for later execution. This then completes the LENS processing.

In summary, the SIMS system is composed of generalized relocatable routines such as an EDIT print routine, a cross-tabulation subroutine etc., and LENS macros and nets which describe the source and target languages or SIMS statements and generated job stream, respectively. Each user has access to the entire system and as such can create his own data base of files, file descriptions and library of SIMS requests unique to his application. If he so chooses he may produce his own version of the SIMS request language and associated generated output. This can be done through alteration of the LENS input. The modular construction of LENS and other SIMS routines plus the paging capability of LENS and the host operating system facilitates many SIMS users to run SIMS simultaneously. Finally, SIMS provides both a novice and experienced computer user with a tool for processing simple, complex, large or small jobs in a manner familiar to him.

SEOSYS: A Generalized System for Extraction from and Analysis of the 1966–1967 Survey of Economic Opportunity Data Files

1. *Logical Structure of the* 1966–1967 *SEO Data Files*

The 1966 and 1967 "Surveys of Economic Opportunity" were conducted by the Bureau of the Census at the request of the Office of Economic Opportunity in order to augment the information regularly collected in the Current Population Surveys (CPS) for February and March of each year. In addition to a number of items common to both surveys (such as age, family status, work experience and income), the SEO also provides information on other characteristics such as housing, marital history, training, assets and liabilities. The main purpose in collecting this information was to provide a base for micro-analytic research in exploring the causes and correlates of poverty. The files have been specially designed, edited and documented to this end.

The 1966 SEO sample consisted of about 30,000 households and was made up of two parts: (1) a national sample (about 18,000) drawn in the same way as the Current Population Survey Sample and (2) a supplementary sample (about 12,000) in areas with a large concentration of nonwhites. The sample was designed in this way to improve estimates of the characteristics of the poor, in particular, the nonwhite poor.

The 1967 SEO sample consisted of reinterviews of the same addresses included in the 1966 SEO. Most of the questions asked in 1966 were asked again in 1967 making some measures of change possible for persons interviewed in both years.

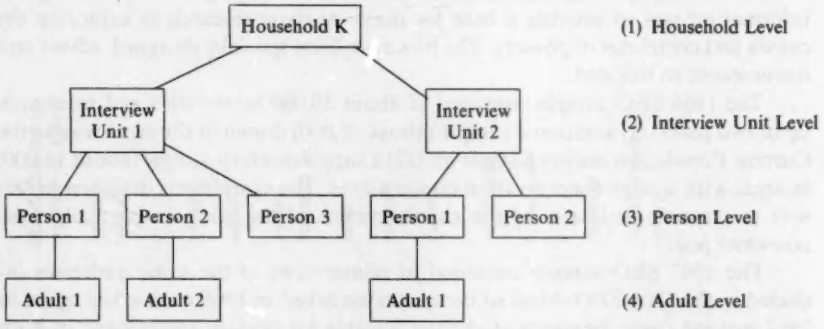
Each SEO file is described in detail by a codebook which contains a complete description of the substantive content of the file, a file layout, a list of the file attributes and their possible values, and extensive textual documentation relevant for users of the file. The codebook describes the four types of "segments" contained in the file. The index provides a cross-reference table which lists all the items on both the 1966 and 1967 files, their question numbers, the "segment" in which they will be found, and the character positions in which they are located on the magnetic tape file. If an "attribute" is the same from year to year, it will occupy the same place in each file; in fact, all the physical characteristics (segment length, etc.) are identical for both files.

The four segments contained in each SEO file are called Household (HHLD), Interview (INTV), Person (PERS), and Adult (ADLT). A household segment exists for every address included in the SEO sample regardless of whether an interview was obtained there. This segment contains all the questions or attributes common to a sample address, e.g., housing information and the geographic and sample codes.

For addresses at which an interview was obtained, the household segment is followed by one or more interview segments. An interview unit is a family or an individual not a member of a family. The INTV segment contains all the information which pertains to the whole interview unit, such as responses to questions on assets, liabilities, monthly rent, and income other than earnings.

Within each interview unit there is a person segment for every person in the unit containing information on age, sex, race, educational attainment and family relationship. For adults (persons 14 years or older) there is also an adult segment containing information on last year's work experience and earnings and other special information collected for that year. In the 1967 file there is also a person segment for people interviewed in 1966 who were not in the interview unit when it was reinterviewed in 1967 (although there was no indication that they had left the unit during 1966). Only a limited amount of information is available for these people in the 1967 file.

The enumeration unit in each of the Survey of Economic Opportunity files is the household, or address. Each household interviewed contains information about the household, the interview units (families) within it, the people in each



SAMPLE TWO-FAMILY ENUMERATION UNIT

interview unit, and "adult" information for some of these people. It may be useful to think of the information for each SEO household or address as organized within a 4-level structure with the segments of information for the household connected by a simple "tree structure" as illustrated opposite. The four levels implicit in the structure each contain one of the four segment types within the file.

2. *Physical Characteristics of the* 1966–1967 *SEO Data Files*

Although the tree structure is conceptually useful for describing the organization of the file, the organization of the file on magnetic tape is sequential. Segments for each household appear on the tape file in "left list" order, i.e.. that sequence in which they occur when the tree structure is traversed from left to right along its branches. For the above example, the segments would appear on magnetic tape in the following order:

Segment	Level	Content
HHLD	1	Household data
INTV	2	Interview Unit No. 1 data
PERS	3	Person 1 data
ADLT	4	Adult 1 data
PERS	3	Person 2 data
ADLT	4	Adult 2 data
PERS	3	Person 3 data
INTV	2	Interview Unit No. 2 data
PERS	3	Person 1 data
ADLT	4	Adult 1 data
PERS	3	Person 2 data

Segments describing a given household are contiguous on the file. Non-interview households are represented by a household segment only.

Input to SEOSYS must be either the 1966 or 1967 SEO as produced by the Data and Computation Center. These versions of the SEO files contain fixed length physical records or blocks with a record being 96 numeric BCD (Binary Coded Decimal) characters. Blocks contain 30 records each and records of a household may continue over more than one block.

3. *Data Retrieval*

SEOSYS provides an efficient means for retrieval, extraction and analysis of information from the SEO data files. Most standard statistical programs or systems are not capable of directly processing files with complex structures such as SEO. They usually require the data to be of a "rectangular" or matrix structure, in which the columns of the matrix are the variables and the rows the observations. Most often an observation is synonymous with a tape record or card. SEOSYS bridges this gap by retrieving information from the hierarchical tree structure (as illustrated in the sample household) and creating a rectangular file for analysis. This reformatting or structure change may be combined simultaneously with

analysis, or may be done separately by producing an extract or work file which is a subpopulation of SEO to be analyzed at a later date.

Pertinent physical characteristics of the SEO tapes are provided to SEOSYS via an abbreviated machine readable version of the SEO codebook. Using this information and "knowing" the possible tree structures of households in the file, SEOSYS is capable of retrieving attributes from any of the four levels, household, interview, person or adult. The user specifies at what level his analysis will be. SEOSYS then searches a "household tree," "remembering" at what level the analysis will be based and retrieves the attributes or variables to be selected from *any* level. A fixed length observation vector containing these data items from any level is then created, one observation for the level of analysis.

Consider a study of all persons in the survey who are black, have incomes less than $3,000 and who live in multi-family dwellings. The *unit of analysis* or level of analysis in this case is the *person*. Therefore an observation possibly containing information from all levels (e.g., HOUSE SIZE from the HHLD, RACE from the INTV, AGE from the PERS and INCOME from the ADLT) would be created for every person who satisfies the selection criteria. SEOSYS, as it is traversing a household, "saves" attributes or variables from higher levels (e.g., HHLD and INTV) if need be and "looks ahead" for data from lower levels (e.g., ADLT). During this retrieval process searching is terminated immediately if the data interrogated do not satisfy the selection criteria, thereby minimizing retrieval time. For the example request mentioned above, if the household being queried consisted of only one family, the attribute #FAM (number of families) of the household record or segment being equal to 1 would indicate to SEOSYS that persons in this household should not be included in the sample. Any further checking of race or income etc., would be omitted and SEOSYS would then search for the beginning of the next household.

Most analyses performed on survey type data files require some transformation of the data in the master file, creation of new variables or conditional extraction or selection of a sample population. The SEO files are no exception. Because of the extensive amount of information for a household and the complex structure of the files, users of the SEO data will almost always require some form of data transformation to create a subpopulation analysis. SEOSYS allows a user complete interaction with the system through user supplied Fortran subroutines. Such routines facilitate transgeneration of variables at all levels and selection of observations. A user may also perform his own analysis in these supplied routines.

SEOSYS has been developed specifically for the purpose of providing a researcher with a user-oriented system for accessing, extracting, and analyzing data of the Surveys of Economic Opportunity. Since SEOYS has been custom designed for these data files, the retrieval algorithm in SEOSYS provides efficient access to the data while giving users a general system for processing the data. The general features of SEOSYS allow almost any request to be handled with minimal computer time and little or no programming time.

4. *Documentation Available*

The following documents are available free through the University of Wisconsin Data and Computation Center:

—1966 Survey of Economic Opportunity Codebook
—1967 Survey of Economic Opportunity Codebook
—1966 and 1967 Survey of Economic Opportunity Sample Design and Weighting
—The Comparison of Selected Economic and Demographic Characteristics from the 1966 and 1967 Surveys of Economic Opportunity and the Comparable Current Population Surveys
—1966 Survey of Economic Opportunity Unweighted Counts (Including weighted estimates of Income, Asset and Liability items)
—1967 Survey of Economic Opportunity Unweighted Counts (Including weighted estimates of Income, Asset and Liability items)
—1966 and 1967 Survey of Economic Opportunity Sample Variance Estimates
—1966 and 1967 Survey of Economic Opportunity Cross-Year Tabulations
—SEO Data Files—Fixed Length Format
—SEOSYS: A Generalized System for Extraction from and Analysis of the 1966–1967 Survey of Economic Opportunity Data Files—Users Manual

The documents listed above and others have been compiled by E. JoAn Olson into the *Survey of Economic Opportunity Bibliography*. The bibliography is in machine readable form and is printed by the computer via the indexing system, UWIS, developed by the Madison Academic Computing Center at the University of Wisconsin.

The list of documents is indexed by author and documents with more than one author appear once for each author. The entries of the bibliography have been assigned to one of the following categories:

(1) User Guide
(2) Thesis (B.A.)
(3) Thesis (M.A.)
(4) Thesis (PhD)
(5) Forthcoming
(6) Working Paper
(7) Published
(8) Conference
(9) Other

The category name appears on the listing. The bibliography has also been indexed by key title words.

Acknowledgments

The SIMS system has been funded entirely by the National Science Foundation, grant GS-1992, and has been under the faculty direction of Professor Dennis Aigner, with Max Ellis directing the system design and implementation. Significant contributions in development of the system have been made by the following senior staff members of DACC: William Katke, Kenneth Nelson, James Olson and Shou-chuan Yang. These persons with Max Ellis have designed the system, its user interface and programs.

The development of SEOSYS was funded by the Office of Economic Opportunity and the Institute for Research on Poverty. Programming of the system was done by Kenneth Nelson, Linda Werner, and Luise Cunliffe. Nancy Williamson and Ronald Sepanik contributed significantly to the design of the system and

assisted the programmers in the testing of SEOSYS. The portion of this paper pertaining to the Survey of Economic Opportunity includes contributions from Ronald Sepanik and David Richardson. Descriptions of the SEO data files have been reproduced in part from *The 1966 and 1967 Survey of Economic Opportunity Files and Related Software*, Brookings Computer Center, Memo #48, June 30, 1969 by George Sadowsky and Marjorie Reed.

University of Wisconsin

References

[1] Ellis, Max E. *Fortran Coding Standards*, Data and Computation Center Technical Paper (TP-15), University of Wisconsin, Madison, Wisconsin., December, 1970.

[2] Ellis, Max E. and K. H. Nelson. *A Data Description Language for Hierarchical Data Files*, Presented at ACM SICFIDENT workshop on Data Description and Access, Reprinted as Data and Computation Center Paper (TP-11), University of Wisconsin, Madison, Wisconsin, August, 1970.

[3] Control Data Corporation, 3400/3600/3800 *Computer Systems Fortran Reference Manual*, Publ. No. 60132900, A, 1965.

[4] Katke, William. *LENS Reference Manual-Preliminary*, Data and Computation Center Working Paper, University of Wisconsin, Madison, Wisconsin, August, 1971.

[5] UNIVAC. *Fundamentals of Fortran*, UP-7536, October 14, 1968.