

The Use of External Software to Import Data into the SAS® System

Richard Smith, Data Explorations

Art Carpenter, Data Explorations

ABSTRACT

A brief introduction is presented in the use of external software, such as DBMS/COPY® and DBMS/Engines® by Conceptual Software, Inc., to import external data into the SAS System. DBMS/Engines can be used in libname statements to allow easy access to external data in a method that makes the external data appear as SAS data sets to the user. A summary of tools available within the SAS System will also be discussed.

KEY WORDS

DBMS/COPY, DBMS/Engines, engines, SAS/ACCESS, Stat/Transfer, translate, transfer, import, export, third-party, wizard

INTRODUCTION

Internal tools that have allowed SAS to communicate to non-SAS database systems have long been available as part of the SAS System. Initially the primary means of transferring data was by creating and reading flat files (usually through the DATA step). Early in Version 6 the procedures DBF and DIF were introduced to translate between these to common file formats on the PC. With the recent explosion of database systems, and the increased power of the desktop workstations the need for the ability to move data into and out of the SAS System has greatly increased.

The software industry has responded to this demand with a number of solutions. The availability of internal SAS tools has greatly increased in both number and sophistication. Third party software manufactures now also provide products with the specific purpose of moving data between database systems.

This paper will briefly discuss some of the internal SAS data transfer capabilities and will highlight third party products that have proved to be useful. We will discuss the advantages and disadvantages of these products and compare them to the internal SAS counterparts.

TOOLS AVAILABLE WITHIN SAS SOFTWARE

SAS software includes a variety of options to import and export data. These include:

- DATA_NULL_
- IMPORT/EXPORT WIZARD
- ODBC

- SAS/ACCESS®
- DDE
- OLE

A detailed description of SAS System import tools was presented by Andrew Kuligowski and Nancy Roberts at SUGI 23 (Kuligowski and Roberts, 1998). Andy presented a tutorial on this subject at WUSS '98.

A brief summary of these tools is included below.

DATA_NULL_

The DATA_NULL_ step has long been used to translate data into and out of the SAS System. Through the use of the INPUT and PUT statements it has been possible to move data between SAS and any other database system that can read or write a flat file. While highly flexible this method can be time consuming, and because it requires manual intervention, it can also be prone to errors. For these reasons other data transfer tools have evolved and are now available for general use.

IMPORT/EXPORT WIZARD

The first SAS procedures specifically designed to translate into and out of the SAS format were DBF and DIF. Although still available, these have evolved into the more sophisticated IMPORT and EXPORT Wizard

The import/export wizard is activated in SAS when you choose either the import or export selection from the File pull-down menu. The base SAS module includes the ability to interactively import or export ASCII, comma delimited, and dbase files. If you have one of the SAS/ACCESS products such as SAS/ACCESS for PC File Formats, you will be able to import and export to many additional file formats.

The SAS code generated by an interactive session of the wizard can be retrieved and reused in batch programs. After execution recall the code from the program stack into the program editor. You may need to RECALL twice to retrieve both the PROC DATASETS statements and the prior PROC ACCESS statements. The code can then be added to batch jobs or edited for use with macros.

ODBC

Open Database Connectivity, ODBC, provides a standardized means of data transfer between data systems that are ODBC compliant. You will need to have an ODBC driver for each of the data systems that are to share data. SAS ships a ODBC driver for users operating under Windows and OS2. Making ODBC work is a fairly

complex process, however once the links are properly established, the transfer of information is reasonably seamless.

SAS/ACCESS

SAS/ACCESS modules are selected for the database(s) that are to be accessed. Typically the available components will vary by operating environment. Usually those running under Windows will have SAS/ACCESS for PC File Formats. This module provides the ability to read and write most of the data forms commonly encountered on standard PC systems. These include EXCEL and LOTUS spreadsheets. SAS/ACCESS is a leased module and must be requested. Access to additional databases may require the lease of additional SAS/ACCESS modules.

DDE

Dynamic Data Exchange, DDE, permits the exchange of information between two active processes in a Windows or OS2 environment. A link or connection is established between the two processes. Known as a triplet, this connection allows information that has been created in another application to be immediately available to your SAS application. Because both applications must be active DDE is less useful in a batch environment.

OLE

Like DDE, Object Linking and Embedding, OLE, allows the transfer of information between SAS and other applications. Unlike DDE, OLE is more graphical in nature and in SAS is most often used by FRAME catalog entries (from within a SAS/AF application). The OLE methodology allows information to be shared bi-directionally.

EXTERNAL SOFTWARE

A number of software products are available for PC based systems that provide data translation capabilities. As a general rule these have the advantage that they are purchased (not leased as is SAS/ACCESS), and usually they will translate between two non-SAS products directly. Rather than cover all of these products this paper will concentrate on three that directly translate SAS data sets, as well as a more direct method.

- Stat/Transfer™
- DBMS/COPY
- DBMS/Engines
- Database software in COMMAND line mode

The authors have been using DBMS/COPY and DBMS/Engines for years. These tools have proved to be very useful in translating external databases to SAS data sets or other databases. Stat/Transfer by Circle Systems, Inc. has recently added the capability to import and

export external databases to the SAS system and is discussed briefly below.

Stat/Transfer

Stat/Transfer runs under Windows95 and Windows NT. It has the capability to translate data stored in one database system to another, and supports about 12 types of data formats.

Circle Systems has recently added the capability to import and export SAS data sets with the Stat/Transfer software package. It can be used both interactively and as a batch operation. Although Stat/Transfer does not have a great many options to handle special data cases, it should handle most typical file layouts. The installed software requires much less storage space than DBMS/COPY, and although we have not yet done any benchmark studies Circle Systems reports that Stat/Transfer may be faster than DBMS/COPY in some transfers. For simple data transfers, an engines version (see DBMS/Engines) also available.

Unlike DBMS/COPY, Stat/Transfer is able to create MS Access databases, even if none already exists, and is able to create a table in the database even if a prior one already exists. Although it still uses ODBC, it has eliminated the need for the user to edit the ODBC manager.

This product is useful for general transfer operations that do not need all of the options present in DBMS/COPY.

DBMS/COPY®

DBMS/COPY has had the capability to import and export SAS data sets for many years, and it supports the transfer of over 100 types of file formats. File transfer can be between any two of the formats and the translation operation can include a programming step (similar both syntactically and in scope and to the SAS DATA step) that performs data manipulations. DBMS/COPY can be used in both interactive and batch modes to perform the translations.

DBMS/COPY has a very extensive list of options and commands that allow the programmer to control the variable and database attributes before translating to SAS data sets. Because it is possible to execute external applications from within a SAS program, a batch DBMS/COPY program can easily be executed from within a SAS program. This allows the easy import or export of SAS data sets to other database formats.

The following is a statement from SAS code used by the authors to access a Dbase® database, modify a 255 character length variable into 2 smaller variables and translate the resulting Dbase file to a SAS data set:

```
* change long dbase variable to 2 smaller  
ones and import into SAS;  
x "d:\dbmscopy\dbmswin PLUS hmis1.prg";
```

The X statement starts DBMS/COPY in batch mode and executes the following DBMS/COPY program:

```
/*hmims1.prg
dbmscopy plus program to modify too-long
dbase III vars before outputting to a SAS
database.*/

compute;
  in=f:\hmis\hmis1.dbf(modify=hlthhaz2-120)
  out=f:\hmis\hmis1.ssd;
  hlthhaz2=substr(hlthhaz1,201)||hlthhaz2;
  format hlthhaz2 $120;
  format hlthhaz1 $200;
run;
```

The COMPUTE block in the above step is similar to the DATA step and has most of the DATA steps capabilities.

To translate to and from a MS Access database DBMS/COPY uses ODBC. There are some current limitations with this process, however, a recent discussion with the software developers revealed that they were currently working on an easier implementation of translation for MS Access.

DBMS/Engines[®]

DBMS/Engines is a second product by Conceptual Software, Inc. This software is especially geared for use with the SAS System. Once installed the user simply adds a database specific engine to the LIBNAME statement and all read write access to that library will be appropriately translated. This makes external data sets, such as, EXCEL tables appear as SAS data sets to the user.

This means that the use of DBMS/Engines is very transparent to the user. The translation engines are provided in the form of DLLs, and this is what allows the programmer to use a simple libname statement to access external databases. The DBMS/Engines software also has several options but the simplicity of the simple libname use makes it easy for the programmer. The following is a simple example using an engine in a libname statement (Carpenter, 1998, p. 106):

```
libname sasdata 'd:\proj1\sasdata';
libname xldat dbexcel4 'd:\proj1\excel4';

proc datasets;
  copy in=sasdata out=xldat;
run;
```

In the above example, all SAS data sets in the directory indicated by the *libref* SASDATA are copied to the directory indicated by XLDAT as EXCEL tables (Version 4). The tables will have the same name as the SAS data set and the first row of the table will contain the SAS data set variable names. Since formats are utilized in the

conversion process, SAS dates can be correctly converted to EXCEL dates making the process Y2K compliant.

DATABASE SOFTWARE IN COMMAND-LINE MODE

So you don't wish to use any specialized translation software packages? You don't need to if you have a database software program that can be used in command-line mode to run macros or to use options to do its own translation to either SAS or to comma delimited or fixed text data sets. A good example of software with this capability is Microsoft Access.

Using the SAS X statement, the user can run MS Access with the /x option followed by a macro name as shown in the following statement:

```
x "c:\msoffice\msaccess\msaccess /x x2csv";
```

This example uses a simple MS Access macro (x2csv) that saves a specific MS Access table to a specific comma delimited file with the variable names in the first observation. This file can then be easily imported into a SAS data set using the INFILE and INPUT statements:

SUMMARY

It is very common to be in a situation where it is necessary to move data from one database format to another. Fortunately there are many ways to accomplish the translations. SAS software has a several translation options built into the base system and others can be added by leasing one of several versions of SAS/ACCESS.

Several third party solutions are also available. These include Stat/Transfer, DBMS/COPY and DBMS/Engines. These solutions range from simple to complex and from simplistic to highly flexible. DBMS/Engines allows the SAS developer to create an operating environment that will seamlessly and transparently read and write SAS data sets using over 100 different file formats.

REFERENCES

Carpenter, Arthur L., *Carpenter's Complete Guide to the SAS[®] Macro Language*, Cary, NC:SAS Institute, Inc., 1998, 242 pp.

Kuligowski, Andrew and Nancy Roberts. (1998) "Basic Methods to Introduce External Data into the SAS System" in *Proceedings of the Twenty-Third Annual SAS Users Group International Conference*. 23: 300-309.

Conceptual Software, Inc. (1998) Conceptual Software web page address: <http://www.conceptual.com>

Circle Systems, Inc. (1998) Circle Systems web page address: <http://www.stattransfer.com>

ABOUT THE AUTHORS

Richard Smith and Art Carpenter are senior partners at Data Explorations. Data Explorations provides data management, analyses, and programming services nationwide and is a SAS Quality Partner™.

Richard O. Smith

Richard Smith has a masters in Biology/Ecology and has provided complete data management and analysis services for numerous environmental research projects as a senior biologist, SAS programmer, and project manager. He has also provided programming and management services for the health related industries. He has been using SAS extensively since 1981 and is a SAS Quality Partner™.

Arthur L. Carpenter

Art Carpenter's publications list includes the books *Quick Results with SAS/GRAPH® Software*, *Carpenter's Complete Guide to the SAS® Macro Language*, and two chapters in *Reporting from the Field*. In addition he has presented or published over three dozen papers and posters. Art has taken numerous leadership positions in local, regional, and national SAS user groups and is a SAS Quality Partner™. He teaches SAS courses and provides contract SAS programming support nationwide.

Art has a masters in applied mathematics with an emphasis in statistics. He has held positions as senior statistician, programmer, and project manager in various industries and research settings.

AUTHOR CONTACTS

Richard Smith
Data Explorations
2270 Camino Vida Roble, L
Carlsbad, CA 92009
Phone: (760) 438-1762
Email: ROSmith@SciX.com



Art Carpenter
Data Explorations
2270 Camino Vida Roble, L
Carlsbad, CA 92009
Phone: (760) 945-0613
Email: Art@Caloxy.com



TRADEMARK INFORMATION

SAS, SAS/AF and SAS/ACCESS are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries.

DBMS/COPY and DBMS/Engines are registered trademarks of Conceptual Software, Inc.

Stat/Transfer is a trademark of Circle Systems, Inc.

Other brand and product names are registered trademarks or trademarks of their respective companies.

® Indicates USA registration