

Automation of Conversion of SAS® Programs to Text files for submissions to the FDA

Sapan Shah
Sachin Agarwal

ABSTRACT

During Clinical Trials Statistical Programmers, works on creation of many SAS® programs to produce SDTM, ADaM datasets, Tables, Listings and Graphs as per the study requirements and various ad-hoc and post-hoc request also, which ends with “.sas” extension. The “.sas” extension SAS® programs are required for working on different SAS® software but for NDA submission process, one of the requirement by FDA is that all the submitted SAS® programs have to be in “.txt” format. In this paper, we will show how to utilize PC SAS® software and MS command prompt in order to change extension from “.sas” to “.txt” for submission to the FDA. This “.sas” to “.txt” conversion Macro helps Statistical Programmer in saving their valuable time by eliminating manual conversion as well as eliminate the error possibility, which might happen during manual conversion.

INTRODUCTION

In this paper, we are going to explain how to utilize PC SAS® software and MS command prompt in order to change extension from “.sas” to “.txt” for submission to the FDA without losing the indentation of the original program(s).

We will use an example to demonstrate how to utilize to different software’s in order to change the extension of the file.

THE EXAMPLE

This paper uses a PC SAS® 9.4 software program macro and embedded Windows SERVER 2008 command prompt within the SAS® program to demonstrate the automation process to change extension from “.sas” to “.txt”. In order to utilize this code, we will be required to make the copies of all the original “.sas” programs from the original directory to a temporary directory where we plan to save the converted “.txt” program files, which happens automatically using this code itself:

```
%macro Convert_SAS_txt(source=%nrstr(), prod=%nrstr());  
proc datasets lib=work nolist kill memtype=data;  
run;  
quit;  
%put &source.;  
%put &prod.;
```

```
/* THE DATA STEP MENTIONED BELOW IS GOING TO READ ALL THE FILES PRESENT IN THE
FOLDER WHOSE LOCATION IS DEFINED IN THE MACRO VARIABLE "SOURCE" AND CREATES A
DATASET WITH THE NAME SOURCE IN THE WORK LIBRARY */
```

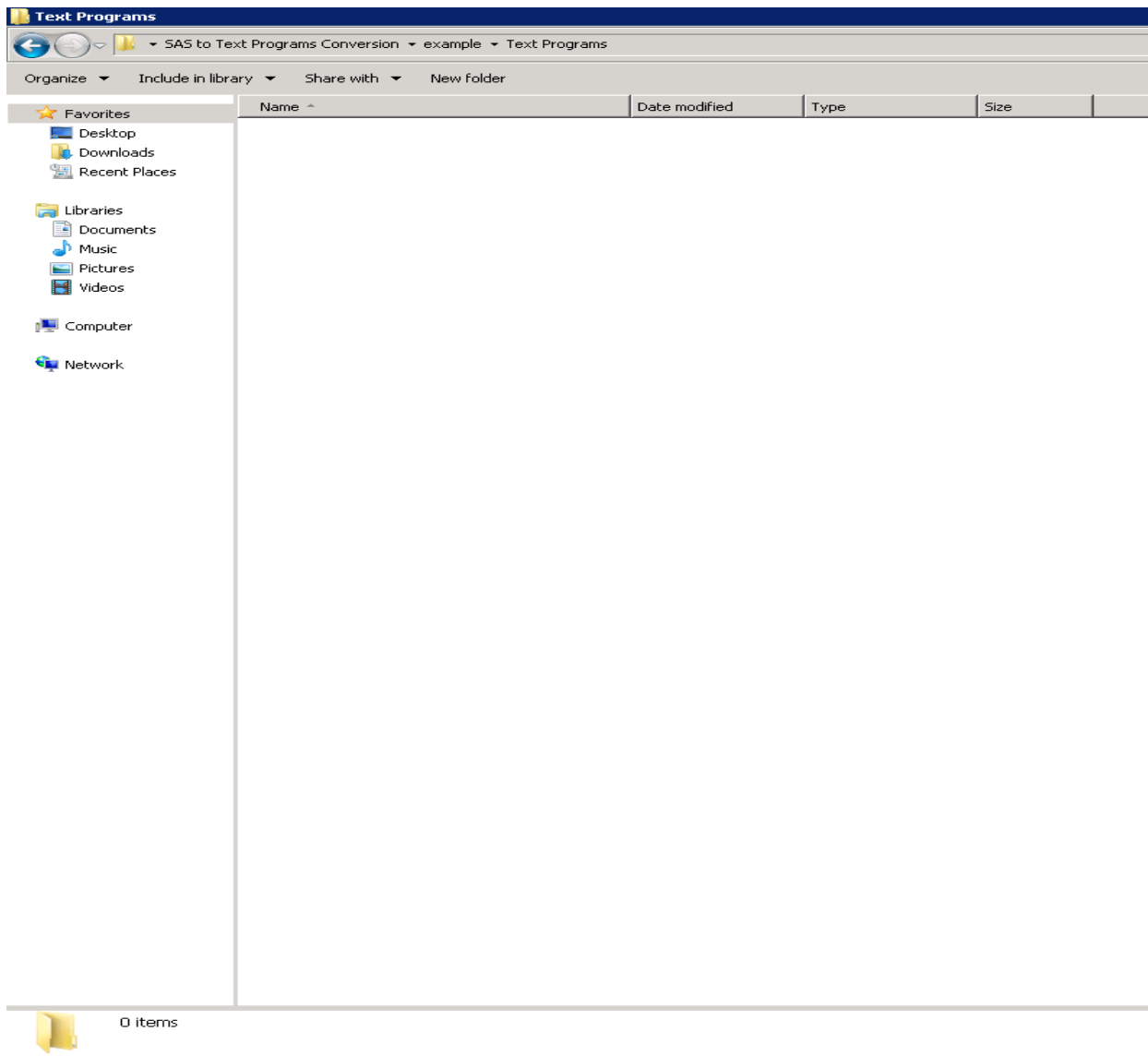
```
data source ;
infile "dir /b ""&source.\\" pipe trunccover;
input name $1000. ;
length name_ $1000.;
name_ =lowercase(name);
drop name;
rename name_ =name;
run;
```

Name	Date modified	Type	Size
adae	7/24/2018 11:16 AM	SAS System Program	9 KB
adcm	7/2/2018 11:42 AM	SAS System Program	8 KB
addv	7/23/2018 12:13 PM	SAS System Program	6 KB
adeg	7/24/2018 11:19 AM	SAS System Program	7 KB
adex	8/2/2018 10:58 AM	SAS System Program	10 KB
adnh	7/3/2018 8:58 AM	SAS System Program	3 KB
adneuro	7/17/2018 2:02 PM	SAS System Program	4 KB
adpe	7/3/2018 9:06 AM	SAS System Program	3 KB
adsl	7/19/2018 2:48 PM	SAS System Program	16 KB
adv5	7/24/2018 11:21 AM	SAS System Program	13 KB
t-14-01-01-01	5/4/2018 12:34 AM	SAS System Program	11 KB
t-14-01-01-02	5/4/2018 12:34 AM	SAS System Program	11 KB
t-14-01-01-03	5/4/2018 12:34 AM	SAS System Program	11 KB
t-14-01-01-04	5/4/2018 12:34 AM	SAS System Program	7 KB
t-14-01-01-05	5/4/2018 12:34 AM	SAS System Program	7 KB
t-14-01-02-01	5/4/2018 12:34 AM	SAS System Program	8 KB
t-14-01-03-01-01	5/4/2018 12:34 AM	SAS System Program	17 KB
t-14-01-03-01-03	5/4/2018 12:34 AM	SAS System Program	21 KB
t-14-01-03-01-10	5/4/2018 12:34 AM	SAS System Program	17 KB
t-14-01-03-02-01	5/4/2018 12:34 AM	SAS System Program	16 KB
t-14-01-03-02-03	5/21/2018 2:18 PM	SAS System Program	20 KB
t-14-01-03-02-10	5/4/2018 12:34 AM	SAS System Program	16 KB
t-14-01-03-02-11	5/4/2018 12:34 AM	SAS System Program	7 KB
t-14-01-03-02-13	5/4/2018 12:34 AM	SAS System Program	7 KB
t-14-01-03-02-14	5/4/2018 12:34 AM	SAS System Program	8 KB
t-14-01-03-03	5/4/2018 12:34 AM	SAS System Program	9 KB
t-14-01-04-01	5/4/2018 12:34 AM	SAS System Program	12 KB
t-14-01-04-03	5/4/2018 12:34 AM	SAS System Program	10 KB
t-14-02-01-01	4/27/2018 12:09 PM	SAS System Program	10 KB
t-14-02-01-02	4/27/2018 12:09 PM	SAS System Program	16 KB
t-14-02-02-01-01	4/27/2018 12:09 PM	SAS System Program	11 KB
t-14-02-02-01-02	4/27/2018 12:09 PM	SAS System Program	9 KB
t-14-02-02-01-03	8/22/2018 5:35 PM	SAS System Program	18 KB
t-14-02-02-02-01	4/27/2018 12:09 PM	SAS System Program	11 KB
t-14-02-02-03-01	4/27/2018 12:09 PM	SAS System Program	11 KB
v_adeq5d5l	7/24/2018 1:19 PM	SAS System Program	5 KB
v_ades	7/24/2018 2:12 PM	SAS System Program	5 KB
v_adfq	7/24/2018 3:23 PM	SAS System Program	5 KB
v_adnms	7/24/2018 3:23 PM	SAS System Program	7 KB

39 items

/* THE DATA STEP MENTIONED BELOW IS GOING TO READ ALL THE FILES WHICH ARE ALREADY PRESENT IN THE FINAL FOLDER WHERE ALL THE CONVERTED SAS PROGRAMS IN TEXT FORMAT ARE GOING TO BE PLACED */

```
data prod ;  
  infile "dir /b ""&prod.\"" " pipe trunccover;  
  input name $1000. ;  
  length name_ $1000.;  
  name_=lowercase(name);  
  drop name;  
  rename name_=name;  
run;
```

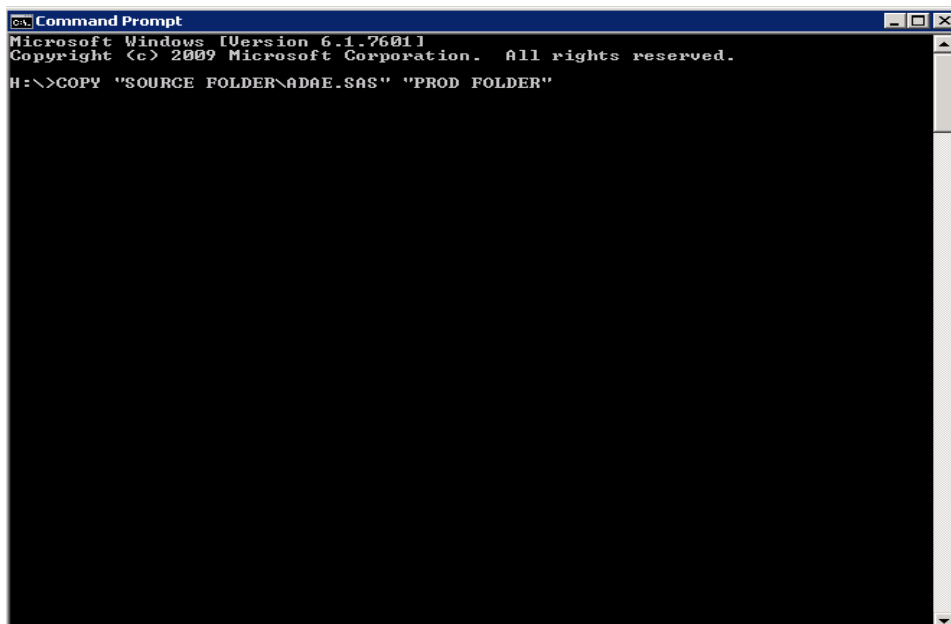


```
/* THIS PROC SQL STEP WILL CREATE A DATASET WITH THE NAMES OF ALL SAS PROGRAMS  
PRESENT IN THE SOURCE FOLDER FOR WHICH THERE ARE NO CORRESPONDING NAME TEXT  
FILE PRESENT IN PROD FOLDER: FOR EXAMPLE IF THERE IS A SAS PROGRAM WITH THE NAME  
ADSL IN SOURCE FOLDER AND THERE IS A CORRESPONDING ADSL TEXT PROGRAM IN PROD  
FOLDER THEN THE DATASET CREATED BELOW WILL NOT CONTAIN ADSL */
```

```
proc sql noprint ;  
  create table newfiles as  
  select * from source  
  where not (upcase(name) in (select upcase(name) from prod))  
  and upcase(scan(name,-1,'.'))= 'SAS';  
quit;
```

```
/* THE STEP MENTIONED BELOW WILL USE THE WINDOWS COMMAND PROMPT TO COPY SAS  
PROGRAMS PRESENT IN THS SOURCE FOLDER BUT NOT IN PROD FOLDER AS SPECIFIED ABOVE  
TO THE PROD FOLDER ONE BY ONE USING COPY COMMAND */
```

```
data _null_ ;  
  set newfiles ;  
  length cmd $2000. ;  
  cmd = catx(' ', 'copy', quote(catx('\', "&source.", name)), quote("&prod."));  
  infile cmd pipe filevar=cmd end=eof ;  
  do while (not eof);  
    input;  
    put _infile_ ;  
  end;  
run;
```



The image shows a screenshot of a Windows Command Prompt window. The title bar reads "Command Prompt". The window content displays the following text: "Microsoft Windows [Version 6.1.7601] Copyright (c) 2009 Microsoft Corporation. All rights reserved. H:\>COPY "SOURCE FOLDER\ADAE.SAS" "PROD FOLDER". The rest of the window is black, indicating that the command has been executed and the output is not visible.

```
/* THE DATA STEP BELOW WILL CREATE A DATASET CONTAINING THE NAMES OF ALL THE SAS
PROGRAMS WHICH ARE NOW PRESENT IN PROD FOLDER: THIS WILL INCLUDE ALL THE SAS
PROGRAMS WHICH ARE COPIED FROM SOURCE FOLDER AS WELL AS THE SAS PROGRAMS
WHICH WERE PRESENT IN THE PROD DOLDER BEFORE ANY SAS PROGRAM WAS COPIED BY
WINDOWS COPY COMMAND FROM SOURCE FOLDER */
```

```
filename prod1 pipe "dir ""%unquote(&prod.)"" /b" lrecl=32767;
```

```
data prod_set (where=(scan(name,-1,'.')= 'sas'));
    infile prod1 trunccover;
    input name $char1000.;
    prod_ord = _n_;
    nam_ = upcase(scan(name,1,"."));
run;
```

```
proc sort data=prod_set;
    by nam_;
run;
```

```
/* THE DATA STEP BELOW AND PROC SQL COMBINE WILL CREATE A MACRO VARIABLE WHICH
WILL CONTAIN THE NAMES OF ALL THE SAS PROGRAMS WITH ".sas" EXTENSION PRESENT IN
PROD FOLDER */
```

```
data prod_set2;
    length third_nam $4000;
    set prod_set end=eof;
    sec_nam= cats("","nam_", "");
    retain third_nam "";
    third_nam=catx(" ",sec_nam, third_nam);
    keep third_nam;
    if eof;
run;
```

```
proc sql noprint;
    select third_nam into :set_both from prod_set2;
quit;
```

```
%put &set_both;
```

```
/* THIS DATA STEP WILL CREATE MULTIPLE MACRO VARIABLES WITH EACH MACRO VARIABLE  
HAVING THE VALUE OF SAS PROGRAM NAME STORED INTO IT */
```

```
data _null_ ;  
    set prod_set;  
    call symput("newset" || strip(put(_n_,best12.)),strip(lowercase(scan(name,1,'.'))));  
run;
```

```
%put &newset1. , &newset2.;
```

```
/* THE PROC SQL STEP MENTIONED BELOW WILL COUNT THE NUMBER OF SAS PROGRAMS  
PRESENT IN THE PROD FOLDER */
```

```
proc sql noprint;  
    select count(*) into :numb_prod from prod_set;  
run;  
quit;
```

```
%let numb_prod=%left(&numb_prod.);
```

```
%put &numb_prod.;
```

```
/* THIS DATA STEPS MENTIONED BELOW WILL WORK IN FOLLOWING WAY:
```

1. EVERY TIME DO LOOP IS ITERATED A TEMPDATA WILL BE CREATED FOR EACH SAS PROGRAM NOW PRESENT IN THE PROD FOLDER. IT WILL CONTAIN THE CONTENT OF THE SAS PROGRAM IN IT.

2. DATA _NULL_ STEP WILL READ THE TEMPDATA CREATED IN THE 1ST STEP AND CREATE A CORRESPONDING SAS PROGRAM IN THE TEXT FORMAT IN THE PROD FOLDER.

3. THE SECOND DATA _NULL_ STEP IS THEN USED TO DELETE THE SAS PROGRAM WITH .SAS EXTENSION IN THE PROD FOLDER SO AS THE IN THE END PROD FOLDER ONLY CONTAINS THE SAS PROGRAM IN THE TEXT FORMAT (THIS IS AN OPTIONAL STEP, THESE TEMP PROGRAMS CAN BE DELETED DIRECTLY FROM THE WINDOWS FOLDER USING RIGHT CLICK METHOD)*/

```
%do i = 1 %to &numb_prod.;
```

```
    filename progs "&prod.\&&newset&i...sas";
```

```
    data tempdata ;  
        length filename $2000;  
        infile progs filename=filename eov=eov length=L;  
        input line $varying2000. l;  
        retain=_n_ ;  
        if _n_ eq 1 or eov then do;  
            eov=0;
```

```

        end;
run;

data _null_ ;
    set tempdata (keep=LINE);
    FILE "&prod.\&&newset&i...txt" ;
    PUT LINE;
run;

data _null_ ;
    fname="tempfile";
    rc=filename(fname,"&prod.\&&newset&i...sas");
    if rc = 0 and fexist(fname) then
        rc=fdelete(fname);
    rc=filename(fname);
run;

%end;
%mend Convert_SAS_txt;

```

Name	Date modified	Type	Size
adae	3/14/2020 11:38 PM	Text Document	9 KB
adcm	3/14/2020 11:38 PM	Text Document	8 KB
addv	3/14/2020 11:38 PM	Text Document	6 KB
adeg	3/14/2020 11:38 PM	Text Document	7 KB
adex	3/14/2020 11:38 PM	Text Document	10 KB
admh	3/14/2020 11:38 PM	Text Document	3 KB
adneuro	3/14/2020 11:38 PM	Text Document	4 KB
adpe	3/14/2020 11:38 PM	Text Document	3 KB
adsl	3/14/2020 11:38 PM	Text Document	15 KB
adv	3/14/2020 11:38 PM	Text Document	13 KB
t-14-01-01-01	3/14/2020 11:38 PM	Text Document	10 KB
t-14-01-01-02	3/14/2020 11:38 PM	Text Document	10 KB
t-14-01-01-03	3/14/2020 11:38 PM	Text Document	9 KB
t-14-01-01-04	3/14/2020 11:38 PM	Text Document	6 KB
t-14-01-01-05	3/14/2020 11:38 PM	Text Document	6 KB
t-14-01-02-01	3/14/2020 11:38 PM	Text Document	7 KB
t-14-01-03-01-01	3/14/2020 11:38 PM	Text Document	15 KB
t-14-01-03-01-03	3/14/2020 11:38 PM	Text Document	19 KB
t-14-01-03-01-10	3/14/2020 11:38 PM	Text Document	15 KB
t-14-01-03-02-01	3/14/2020 11:38 PM	Text Document	15 KB
t-14-01-03-02-03	3/14/2020 11:38 PM	Text Document	18 KB
t-14-01-03-02-10	3/14/2020 11:38 PM	Text Document	15 KB
t-14-01-03-02-11	3/14/2020 11:38 PM	Text Document	6 KB
t-14-01-03-02-13	3/14/2020 11:38 PM	Text Document	6 KB
t-14-01-03-02-14	3/14/2020 11:38 PM	Text Document	7 KB
t-14-01-03-03	3/14/2020 11:38 PM	Text Document	8 KB
t-14-01-04-01	3/14/2020 11:38 PM	Text Document	10 KB
t-14-01-04-03	3/14/2020 11:38 PM	Text Document	8 KB
t-14-02-01-01	3/14/2020 11:38 PM	Text Document	9 KB
t-14-02-01-02	3/14/2020 11:38 PM	Text Document	15 KB
t-14-02-02-01-01	3/14/2020 11:38 PM	Text Document	10 KB
t-14-02-02-01-02	3/14/2020 11:38 PM	Text Document	8 KB
t-14-02-02-01-03	3/14/2020 11:38 PM	Text Document	16 KB
t-14-02-02-02-01	3/14/2020 11:38 PM	Text Document	10 KB
t-14-02-02-03-01	3/14/2020 11:38 PM	Text Document	10 KB
v_adeq5d5l	3/14/2020 11:38 PM	Text Document	5 KB
v_ades	3/14/2020 11:38 PM	Text Document	5 KB
v_adfq	3/14/2020 11:38 PM	Text Document	5 KB
v_adhmss	3/14/2020 11:38 PM	Text Document	6 KB

PROGRAMMING TIPS

1. %NRSTR quoting function:

Quotation marks are required in the SOURCE and PROD key parameter macro since the directory folder names may contain "-" in the name. SAS® software will hang while execution without the quotation marks.

2. SOURCE = The macro variable SOURCE specifies the name of the folder where original SAS® programs with ".sas" extension are present. %NRSTR is must while using this keyword parameter, (Required).

3. PROD = The macro variable SOURCE specifies the name of the folder where original SAS® programs with ".sas" extension are present. %NRSTR is must while using this keyword parameter, (Required).

CONCLUSION

In this paper, we chose to demonstrate the use of Windows command prompt in order to copy the SAS® programs from one windows folder to another windows folder using the SAS® software as well as changing the extension of the program from ".sas" to ".txt". This will reduce the manual work load as well as margin of error of the statistical programmer during the time of FDA submission. The restriction this code is, it works fine with Windows based SERVER PC SAS® software and the Windows com