



Customizable Smart Search Tool for Project Leads

Paper: CT07

Bibas Timilsina

Principal Statistical Programmer (FSP)



5-8 November
ICC Birmingham

EU 2023

The Clinical Data
Science Conference



parexel[®]
FSP

AGENDA

- **Introduction**

- Various real time **scanning** scenarios
- SAS FILENAME statement

- **Script template**

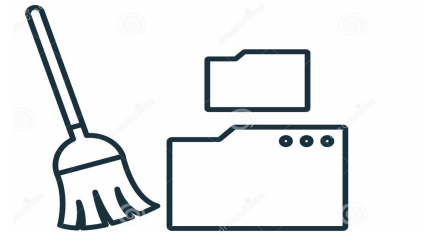
- Logic
- Program

- **Examples/Uses**

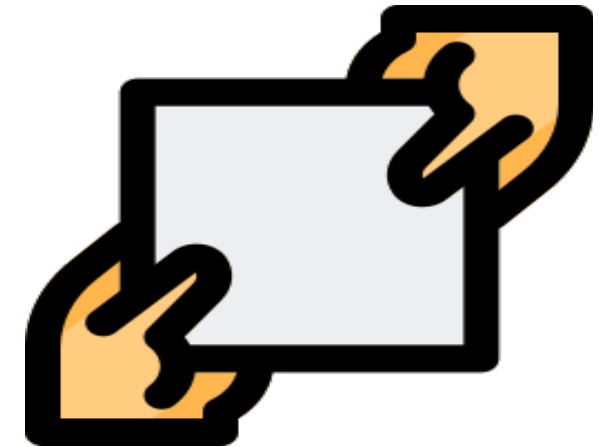
- Token scan
- Quality control tools

- **Conclusion**

- **Questions**



DATA CLEANING



Introduction

Various real time scanning scenarios

- Identifying the
 - ✓ Reference output programs using specific parameters, datasets, and/or analysis
 - ✓ Possible updates based on known scenarios for existing programs
 - ✓ Common data issues in SDTM level
 - ✓ Run status
 - ✓ Recent external files and adding automation on SDTM and ADaM programming
 - ✓ Log issues
 - ✓ QC status by scanning compare output files
 - ✓ Good programming practice related issues
 - ✓ List of program files for batch processing



SAS FILENAME statement

“The FILENAME statement has been *around a long time*. It has been traditionally used to *reference an external file on disk*. However, there are many “*not so well-known*” functionalities, for which the FILENAME statement is a *boon*. The FILENAME statement specifies a file reference (commonly called as fileref), that *serves as a link to an external file, device or an access method*. It *doesn’t do any type of data processing but still in the world of SAS, this is one of the most commonly and frequently used statements.*”

1. Reading data from an external file: fileref to the external file and then use the fileref in subsequent DATA or PROC steps to read the data from that file into SAS.

2. Writing data to an external file: fileref to an external file and then use the fileref in a DATA step to write data from SAS to that file.

3. Accessing system files: fileref to a system file, such as a text file containing configuration settings or parameters. This allows you to read or modify the contents of the system file within your SAS program.

4. Importing and exporting data: fileref to an input or output file. This helps to establish a connection between SAS and the external software for seamless data transfer.

5. Accessing external databases: fileref to an external database, allowing you to read or write data directly from/to that database within your SAS program.

FILENAME statement was described in The SAS Language Guide for Personal Computers (Release 6.03 Edition) (SAS, 1988):

- The FILENAME statement associates a SAS fileref (a file reference name) with an external file's complete name (directory plus file name). The fileref is then used as a shorthand reference to the file in the SAS programming statements that access external files (INFILE, FILE, and %INCLUDE). Associating a fileref with an external file is also called defining the file. Use the FILENAME statement to define a file before using a fileref in an INFILE, FILE, OR %INCLUDE statement.

Syntax: *FILENAME fileref <device-type/access methods> 'external-file' <ENCODING='encoding-value'> <options><operating-environment- options>;*

Syntax detail in associated paper CT07

Script template

Logic: Use of **FILENAME** statement with **PIPE** combined with **GREP/LS/DIR** options

Program

```
%macro getfilename ( path          =,
                    Out            =,
                    Keyword        = ,      /*blank for LS option*/
                    filenameoptions= grep,   /*LS, GREP, DIR... */
                    filetype       =sas      /* file extension sas, log, lst, txt, out, sas7bdat, ... */
                    );
%let primarypath=&path.;
filename check pipe " &filenameoptions. -i &keyword. &primarypath.*.&filetype.";
data &out.;
  length name $1000.;
  infile check dlm='??'; /*can be any special character which is not expected in the files that we are scanning*/

  input name;

  /*scenario based conditions*/

  filename=uppercase(scan(scan(name,10,'/'),1,'.'));
run;
%mend getfilename;

/**** example call****/

%getfilename(path=dummy1/dummy2/dummy3/dummy4/dummy5/share/data_analysis/prod/program/, /*physical path*/

             out=check_HISTO1, /* output dataset name*/

             keyword= HISTO1 , /*blank for ls*/

             filenameoptions= grep, /*ls, grep, Dir... */

             filetype=sas /*file extension sas, log, lst , sas7bdat, out, txt,..... */

             );
```

Examples/Uses

Token scan

```
%getfilename(path=dummy1/dummy2/dummy3/dummy4/dummy5/share/data_analysis/prod/program/,  
  out=test,  
  keyword= HISTO1,  
  filenameoptions= grep,  
  filetype=sas  
);
```

name	filename
dummy1/dummy2/dummy3/dummy4/dummy5/share/data_analysis/prod/program/ad_adsl.sas: data subgroups (keep=usubjid histo histo1);	AD_ADSL
dummy1/dummy2/dummy3/dummy4/dummy5/share/data_analysis/prod/program/g_ae_crs_sankey.sas: /* histo1 in ('DLBCL' 'trFL' 'FL 3B') */	G_AE_CRS_SANKEY
dummy1/dummy2/dummy3/dummy4/dummy5/share/data_analysis/prod/program/g_ef_swim.sas: * 22SEP2023 timilsib: Updated HISTO1 to HISTO2	G_EF_SWIM
dummy1/dummy2/dummy3/dummy4/dummy5/share/data_analysis/prod/program/t_dm_subage65.sas: keep studyid usubjid saffl iapop agegrp65 histo1;	T_DM_SUBAGE65
dummy1/dummy2/dummy3/dummy4/dummy5/share/data_analysis/prod/program/t_ef_concord.sas: adsl (in=c keep= studyid usubjid histgroup histo histo1 actarm itfl saffl actarmcd respm3);	T_EF_CONCORD

```
%getfilename(path=dummy1/dummy2/dummy3/dummy4/dummy5/data_analysis/dummy_reporting/prod/output/,  
  out=test,  
  keyword= unequal,  
  filenameoptions= grep,  
  filetype=lst  
);
```

name	filename
dummy1/dummy2/dummy3/dummy4/dummy5/data_analysis/dummy_reporting/prod/output/qc_ad_adae.lst: Number of Observations with Some Compared Variables Unequal: 5.	QC_AD_ADAE
dummy1/dummy2/dummy3/dummy4/dummy5/data_analysis/dummy_reporting/prod/output/qc_ad_adce.lst: Number of Observations with Some Compared Variables Unequal: 5.	QC_AD_ADCE
dummy1/dummy2/dummy3/dummy4/dummy5/data_analysis/dummy_reporting/prod/output/qc_ad_adex.lst: Number of Observations with Some Compared Variables Unequal: 0.	QC_AD_ADEX
dummy1/dummy2/dummy3/dummy4/dummy5/data_analysis/dummy_reporting/prod/output/qc_ad_adtte.lst: Number of Observations with Some Compared Variables Unequal: 0.	QC_AD_ADTTE

```
%getfilename(path=dummy1/dummy2/dummy3/dummy4/dummy5/data_analysis/dummy_reporting/prod/output/,  
  out=test,  
  keyword= ,  
  filenameoptions= ls,  
  filetype=lst  
);
```

name	filename
26284961 dummy1/dummy2/dummy3/dummy4/dummy5/data_analysis/dummy_reporting/prod/output/qc_ad_adae.lst	QC_AD_ADAE
26153418 dummy1/dummy2/dummy3/dummy4/dummy5/data_analysis/dummy_reporting/prod/output/qc_ad_adce.lst	QC_AD_ADCE
26219161 dummy1/dummy2/dummy3/dummy4/dummy5/data_analysis/dummy_reporting/prod/output/qc_ad_adex.lst	QC_AD_ADEX
26350699 dummy1/dummy2/dummy3/dummy4/dummy5/data_analysis/dummy_reporting/prod/output/qc_ad_adtte.lst	QC_AD_ADTTE

```
%getfilename(path=dummy1/dummy2/dummy3/dummy4/dummy5/data_analysis/dummy_reporting/data/sdtm/,  
  out=test,  
  keyword= RE-Treat,  
  filenameoptions= grep,  
  filetype=sas7bdat  
);
```

name	filename
Binary file dummy1/dummy2/dummy3/dummy4/dummy5/data_analysis/dummy_reporting/data/sdtm/eg.sas7bdat matches	EG
Binary file dummy1/dummy2/dummy3/dummy4/dummy5/data_analysis/dummy_reporting/data/sdtm/lb.sas7bdat matches	LB
Binary file dummy1/dummy2/dummy3/dummy4/dummy5/data_analysis/dummy_reporting/data/sdtm/rs.sas7bdat matches	RS
Binary file dummy1/dummy2/dummy3/dummy4/dummy5/data_analysis/dummy_reporting/data/sdtm/tv.sas7bdat matches	TV
Binary file dummy1/dummy2/dummy3/dummy4/dummy5/data_analysis/dummy_reporting/data/sdtm/vs.sas7bdat matches	VS

More options in paper CT07

Quality control tools

- Log checker

- ✓ Scan log files for

ERROR|WARNING|CRITICAL|ALERT|USER_PROGRAM_(ERRORS|WARNINGS)

NOTE: +(uninitial|invalid|MERGE statement|not executed|format was too small|converted |not found|not be loaded|missing|outside the axis range|not be performed|Division by zero

- ✓ Collect based on filename and combine with validation log to get first line and QC'er details
 - ✓ Clean up and generate the report in suitable format and share

- QC status checker

- ✓ Scan compare report files for

Unequal|Differing Attributes|not found|but not in|warning

- ✓ Identify the file modification dates for first line and QC files and compare
 - ✓ Collect based on filename and combine with validation log to get first line and QC'er details
 - ✓ Clean up and generate the report in suitable format and share

Sample program in paper CT07

Conclusion

Make use of FILENAME statement with PIPE option and scan around.

Say **NO** to manual scan where possible, Time matters!

Do not scroll, scan it!



Thank you

CONTACT INFORMATION

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

Author Name: Bibas Timilsina

Company: Parexel

Address: Parexel International, Navigation House, 1 South Quay Drive, Sheffield, S2 5SU, United Kingdom

Email: Bibas.timilsina@parexel.com

LinkedIn: [linkedin.com/in/bibas-timilsina-2075775b](https://www.linkedin.com/in/bibas-timilsina-2075775b)

Web: <https://www.parexel.com/>