**THE SURVEY OF INCOME AND
PROGRAM PARTICIPATION**

**SURVEYS-ON-CALL:
ON-LINE ACCESS TO
SURVEY DATA**

No. 229

**Stacey Furukawa
Enrique Lamas
Judith Eargle
Census Bureau**

U.S Department of Commerce BUREAU OF THE CENSUS

SURVEYS-ON-CALL:
ON-LINE ACCESS TO SURVEY DATA

by

Stacy Furukawa, Enrique Lamas, and Judith Eargle
Housing and Household Economic Statistics Division
U.S. Bureau of the Census

August 9, 1995

Surveys-On-Call: On-Line Access to Survey Data

Surveys-On-Call is an on-line system for creating extracts from selected micro-data files available from the Census Bureau. Using Surveys-On-Call researchers can download public use data via the Internet or through dial-in access. Traditionally, the Census Bureau has made public-use micro-data files available only in tape or cartridge format. These files are typically large, requiring relatively large computer capacity and relatively extensive programming experience to process these files and prepare analytical data files. Surveys-on-Call is the first Census Bureau software through which public-use data are made available to users via on-line computer systems. The system allows users to create extracts of public-use data files and then download the extracted data onto their own computer systems for further processing and analysis. Surveys-On-Call is a menu- driven user-friendly system that allows users to easily extract data, thereby reducing the learning curve previously necessary for processing tape or cartridge files. Users can download the datasets and use software with which they are familiar on their own computer systems to analyze the extracted data.

Surveys-On-Call evolved from an earlier dial-in stand-alone computer system called SIPP-On-Call. The SIPP-On-Call system was developed under a contract with the Food and Consumption Service (FCS) and was intended to provide FCS and other users with an online computer system to access Survey of Income and Program Participation (SIPP) data files. The SIPP-On-Call system had only a dial-in capacity and had only SIPP public-use data files available. As the use of this system expanded, it was moved to the Internet and other household surveys were added to the system, specifically the March Demographic Supplement to the Current Population Survey (CPS), the American Housing Survey (AHS), and the Consumer Expenditure Survey (CES). In addition, integrating the software as part of the Internet access made the system more widely available and the transfer of large data sets much easier. Concurrent with the expansion of the data sets to other surveys and to its integration as part of Internet products, the name of the system was changed to Surveys-on-Call.

The purpose of this paper is to describe the Surveys-On-Call system for users and to present a simple users' guide that will assist in the use of the system.[1] The following sections provide a description of the system and general instructions on use of the system.

I.   Connecting to Surveys-On-Call

Access to Surveys-On-Call can be either through the user's Internet account or through dial-in with a modem.[2] The system is designed to be accessed through a character based interface

---

[1]For detailed step-by-step instructions for the Surveys-On-Call system, contact the Surveys-On-Call support staff on (301) 763-8375.

[2]The SIPP-On-Call dial-in system required users to apply for an account before accessing the data. This is no longer necessary. Dial-in users now have access to Surveys-On-Call through what is in effect a temporary Internet connection.

(Telnet or dial-in) or through a World Wide Web (WWW)/Hyper Text Mark-Up Language (HTML) browser interface (NetScape or Lynx). For a character interface, Internet users can telnet to the Bureau's site using the command:[3]

telnet gateway.census.gov

Users will then be asked to log in. At the prompt, the user designates "desuser" as a login name. The system then queries the user's full Internet address. This address is then used to e-mail notification of job completion and where the dataset is stored. Having responded with an e-mail address the user returns to the main Surveys-on-Call menu "DES - Data Source Selection Menu" (see Screen 1).

For access through a WWW/HTML browser interface, Surveys-On-Call can be accessed using Mosaic, Lynx, or Netscape. Internet users with Lynx software can log in using the command:

lynx http://www.census.gov/

This provides users with a non-graphics version of the Census Bureau's WWW home page. The user can select "Data Access Tools" and connect to Surveys-On-Call using the HTML access.

Internet users with Netscape[4] can use the general Census Bureau address:

http://www.census.gov/

Dial-in users can access the system through a telephone modem.[5] After dialing in, there will be a screen with three options:

1 - Connect to the data extraction system
2 - View the file storage directory
3 - Transfer files

To create an extract, the user would select item number 1 which would bring the main data source selection menu.

---

[3]Please note that the Bureau's Internet staff recently posted a version 4.2 of the extraction system software, in addition to their first system version on World Wide Web. Both systems are currently being tested.

[4]Netscape 1.1 or a comparable browser with HTML forms capability is recommended. If the browser lacks HTML forms capability, a more recent version must be installed or the user can access SOC through Telnet.

[5]The telephone number for dial-in users is 301-457-2341.

II.   specifying an Extract
II.   A Telnet Access

        The first step in the creation of an extract is to select the file to use. At the data source selection menu, the user selects the number corresponding to the specific file. The actual datasets available on Surveys-on-Call are identical to the public use files available on tape, cartridge, or CD-ROM through the Bureau's Data Users Services Division. The computer system used for Surveys-On-Call has safeguards built in so that confidential data cannot be accessed through this system.

        The next screen "DES - Operations Menu for: dataset" (see Screen 2) is the main instruction menu for the dataset selected. Item 1 ("User Notes/Caution Statements") contains the public use data dictionary, notices of corrections/deletions/additions (user notes), and other information specific to this file.[6]

        To request an extract, the following items on the operations menu must be specified:

    2.   Set Parameters - used to select the number of records (for example, 10, 100 or MAX for all), the output file format(SAS, ASCII), plus two other choices for compression type and selection logic;
    3.   Set Input File - used to select the files for extract creation;
    4.   Set Output Variables - used to select the variables to be extracted;
    5.   Selection Criteria - (Optional) - used to subset the records included in the extract; and
    6.   Perform Data Extraction.

        Additional details on the steps that are necessary to create an extract are provided below. The same steps are followed whether the extracts are from the SIPP (core cross-sectional, topical module, or longitudinal), CPS (person, family, or household), AHS, or CES files.[7]

II.A.1 Setting Parameters

        Four parameters are specified in this section (Screen 3-"DES-Parameters for: dataset"): the number of observations to be extracted (parameter MR), the output file compression (parameter OC), the format of the output file (parameter OF), and selection criteria processing logic (parameter VS). The system defaults are as follows: 10 observations, UNIX compressed form, the output file type to an ASCII "BOX" file, and the selection criteria processing as

---

[6]Menu item 11 is labeled "Documentation - File Content". This item contains a truncated version of the dictionary used by the Surveys-On-Call code to load files and does not contain all the information necessary for documentation.

[7]Each time a job is specified all the information must be entered. The system reverts to default values at the completion of an extract.

"AND".

To change any of the parameter defaults, menu item 2 "Modify Parameters" can be used. The parameter to change is entered; then the new parameter setting is selected. For example, to request that all observations be used, the parameter "MR" is selected and then the value of this parameter is entered as "MAX". For testing extracts, it is recommended that users extract a small file to assure that the output file type is compatible with their systems and that file variables are suitable for user analysis.

Menu item 4 "Documentation of Parameters" is used to view what options are available for each of the parameters. For example, the following is additional information on the values listed under OF:

SAS1 A UNIX System 5 Sun 4 Version SAS file
SAS2 A SAS transport file (CPORT, Version 6.09)
SAS3 A SAS transport file (XPORT, Version 6.09)
BOX1 An ASCII "BOX" file
      This is a specially formatted ASCII file which contains data and all descriptive information required for processing and documentation in one file. The BOX format allows data to be moved in ASCII text form across diverse hardware and operating systems, software applications, and programming languages, without losing descriptive information about the data.[8]
SQL1 A tab-delimited ASCII text file with Structured Query Language (SQL) program file. This 2-file output combination consists of an ASCII text data file delimited by "TAB" characters and an ASCII text SQL program file. The data file is compatible with most spreadsheet and database import/export features. The SQL program, though generated specifically for Ingress Database software, could be modified to load other SQL database software as well.

The user can also specify the compression type performed on the output extract data file. Compression reduces the size of the file significantly (to approximately one-tenth) and has benefits in storing and transferring data files. The user can specify UNIX, PKZIP, GZIP, or no compression.

The last parameter, the VS parameter, affects how the selection criteria (an optional step, operations menu item 5) are processed. The default is "AND"--that is, each criteria is linked by "and" logic. For example, if two subsetting commands were entered--age=65 or older, sex=female--the "AND" logic would output all women 65 or older. If the VS parameter were changed to "OR", then the output file would contain men 65 or older and all women regardless of age.

The menus for the other functions all share the same general structure. Item 1 shows you

---

[8]Additional documentation on this format type is available on request from the SIPP staff (301-763-8375).

the current settings, item 2 is where you request changes, followed by delete and reset options and documentation associated with the parameters.

II.A.2 Selecting the Input File

Item 3 "Set input file" in the Operations menu is used to specify the input datasets within a chosen survey. A complete list of files currently available for extraction is stored in menu item 10 "Documentation of Files". SIPP data files (Screen 4) can be classified into 3 types:

o   Cross-sectional Core Files - which are cross-sectional files with core labor force, income, and program participation data available from each interview.

o   Topical Module Files - which are cross-sectional data for the supplements or topical modules asked in an interview.

o   Longitudinal files - which are files containing core data from all the interviews matched together for each person. These files include selected information on labor force participation, family changes, health insurance coverage, income received, and participation in assistance programs on a monthly basis for the length of the panel.

All topical modules consist of only one file. For core files, all waves available are listed. For core extractions multiple core input files can be specified. The extract will consist of a single output file with observations from all the waves specified.

II.A.3 Selecting the Output Variables

Item 4 "Set Output Variables" in the operations menu is used to select output variables. The full documentation on variables is contained in the User Notes section. For SIPP files, the current default is set to "primary" variables, a group of less than a dozen variables which usually includes weights, wave, panel, rotation, and person identifiers.

Menu item 2 "Add output Variables" (on Screen 5 - "DES - Output Variables for: dataset") is used to select additional variables to extract. A shortened version of the variables is listed under menu item 5 and menu item 6 allows for a key word search of the file contents. Once all variable names are entered, the system scans your list, and accepts all variables that it recognizes. If there were no errors, the system returns to the "Output Variables" menu. If errors were detected, the list of unrecognized variables will be displayed and variable names can be reentered using menu item 2 "Add Output Variables". These additions are scanned again and, if correct, added to the list.[9]

---

[9]Menu item 4 "Reset variables" appears to give the option of selecting all variables. Currently, due to space considerations, Surveys-On-Call does not allow extracts of the entire dataset. Each variable must be entered individually.

To check the full variable list selected, item 1 "View Output Variables" will list variables included which can be compared against the user's list. If variables are missing, menu item 2 "Set output variables" can be used to add them. They will be appended to your previous list.

If after an extract is created additional variables are needed, another extract can be specified and merged to the original extract. The sample unit ID, entry address ID, and person number variables are particularly important because they are the unique identifiers that will allow files to be merged.

II.A.4 Selection Criteria

A selection criteria can be used to subset the output dataset by any of the variables specified in the "Output Variables" menu.   Multiple conditions can be specified. This selection can help reduce considerably the size of the output dataset. For example, core files are in person-month format. By using the selection criteria (Screen 6 - "DES - Selection Criteria for: dataset"), a dataset can be constructed for only a particular month of data.

By using item 2 "Add Selection Criteria", the system will prompt for the name of the variable for subsetting. The system will ask whether it should be treated as a numeric variable. The system will generate two menus to indicate very simply how to subset on this variable. The system does allow multiple criteria to be used in creating extract files.

As in the other sections, the request can be viewed by selecting menu item 1 "View Selection Criteria". Additional subsetting instructions can be added by selecting menu item 2 "Add Selection Criteria", can be canceled using menu item 3 "Delete Selection Criteria", or can be completely removed using menu item 4 "Reset Selection Criteria to None".

II.A.5 Performing the Data Extraction

Once all the above steps have been completed, the job is submitted using item 6 "Perform data extraction" from the Operations menu (Screen 2).

An identification number for the job (10-11 digits, including a dash) will be displayed. If an Internet account is used, a message with the identification number will be sent to that mailbox address on the completion of the extraction. The job ID number is the name of the location where the dataset will be stored. Since there are a limited number of storage spaces, it is important to retrieve your datasets within twenty-four hours,
otherwise it may be deleted.

II.B   WWW/HTML Browser Interface Access

Surveys-on-Call can be accessed on an HTML browser interface. This is a point and click interface to be used with browsers such as Mosaic, Netscape, or Lynx. It consists of 6 main

screens or "levels":

Level 1:	Enter your Internet address and select a survey
Level 2:	Select the survey year
Level 3:	Select the input dataset
Level 4:	Set a variety of parameters, including output file format, compression type, variables
Level 5:	Specify selection criteria (optional)
Level 6:	Confirmation screen displaying the job identification number and FTP instructions

To request an extract, information must be submitted in Levels 1 through 5. The Level 5 screen is displayed even if selection criteria variables were not requested in Level 4.

Level 1

On this screen (Screen 7) users are prompted for two items: a full e-mail address; and the survey to perform the extract. The Internet address is used to notify the user that the job is finished and where your dataset is stored. The default survey is the Survey of Income and Program Participation, but it can be changed by clicking on the button preceding any of the other surveys. Once you have completed these two steps, click on the SUBMIT button at the bottom of the screen.

There are several features that are standard on all Level screens:

o   By clicking on the "Level" title at the top of the screen, the system will link to additional help documentation.

o   By selecting the "Help by E-Mail" title, an e-mail template is shown that is a preaddressed to a survey-specific contact who can help you with subject matter questions.

Level 2

This screen (Screen 8) is used to select the year of the survey.

Level 3

This screen (screen 9) is used to select the dataset for the extract. For the SIPP files, the default is set to the core files.

Level 4

This screen (Screen 10) is used to specify three different sets of information:
o   General extraction parameters;
o   The input dataset; and
o   A list of variables to be output and/or used to subset your dataset.

There are four extraction parameters that need to be specified: the number of observations to be extracted, the format of your output file, the compression format of your output file, and the selection criteria processing logic. Currently, defaults are set as follows: 10 observations, the output file type to an ASCII "BOX" file, standard "UNIX" compression, and the selection criteria processing as "AND". Changes to the defaults are made by clicking on the corresponding box, pulling down a menu, and selecting another category.

To select your output variables, the system will list all the variables on the dataset with two boxes preceding each variable. Users can select a variable by clicking on the first of the two boxes next to the variable. Primary variables have been preselected. The second box is used to select variables to subset the dataset.

Level 5

Level 5 (Screen 11) is used to subset the output dataset on any of the variables specified for selection criteria. The variables selected for subsetting are displayed, one-by-one. All variables will be shown with two options: selection by category, and
selection by numeric comparison.

Level 6

This screen (Screen 12) will display the job identification number. The job ID number is the name of the location where the dataset will be stored.

III   Downloading your Dataset

Datasets can be downloaded onto users' computer systems in three ways: Internet users with Telnet access; Internet users with WWW access; and dial-in users.

Internet users with Telnet access can use FTP to transfer their files. They can use the following commands to download their files:

    ftp ftp.census.gov

The system will prompt for a user name. The user can enter:

    ftp

and then enter their full Internet address at the password prompt. At the ftp prompt users can
then enter:

      cd /tmp/job-id

where the job-id includes the 8 digit number, dash, and the trailing one or two digits. This brings
users to the storage directory where the following files are stored: the program used to create
your file, the program log, the dataset, the format library, and additional documentation for the
SQLI and BOXI files. The files are downloaded individually by using the command for each file:

      get filename.filetype.Z

When all files are downloaded the users can exit the system. Internet users with WWW
access can connect to the FTP storage site by selecting "PICKUP" on the Surveys-on-Call home
page. The user can select the job identification number and files to
download.

Dial-in users must use a second way to download files. Dial-in users need to re-connect
to Surveys-On-Call to see whether the extracts have been completed. After logging in, users will
see a screen that lists the following choices:

              1 - Extract data
              2 - View files
              3 - Download files

Item 2 "View files" is used to check the status of the job. This is a listing of all the subdirectories
in the tmp directory, the storage space for all Surveys-On-Call extracts. If the job ID number is
listed, the extract is ready to be downloaded.

Item-3 "Download files" is used to download files. Users will be asked to choose one of four
options: Kermit, FTP, X-modem, or Y-modem. Dial-in users can choose FTP and the system
will automatically connect them as an anonymous FTP user. Before users can download their
files they must change directories to the storage space of the data using the command:

      cd /tmp/job-id

IV.   Conclusion

Surveys-On-Call provides users with a system to download publicuse data files directly to
their own computer system. The system provides several capabilities including the ability to
subset the variables and cases extracted for the output dataset. The Census Bureau is continuing
to enhance the system, making it more user friendly and providing users with additional
capabilities and greater flexibility. Recently, the Census Bureau added Internet access capabilities

and a graphic interface through the WWW. With the connection to the Internet, the use of the system has increased. As an example of the level of use, below are two tables showing the number of extracts created during June and July 1995. There were a total of 359 and 342 extracts created using the system in the two months. SIPP was the most frequently used survey followed by the March supplement to the CPS, the AHS, and the CES. The most frequent users were researchers in Federal agencies (such as the Congressional Budget office, Department of Agriculture, and Social Security Administration) followed by those in educational institutions (University of Wisconsin, University of Chicago, UCLA, and Bowdoin College) and those in the private sector (Battelle Institute, Urban Institute, RAND, Institute for Women's Policy Research).

TABLE 1. NUMBER OF EXTRACTS BY SURVEY		
	JUNE 95	JULY 95
SIPP	192	259
CPS	130	70
AHS	26	2
CES	11	11
TOTAL	359	342

TABLE 2. NUMBER OF EXTRACTS BY TYPE OF USER		
	JUNE 95	JULY 95
Government	94	100
Education	163	163
Commercial	102	79
TOTAL	359	342

MENU: <DES> DATA SOURCE SELECTION
PATH: <:DES>

OPTIONS: (DES Ver. 4.2) - Your Current Session ID is: 95060209-10

DATA SOURCES:

1. (SIPP) Survey of Income and Program Participation
2. (CPS) Current Population Survey
3. (AHS) American Housing Survey
4. (CES) Consumer Expenditure Survey [Coming Soon!]

Enter Option Number (or Q=Quit H=Help R=Reshow)and press <RETURN>.

 >>

                                                    hhesoas 09:42:14

MENU: <desl> DATA EXTRACTION SYSTEM - OPERATIONS for: sipp90
PATH: <:DES:SIPP:SIPP90:desl>

OPTIONS: (DES Ver. 4.2) - Your Current Session ID is: 95060209-10

1. USER NOTES / CAUTION STATEMENTS (And Other Documentation)
    (WARNING - READ ALL USER NOTES FIRST!)
2. Set PARAMETERS   (Extraction Specifications)
3. Set INPUT FILE   (Required Extraction Specifications)
4. Set OUTPUT VARIABLES   (Extraction Specifications)
5. Set SELECTION CRITERIA   (Extraction specifications)
6. Perform DATA EXTRACTION (Based on Specifications, items 1-4 above)
7. General Data Collection Information
8. Documentation - Data Extraction Procedure (How to)
9. Documentation - Parameters
10. Documentation - Files
11. Documentation - File Content (Variables,Cateqories,etc.)
    (More Options Available, use D to move DOWN and U to move UP the List.)

Enter Option Number (or Q=Quit H=Help R=Reshow) and press <RETURN>

  >> D

MENU: <desp> DATA EXTRACTION SYSTEM - PARAMETERS for: sipp90
PATH: <:DES:SIPP:SIPP90:desl:desp>

OPTIONS: (DES Ver. 4.2) - Your Current Session ID is: 95060209-10

1.   VIEW   Parameters
2.   MODIFY   Parameters
3.   RESET ALL   Parameters to Default Values
4.   DOCUMENTATION   of Parameters

Enter Option Number (or Q=Quit H=Help R=Reshow) and press <RETURN>.

 >>

                                                        hhesoas 09:50:38

MENU: <desf> DATA EXTRACTION SYSTEM - INPUT FILES for: sipp90
PATH: <:DES:SIPP:SIPP90:desl:desf>

OPTIONS: (DES Ver. 4.2) - Your Current Session ID is: 95060209-10

1.   VIEW   Input Files (by File Number)
2 .   ADD   Input Files (by File Number)
3.   DELETE   Input Files (by File Number)
4.   RESET   Input Files
5 .   DOCUMENTATION of Files

Enter Option Number (or Q=Quit H=Help R=Reshow) and press <RETURN>.

  >>

MENU: <desv> DATA EXTRACTION SYSTEM - OUTPUT VARIABLES for: sipp90
PATH: <:DES:SIPP:SIPP90:desl:desv>

OPTIONS: (DES Ver. 4.2) - Your Current Session ID is: 95060209-10

1. VIEW   Output Variables (by Name)
2. ADD   Output Variables (by Name)
3. DELETE   Output Variables (by Name)
4. RESET   Output Variables
5. DOCUMENTATION   of File Contents (Variables,Cateqories,etc.)
6. DOCUMENTATION   File Contents SEARCH

Enter Option Number (or Q=Quit H=Help R=Reshow) and press <RETURN>

 >>

hhesoas 09:53:44

MENU: <desc> DATA EXTRACTION SYSTEM - SELECTION CRITERIA for: sipp90
PATH: <:DES:SIPP:SIPP90:desl:desc>

OPTIONS: (DES Ver. 4.2) - Your Current Session ID is: 95060209-10

1. VIEW   Selection Criteria
2. ADD   Selection Criteria
3. DELETE   Selection Criteria
4. RESET   Selection Criteria to NONE
5. DOCUMENTATION   of File Contents (Variables,Categories,etc.)
6. DOCUMENTATION   File Contents SEARCH

Enter Option Number   (or Q=Quit H=Help R=Reshow) and press <RETURN>.

  >>