SAS® GLOBAL FORUM 2020 #SASGF

Paper 4991-2020

A Program to Compare Two SAS® Format Catalogs

Michael A. Raithel, Westat

Abstract

SAS programming professionals are sometimes faced with the task of determining the differences between two SAS format catalogs. Perhaps they received an updated format catalog from a collaborating organization; or maybe a colleague updated a format catalog to reflect changes in the underlying data. Either way; how can programmers tell which catalog entries and value/label pairs have been modified? If the two catalogs being compared are relatively small, then the tried-and-true method of outputting each of them via the FMTLIB option of PROC CATALOG and then manually comparing the listings may suffice. But, this method is laborious and error-prone when there are a large number of formats and format value/label pairs.

This paper presents a SAS program that compares two SAS format catalogs and reports the differences between them. It identifies mismatches in the format name, start value, end value, and label between the two catalogs being compared. Because the comparisons are done programmatically, this method eliminates tedious manual reviews and directly identifies all differences. Readers can immediately begin using this program to compare their own SAS format catalogs.

Introduction

This paper describes the *Compare Two SAS Format Catalogs Utility* program; a macro program that can be used to discover mismatches between two SAS format catalogs. The program differentiates between the two catalogs by labeling them as BASE and COMP. BASE is the catalog used as the basis for the comparison; the one the other catalog will be matched against. COMP is the catalog that is to be compared against the BASE catalog to determine if there are any differences between the two.

The program identifies these possible mismatch outcomes:

- Catalog entry is in BASE but not in COMP
- Catalog entry is in COMP but not in BASE
- Catalog entry is in both BASE and COMP catalogs, but there is a:
  - Mismatch on END
  - Mismatch on LABEL

If there is a perfect match between the two SAS format catalogs, the program produces a report stating that fact and listing all of the values for FMTNAME, START, END, and LABEL found in them.

The *Compare Two SAS Format Catalogs Utility* program can be found in its entirety in **Appendix A**. It is designed to be placed in a directory and invoked by the calling program. The calling program, *Execute Compare Two SAS Format Catalogs Utility*, can be found in **Appendix B**. Configure the macro calls at the bottom of the calling program to execute the *compare* program against as many pairs of format catalogs as you would like. Also, change the %INCLUDE to point to where you store the utility program in you computing environment. **Appendix C** contains the *Create Test Format Catalogs* SAS program. This program creates a couple of format catalogs you can use to test and tune the *Compare Two SAS Format Catalogs Utility* program to work in your computing environment.

The *Compare Two SAS Format Catalogs Utility* program is made up of six separate steps. The next six sections of this paper provide detailed explanations of what is done in each step.

Step 1 – LIBNAME Allocations

This step holds two LIBNAME statements that specify the SAS data libraries that hold each of the two catalogs. The BASELIB libref is where the BASE catalog is located; the COMPLIB libref, where the COMP catalog is located. The full path name of each directory is passed to the program via the &BASELIB and &COMPLIB macro parameters.

Step 2 – Dump the Contents of the BASE SAS Catalog

Step 2 uses PROC FORMAT to dump the contents of the BASE format catalog into a SAS data set named FormatBase. This is done by specifying the CNTLOUT= option. Only four variables are kept: FMTNAME, START, END, and LABEL; and the last two variables are renamed to BaseEnd and BaseLabel, respectively. The FormatBase SAS data set is sorted by FMTNAME and START for future merging.

Step 3 – Dump the Contents of the COMP SAS Catalog

Step 3 uses PROC FORMAT to dump the contents of the COMP format catalog into a SAS data set named FormatComp. This is done by specifying the CNTLOUT= option. Only four variables are kept: FMTNAME, START, END, and LABEL; and the last two variables are renamed to CompEnd and CompLabel, respectively. The FormatComp SAS data set is sorted by FMTNAME and START for future merging.

Step 4 – Match Merge BASE and COMP

Step 4 performs a match merge of the FormatBase and FormatComp SAS data sets by variables: FMTNAME and START. It may produce up to three SAS data sets:

- **Bothcats** – This data set contains catalog entries found in both the Base SAS catalog and the Comp SAS catalog, based on matching the values of FMTNAME and START.
- **Basecatonly** – This data set contains catalog entries found in the Base SAS catalog, but not in the Comp SAS catalog, based on attempting to match the values of FMTNAME and START.
- **Compcatonly** – This data set contains catalog entries found in the COMP SAS catalog, but not in the Base SAS catalog, based on attempting to match the values of FMTNAME and START.

…depending upon the matches between the two data sets.

In *Basecatonly*, The CompEnd and CompLabel variables are dropped since they only contain missing values. The same is done for BaseEnd and BaseLabel in the Compcatonly SAS data set.

Step 5 – Create Report Data Sets and Macro Flags to Identify Empty Data Sets

This step begins by creating two mismatch data sets from the Bothcats SAS data set. Data set *EndMismatch* contains observations where variable BaseEnd does not equal variable CompEnd. Data set *LabelMismatch* contains observations where variable BaseLabel does not equal variable CompLabel. If there are no mismatches between the corresponding END and LABEL variables, then the EndMismatch and LabelMismatch data sets will be empty.

The second part of Step 5 determines if any of the mismatch data sets that were created in Step 4 and previously in Step 5 have observations in them. If so, then a macro flag variable is set to “1” so that those data sets will be reported upon in Step 6. If not, a macro flag is set to “0” so that no report is created. The macro flag variables are:

- **BASECATONLY** – A “1” means that there are entries in the BASE catalog that are not in the COMP catalog.
- **COMPCATONLY** – A “1” denotes that there are entries in the COMP catalog that are not in the BASE catalog.
- **ENDMISMATCH** – A “1” means that there is a mismatch in the END value for entries in the COMP and BASE catalogs for entries that matched on FMTNAME and START.
- **LABELMISMATCH** – A “1” means that there is a mismatch in the LABEL value for entries in the COMP and BASE catalogs for entries that matched on FMTNAME and START.

The mechanism for setting the macro flags is to access the SASHELP.VTABLE view, subset to observations where LIBNAME equals "WORK", and then access the observations with the following *memnames*: BASECATONLY, COMPCATONLY, ENDMISMATCH, and LABELMISMATCH. Once it has the aforementioned observations, the program determines if there are any observations in that data set by inspecting the value of the NOBS variable. If NOBS is greater than zero, then the appropriate macro variable is set to "1" via a CALL SYMPUT.

Step 6 – Create Reports of the Mismatches

Step 6 uses the macro flag variables created in Step 5 to produce the appropriate mismatch reports. The reports are written to Word documents wherein the names of the BASE and COMP format catalogs are a part of the document name. Depending upon the macro flag values, one or more of these discrepancy report document files may be produced:

- Catalog Entries Found in BASE (&BASECAT) But Not Found In COMP (&COMPCAT).docx
- Catalog Entries Found in COMP (&COMPCAT) But Not Found In BASE (&BASECAT).docx
- Catalog Entries Found in BASE (&BASECAT) and COMP (&COMPCAT) With Mismatched Values for END.docx
- Catalog Entries Found in BASE (&BASECAT) and COMP (&COMPCAT) With Mismatched Values for LABEL.docx

Note that &BASECAT and &COMPCAT resolve to the actual names of the BASE format catalog and COMP format catalog, respectively.

If all four of the macro flag variables are set to "0", then Step 6 creates this report document:

- No Differences Found Between Entries in the BASE (&BASECAT) and COMP (&COMPCAT) Catalogs.docx

Consequently, Step 6 may produce between one and four documents that report on the compatibility of the entries of the two SAS format catalogs.

Conclusion

This paper introduced the *Compare Two SAS Format Catalogs Utility* program; which can be used to discover mismatches between two SAS format catalogs. The program uses the CNTLOUT= option of PROC FORMAT to dump the catalog entries of the two format catalogs into separate SAS data sets. Then, it performs a match-merge of the data sets by FMTNAME and START to find matches and mismatches. The program goes on to check for other types of mismatches and then creates reports specifying the results.

The *Compare Two SAS Format Catalogs Utility* program is ideal for comparing large format catalogs to find the differences between them. Consider how this program can be of help when you need to compare to SAS format catalogs in your own programming environment.

Disclaimers

The contents of this paper are the work of the author and do not necessarily represent the opinions, recommendations, or practices of Westat.

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.

References

Raithel, Michael A. 2017. Did You Know That? Essential Hacks for Clever SAS Programmers: Over 100 Essential Hacks to Make Your Programs Leaner, Cleaner, and More Competitive. Bethesda, Maryland: Michael A. Raithel
Available: http://www.amazon.com/Michael-A.-Raithel/e/B001K8GG90/ref=ntt_dp_epwbk_0

Raithel, M.A. (2017). PROC DATASETS; The Swiss Army Knife of SAS Procedures. *Proceedings of the SAS Global Forum 2017 Conference*.
Available: http://support.sas.com/resources/papers/proceedings17/0963-2017.pdf

SAS Institute Inc. 2015. *Base SAS® 9.4 Procedures Guide, Fifth Edition*. Cary, NC: SAS Institute Inc.
Available: http://support.sas.com/documentation/cdl/en/proc/68954/PDF/default/proc.pdf

SAS Institute Inc. 2015. *SAS® 9.4 Language Reference: Concepts, Fifth Edition*. Cary, NC: SAS Institute Inc.
Available: http://support.sas.com/documentation/cdl/en/lrcon/68089/PDF/default/lrcon.pdf

Acknowledgements

The author would like to thank Westat management for supporting his participation in SAS Global Forum 2020.

Contact Information

I would love to get your feedback on this paper; especially if you found it helpful. You can contact me at the following email address:
michaelraithel@westat.com

Appendix A – Compare Two SAS Format Catalogs Utility

```
/******************************************************************************************************/
/*  Program: Compare Two SAS Format Catalogs Utility.sas                                              */
/*                                                                                                    */
/*  Author: Michael A. Raithel                                                                        */
/*                                                                                                    */
/*  Created: 9/6/2019                                                                                 */
/*                                                                                                    */
/*  Purpose: This SAS macro compares two SAS format catalogs to determine if there are mismatches in  */
/*           fmtname, start, end, and label. It creates reports of any discrepancies it finds.        */
/*                                                                                                    */
/*  Note:    The program differentiates between the two catalogs by labeling them BASE and COMP.     */
/*                                                                                                    */
/*           BASE - Is the catalog used as the basis for the comparison; the one the other catalog    */
/*                  will be matched against.                                                          */
/*           COMP - Is the catalog that is to be compared against the BASE catalog to determine if    */
/*                  there are any differences between the two.                                        */
/*                                                                                                    */
/*  Parameters: Users must specify the following five parameters:                                     */
/*                                                                                                    */
/*             BASELIB    - The directory holding the SAS format library you want to use as the BASE  */
/*                          for the comparison.                                                       */
/*             BASECAT    - The catalog that you want to use as the BASE for comparison.              */
/*                                                                                                    */
/*             COMPLIB    - The directory holding the SAS format library you want to COMPare          */
/*             COMPCAT    - The catalog that you want to COMPare to the base catalog.                 */
/*                                                                                                    */
/*             REPORTS    - The directory where reports are to be written.                            */
/*                                                                                                    */
/*  Outputs: This program may create one or more of the following reports, depending upon mismatches: */
/*                                                                                                    */
/*            - Catalog entries found in BASE (&BASECAT) But Not Found In COMP (&COMPCAT)             */
/*            - Catalog entries found in COMP (&COMPCAT) But Not Found In BASE (&BASECAT)             */
/*            - Catalog entries found in BASE (&BASECAT) and COMP (&COMPCAT) With Mismatched          */
/*              Values for END                                                                        */
/*            - Catalog entries found in  BASE (&BASECAT) and COMP (&COMPCAT) With Mismatched Values  */
/*              for LABEL                                                                             */
/*                                                                                                    */
/*            ...or if there are no mismatches, it will create this report:                           */
/*                                                                                                    */
/*            - No Differences Found Between Entries in the BASE (&BASECAT) and COMP (&COMPCAT)       */
/*              Catalogs                                                                              */
/*                                                                                                    */
/*  Change Log:                                                                                       */
/*                                                                                                    */
/******************************************************************************************************/

%MACRO CATCOMPR(BASELIB=, BASECAT=, COMPLIB=, COMPCAT=, REPORTS=);

options symbolgen mprint mlogic source2 nonumber;


    /*******************************************/
    /* Part 1 - LIBNAME allocations            */
    /*******************************************/
```

```
/* Specify the directory holding the BASE format catalog */
libname BASELIB "&BASELIB";

/* Specify the directory holding the COMP format catalog */
libname COMPLIB "&COMPLIB";


    /*********************************************************************************/
    /* Part 2 - Dump contents of the BASE SAS catalog, then sort for merge in Part 4 */
    /*********************************************************************************/

proc format library=BASELIB.&BASECAT
            cntlout=FormatBase(keep=fmtname start end label rename=(end=BaseEnd label=BaseLabel));
run;

proc sort data=FormatBase;
      by fmtname start;
run;


    /*********************************************************************************/
    /* Part 3 - Dump contents of the COMP SAS catalog, then sort for merge in Part 4 */
    /*********************************************************************************/

proc format library=COMPLIB.&COMPCAT
            cntlout=FormatComp(keep=fmtname start end label rename=(end=CompEnd label=CompLabel));
run;

proc sort data=FormatComp;
      by fmtname start;
run;

    /************************************************************************************/
    /* Part 4 - Match merge BASE and COMP SAS data sets to find matches and differences */
    /************************************************************************************/

data bothcats
     basecatonly(drop=CompEnd CompLabel)
     compcatonly(drop=BaseEnd BaseLabel);

merge FormatBase(in=a) FormatComp(in=b);
      by fmtname start;

          if a and b then output bothcats;
          if a and not b then output basecatonly;
          if b and not a then output compcatonly;

run;

    /*********************************************************************************/
    /* Part 5 - Create report data sets and macro flags to identify empty data sets  */
    /*********************************************************************************/

/* Create END mismatch data set */
data EndMismatch;
set bothcats(where=(BaseEnd ne CompEnd));
run;

/* Create LABEL mismatch data set */
data LabelMismatch;
set bothcats(where=(BaseLabel ne CompLabel));
run;


/* Create macro flags to identify non-empty data sets */

data _null_;
set sashelp.vtable(where=(libname="WORK"));

select(memname);
       when("BASECATONLY") do;
               if nobs > 0 then call symput("BASECATONLY","1");
                      else call symput("BASECATONLY","0");
       end;
       when("COMPCATONLY") do;
```

```
               if nobs > 0 then call symput("COMPCATONLY","1");
                      else call symput("COMPCATONLY","0");
        end;
        when("ENDMISMATCH") do;
               if nobs > 0 then call symput("ENDMISMATCH","1");
                      else call symput("ENDMISMATCH","0");
        end;
        when("LABELMISMATCH") do;
               if nobs > 0 then call symput("LABELMISMATCH","1");
                      else call symput("LABELMISMATCH","0");
        end;
        otherwise;
end;

run;

    /***********************************************/
    /* Part 6 - Create reports of the mismatches. */
    /***********************************************/

ODS RESULTS Off;

/* Report #1 - Catalog entries found in BASE but not in COMP */

%IF &BASECATONLY = 1 %THEN %DO;

        ods WORD file="&REPORTS\Catalog entries found in BASE (&BASECAT) But Not Found In COMP (&COMPCAT).docx";

        Proc print data=basecatonly noobs;
            var start baseLABEL baselabel;
               by fmtname;
               id fmtname;
        title1 "Catalog Entries Found in the BASE Catalog That Are Not in the COMP Catalog";
        title2 "When Attempting to Match on: FMTNAME and START";
        footnote1 "BASE Catalog is: &BASECAT in &BASELIB";
        footnote2 "COMP Catalog is: &COMPCAT in &COMPLIB";
        run;

        ods word close;

%END;

/* Report #2 - Catalog entries found in COMP but not in BASE */

%IF &COMPCATONLY = 1 %THEN %DO;

        ods WORD file="&REPORTS\Catalog entries found in COMP (&COMPCAT) But Not Found In BASE (&BASECAT).docx";

        Proc print data=compcatonly noobs;
            var start compLABEL complabel;
               by fmtname;
               id fmtname;
        title1 "Catalog Entries Found in the COMP Catalog That Are Not Found in the BASE Catalog";
        title2 "When Attempting to Match on: FMTNAME and START";
        footnote1 "COMP Catalog is: &COMPCAT in &COMPLIB";
        footnote2 "BASE Catalog is: &BASECAT in &BASELIB";
        run;

        ods word close;

%end;


/* Report #3 - Catalog entries with a mismatch on END */

%IF &ENDMISMATCH = 1 %THEN %DO;

        ods WORD file="&REPORTS\Catalog entries found in BASE (&BASECAT) and COMP (&COMPCAT) With Mismatched
                      Values for END.docx";

        Proc print data=EndMismatch noobs;
            var start BaseEnd CompEnd;
               by fmtname;
               id fmtname;
        title1 "Catalog Entries Found in Both the BASE and the COMP Catalog";
        title2 "That Have Mismatched Values for END";
        footnote1 "BASE Catalog is: &BASECAT in &BASELIB";
        footnote2 "COMP Catalog is: &COMPCAT in &COMPLIB";
        run;

        ods word close;

%END;

/* Report #4 - Catalog entries with a mismatch on LABEL */

%IF &LABELMISMATCH = 1 %THEN %DO;

        ods WORD file="&REPORTS\Catalog entries found in  BASE (&BASECAT) and COMP (&COMPCAT) With Mismatched
                        Values for LABEL.docx";

        Proc print data=LabelMismatch noobs;
            var start BaseLabel CompLabel;
                by fmtname;
                id fmtname;
        title1 "Catalog Entries Found in Both the BASE and the COMP Catalog";
        title2 "That Have Mismatched Values for LABEL";
        footnote1 "BASE Catalog is: &BASECAT in &BASELIB";
        footnote2 "COMP Catalog is: &COMPCAT in &COMPLIB";
        run;

        ods word close;

%END;

/* Report #5 - Report for when there are no mismatches */

%IF &BASECATONLY = 0 %THEN
        %IF &COMPCATONLY = 0 %THEN
                %IF &ENDMISMATCH = 0 %THEN
                        %IF &LABELMISMATCH = 0 %THEN %DO;

        ods WORD file="&REPORTS\No Differences Found Between Entries in the BASE (&BASECAT) and COMP (&COMPCAT)
                        Catalogs.docx";

        Proc print data=bothcats noobs;
            var start BaseEnd CompEnd BaseLabel CompLabel;
                by fmtname;
                id fmtname;
        title1 "No Mismatched Catalog Entries Found Between the BASE and the COMP Catalogs";
        title2;
        footnote1 "BASE Catalog is: &BASECAT in &BASELIB";
        footnote2 "COMP Catalog is: &COMPCAT in &COMPLIB";
        run;

        ods word close;

%END;

%MEND CATCOMPR;
```

Appendix B – Execute Compare Two SAS Format Catalogs Utility

```
/******************************************************************************************************/
/*  Program: Execute Compare Two SAS Format Catalogs Utility.sas                                      */
/*                                                                                                    */
/*  Author: Michael A. Raithel                                                                        */
/*                                                                                                    */
/*  Created: 9/6/2019                                                                                 */
/*                                                                                                    */
/*  Purpose: This is the driver program for the "Execute Compare Two SAS Format Catalogs Utility.sas" */
/*           %CATCOMPR macro program that compares two SAS format catalogs to determine if there are  */
/*           mismatches in fmtname, start, end, and label. The macro creates reports of any           */
/*           discrepancies it finds.                                                                  */
/*                                                                                                    */
/*  Note:    %CATCOMPR differentiates between the two catalogs by labeling them BASE and COMP.        */
/*                                                                                                    */
/*           BASE - Is the catalog used as the basis for the comparison; the one the other catalog    */
/*                  will be matched against.                                                          */
/*           COMP - Is the catalog that is to be compared against the BASE catalog to determine if    */
/*                  there are any differences between the two.                                        */
/*                                                                                                    */
/*  Parameters: You must specify the following five parameters, below                                */
/*                                                                                                    */
/*             BASELIB    - The directory holding the SAS format library you want to use as the BASE  */
/*                          for the comparison.                                                       */
/*             BASECAT    - The catalog that you want to use as the BASE for comparison.              */
/*                                                                                                    */
/*             COMPLIB    - The directory holding the SAS format library you want to COMPare          */
/*             COMPCAT    - The catalog that you want to COMPare to the base catalog.                 */
/*                                                                                                    */
/*             REPORTS    - The directory where reports are to be written.                            */
/*                                                                                                    */
/*  Outputs: This program may create one or more of the following reports, depending upon mismatches: */
/*                                                                                                    */
/*            - Catalog entries found in BASE (&BASECAT) But Not Found In COMP (&COMPCAT)             */
/*            - Catalog entries found in COMP (&COMPCAT) But Not Found In BASE (&BASECAT)             */
/*            - Catalog entries found in BASE (&BASECAT) and COMP (&COMPCAT) With Mismatched          */
/*              Values for END                                                                        */
/*            - Catalog entries found in  BASE (&BASECAT) and COMP (&COMPCAT) With Mismatched Values  */
/*              for LABEL                                                                             */
/*                                                                                                    */
/*            ...or if there are no mismatches, it will create this report:                           */
/*                                                                                                    */
/*            - No Differences Found Between Entries in the BASE (&BASECAT) and COMP (&COMPCAT)       */
/*              Catalogs                                                                              */
/*                                                                                                    */
/*  Change Log:                                                                                       */
/*                                                                                                    */
/******************************************************************************************************/

%include "\\Production\Directory\Compare Two SAS Format Catalogs Utility.sas";

%CATCOMPR(BASELIB = XXXX,
         BASECAT = XXXX,
          COMPLIB = XXXX,
          COMPCAT = XXXX,
          REPORTS = XXXX
          );
```

Appendix C –Create Test Format Catalogs

```
/***************************************************************************************/
/*  Program: Create Test Format Catalogs.sas                                           */
/*                                                                                     */
/*  Author: Michael A. Raithel                                                         */
/*                                                                                     */
/*  Created: 9/4/2019                                                                  */
/*                                                                                     */
/*  Purpose: This program creates two SAS format catalogs to test the format comparison*/
/*           program.                                                                  */
/*                                                                                     */
/***************************************************************************************/

options nodate nonumber linesize=200 dlcreatedir;

libname ctempcat "c:\temp\mycats";

/* Create potato chip format */
proc format library=ctempcat.chipsfmt;
      value chiptype
      1 = "Ruffles"
      2 = "Lay's"
      3 = "Kettle Chips"
      4 = "Miss Vickies"
      5 = "Pringles"
      other = "The subject does not eat potato chips very often and consequently has no
               stated preference"
      ;
      value agecat
      low - 12  = "Pre Teen"
      12 - 14   = "Young Teen"
      15 - high = "Older Teen"
      ;
```

```
       value $bynary
       "0" = "zero"
       "1" = "one"
         ;
run;

/* Create MODIFIED potato chip format */
proc format library=ctempcat.chipsfm2;
      value chiptype
      1 = "Ruffles"
      2 = "Michael's"
      3 = "Kettle"
      4 = "Miss Vickies"
      5 = "Pringles"
      other = "The subject does not eat potato chips very often and consequently has no
               stated preference"
      ;
      value agecat
      low - 12  = "Pre Teen"
      12 - 14   = "Young Teen"
      15 - 16  = "Medium Teen"
      17 - high = "Older Teen"
      ;
       value $mikebday
       "0427" = "Correct Month Day!"
       other  = "Incorrect Month/Day"
         ;
run;
```