This tutorial provides SAS System® and SYSTEM 2000® DBMS users with the fundamental tools needed for using the interfaces between SYSTEM 2000 and the SAS System. This tutorial is geared to experienced SAS users, who are familiar with the concepts of the SYSTEM 2000 data base.

SYSTEM 2000 Data Base Management System (DBMS) is a product of SAS Institute Inc. It is a hierarchical data base that runs on IBM®, CDC®, and Sperry systems.

In a hierarchical data base, data records are related logically into data trees. There is one record at the top of each entry and usually one or more subtrees extending down. Data coming from a hierarchical structure to a SAS data set must first be flattened out. The SAS data set will contain this flattened data. The same is true for data moving in the opposite direction; data coming from a SAS data set to a SYSTEM 2000 data base must first be flattened.

Three SAS procedures are currently available that interface with SYSTEM 2000 DBMS. These are the S2K, S2KLOAD, and QUEST procedures. PROC S2K and PROC S2KLOAD execute in full-screen, line and batch modes. PROC QUEST executes in full-screen and line modes.

PROC S2K creates a SAS data set containing either data extracted from a SYSTEM 2000 data base or items to be extracted from the data base in a later execution.

To invoke PROC S2K in full-screen mode, type

```
PROC S2K;
RUN;
```

The Data Base Access Panel (FIGURE 1) is the first screen displayed. This panel requests information about the data base and the SAS data set.

The PASSWORD field specifies the SYSTEM 2000 access password for the data base specified. The data base is entered in the SYSTEM 2000 DATA Base NAME field. Valid entries for DATA BASE ACCESS field are S for single-user access and M for Multi-User™ access. Single-user access is used when the data base data sets are allocated exclusively to your TSO or CMS session and you are the only user of them. Multi-User access is used when the data base data sets are allocated to a control region in TSO or to a SYSTEM 2000 batch machine in CMS. The OUTPUT SAS DATA SET field is the SAS data set where you want your extracted data to be written. If you do not specify a first-level name, the default is WORK and the data set is temporary.

When all information is correctly entered on the Data Base Access Panel, the Extract Data Selection Panel (FIGURE 2) will be displayed. This screen will only display the items that your password permits you to read. The procedure accesses the SYSTEM 2000 data dictionary to obtain information on this panel. This panel displays the SYSTEM 2000 name, the SYSTEM 2000 C-number, and the format. The format has been converted to the most appropriate SAS format.

To select items to be extracted from the data base, type an S in the FUNC field next to that item’s SYSTEM 2000 name. A SAS name can be entered in the column labeled SAS NAME, or one will be generated from the first eight nonblank characters of the SYSTEM 2000 name. The generated format can be changed by typing over it. System 2000 DBMS accepts character and text fields longer than the defined size and places the overflow portion in the Extended Field Table. By expanding the default format size that appears on the screen, you can extract character or text data from the Extended Field Table up to a limit of 200 characters.

At the bottom of the screen is room for a where clause. User-defined conditions that specify criteria for selecting certain records from the data base can be entered here. The C-number must be used in the where-clause condition. All SYSTEM 2000 where-clause options are supported except HAS, NOT, NON-KEY, and SAME.

There are several restrictions that apply to the selection of data. Selections can be extracted from only one path. Next to the RECORD LABEL NAME field, in the right-hand margin, are three asterisks. When you are selecting data from more than one level, the word PATH will appear next to the asterisks to validate path selections. Selections into the data base are restricted to sixteen levels. The number of levels is counted from the highest level extracted from, down to the lowest. This count includes all levels in between, whether or not selections were made on these levels.

When you press the ENTER key, SAS names will be generated for selections where SAS names were not entered. The SAS formats, the SAS names and the one path restriction will be validated at this time. You can make corrections and additions and press the ENTER key again.

After you have entered all corrections and additions, press the END key. This causes the SAS data set to be built with data selected from the SYSTEM 2000 data base. The Data Base Access Panel will be displayed. To browse the SAS data set that was created, ENTER the BROWSE command on the command line. Now the SAS data set is ready to use with other SAS procedures or in the DATA step.

The extraction process can be broken down into a two-step process by using the SAVESEL option in the PROC S2K statement. When this option is used, an intermediate SAS data set is created in place of the SAS data set containing the extracted data. This intermediate SAS data set contains information that can be used as input to a subsequent batch execution of PROC S2K. This subsequent execution creates the SAS data set containing the extracted data. Using this option can be advantageous when the data base is very large and when it would not be cost effective to use on-line resources to do the extraction. This option can also be used if the same items are frequently selected for extraction.

PROC S2K supports interactive line mode as well as batch mode execution. Everything done in full-screen mode can be done in interactive line mode or in batch. There are several other statements used with PROC S2K to duplicate the full-screen features. With the SELECT statement, C-numbers are selected and optionally given SAS names. The FMT statement can be used to override the default formats. The LIST statement will list all your selections, all the SYSTEM 2000 data base items, or the attributes of any C-number that the password has read-access to. A where clause can be entered with the WHERE statement.
PROC S2KLOAD populates a SYSTEM 2000 data base that has already been defined with data from a SAS data set. Empty data bases can be loaded, or new entries an be inserted into data bases that already contain data.

To invoke PROC S2KLOAD in full-screen mode, type

```
PROC S2KLOAD;
RUN;
```

The first screen displayed is the Data Access Panel (FIGURE 3), which requests information about the SYSTEM 2000 data base and the SAS data sets.

The PASSWORD field requests the SYSTEM 2000 access password for the data base specified. The next field on the Data Access Panel, the S2K DATA BASE NAME field, specifies the SYSTEM 2000 data base to be loaded. Following the S2K DATA BASE NAME field is a DATA BASE ACCESS field. This field specifies the data base access mode. You must enter an S or an M. If you enter an S, single-user access is requested. If you enter M, Multi-User access is requested. (For more details on single-user and Multi-User access, see the explanation under PROC S2K.) The SAS INPUT DATA SET field requested on the Data Access Panel specifies the SAS data set that contains the data to load into the SYSTEM 2000 data base.

You must enter correct information in all of the fields mentioned above to continue. You may optionally enter in one additional field found on the Data Access Panel. This is the SAS OUTPUT MAP DATA SET field. The SAS OUTPUT MAP DATA SET field specifies a SAS data set where the mapping specifications are to be written. This can be a permanent or temporary (WORK) SAS data set.

After you have entered the correct information in all the required fields on the Data Access Panel, a Mapping Specifications Panel (FIGURE 4) is displayed. This panel lists all the variables contained in the specified input SAS data set and provides fields for the C-numbers to be entered next to the variables. Assigning C-numbers to the variables creates the map that PROC S2KLOAD uses to load the specified SYSTEM 2000 data base.

You can enter a D in the FUNC field on the PROC S2KLOAD Mapping Specifications Panel to remove the variable from the list of variables available for mapping and from the display. You can enter an H to mark the variable with the HOLD function, so its values will be held for matching multiple observations for the same entity.

Enter the SYSTEM 2000 C-number for the data base item that the SAS variable should map to in the MAP TO field. The SAS NAME field displays the SAS variable name of a variable in the input SAS data set and provides fields for the C-numbers to be entered next to the variables. Assigning C-numbers to the variables creates the map that PROC S2KLOAD uses to load the specified SYSTEM 2000 data base.

As an aid to the map specifications, a SYSTEM 2000 Data Base Reference Panel (FIGURE 5) can be displayed. This panel lists only the items PROC S2KLOAD is allowed to load as determined by the password specified on the Data Access Panel. The data base schema information as described during the SYSTEM 2000 DEFINE process is displayed. This panel is for viewing only.

After you have entered the C-numbers on the Mapping Specifications Panel, you can enter the MAP command to create the map data set. This map data set can be browsed and re-created if necessary. The SYSTEM 2000 data base can be loaded by entering the LOAD command on the Mapping Specifications Panel command line. After the data base is loaded, the Data Access Panel is displayed again with load summary information included.

For PROC S2KLOAD batch mode and interactive line-mode, additional statements are used. The ERASE C-number statement specifies the C-number in a map entry previously entered that is not to be included in the map for the data base load. A LIST statement is available for listing all the mapping specifications. If you enter the ALL option on the LIST statement, all of the SYSTEM 2000 data base items that can be accessed are listed. If you enter the SASDATA option, all the variables in the input SAS data set are listed. If you enter the IDICT option or the C-number, the variable that the C-number is applied to, the SAS name, and the HOLD option (if used) are listed. The MAPTO statement allows you to specify the item or items to be loaded with data from the input SAS data set and the variable or variables to be the source of the data values for the specified item or items. No duplicates are allowed. Therefore, a given SAS variable can map its values to only one SYSTEM 2000 item, and only one SAS variable can map values to a given SYSTEM 2000 item.

There are two other statements available for line and batch mode. The REPLACE statement replaces a C-number in a map entry previously specified. The RESET statement clears all previously specified map entries.

For a very large data base load, a permanent map data set can be created by executing PROC S2KLOAD interactively. The procedure can then be executed in batch to do the actual load. This will avoid using on-line resources to load the data.

PROC QUEST is the connection between the SYSTEM 2000 Self-Contained Facility (SCF) and the SAS System. The procedure allows SCF commands to be executed while you are running a SAS session. The Self-Contained Facility encompasses the Control, Define, Access, Queue, and Report languages. The Control language is used for administration functions such as assigning passwords. The Define language is used to define and redefine data bases. The Access language accesses a data base for retrievals and updates. The Queue language groups updates and retrievals into batches of commands to process large runs more efficiently. Finally the Report language is used to produce reports easily.

To invoke the QUEST procedure type

```
PROC QUEST;
RUN;
```

If you are running under a single-user environment, the ACCESS=S option must be used in the procedure statement. The default access is M for Multi-User. (For more details on single-user and Multi-User access see the explanation under PROC S2K.)

Once you have invoked the procedure, the SAS System responds with the message 'S2K3212/OO - SYSTEM 2000 INTERACTIVE INTERFACE READY'. When you invoke the procedure with single-user access, the copyright information, CPU ID, and SYSTEM 2000 version number are also provided by the SAS System. Now enter the usual QUEST session statements. For example to open a SYSTEM 2000 data base, enter:

```
USER, DEMO;
DBN IS EMPLOYEE;
```

Notice that the statements terminate with the usual SAS semicolon, rather than the SYSTEM 2000 colon.

When you are running PROC QUEST under Multi-User access, there are some considerations. Some types of SYSTEM 2000 SCF commands must be grouped together and submitted at one time. This can be done by using Multiple Command Submission (MCS). MCS puts PROC QUEST into command-queuing mode.

148
Commands are written to a 1200-byte buffer and held until submitted to SYSTEM 2000. To submit queued commands, enter two semicolons from the program editor screen. To end MCS enter the command SCS (for single command submission).

If you are using TSO and running Multi-User access, a temporary output file should be allocated to your session. The libref for this file is S2KOUTP. The disk file is used by PROC QUEST to store output temporarily. This file is not used in single-user access.

To terminate PROC QUEST enter the END or EXIT statement. If the ENDSAS statement, /, or the BYE command is used, both the SAS System and SYSTEM 2000 will terminate.

Work is currently under way to develop a SYSTEM 2000 option for SAS/FSP software. This option will give SAS/FSP software users the ability to access SYSTEM 2000 databases in order to create, display, or update. There will be new commands for the PROC FSEDIT/SYSTEM 2000 option. A WHERE command will allow a where clause to be entered. An ORDER command will sort the data base records in the specified order. The new TREE command will create a graphic picture of the hierarchical structure of the data base. Finally, the DESCRIBE and CONTENTS commands will display the data base definition.

The interfaces between the SAS System and SYSTEM 2000 DBMS give SAS System or SYSTEM 2000 user powerful and flexible capabilities. The S2K procedure creates a SAS data set from a SYSTEM 2000 data base. This data set can then be used with other SAS procedures to produce reports, graphic output or data sets, to perform statistical analysis, and so on. SAS data sets created from other SAS procedures or from the SAS DATA step can be loaded into a SYSTEM 2000 data base using PROC S2KLOAD. And PROC QUEST brings a variety of capabilities from SAS software to a SCF user.

SAS, SAS/FSP, Multi-User, and SYSTEM 2000 are registered trademarks of SAS Institute Inc., Cary, NC, USA. IBM is a registered trademark of IBM Corporation. CDC is a registered trademark of Control Data Corp.
<table>
<thead>
<tr>
<th>FIELD</th>
<th>TYPE</th>
<th>LENGTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYSTEM 2000 NAME</td>
<td>CHAR</td>
<td>15</td>
</tr>
<tr>
<td>LAST NAME</td>
<td>CHAR</td>
<td>20</td>
</tr>
<tr>
<td>SYSTEM</td>
<td>CHAR</td>
<td>6</td>
</tr>
<tr>
<td>ADDITIONAL HISTORY</td>
<td>CHAR</td>
<td>15</td>
</tr>
<tr>
<td>ADDRESS</td>
<td>CHAR</td>
<td>30</td>
</tr>
<tr>
<td>CITY</td>
<td>CHAR</td>
<td>10</td>
</tr>
<tr>
<td>STATE</td>
<td>CHAR</td>
<td>2</td>
</tr>
<tr>
<td>ZIP CODE</td>
<td>CHAR</td>
<td>5</td>
</tr>
<tr>
<td>SOCIAL SECURITY NUMBER</td>
<td>CHAR</td>
<td>9</td>
</tr>
<tr>
<td>EDUCATIONAL HIST</td>
<td>CHAR</td>
<td>15</td>
</tr>
<tr>
<td>DATE OF BIRTH</td>
<td>CHAR</td>
<td>8</td>
</tr>
<tr>
<td>MARRIAGE STATUS</td>
<td>CHAR</td>
<td>1</td>
</tr>
<tr>
<td>MAJOR FIELD</td>
<td>CHAR</td>
<td>16</td>
</tr>
<tr>
<td>MINOR FIELD</td>
<td>CHAR</td>
<td>12</td>
</tr>
<tr>
<td>EDUCATION LEVEL</td>
<td>CHAR</td>
<td>7</td>
</tr>
</tbody>
</table>

Figure 5  Data Base Reference Panel  PROC S2KLOAD