

Speedily Process Multiple Data Files Within SAS

Edgar Adams Keane Inc Frankfort, Ky

ABSTRACT

Using knowledge of the windows file naming and folder structure to rapidly process multiple data files using SAS[®] Macro code.

INTRODUCTION

When presented with a need to process many sets of similar data, it became apparent, that if you could use a filename structure that had meaning within the window folder structure it would make a convenient and rapid way to process data and create individual specially named reports. Combining the folder / file naming abilities of windows and the SAS Macro code processing of variable / filename creations could be a perfect marriage.

In the environment of my employment, there are 175+ sets of similar data to be processed on a regular schedule. Each file represents a district's worth of information. Each district has been assigned a unique code number so that each district's data file can contain the district's code. For example, let's say that the unique code for a district named APPLE was 001, then the data file for that district could be named Data_001_Oct2005.txt Where the 001 represents the district code number, the other parts of the file name provides descriptive information about its creation date or other descriptive information as needed. Our current focus will be only on the unique district code contained in the file name.

To make this an easy system for maintenance, the district code and the district name are maintained in a simple file structure, an Excel spreadsheet is used. This file is in an available form for SAS to read for actual processing.

Here is an overview of the process:

Files to be processed are placed into a "staging folder".

In the SAS program

- Read simple file structure that contains district code and name, and create a format of the information.

- Read "staging folder" and determine how many files to process.

- Loop through files in the "staging folder", one at a time, read in filename to extract district code.

- Using the extracted district code and format, create a macro variable for the district name.

- Read district data file and process information, create report.

- Using macro variable of district name, and number, make a file name similar to the form:

 - REPORT_001_APPLEdistrict.PDF

- Go back and get next file name, and process again.

CONCLUSION

The process described can be ported to other operating systems. Whenever there is a need for repetitive processing, this process can be used in one form or another. Keys are using the SAS Macro code language to generate meaningful variables that can be used in identifying the resulting reports and having an input filename structure that can identify the source.

ACKNOWLEDGMENTS

It is only with the help of John Wickizer, Department of Assessment and Implementation, Kentucky Department of

Education, who led me toward the correct solution, that I was able to develop and accomplish this process.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Edgar Adams
Keane, Inc.
9 Fountain Place
Frankfort, Ky 40601
Work Phone: (502) 564-4394
Fax: (502) 564-7749
Email: edgar.adams@kde.state.ky.us

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

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

PROGRAM SOURCE

```
libname mylib 'c:\SAS\temp';
ods path mylib.tmplmst(read) sashelp.tmplmst(read);
/*****
/* Purpose of this routine
/*****
/* To be able to automate the processing of creating reports for the multiple
sources
/*****
/*
/* Created by Edgar Adams October 2005
/* For MWSUG Cincinnati, Ohio
/*
/* Process is as follows....
/*
/* Read in Format statements for District / Number table
/*
/* Create macro variables that contain
/* District Name
/* District Number
/*
/* Read CSV file into a SAS dataset and do report
/* Use District Number and Name to label report
/*

*/
options Linesize = 80
        Center
        date
        nonumber
        ;

Filename DIR 'C:\MWSUG\DATA\DistrictFiles\';

%let RptDate = %sysfunc(date(), mmdyy6.);

/*****
/* Format for District Number and Names
/* Create District Name Format
*/
        %Include "C:\MWSUG\PROGRAMS\DistrictFormat.txt";

/*****
/* Begin processing of information by setting up the macros...
/*****
*/
ODS Listing Close;

%macro CreateRpts(FileToRead);
        /*Associate District Number to District Name for report writing */
        Data _Null_;
                District_Number = &Dist;
                strng = Trim(put(District_Number,Di_TBL.));
                call symput('DName', Trim(strng));
                strng2 = Trim(put(District_Number,z3.));
                call symput('dinum',Trim(Strng2));
        run;
```

PROGRAM SOURCE

```
/* **** */
/* Create individual district reports **** */
/* **** */

ODS PDF style=oaartf
file="C:\MWSUG\REPORT\Rpt_&dinum._%Trim(&DName)_&RptDate..PDF";

/* Begin by build sas dataset from the CSV
*/
data workfile;
    %let _EFIERR_ = 0; /* set the ERROR detection macro variable */
    infile "C:\MWSUG\DATA\DistrictFiles\&FileToRead"
        MISOVER DSD lrecl=32767 firstobs=2 ;
    informat District_Number 3. ;
    informat Count 4. ;

    format District_Number 3. ;
    format Count 4. ;

    input
        District_Number
        Count ;

    if _ERROR_ then call symput('_EFIERR_',1); /* set ERROR detection
macro variable */
run;

/* **** */
/* Create report */
/* **** */
Proc Freq data= workfile;
    title1 "Report";
    title2 "=====";
    title3 "District: &DName";
    table Count / nocol nopercent norow;
run;
ODS PDF Close;
%Mend; *CreateRpts(FileToRead);

%Macro CkDirectory(DirRef);
    * Open directory to find out how many files to process;
    %let drc = %sysfunc(dopen(&DirRef));
    %let NumFiles = %sysfunc(dnum(&drc));
    * Log entry for trouble shooting :) ;
    %if (&NumFiles = .) %then %Put The directory is empty;
    %else %Put The folder contains &NumFiles files;

    %do I = 1 %to &NumFiles;
        * Strip out District Number and Read filename;
        %let Dist = %substr(%sysfunc(dread(&drc,&I)),4,3);
        %let LEFfileName = %sysfunc(dread(&drc,&I));
        %CreateRpts(&LEFfileName);
    %end;
    %let rc=%sysfunc(dclose(&drc));
%Mend; * CkDirectory(DirRef);

%CkDirectory(DIR);
```

PROGRAM SOURCE - DISTRICT FORMAT

```
filename OfflList "C:\MWSUG\DATA\District_names_numbers_2004-2005.xls";

proc import datafile=OfflList
    out=DistFILE
    dbms=excel REPLACE;
run;
DATA DistFORMAT; SET DistFILE;
    fMtname="Di_TBL";
    LABEL=DistrictNA;
    START=DistrictNO;
    END=DistrictNO;
    KEEP FMTNAME LABEL START END;
RUN;
PROC FORMAT CNTLIN=DistFORMAT;
RUN;
```

SAMPLE FORMAT DATA FILE

	A	B
1	DISTRICTNA	DISTRICTNO
2	Adair	001
3	Allen	002
4	Anderson	004
5	Ballard	010
6	Barren	013
7	Bath	014
8	Bell	015
9	Boone	016
10	Bourbon	017
11	Boyd	018
12	Boyle	019
13	Bracken	020
14	Breathitt	021
15		

SAMPLE DATA FILE

	A	B	
1	DistrictNo	Count	
2	1	1	
3	1	2	
4	1	4	
5	1	6	
6	1	8	
7	1	8	
8	1	12	
9	1	18	
10	1	22	
11	1	29	
12			

SAS LOG

Number of files to be processed

```
113 %Macro CkDirectory(DirRef);
114     * Open directory to find out how many files to process;
115     %let drc = %sysfunc(dopen(&DirRef));
116     %let NumFiles = %sysfunc(dnum(&drc));
117     * Log entry for trouble shooting :) ;
118     %if (&NumFiles = .) %then %Put The directory is empty;
119     %else %Put The folder contains &NumFiles files;
120
121     %do I = 1 %to &NumFiles;
122         * Strip out District Number and Read filename;
123         %let Dist = %substr(%sysfunc(dread(&drc,&I)),4,3);
124         %let LEFfileName = %sysfunc(dread(&drc,&I));
125         %CreateRpts(&LEFfileName);
126     %end;
127     %let rc=%sysfunc(dclose(&drc));
128 %mend; * CkDirectory(DirRef);
129
130 %CkDirectory(DIR);
The folder contains 13 files
```

SAMPLE SAS LOG – ONE DISTRICT PROCESSING

NOTE: Writing ODS PDF output to DISK destination
"C:\MWSUG\REPORT\Rpt_001_Adair_080505.PD", printer "PDF".

NOTE: Numeric values have been converted to character values at the places given by: (Line):(Column).
6:109

NOTE: The infile "C:\MWSUG\DATA\DistrictFiles\Dr_001.csv" is:
File Name=C:\MWSUG\DATA\DistrictFiles\Dr_001.csv,
RECFM=V,LRECL=32767

NOTE: 10 records were read from the infile "C:\MWSUG\DATA\DistrictFiles\Dr_001.csv".
The minimum record length was 5.
The maximum record length was 6.

NOTE: The data set WORK.WORKFILE has 10 observations and 2 variables.

NOTE: DATA statement used:
real time 0.07 seconds
cpu time 0.00 seconds

NOTE: There were 10 observations read from the data set WORK.WORKFILE.

NOTE: PROCEDURE FREQ used:
real time 0.86 seconds
cpu time 0.07 seconds

NOTE: ODS PDF printed 1 page to C:\MWSUG\REPORT\Rpt_001_Adair_080505.PDF.

NOTE: DATA statement used:
real time 0.01 seconds
cpu time 0.01 seconds

SAMPLE REPORT

Report

=====

District: Adair

The FREQ Procedure

Count	Frequency	Cumulative Frequency
1	1	1
2	1	2
4	1	3
6	1	4
8	2	6
12	1	7
18	1	8
22	1	9
29	1	10