



A Dynamic Approach to File Organisation using DOS within SAS

Michael Tang

Overview

- Setting the scene
- SAS System Options
- Obtaining the Windows directory listing
- Modifying *dir* DOS command result
- Utilising Tracker information
- Preparing outputs for packaging
- Packaging the outputs
- Housekeeping
- Conclusion

Setting the scene

Name	Date modified	Type
 L1102120102	19/08/2014 21:40	Rich Text Format
 L1102120201	19/08/2014 21:38	Rich Text Format
 L1102120202	19/08/2014 21:38	Rich Text Format
 T11010101	19/08/2014 21:17	Rich Text Format
 T11010102	19/08/2014 21:18	Rich Text Format
 T11010201	19/08/2014 21:17	Rich Text Format
 T11010202	19/08/2014 21:17	Rich Text Format
 T11010301	19/08/2014 21:19	Rich Text Format
 T11010302	19/08/2014 21:18	Rich Text Format
 T11010401	19/08/2014 21:19	Rich Text Format

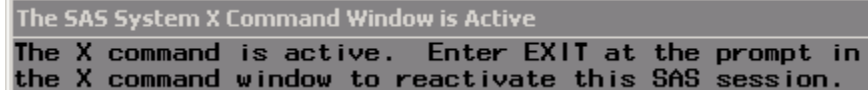
- All TLFs supporting the CSR stored in one directory.
- TLFs to be delivered in batches.
- Batch information stored in a tracker.
- Goal to organise and package TLFs automatically in SAS.

Type	TLF Number	Batch ID
Listing	11.1.7.1	Batch 3
Listing	11.1.8.1	Batch 3
Listing	11.1.9.1	Batch 3
Table	11.2.1.1.1	Batch 4
Table	11.2.1.1.2	Batch 4
Table	11.2.1.1.3	Batch 4
Table	11.2.1.1.4	Batch 4
Table	11.2.1.1.5	Batch 4
Table	11.2.1.1.6	Batch 4
Table	11.2.1.1.7	Batch 4
Table	11.2.1.1.8	Batch 4
Table	11.2.1.1.9	Batch 4
Listing	11.2.1.2.1	Batch 4
Listing	11.2.1.2.2	Batch 4

SAS System Options

XWAIT

Results in the following prompt each time a DOS command is executed.

A screenshot of a SAS system message box. The title bar reads "The SAS System X Command Window is Active". The main text area contains the message: "The X command is active. Enter EXIT at the prompt in the X command window to reactivate this SAS session."

The SAS System X Command Window is Active
The X command is active. Enter EXIT at the prompt in
the X command window to reactivate this SAS session.

SAS session remains locked until the command window is closed.
Use NOXWAIT to deactivate this prompt.

XSYNC

Commands execute synchronously with SAS.

SAS cannot be accessed until the command completes execution.

Obtaining the Windows directory listing

```
FILENAME [fileref] PIPE "dir [file location]";  
e.g: FILENAME rtfdir PIPE "dir C:\Study\Output\";
```

- FILENAME statement executed in SAS calls the DOS command *dir*.
- PIPE option allows SAS to invoke an external command and receive the resulting messages.
- *Fileref* is a custom value which stores the result in SAS.

Obtaining the Windows directory listing

```
C:\Windows\system32\cmd.exe
C:\>dir C:\Study\Output
Volume in drive C is OSDisk
Volume Serial Number is 0CC4-4823

Directory of C:\Study\Output

31/08/2014  14:17    <DIR>          .
31/08/2014  14:17    <DIR>          ..
19/08/2014  21:22             5,503 F1101060101.rtf
19/08/2014  21:23             5,503 F1101060102.rtf
19/08/2014  21:32             5,503 F1102010205.rtf
19/08/2014  21:32             5,503 F1102010206.rtf
19/08/2014  21:37             5,503 F1102110201.rtf
19/08/2014  21:37             5,503 F1102110202.rtf
19/08/2014  21:23             5,503 L1101060201.rtf
19/08/2014  21:23             5,503 L1101060202.rtf
19/08/2014  21:24             5,503 L11010701.rtf
19/08/2014  21:24             5,503 L11010801.rtf
19/08/2014  21:24             5,503 L11010901.rtf
19/08/2014  21:27             5,503 L1102010201.rtf
19/08/2014  21:29             5,503 L1102010202.rtf
19/08/2014  21:31             5,503 L1102010203.rtf
19/08/2014  21:31             5,503 L1102010204.rtf
19/08/2014  21:33             5,503 L1102010207.rtf
19/08/2014  21:33             5,503 L1102010208.rtf
```

Example of result when executing *dir* within the command window.

Using *[fileref]*, the result can be output into a SAS dataset.

```
data rtffiles;
  infile [fileref] length=reclen;
  input rtfname $varying200. reclen;
run;
```

	rtfname
1	Volume in drive C is Data
2	Volume Serial Number is 7E9F-F44D
3	
4	Directory of C:\Study\Output
5	
6	08/19/2014 04:22 PM 5,503 F1101060101.rtf
7	08/19/2014 04:23 PM 5,503 F1101060102.rtf
8	08/19/2014 04:32 PM 5,503 F1102010205.rtf
9	08/19/2014 04:32 PM 5,503 F1102010206.rtf
10	08/19/2014 04:37 PM 5,503 F1102110201.rtf
11	08/19/2014 04:37 PM 5,503 F1102110202.rtf
46	08/19/2014 04:34 PM 5,503 T1102100101.rtf
47	08/19/2014 04:35 PM 5,503 T1102100102.rtf
48	08/19/2014 04:36 PM 5,503 T1102110101.rtf
49	08/19/2014 04:36 PM 5,503 T1102110102.rtf
50	44 File(s) 242,132 bytes
51	0 Dir(s) 245,105,889,280 bytes free

Modifying *dir* DOS command result

- Use the * wildcard operator to select file types of interest, e.g. *.rtf.
- A bare format (/B) can be applied to the result to remove the following:
 - Header and Footer information
 - File attributes, i.e. Last Modified Date & Time and File Size

The result of the bare format simply comprises the file name and type and is run as below:

```
FILENAME rtfdir PIPE
"dir C:\Study\Output\*.rtf /B"
```

	rtfname		
1	Volume in drive C is Data		
2	Volume Serial Number is 7E9F-F44D		
3			
4	Directory of C:\Study\Output		
5			
6	08/19/2014 04:22 PM	5,503	F1101060101.rtf
7	08/19/2014 04:23 PM	5,503	F1101060102.rtf
8	08/19/2014 04:32 PM	5,503	F1102010205.rtf
9	08/19/2014 04:32 PM	5,503	F1102010206.rtf
10	08/19/2014 04:37 PM	5,503	F1102110201.rtf
11	08/19/2014 04:37 PM	5,503	F1102110202.rtf
46	08/19/2014 04:34 PM	5,503	T1102100101.rtf
47	08/19/2014 04:35 PM	5,503	T1102100102.rtf
48	08/19/2014 04:36 PM	5,503	T1102110101.rtf
49	08/19/2014 04:36 PM	5,503	T1102110102.rtf
50	44 File(s) 242,132 bytes		
51	0 Dir(s) 245,105,889,280 bytes free		

Utilising Tracker information

- Tracker in Excel format can be read into SAS using PROC IMPORT.

Type	TLF Number	Batch ID
Figure	11.1.6.1.1	Batch 3
Figure	11.1.6.1.2	Batch 3
Figure	11.2.1.2.5	Batch 4
Figure	11.2.1.2.6	Batch 4
Figure	11.2.11.2.1	Batch 5
Figure	11.2.11.2.2	Batch 5
Table	11.2.10.1.1	Batch 5
Table	11.2.10.1.2	Batch 5
Table	11.2.11.1.1	Batch 5
Table	11.2.11.1.2	Batch 5

TLF Numbers from the tracker require a transformation to match file naming convention to merge Batch ID.

1	F1101060101.rtf
2	F1101060102.rtf
3	F1102010205.rtf
4	F1102010206.rtf
5	F1102110201.rtf
6	F1102110202.rtf
7	T1102100101.rtf
8	T1102100102.rtf
9	T1102110101.rtf
10	T1102110102.rtf

- Filename contains no periods in the TLF number.
- Values below 10 are given a prefix of '0'.

Preparing outputs for packaging

- A staging area is set up to segregate outputs to be packaged.
- The *mkdir* DOS command is run using the X statement in SAS.

```
X mkdir "[file location]";
```

- Provided the user has permission, the new directory will be created.

Preparing outputs for packaging

- The *copy* DOS command is used to copy outputs to the staging area.
- These commands are compiled within a data step for each output to be copied over.
- Basic *copy* command structure:

```
copy "[current file location]\[filename including extension]"  
      "[new file location]"
```

e.g. `copy "C:\Study\Output\Table.rtf" "C:\Study\Output\Stage"`

- To execute DOS commands in a data step:

```
call system(var)
```

Preparing outputs for packaging

- Variable *DOScmd* is created compiling the *copy* DOS command.

Batch ID	newnum	rtfname	DOScmd
Batch 3	F1101060101	F1101060101.rtf	copy "U:\PhUSE\Output\F1101060101.rtf" "U:\PhUSE\Output\Batch 3"
Batch 3	F1101060102	F1101060102.rtf	copy "U:\PhUSE\Output\F1101060102.rtf" "U:\PhUSE\Output\Batch 3"
Batch 3	L1101060201	L1101060201.rtf	copy "U:\PhUSE\Output\L1101060201.rtf" "U:\PhUSE\Output\Batch 3"

- Takes the RTF file name from each row to instruct DOS to copy that file to a new area.
- Code construction:

```
%let fileout = U:\PhUSE\Output\;  
%let batch = Batch 3;  
data rtf_copy;  
    set rtf_track(where=(batch_id="&batch."));  
    DOScmd = 'copy "' || strip("&fileout.") || strip(rtfname) ||  
            '" "' || strip("&fileout.&batch.")) || '";  
    call system(DOScmd);  
run;
```

Packaging the outputs

- X statement used to call WinZip.

```
X call "C:\Program Files\WinZip\WINZIP32.EXE" -a -r  
      "C:\Output\Packaged File.ZIP"  
      "C:\Output";
```

- The three file paths point to these locations:

1. WinZip program.
2. Name of packaged file.
3. Folder(s) contents to be packaged.

- WinZip options:

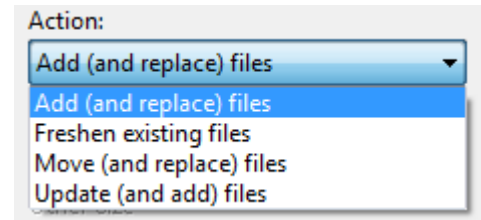
- *-a* adds files to the package, this is one of the four actions in WinZip, one must be specified for WinZip to execute.
- *-r* tells WinZip to include subfolders and their contents.

Packaging the outputs

- Other WinZip options:

- WinZip actions (one must be specified):

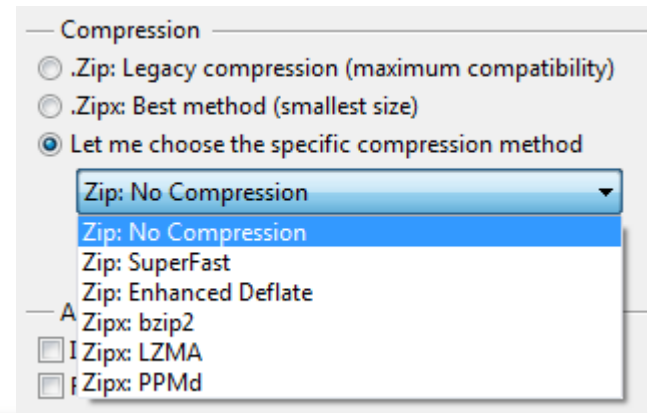
- *-f* - freshens existing files,
- *-m* - moves (and deletes) files,
- *-u* - updates (and adds) files.



- *-min* runs WinZip in the background.
- *-hs* includes hidden and system files.
- *-s"Password"* adds a case-sensitive password.

- Compression options:

- *-e0* - no compression,
- *-es* - super-fast,
- *-ef* - fast,
- *-en* - normal,
- *-ex* - extra.



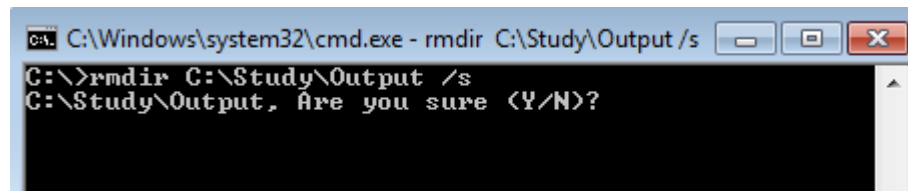
Housekeeping

- Once the outputs are packaged, the staging area can be removed.

```
X rmdir [file location] /s /q;
```

- Directory removal options:

- /s removes a non-empty directory.
Required since the purpose is to remove the packaged directory.
- /q suppresses the 'Are you sure (Y/N)?' dialog when removing non-empty directories.



Conclusion

- Run this inside a macro for minimal program modification.
- Can be executed at various stages of a study, or to create a number of packages at the same time.
- Method can be adapted to assist top level review prior to delivery.
- Consistency checks against the tracker.
- Presence of outputs.
- Using file attributes such as date and time against recorded dates.

Questions

