

THE BENEFITS OF UTILIZING SAS/ACCESS TO ACCESS DIF FILES

George Matthews
The University of Georgia

ABSTRACT

As an avid user of proc dif in PC SAS, I was somewhat dismayed, when I migrated to SAS for Windows, that this procedure would no longer be available. Although SAS Institute has since made this procedure available to licensed users of the SAS/ACCESS product, this paper takes a look at the benefits of utilizing the SAS/ACCESS product to access dif files. The techniques for defining the access and view descriptors for dif files created with Lotus 1-2-3 Release 3 will be examined. Also, the flexibility of the descriptor files in comparison to the previous methods of accessing dif files will be discussed.

INTRODUCTION

In the late '70s Lotus Development Corporation utilized DIF (Data Interchange Format) as a data interchange tool to move data between Visicalc and Lotus 1-2-3. Although Lotus paved the way for world-wide use of this format, DIF is a trademark of Software Arts, Inc. of Sunnyvale, California. A dif file is an ascii text file with a file header section and a data section. The dif format allows other software packages to retrieve the data independently of the software the data was originally created with. Most spreadsheet packages can utilize this format. In order for spreadsheet data to be processed with SAS software, the data must be converted to dif format. Since the earliest release of PC-based SAS, SAS Institute has provided a means for retrieving data in dif format. The dif procedure was provided, as means of accessing dif files, with the first production version of PC SAS (6.03). It continues to be utilized as part of the current version of PC SAS 6.04. With the 1993 release of SAS for Windows, a separate product, called SAS/ACCESS, was developed as a means of accessing data in DIF format. The SAS/ACCESS product provides an interface to PC file formats like dif. The benefits of utilizing SAS/ACCESS batch command language to access dif files will be

examined. The execution of PROC ACCESSs in batch mode was made available by SAS Institute with maintenance release TS407 of the SAS for Windows 6.08 product.

DESCRIPTION OF THE DATA

The Lotus 1-2-3 Translate Utility was utilized to translate a Lotus 1-2-3 release 3 file (survey.wk3) into a dif file (survey.dif). The dif file consisted of 75 rows and 79 columns. Twenty-Seven columns were utilized to describe the demographic information of the participants. These variables are listed in Table #1. Eighteen columns were utilized to measure how much emphasis the participants placed on material wealth. These columns are listed in Table #2. Seventeen columns were utilized to measure participants involvement with religious and social organizations. These columns are listed in Table #3. Seventeen columns were utilized to measure participants opinions about American-made products. These columns are listed in Table #4.

Table 1:Demographics

<u>Column</u>	<u>Default Type</u>
id	numeric
age	numeric
sex	numeric
race	numeric
ethnic	numeric
relig	numeric
marital	numeric
educ	numeric
born	numeric
uscity	numeric
usa	numeric
outus	numeric
vehicle	numeric
brand	numeric
carpurch	numeric

newused	numeric
paidfor	numeric
time	numeric
prevcar	numeric
influenc	numeric
person	numeric
nexttime	numeric
nextnew	numeric
import	numeric
expect	numeric
dream	numeric
reality	numeric

Table 2:Material Wealth

<u>Column</u>	<u>Default Type</u>
admire	numeric
acqmat	numeric
empsucc	numeric
ownlot	numeric
impress	numeric
payatten	numeric
ineed	numeric
lsimple	numeric
notimp	numeric
notprac	numeric
buypls	numeric
luxury	numeric
lessemp	numeric
realneed	numeric
lbetter	numeric
hapnice	numeric
happy	numeric
lktohave	numeric

Table 3:Religious Involvement

<u>Column</u>	<u>Default Type</u>
worth	numeric
carpref	numeric
racedriv	numeric
cfun	numeric
wrapup	numeric
applianc	numeric
attach	numeric
daypress	numeric

caradv	numeric
bored	numeric
norace	numeric
recharg	numeric
natural	numeric
talk	numeric
notord	numeric
satisfy	numeric
discuss	numeric

Table 4:American-Made Products

<u>Column</u>	<u>Default Type</u>
always	numeric
importcs	numeric
buyamer	numeric
first	numeric
unamer	numeric
notrt	numeric
real	numeric
rich	numeric
best	numeric
little	numeric
hurts	numeric
curbs	numeric
longrun	numeric
foreign	numeric
tax	numeric
only	numeric
nowork	numeric

ACCESS METHODS

The SAS System for Windows interacts with a dif file using SAS/ACCESS descriptor files. There are two types of descriptor files:

- access descriptors
- view descriptors

An *access descriptor* describes the data that is in the dif file and has a member type of ACCESS. The information includes:

- dif filename
- column names
- column types
- default sas variable names
- default sas formats

The access file can be used as a master to create several view descriptors.

A *view descriptor* defines the data that was described by the access descriptor. A view descriptor can define all or just a subset of the data that is in the dif file. A view descriptor is a SAS data set with a member type of VIEW.

This section will describe the following steps:

- How to create access descriptor named DFALIB.SURVEY
- How to create view descriptor named DIFVLIB.USASURV

The quickest and sometimes most efficient way of creating the access and view descriptors is in a single PROC ACCESS execution. The following example provides the SAS code that was utilized to create the descriptor for the data described in the previous section.

```
libname dfalib 'c:\saswin\saslib\';
libname dfvlib 'c:\saswin\saslib\';
```

```
proc access dbms=dif;
  create difalib.survey.access;
    path = "c:\123r3\survey";
    diflabel;
    skiprows=2;
    assign=yes;
  create difvlib.usasurv.view;
  select all;
```

run;

An explanation of each statement follows:

libname

utilized to establish the librefs with their respective SAS data libraries.

proc access

invokes the access procedure.

dbms=

specifies the interface engine to be utilized (dif,dbf,etc.).

create

identifies the access descriptor or view descriptor to be created.

path=

identifies the complete path and filename of the file that contains the data.

diflabel

is a valid statement only when utilized for accessing dif files. It indicates that the DIF file columns are in the first row and sas variable names should be created based on the dif column names.

skiprows=

tells sas to skip the first rows of the dif file and begin reading the data in the first row which contains the data. This statement is only valid when accessing dif files.

assign=

indicates whether or not the variable names and formats can be changed. A value of yes indicates that they can be modified. A value of no indicates that they cannot be modified.

select

selects some or all of the columns to be included in the view descriptor. It is required if the view is to be created.

run

processes the last view descriptor.

Some additional statements that can be utilized as part of the PROC ACCESS are as follows:

rename

renames the sas variables that were created from the DIF column names.

format

assign sas formats to the sas variables.

list all

lists the variables, columns, and formats in the SAS log.

subset where

provides where clause processing.

The PROC ACCESS statements can be executed from the program editor or in batch mode.

THE FLEXIBILITY OF DESCRIPTOR FILES

The flexibility of the descriptor files, in comparison to previous methods of accessing dif files, is obvious when we focus on the power of the view descriptor. The view descriptor can be utilized in any SAS procedure just as you would utilize other SAS data sets. For example, the following code could be used to print the data that was described by the descriptor files in the previous section.

```
proc print data=difvlib.usasurv;
title '1994 Survey Data';
run;
```

The data is not retrieved until it is needed by the SAS procedure. The flexibility of the descriptor files is further demonstrated by the fact that you can change the

variable type. Also, several view descriptors can be created with a single execution of the access procedure. If utilized as described by this paper, the descriptor files provides the SAS programmer with a flexible and efficient tool.

CONCLUSION

The dif procedure offers a valuable means of accessing spreadsheet data, but the SAS/ACCESS method offers much more flexibility as demonstrated in this paper. Since SAS/ACCESS is available for other platforms, the techniques employed may be useful for reading other types of files as well.

REFERENCES

1. Getting Started with SAS/ACCESS Software, Version 6, First Edition, SAS Institute.
2. SAS/ACCESS Interface to PC File Formats, Usage and Reference, Version 6, First Edition, SAS Institute.

ACKNOWLEDGEMENT

The author wishes to thank Jeannie McElhannon for word processing support and the PC Interface Support Group at SAS Institute for technical support. The author thanks Joe Kelley, Greg Nelson, and Barbie Matthews for comments, suggestions, and support throughout the development of this paper. The author can be contacted at:

The University of Georgia
UCNS
Computer Services Annex
Athens, Ga. 30602-1911
(706)542-5359
gmatthew@uga.cc.uga.edu

SAS is a registered trademark of SAS Institute, Inc., in the USA and other countries

DIF is a registered trademark of Software Arts, Inc.

Lotus 1-2-3 is a registered trademark of Lotus Development Corp.