PharmaSUG China 2017 - 57

Creating a list of files within a folder and inserting EXCEL hyperlinks to open each file

Vincent Fan, SANOFI Corporation, Beijing, China
Qingan Chang, SANOFI Corporation, Beijing, China

ABSTRACT

It is a common requirement to generate a list of files contained within a folder. This paper will demonstrate how to access the directory information, how to create a Microsoft EXCEL file of the list, and especially, how to insert hyperlinks in the EXCEL list to open these files.

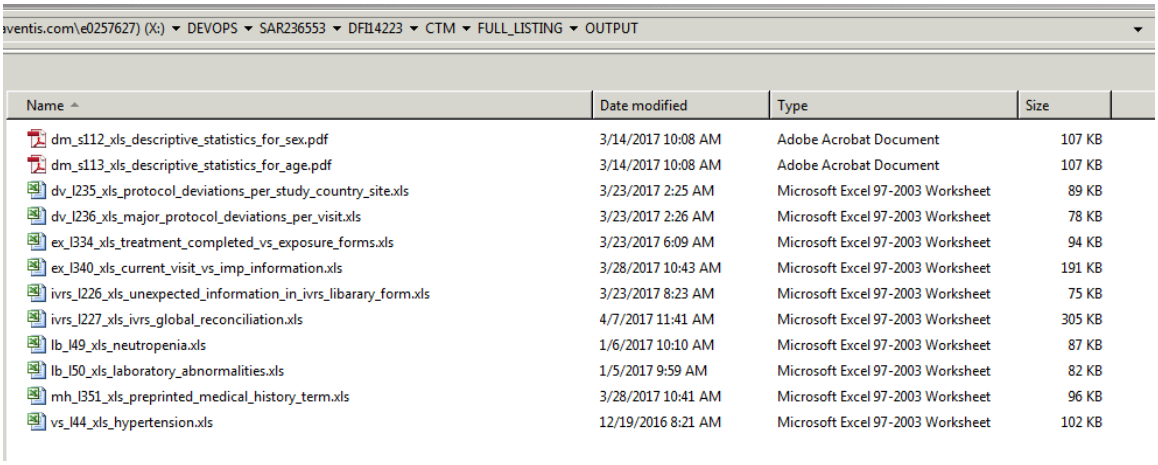
INTRODUCTION

For a clinical study, dozens of SAS reports need to be developed for data review or data validation purpose. A directory listing of all these reports, displaying basic information such as file name, format and size, could be useful in documenting the work performed on a project. Microsoft EXCEL format would be a good choice for the directory listing. Then users could annotate or comment on the listing file to track the review status of each report. At the meantime, if we could create a hyperlink for directly opening each report, it would be very convenient for users.

ACCESS THE DIRECTORY INFORMATION

This can be achieved in SAS using an UNNAMED PIPE statement to submit native commands directly to the operating system. After invoking the system command, SAS then process the records returned in a DATA step.

For example, as shown as **Display 1**, there are a dozen of reports within a directory of “xxx”.



Display 1. Reports in a directory

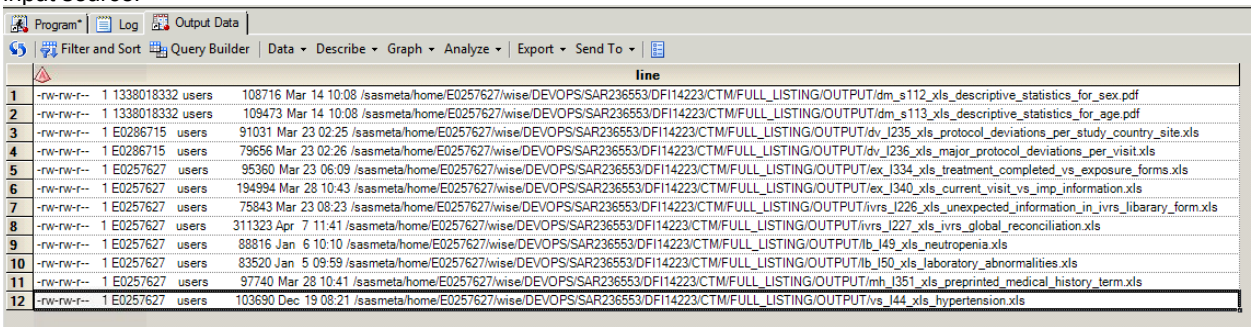
The following **SAS code** shows how to access the directory information. All SAS codes in this paper are based on a UNIX operating system.

```
*** Access the directory information to get the list of files contained within the
target folder;

    %LET target_folder= xxx;

    FILENAME pipedir PIPE " ls -l &target_folder./*.* ";

    DATA _sm_filenames;
          INFILE pipedir TRUNCOVER LRECL=5000;
          INPUT line $char1000.;
    RUN;
```

ls is a Linux/Unix shell command that lists directory contents of files and directories. The *-l* option is chosen, so the following information will be displayed for each file: file mode, number of links, owner name, group name, number of

bytes in the file, abbreviated month, day-of-month file was last modified, hour file last modified, minute file last modified, and the pathname.

The FILENAME statement connects the output of the *ls -l* command to the fileref *pipedir*. The following DATA step then creates a data set named _sm_filenames (as shown as **Display 2**) from the INFILE statement that points to the input source.



	line
1	-rw-rw-r-- 1 1338018332 users 108716 Mar 14 10:08 /sasmeta/home/E0257627/wise/DEVOPS/SAR236553/DFI14223/CTM/FULL_LISTING/OUTPUT/dm_s112_xls_descriptive_statistics_for_sex.pdf
2	-rw-rw-r-- 1 1338018332 users 109473 Mar 14 10:08 /sasmeta/home/E0257627/wise/DEVOPS/SAR236553/DFI14223/CTM/FULL_LISTING/OUTPUT/dm_s113_xls_descriptive_statistics_for_age.pdf
3	-rw-rw-r-- 1 E0286715 users 91031 Mar 23 02:25 /sasmeta/home/E0257627/wise/DEVOPS/SAR236553/DFI14223/CTM/FULL_LISTING/OUTPUT/dv_l235_xls_protocol_deviations_per_study_country_site.xls
4	-rw-rw-r-- 1 E0286715 users 79656 Mar 23 02:26 /sasmeta/home/E0257627/wise/DEVOPS/SAR236553/DFI14223/CTM/FULL_LISTING/OUTPUT/dv_l236_xls_major_protocol_deviations_per_visit.xls
5	-rw-rw-r-- 1 E0257627 users 95360 Mar 23 06:09 /sasmeta/home/E0257627/wise/DEVOPS/SAR236553/DFI14223/CTM/FULL_LISTING/OUTPUT/ex_l334_xls_treatment_completed_vs_exposure_forms.xls
6	-rw-rw-r-- 1 E0257627 users 194994 Mar 28 10:43 /sasmeta/home/E0257627/wise/DEVOPS/SAR236553/DFI14223/CTM/FULL_LISTING/OUTPUT/ex_l340_xls_current_visit_vs_imp_information.xls
7	-rw-rw-r-- 1 E0257627 users 75843 Mar 23 08:23 /sasmeta/home/E0257627/wise/DEVOPS/SAR236553/DFI14223/CTM/FULL_LISTING/OUTPUT/ivrs_l226_xls_unexpected_information_in_ivrs_libarary_form.xls
8	-rw-rw-r-- 1 E0257627 users 311323 Apr 7 11:41 /sasmeta/home/E0257627/wise/DEVOPS/SAR236553/DFI14223/CTM/FULL_LISTING/OUTPUT/ivrs_l227_xls_ivrs_global_reconciliation.xls
9	-rw-rw-r-- 1 E0257627 users 88816 Jan 6 10:10 /sasmeta/home/E0257627/wise/DEVOPS/SAR236553/DFI14223/CTM/FULL_LISTING/OUTPUT/lb_l49_xls_neutropenia.xls
10	-rw-rw-r-- 1 E0257627 users 83520 Jan 5 09:59 /sasmeta/home/E0257627/wise/DEVOPS/SAR236553/DFI14223/CTM/FULL_LISTING/OUTPUT/lb_l50_xls_laboratory_abnormalities.xls
11	-rw-rw-r-- 1 E0257627 users 97740 Mar 28 10:41 /sasmeta/home/E0257627/wise/DEVOPS/SAR236553/DFI14223/CTM/FULL_LISTING/OUTPUT/mh_l351_xls_preprinted_medical_history_term.xls
12	-rw-rw-r-- 1 E0257627 users 103690 Dec 19 08:21 /sasmeta/home/E0257627/wise/DEVOPS/SAR236553/DFI14223/CTM/FULL_LISTING/OUTPUT/vs_l44_xls_hypertension.xls

Display 2. Contents of _sm_filenames data set

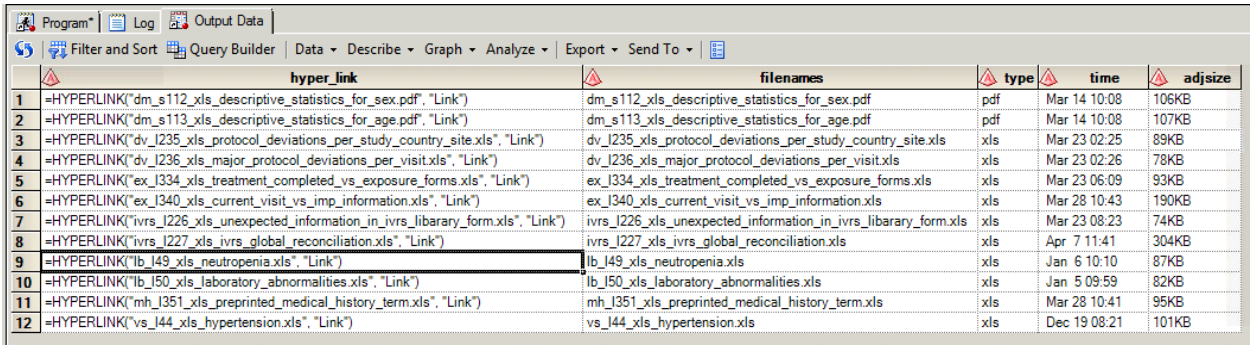
RETRIEVE ATTRIBUTE INFORMATION AND CREATE A HYPERLINK FOR EACH FILE

The following **SAS code** are aimed to retrieve attribute information of each file from the converged variable "*line*", including file name, file type, file size, date and time when file last modified. In addition, Microsoft EXCEL function HYPERLINK here is used to provide the linking mechanism from one cell to another file within the same folder. It should be like *=HYPERLINK ("A", "B")* in EXCEL cell. *A* stands for the name of the target file to be linked, *B* stands for the text that will be shown in the EXCEL cell.

```
*** Retrieve valuable information as separate variables and create the Microsoft EXCEL
hyperlinks;
      DATA filelist;
            SET _sm_filenames;
            FORMAT hyper_link filenames type time size adjsize $200. pos len RE 6.0;

            IF _N_ = 1 THEN
                  RE = PRXPARSE('/(?<=\d)\s+\w{3}\s+\d{1,2}\s+[0-
9\:]{4,5}\s+(?=\/sasmeta)/');
            RETAIN RE;
            CALL PRXSUBSTR(RE,strip(line),pos,len);

            time=substr(strip(line),pos,len);
            filenames=scan(strip(line),-1,'/');
            type= scan(strip(filenames),-1,'.');
            size=scan(strip(line),5, ' ');
            adjsize=strip(put(input(size,best.)/1024,20.0))!!'KB';

            hyper_link='=HYPERLINK('!!'"'!!strip(filenames)!!'"' !! ',
"'!!'Link'!!'")';

            IF strip(filenames) ne '_directory_master_file.xls' AND
upcase(strip(type)) in ('PDF' 'XLS' 'XLSX');
```

```
            KEEP filenames Type time adjsize hyper_link;
            LABEL filenames     =      'File name'
                  type          =      'Type'
                  time          =      'Date of creation or last modification*'
                  adjsize       =      'Size'
                  hyper_link    =      'HyperLink'
            ;
      RUN;

      PROC SORT DATA=filelist;
            BY filenames;
      RUN;
```

Please check the output SAS data set *filelist* as **Display 3**.



Program* | Log | Output Data

Filter and Sort | Query Builder | Data | Describe | Graph | Analyze | Export | Send To

	hyper_link	filenames	type	time	adjsize
1	=HYPERLINK("dm_s112_xls_descriptive_statistics_for_sex.pdf", "Link")	dm_s112_xls_descriptive_statistics_for_sex.pdf	pdf	Mar 14 10:08	106KB
2	=HYPERLINK("dm_s113_xls_descriptive_statistics_for_age.pdf", "Link")	dm_s113_xls_descriptive_statistics_for_age.pdf	pdf	Mar 14 10:08	107KB
3	=HYPERLINK("dv_l235_xls_protocol_deviations_per_study_country_site.xls", "Link")	dv_l235_xls_protocol_deviations_per_study_country_site.xls	xls	Mar 23 02:25	89KB
4	=HYPERLINK("dv_l236_xls_major_protocol_deviations_per_visit.xls", "Link")	dv_l236_xls_major_protocol_deviations_per_visit.xls	xls	Mar 23 02:26	78KB
5	=HYPERLINK("ex_l334_xls_treatment_completed_vs_exposure_forms.xls", "Link")	ex_l334_xls_treatment_completed_vs_exposure_forms.xls	xls	Mar 23 06:09	93KB
6	=HYPERLINK("ex_l340_xls_current_visit_vs_imp_information.xls", "Link")	ex_l340_xls_current_visit_vs_imp_information.xls	xls	Mar 28 10:43	190KB
7	=HYPERLINK("ivrs_l226_xls_unexpected_information_in_ivrs_libarary_form.xls", "Link")	ivrs_l226_xls_unexpected_information_in_ivrs_libarary_form.xls	xls	Mar 23 08:23	74KB
8	=HYPERLINK("ivrs_l227_xls_ivrs_global_reconciliation.xls", "Link")	ivrs_l227_xls_ivrs_global_reconciliation.xls	xls	Apr 7 11:41	304KB
9	=HYPERLINK("lb_l49_xls_neutropenia.xls", "Link")	lb_l49_xls_neutropenia.xls	xls	Jan 6 10:10	87KB
10	=HYPERLINK("lb_l50_xls_laboratory_abnormalities.xls", "Link")	lb_l50_xls_laboratory_abnormalities.xls	xls	Jan 5 09:59	82KB
11	=HYPERLINK("mh_l351_xls_preprinted_medical_history_term.xls", "Link")	mh_l351_xls_preprinted_medical_history_term.xls	xls	Mar 28 10:41	95KB
12	=HYPERLINK("vs_l44_xls_hypertension.xls", "Link")	vs_l44_xls_hypertension.xls	xls	Dec 19 08:21	101KB

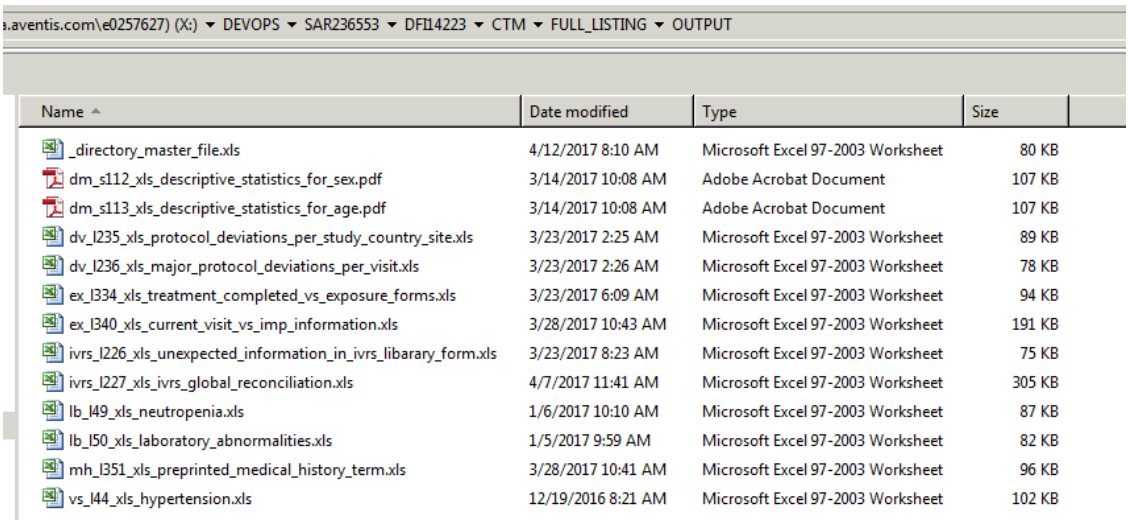
Display 3. Contents of filelist data set

CREATE A MICROSOFT EXCEL FILE

We use ODS tagsets.excelxp together with report procedure to create an EXCEL file from the SAS data set *filelist,* as shown as the following **SAS code**. Then we get an EXCEL file named *_directory_master_file.xls*, which is located within the same folder as other files, as shown as **Display 4** and **Display 5.**

```
*** Create a Microsoft EXCEL file of the list;
    ODS _ALL_ CLOSE;
    ODS tagsets.excelxp options( embedded_titles = 'yes' embed_titles_once = 'yes'
        embedded_footnotes = 'yes' autofit_height = 'yes' skip_space = '1,0,0,1,1'
        autofilter = "all" default_column_width = "15" Absolute_Column_Width= "15"
        sheet_name = "Hyperlink to each file")
        FILE= "&target_folder./_directory_master_file.xls" STYLE=LISTING;

    PROC REPORT DATA =filelist NOWD MISSING SPACING=1 HEADLINE HEADSKIP SPLIT="@";
         COLUMN hyper_link filenames type time adjsize;
         DEFINE hyper_link / STYLE = [COLOR=bib FONTWEIGHT= bold] ;
    RUN;
```

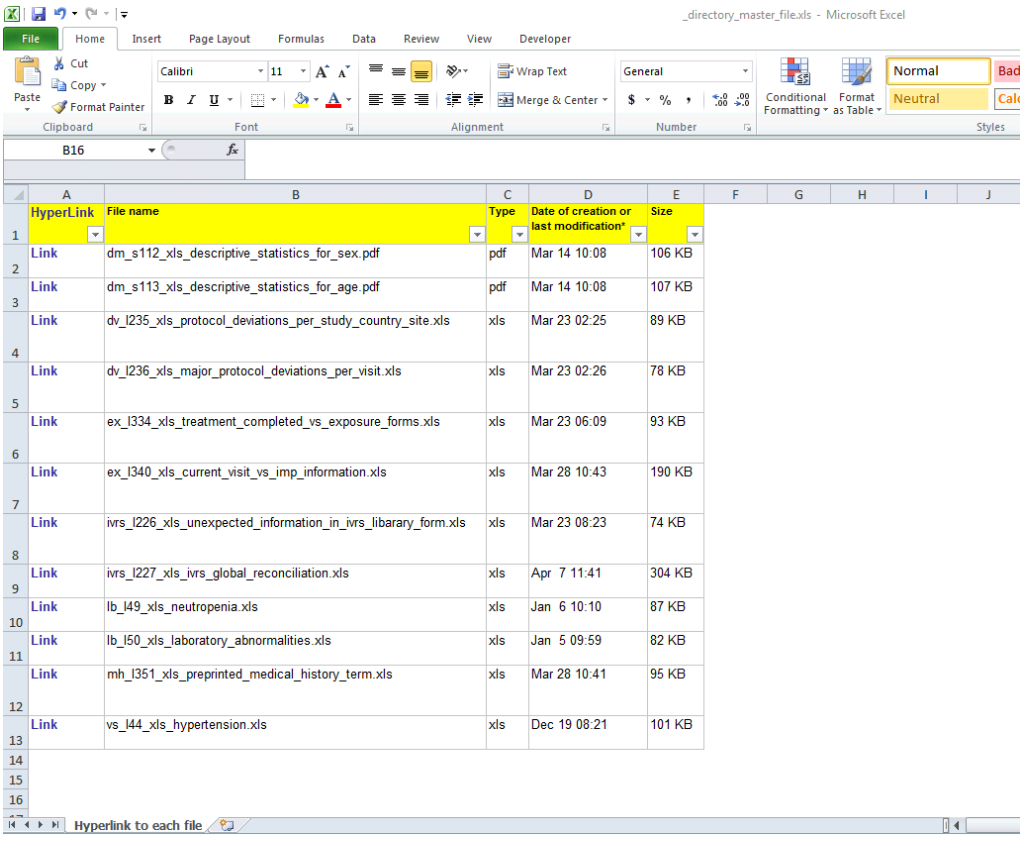
```
ODS tagsets.excelxp CLOSE;
ODS LISTING;
```



.aventis.com\e0257627) (X:) ▸ DEVOPS ▸ SAR236553 ▸ DFI14223 ▸ CTM ▸ FULL_LISTING ▸ OUTPUT

Name	Date modified	Type	Size
_directory_master_file.xls	4/12/2017 8:10 AM	Microsoft Excel 97-2003 Worksheet	80 KB
dm_s112_xls_descriptive_statistics_for_sex.pdf	3/14/2017 10:08 AM	Adobe Acrobat Document	107 KB
dm_s113_xls_descriptive_statistics_for_age.pdf	3/14/2017 10:08 AM	Adobe Acrobat Document	107 KB
dv_l235_xls_protocol_deviations_per_study_country_site.xls	3/23/2017 2:25 AM	Microsoft Excel 97-2003 Worksheet	89 KB
dv_l236_xls_major_protocol_deviations_per_visit.xls	3/23/2017 2:26 AM	Microsoft Excel 97-2003 Worksheet	78 KB
ex_l334_xls_treatment_completed_vs_exposure_forms.xls	3/23/2017 6:09 AM	Microsoft Excel 97-2003 Worksheet	94 KB
ex_l340_xls_current_visit_vs_imp_information.xls	3/28/2017 10:43 AM	Microsoft Excel 97-2003 Worksheet	191 KB
ivrs_l226_xls_unexpected_information_in_ivrs_libarary_form.xls	3/23/2017 8:23 AM	Microsoft Excel 97-2003 Worksheet	75 KB
ivrs_l227_xls_ivrs_global_reconciliation.xls	4/7/2017 11:41 AM	Microsoft Excel 97-2003 Worksheet	305 KB
lb_l49_xls_neutropenia.xls	1/6/2017 10:10 AM	Microsoft Excel 97-2003 Worksheet	87 KB
lb_l50_xls_laboratory_abnormalities.xls	1/5/2017 9:59 AM	Microsoft Excel 97-2003 Worksheet	82 KB
mh_l351_xls_preprinted_medical_history_term.xls	3/28/2017 10:41 AM	Microsoft Excel 97-2003 Worksheet	96 KB
vs_l44_xls_hypertension.xls	12/19/2016 8:21 AM	Microsoft Excel 97-2003 Worksheet	102 KB

Display 4. Created excel file (directory master file.xls) in the same directory



_directory_master_file.xls - Microsoft Excel

File | Home | Insert | Page Layout | Formulas | Data | Review | View | Developer

HyperLink	File name	Type	Date of creation or last modification*	Size
Link	dm_s112_xls_descriptive_statistics_for_sex.pdf	pdf	Mar 14 10:08	106 KB
Link	dm_s113_xls_descriptive_statistics_for_age.pdf	pdf	Mar 14 10:08	107 KB
Link	dv_l235_xls_protocol_deviations_per_study_country_site.xls	xls	Mar 23 02:25	89 KB
Link	dv_l236_xls_major_protocol_deviations_per_visit.xls	xls	Mar 23 02:26	78 KB
Link	ex_l334_xls_treatment_completed_vs_exposure_forms.xls	xls	Mar 23 06:09	93 KB
Link	ex_l340_xls_current_visit_vs_imp_information.xls	xls	Mar 28 10:43	190 KB
Link	ivrs_l226_xls_unexpected_information_in_ivrs_libarary_form.xls	xls	Mar 23 08:23	74 KB
Link	ivrs_l227_xls_ivrs_global_reconciliation.xls	xls	Apr 7 11:41	304 KB
Link	lb_l49_xls_neutropenia.xls	xls	Jan 6 10:10	87 KB
Link	lb_l50_xls_laboratory_abnormalities.xls	xls	Jan 5 09:59	82 KB
Link	mh_l351_xls_preprinted_medical_history_term.xls	xls	Mar 28 10:41	95 KB
Link	vs_l44_xls_hypertension.xls	xls	Dec 19 08:21	101 KB

Hyperlink to each file

Display 5. Contents of directory master file.xls

CONCLUSION

Users could open each file by clicking the link, and of course, they could annotate or comment on this EXCEL file to track the review status of each report as they want. When the folder transferred from one directory to another, even

different operating system, the hyperlinks still work if the *_directory_master_file.xls* and other files are stored in a same folder.