

Paper 1549-2014
Build your Metadata with PROC CONTENTS and ODS OUTPUT
Louise S. Hadden, Abt Associates Inc.

ABSTRACT

Simply using an ODS destination to replay PROC CONTENTS output does not provide the user with attractive, usable metadata. Harness the power of SAS® and ODS output objects to create designer multi-tab metadata workbooks with the click of a mouse!.

INFORMATION OVERLOAD

SAS procedures can convey an enormous amount of information – sometimes more information than is needed. By manipulating procedural output and ODS output objects, we can pick and choose just the information we want to see. We can then harness the power of SAS reporting procedures and ODS to present the information we want to display accurately and attractively. We wanted to create a metadata Excel® workbook with an index of all data sets delivered with separate variable level tabs for each data set to accurately describe a monthly data submission to our client. The challenge was how to accomplish this in a single SAS program without post-editing. This paper and presentation will demonstrate how we met our challenge.

DISCOVERING ODS OUTPUT OBJECTS

Most SAS procedures generate ODS objects behind the scenes. SAS uses these objects in conjunction with style templates that have custom “buckets” for certain types of output to produce the output we see in all destinations (including the SAS listing).

Use the ODS TRACE command to identify ODS output objects. Choose the procedure that you are using, and “surround” the procedure (including any ODS GRAPHICS calls, etc. you may be using) with ODS TRACE and ODS TRACE OFF commands.

```
ODS TRACE ON / LABEL;  
ODS RTF FILE='yourfilename.rtf' PATH=ODSout STYLE=styles.journal2;  
ODS GRAPHICS ON;  
. . . your procedures here . . .  
ODS RTF CLOSE;  
ODS GRAPHICS OFF;  
ODS TRACE OFF;
```

If you are running interactively, you will see the ODS objects generated in the results window on the left, and can select, view and save these objects. In some cases, you can manipulate the objects. ODS trace information will be located in your SAS log. Although the ODS objects can be viewed in the results window, you will not get enough information about the objects to truly customize your output, so the log is essential for both interactive and batch users.

If you are running any SG procedures, use the following command before running the procedures.

```
ODS LISTING SGE=ON;
```

This creates an editable (using the ODS graphics editor) file which allows you to change titles, background colors, etc. in your graphs which you can then save.

If you run in batch mode, output from the ODS TRACE command will be included in your log file.

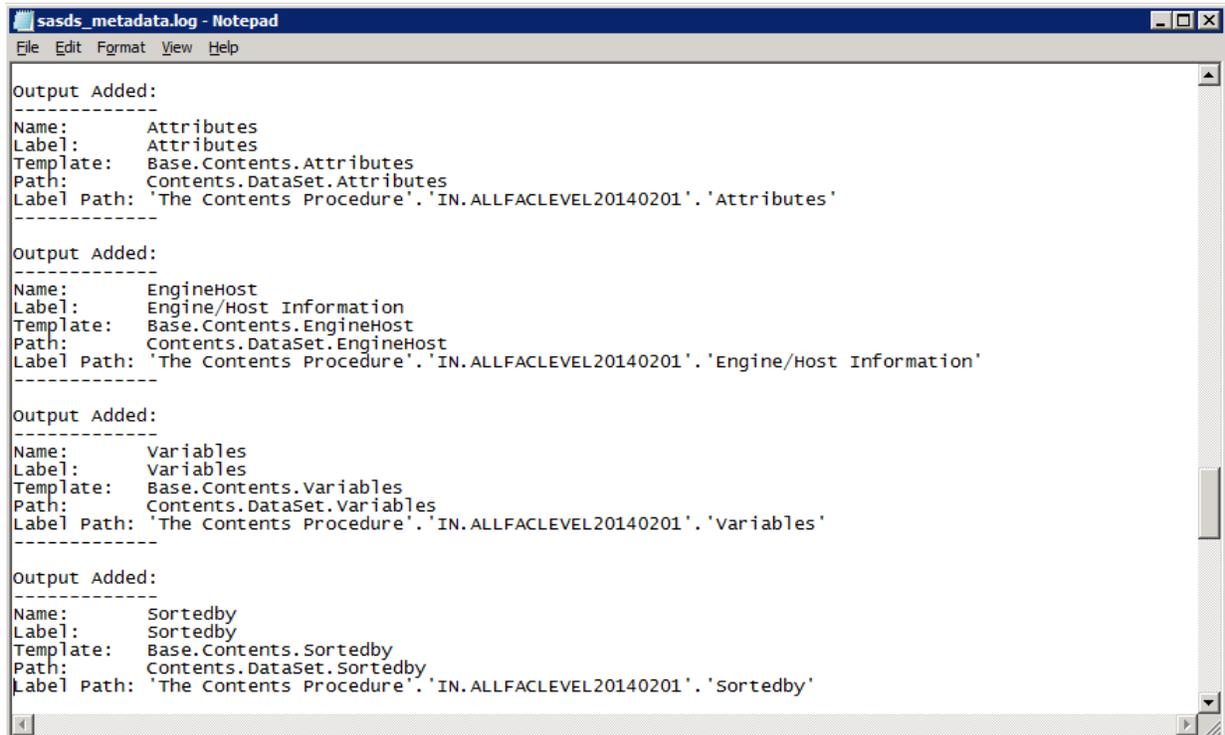
UTILIZING ODS OUTPUT OBJECTS

We wished to create an Excel workbook with an index tab with basic information about all data sets being delivered and documented for a given month in a given domain (files to be uploaded to DATA.MEDICARE.GOV), plus tabs for each data set with selected details on variables. In this case, we will be using the CONTENTS procedure in conjunction with ODS output objects. PROC CONTENTS does create an output data set (OUT=); however, it does not contain one desired piece of information, the variables that a data set are sorted by (if any). It also is rectangular, meaning that many variables on a system level are included on every record. ODS RTF or ODS HTML calls sandwiched around PROC CONTENTS also yields a lot of unnecessary information and would require a lot of post-editing. A macro was designed that produces a PROC CONTENTS for each data set, outputs ODS output objects, manipulates the objects and outputs two temporary data sets (a header line for the index tab, and a variable listing) for each data set.

```
ODS TRACE ON / LABEL;  
  
ODS RTF FILE='contents_allfaclevel.rtf' PATH=ODSout STYLE=styles.journal2;  
  
ODS GRAPHICS ON;  
  
    PROC CONTENTS DATA=dd.allfaclevel20140201;  
  
    RUN;  
  
ODS RTF CLOSE;  
  
ODS GRAPHICS OFF;  
  
ODS TRACE OFF;
```

EXPLORING ODS OUTPUT OBJECTS CREATED BY PROC CONTENTS

After running the code snippet above we can review our log and identify all the output objects generated by the contents procedure and their system names. Note that the output objects may vary! For example, PROC SURVEYFREQ produces different output objects based on whether you are running a one-way table or crosstabs, and may NOT generate chi-squared statistics if there are any 0 cells in a crosstab. In our case, if a data set is not sorted, PROC CONTENTS will not generate ODS output objects relating to sorting (SORTEDBY). If you want to report on sorting, you may need to have conditional processing that will indicated that a dataset is not sorted if there are no ODS output objects related to sorting.



```
sasds_metadata.log - Notepad  
File Edit Format View Help  
-----  
Output Added:  
Name:      Attributes  
Label:     Attributes  
Template:  Base.Contents.Attributes  
Path:     Contents.DataSet.Attributes  
Label Path: 'The Contents Procedure'. 'IN.ALLFACLEVEL20140201'. 'Attributes'  
-----  
Output Added:  
Name:      EngineHost  
Label:     Engine/Host Information  
Template:  Base.Contents.EngineHost  
Path:     Contents.DataSet.EngineHost  
Label Path: 'The Contents Procedure'. 'IN.ALLFACLEVEL20140201'. 'Engine/Host Information'  
-----  
Output Added:  
Name:      Variables  
Label:     Variables  
Template:  Base.Contents.Variables  
Path:     Contents.DataSet.Variables  
Label Path: 'The Contents Procedure'. 'IN.ALLFACLEVEL20140201'. 'Variables'  
-----  
Output Added:  
Name:      Sortedby  
Label:     Sortedby  
Template:  Base.Contents.Sortedby  
Path:     Contents.DataSet.Sortedby  
Label Path: 'The Contents Procedure'. 'IN.ALLFACLEVEL20140201'. 'Sortedby'
```

We can see that PROC CONTENTS generated many temporary ODS output objects (datasets) that go into the “print” output: ATTRIBUTES, VARIABLES, SORTEDBY and ENGINEHOST. These datasets are not necessarily on the same level and in order to create a single data set from the various pieces, more exploration is needed. To save these temporary files to either work or permanent data sets, use the ODS OUTPUT statement.

```

ODS OUTPUT ATTRIBUTES=atr VARIABLES=var SORTEDBY=sortedby;

PROC CONTENTS DATA=&inlib.&infi ;

RUN;

ODS OUTPUT CLOSE;

PROC CONTENTS DATA=var VARNUM;

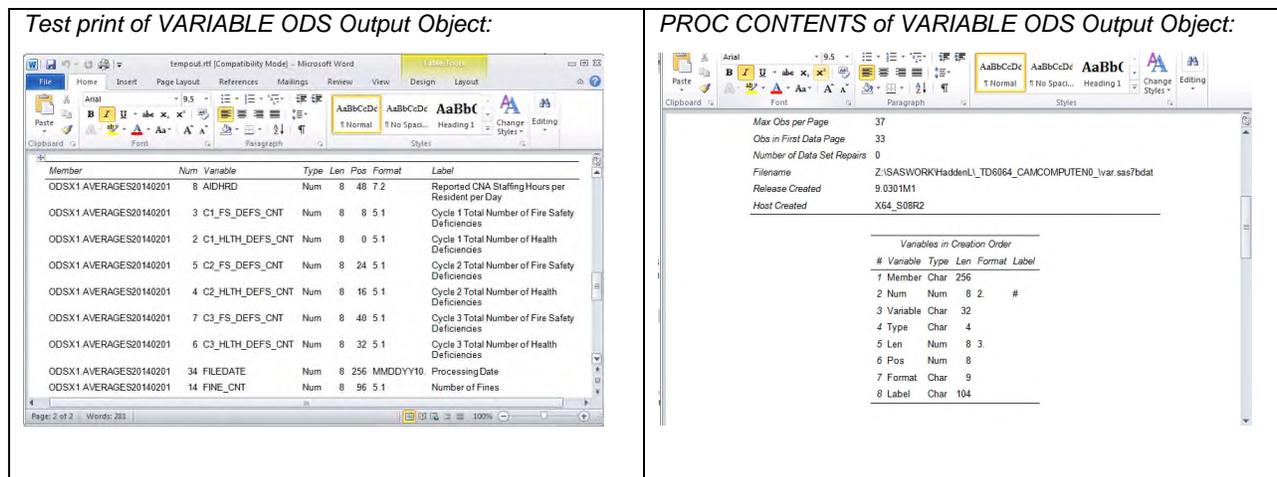
RUN;

PROC PRINT DATA=var (obs=10) NOOBS;

RUN; . . .

```

It is recommended that you do a test print of 10 observations even if you are working with a small data set, otherwise your output could be very large. 10 observations is generally enough to get a good picture. The contents is also very important as fields which might appear to be numeric may be character – SAS creates “print” variables for its output – and formatting give you valuable information for further manipulations. Variables within certain ODS output objects may not exist depending on the input data set(s). For example, in the case of PROC CONTENTS, you will not get the variable “informat” in the VARIABLES output object if no variables in the data set have informat.



Snippet of RTF output produced by an ODS sandwich around PROC CONTENTS statement: Note separate sections, not easy to parse out desired information.

Protection	Compressed	NO
Data Set Type	Sorted	YES
Label	Composite Provider-Level File (2014 0201)	
Data Representation	WINDOWS_64	
Encoding	wlatin1 Western (Windows)	

Engine/Host Dependent Information

Data Set Page Size	16384
Number of Data Set Pages	922
First Data Page	2
Max Obs per Page	17

MANIPULATING ODS OUTPUT OBJECTS

Reviewing all the output above allows us to pick and choose items and records from various output data sets, and merge them for reporting. The reasoning behind the test prints and PROC CONTENTS outputs becomes clear as it is vital to understand how the output objects mesh with one another. To enhance your final output, you can merge a descriptive data set in with longer variable descriptions and formatting variables that you might not want to carry in an analytic data set (for example, variables to create shading, italicizing, or bolding.) You can also create macro variables for printing from some ODS output objects using data steps and CALL SYMPUT. In this case, I used ATTRIBUTES, VARIABLES, and SORTEDBY output objects to construct two designer data sets for each data set to be documented. Note that because one of the input data bases is not sorted, we used conditional macro processing to create SORTEDBY if it does not exist.

```
/* first, set up a macro to (1) collect header level information for each DS and (2)
collect variable level information for each ds */
```

```
%MACRO cont2(runnum,inlib,infi,headfi);
```

```
ODS OUTPUT ATTRIBUTES=atr
            VARIABLES=var
            SORTEDBY=sortedby;
```

```
PROC CONTENTS DATA=&inlib.&infi ;
RUN;
```

```
PROC SORT DATA=var;
  BY num;
RUN;
```

```
DATA &infi;
  SET var (drop=member pos);
  RENAME num=varnum
         variable=name
         len=length;
RUN;
```

```
%IF %SYSFUNC(EXIST(sortedby)) %then %do;
```

```
/* if it exists, go ahead and make our day */
```

```
DATA sortvar (KEEP=sortedby);
  LENGTH sortedby $ 1000;
  SET sortedby (WHERE=(label1='Sortedby'));
  sortedby=cvalue1;
RUN;
```

```
%END;
```

```
%ELSE %DO; /* so it doesn't exist, we make it */
```

```
DATA sortvar;
  LENGTH sortedby $ 1000;
  sortedby='Ordered by Nation, then State';
```

```
RUN;
```

```
%END;
```

```
DATA sorted (KEEP=sorted);
  SET atr (WHERE=(label1='Data Set Type'));
  sorted=cvalue2;
RUN;
```

```
DATA memlabel (KEEP=memlabel);
```

```

        SET atr (WHERE=(label1='Label'));
        memlabel=cvalue1;
RUN;

data nvars (KEEP=nvars);
    SET atr (WHERE=(label1='Member Type'));
    nvars=INPUT(cvalue2,11.);
    FORMAT nvars commall.;
RUN;

data member (KEEP=memname nobs);
    SET atr (WHERE=(label1='Data Set Name'));
    memname=cvalue1;
    nobs=INPUT(cvalue2,11.);
    FORMAT nobs commall.;
RUN;

/* note lack of merge by variable deliberate */
DATA &headfi;
    MERGE member memlabel nvars sorted sortvar;
RUN;

%MEND cont2;

%cont2(1,in,allfaclevel&fileyear.&filedate.,header1);
%cont2(3,ODSx1,qualitymsr&fileyear.&filedate.,header3);
%cont2(2,ODSx1,alldeficiencies&fileyear.&filedate.,header2);
%cont2(4,ODSx1,ownership&fileyear.&filedate.,header4);
%cont2(5,ODSx1,penalties&fileyear.&filedate.,header5);
%cont2(6,ODSx1,averages&fileyear.&filedate.,header6);
%cont2(7,mds,nhqi_website_&QmlastQ_&filedate.,header7);
%cont2(8,mds,nhqi_website_q3_&filedate.,header8);

/* now build the header file */

DATA index;
    LENGTH memlabel sortedby $ 120 sorted $ 3;
    SET header1-header8;
RUN;

/* continued */

```

REPORTING ON DESIGNER DATA SETS BUILT FROM ODS OUTPUT OBJECTS

Our header file and individual data set variable listings have been built, and now it is time to create a multi-tab Excel workbook. The best way to do this and have some control over formatting is to use the ODS TAGSETS.EXCELXP tagset. We can name our tabs and set column widths within SAS for individual worksheets using tagset options: `sheet_name` names each tab, and `Absolute_Column_Width` sets the width for each column. We set and unset the sheet interval as table to generate a tab for each PROC PRINT.

```

/* open an Excel workbook created via ODS TAGSETS.EXCELXP */

ODS TAGSETS.EXCELXP FILE=".\"sasds_metadata_v&SASver..xml"
OPTIONS(orientation="Landscape") STYLE=styles.journal ;

ODS TAGSETS.EXCELXP OPTIONS(Absolute_Column_Width="32,90,10,10,10,60");
ODS TAGSETS.EXCELXP OPTIONS(sheet_interval="Table");
ODS TAGSETS.EXCELXP OPTIONS(sheet_interval="None");
ODS TAGSETS.EXCELXP OPTIONS(sheet_name="Index" );
PROC PRINT DATA=index NOOBS;
    VAR memname memlabel nobs nvars sorted sortedby;
FORMAT nobs nvars commall.;
RUN;

```

```

ODS TAGSETS.EXCELXP OPTIONS(Absolute_Column_Width="10,32,80,10,10,10");
ODS TAGSETS.EXCELXP OPTIONS(sheet_interval="Table");
ODS TAGSETS.EXCELXP OPTIONS(sheet_interval="None");
ODS TAGSETS.EXCELXP OPTIONS(sheet_name="AllFacLevel" );
PROC PRINT DATA=allfaclevel&fileyear.&filedate. NOOBS;
    VAR varnum name label type length format;
RUN;

. . . [repetitive code omitted - ask author for full code sample if needed] . . .

ODS TAGSETS.EXCELXP CLOSE;
QUIT;

```

XML TO XLS(X)

If you looked carefully at the code snippet above, you noticed that the TAGSETS.EXCELXP generates XML and not XLS or XLSX. We use DDE from within our SAS program to "Save As", as we work in a server environment and other options for saving as do not work for us. Once this has successfully completed, then we delete the XML version of the file.

```

/* make sure Excel is open before running program */

OPTIONS NOXWAIT NOXSYNC;

* This next routine starts the Excel application.  *;
FILENAME CMDS DDE 'excel|system';
X "'C:\Program Files\Microsoft Office\Office14\excel.exe'";

DATA _null_;
    x=sleep(3);
RUN;

FILENAME cmds dde 'excel|system';
DATA _null_;
FILE cmds;
X=sleep(10);
PUT                                "[open("S:\Projects\NH-COMPARE\Data_From_CMS\Monthly
Processing\Files&fileyear.&filedate.\5Output\DataMedicareGov\sasds_metadata_v&SASver..
xml" )]";
PUT '[ERROR("FALSE")]';
PUT                                "[SAVE.AS("S:\Projects\NH-COMPARE\Data_From_CMS\Monthly
Processing\Files&fileyear.&filedate.\5Output\DataMedicareGov\sasds_metadata_v&SASver..
xlsx" ,51)]";
X=sleep(2);
PUT '[close("false")]';
PUT '[quit]';
RUN;

X "del sasds_metadata_v&SASver..xml";
X "exit";

```

Screenshot of Index Tab in final workbook:

	A	B	C	D	E
1	memname	memlabel	nobs	nvars	sorted
2	IN.ALLFACLEVEL20140201	Composite Provider-Level File (2014.0201)	15653	123	YES
3	ODSX1.ALLDEFICIENCIES20140201	3 cycles of deficiencies, standard surveys and complaints, deduped – one record per PROVNUM-deficiency combination	488114	15	YES
4	ODSX1.QUALITYMSR20140201	MDS 3.0 Quality Measures for 3 quarters – one record per PROVNUM-quality measure combination	281754	19	YES
5	ODSX1.OWNERSHIP20140201	NH Ownership – one record per PROVNUM-owner-role combination	177002	11	YES
6	ODSX1.PENALTIES20140201	Civil Money Penalties and Payment Denials – one record per PROVNUM-penalty combination	8316	9	YES
7	ODSX1.AVERAGES20140201	State and US averages for selected NHC measures	54	34	NO
8	MDS.NHQI.WEBSITE_2013Q3_0201	Website format file for 2013Q3	281754	4	YES
9	MDS.NHQI.WEBSITE_Q3_0201	Website format file for 2013Q1 - 2013Q3	281754	4	YES

Screenshot of Defs Tab in final workbook:

	A	B	C	D	E	F
1	varnum	name	Label	Type	length	Format
2	1	PROVNUM	Federal Provider Number	Char	10	
3	2	PROVNAME	Provider Name	Char	50	
4	3	STATE	Provider State	Char	2	
5	4	SURVEY_DATE_OUTPUT	Survey Date	Num	8	YYMMDD10.
6	5	SURVEYTYPE	Survey Type	Char	11	
7	6	DEFPREF	Deficiency Prefix	Char	1	
8	7	TAG	Deficiency Tag Number	Char	4	\$CHAR4

CONCLUSION

Using the four basic concepts outlined below you can quickly and easily create your metadata with the click of a mouse.

1. Identify / locate your ODS output object(s) using ODS TRACE or output data set(s) .
2. Analyze your ODS output object(s) or output data set(s) using basic SAS procedures and review.
3. Manipulate your ODS output object(s) or output data set(s) using SAS data steps and/or procedures.
4. Report on your final data set.

Using ODS output objects and output data sets can be an enormous time saver. Application of this technique is not limited to metadata: the screenshot below was created by manipulating a number of different ODS output objects including Chi Squared statistics as well as frequency output from the PROC SURVEYFREQ procedure run on a large complex data set. No post-processing was required.

Table 1. Selected Characteristics by BMI Category

<i>Characteristic</i>	<i>Total 2005</i>	<i>Total 2005</i>	<i>Chi Squared P-Value</i>	<i>Underweight (BMI<18.5)</i>	<i>Normal Weight (BMI 18.5-24.9)</i>	<i>Overweight (BMI 25-29.9)</i>	<i>Obese (BMI 30+)</i>
Total Sample	15,195 (946,008)	100.0±0.00	NA	1.2±0.14	38.3±0.82	47.6±0.68	12.9±0.55
Sex							
Men	11,395 (804,888)	85.1±0.71	<.0001	0.9±0.16	35.2±0.94	50.2±0.79	13.7±0.61
Women	3,800 (141,120)	14.9±0.71	<.0001	2.8±0.30	56.1±1.24	32.9±1.01	8.2±0.63
Age							
17-20	1,187 (130,680)	13.8±0.99	<.0001	1.6±0.40	53.8±1.95	37.9±1.78	6.7±1.35
21-30	6,180 (481,590)	50.9±1.21	<.0001	1.5±0.21	42.4±0.98	44.7±0.96	11.4±0.58
31-39	4,610 (230,112)	24.3±1.05	<.0001	0.6±0.17	26.9±1.22	54.8±0.91	17.7±0.85
40+	3,218 (103,627)	11.0±0.75	<.0001	0.7±0.20	25.2±1.15	57.2±1.22	16.9±0.98
Educational Attainment							
High School or Less	4,409 (352,500)	37.3±1.56	<.0001	1.1±0.26	43.9±1.26	41.8±0.95	13.3±0.93
Some College	6,146 (380,144)	40.2±1.26	<.0001	1.6±0.20	34.5±0.80	49.7±0.79	14.2±0.67

With ODS output objects / output data sets and SAS, the sky is the limit!

REFERENCES AND RECOMMENDED READING

Guan, Calvin and Hadden, Louise. "EXCElIn in DDE: Unlock Useful Tools for Processing Excel®". Proceedings of NorthEast SAS Users Group 2013 Conference. September, 2013.

Hadden, Louise. "With a Trace: Making Procedural Output and ODS Output Objects Work For You". Proceedings of SAS Global Forum 2013 Conference. May 2013.

Zender, Cynthia. "The Greatest Hits: ODS Essentials Every User Should Know". Proceedings of SAS Global Forum 2011 Conference. April 2011.

ACKNOWLEDGEMENTS

The author would like to express her appreciation for the inspiration, creativity and help provided by Cynthia Zender of SAS and colleagues Christianna Williams, Calvin Guan, and Alan White.



CONTACT INFORMATION

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

Louise Hadden louise_hadden@abtassoc.com

Code samples (including for the table snippet shown above) are available upon request.

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.