



CHECK FOR FILE CONSISTENCY

– PhUSE 2007 - Lisboa –

Arnaud DAUCHY, Sanofi Aventis, Paris, France



Agenda

Scope & Perimeter

File Consistency

Script Operations

Programming & Implementation

Script Details

Report File

Example

Improvements



Agenda

Scope & Perimeter

File Consistency

Script Operations

Programming & Implementation

Script Details

Report File

Example

Improvements



Scope

Improvement based on different legacy tools

Necessity for automatic validation checks

Provide tools to ensure:

- Compliance among SAS® files
- Some of the 21 CFR Part 11 requirements
- The respect of department naming conventions



Perimeter

Usual SAS® objects

- .SAS as source
- .LOG as execution proof
- Different outputs
 - 【 LST, RTF, CGM, HTML, ...
 - 【 datasets
 - 【 graphs

Do not consider

- Extraction log
- Files having no link with SAS® execution
 - 【 DOC, ZIP, PDF



Agenda

Scope & Perimeter

File Consistency

Script Operations

Programming & Implementation

Script Details

Report File

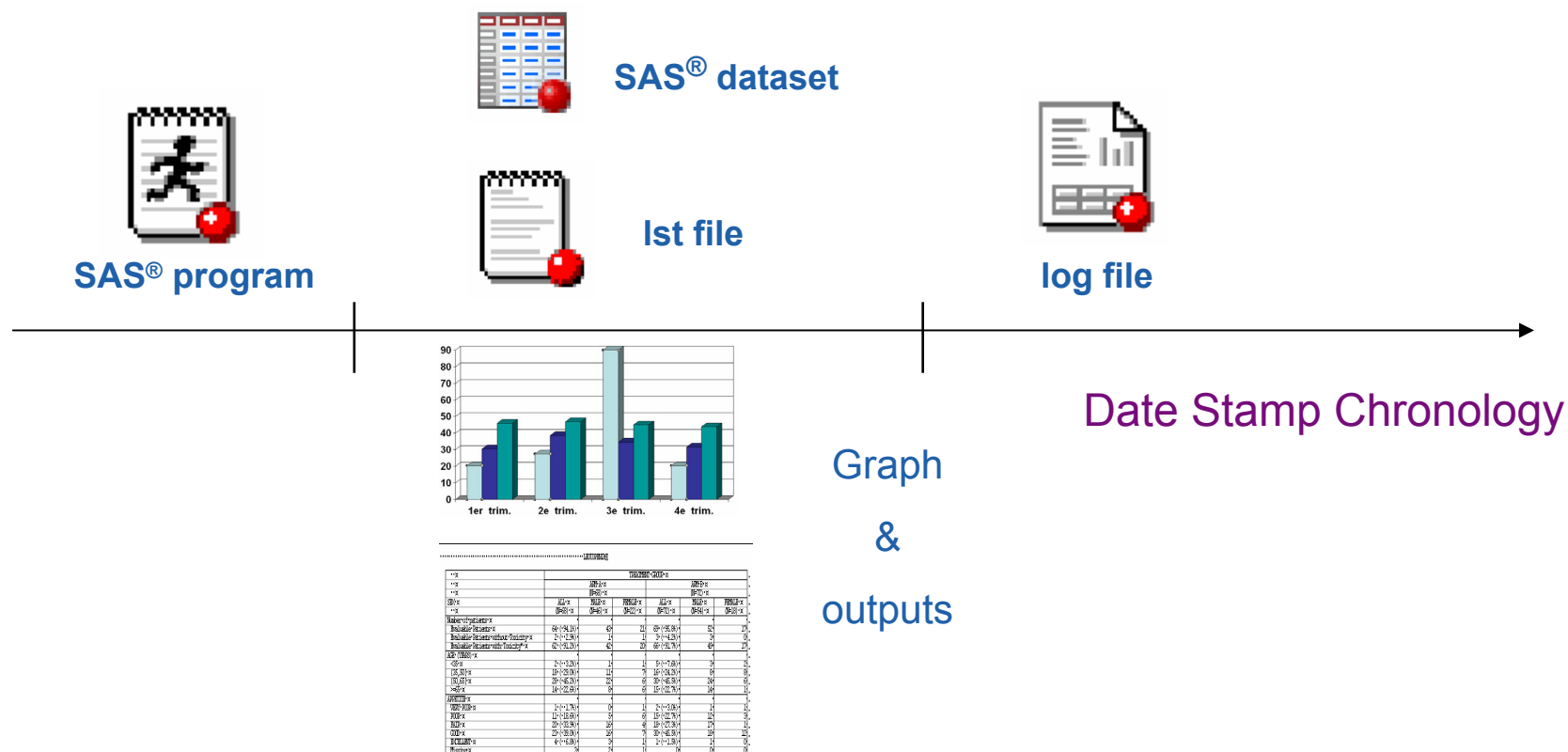
Example

Improvements



File Consistence (1/2)

Date Stamp of program, log and outputs must be consistent





File Consistency (2/2)

Output must be stored in correct standard folders

Names of different outputs & datasets must respect department rules

Computing environment folders:

- DATA_R
 - 【 Raw Datasets
- DATA_A
 - 【 Analysis Datasets
- PGM_ANA
 - 【 Program for analysis datasets
- PGM_RPT
 - 【 Programs for reporting
- LOG
 - 【 All execution log files
- OUTPUT
 - 【 Tables, Listings & Graphs are stored in that folder



Agenda

Scope & Perimeter

File Consistency

Script Operations

Programming & Implementation

Script Details

Report File

Example

Improvements



Script operations (1/4)

Step 1

- **Validate that all .sas have an associated .log file**
 - 【 Each program must have been executed
- **If the log file exists, its date must be later than .sas' date**
 - 【 Track program modification without re-execution
 - 【 If the log date predates the .sas, there will be an issue

Step 2

- **Validate that all .logs have an associated .sas file**
 - 【 PGM_ANA
 - 【 PGM_RPT



Script operations (2/4)

Step 3

- **For each output, a corresponding .sas must exist**

- 【 This .sas must be located in PGM_RPT
- 【 Indexation is taken into account, see below

- **Following tests are performed:**

- 【 `date .sas < date output`
- 【 `date output < date .log`

Step 4

- **For each dataset (DATA_A), a corresponding .sas must exist**

- 【 This .sas must be located in PGM_ANA
- 【 Indexation is taken into account, see below

- **Following tests are done:**

- 【 `date ae.sas < date ae.sas7bdat`
- 【 `date ae.sas7bdat < date ae.log`



Script operations (3/4)

Step 5

- Each .sas from PGM_ANA must generate an SAS® dataset

Step 6

- Raw & derived dataset dates must be consistent
 - 【 Date of most recent DATA_R < Date of first DATA_A
 - 【 Otherwise, list all DATA_A which do not fit this condition

Step 7

- All raw data must have been extracted the same day
 - 【 Otherwise, list all DATA_R whose date is different from the date of oldest DATA_R



Script operations (4/4)

Indexation

- **Derivations :**

- 【 x.sas
- 【 x.log (LOG)
- 【 x.sas7bdat (DATA_A)

- **Reporting :**

- 【 XXX_free(_P)_T.sas
- 【 XXX_free(_P)_T.log (LOG)
- 【 XXX_free(_P)_T_#_L.rtf (OUTPUT)

- **XXX**

- 【 DEM, EFF, LAB, ...

- **T**

- 【 T, L, G

- **L**

- 【 I, X,



Agenda

Scope & Perimeter

File Consistency

Script Operations

Programming & Implementation

Script Details

Report File

Example

Improvements



Programming & Implementation

Unix Script

- Stored on SAS® environment platform
- Short time for execution

Execution through SAS®

- Remote submit access
- Ask for OS execution

```
└ x “..” ;
```

```
└ %sysexec ...
```



Agenda

Scope & Perimeter

File Consistency

Script Operations

Programming & Implementation

Script Details

Report File

Example

Improvements



Script Details (1/3)

List of all objects

```
list1= $HOME/list1.txt  
ls -ld OUTPUT/* > $list1  
cat $list1 | grep -vi "\.pdf" | sort | uniq > $list2  
list_out=`cat $list2`
```

Loop on all objects for different checks

● Remote submit access

```
for file in `echo $list_sas`  
do  
    ... operations...  
done
```



Script Details (2/3)

Existence

- `! -a LOG/$name.log`
- `if [ ! -a PGM_ANA/$name.sas ] && [ ! -a PGM_RPT/$name.sas ]`

Names

- `name=`echo $file | cut -d "/" -f2 | cut -d "." -f1``
- `name=`echo $file | sed -e "s/^(_L_[0-9]_I\.)/_L./" ``

file = PGM_ANA / test . sas





Script Details (2/3)

Comparison of dates

● LOG/\$name.log -ot \$file

Comparison of extract & derived datasets

```
[ recent_DR=`ls -rt ../BS/DATA_R/*.sas7bdat | tail -1`  
[ old_DA=`ls -t DATA_A/*.sas7bdat | tail -1`  
[ if [ $old_DA -ot $last_DR ]  
[ then...
```



Script Details (3/3)

Example

```
for file in `echo $list_sas`  
do  
    name=`echo $file | cut -d "/" -f2 | cut -d "." -f1`  
    #-----  
    if [ ! -a LOG/$name.log ]  
    then  
        echo "-G> $file has no log associated " >> $report  
    else  
        #-----  
        #TEST: if log exist, '.sas' must be older than '.log'  
        #-----  
        if [ LOG/$name.log -ot $file ]  
        then  
            echo "-B> $file is more recent than its log : " >> $report  
            sasfile=`ls -al $file`  
            logfile=`ls -al LOG/$name.log`  
            echo "$sasfile" >> $report  
            echo "$logfile\n" >> $report  
        fi  
    fi  
done
```



Agenda

Scope & Perimeter

File Consistency

Script Operations

Programming & Implementation

Script Details

Report File

Example

Improvements



Report file (1/3)

Keep a report file (text file)

- `echo " ...report text ..." >> $report`

Layout & appearance

- From text to RTF

```
{\rtf
{\fonttbl
{\f1\fnil Courier New;}
}
{\colortbl
;\red255\green0\blue0;\red50\green155\blue50;\red0\green0\blue255
;}
      ...report text ...
}
```



Report file (2/3)

Tagset

● Colors

```
[ \par\\f1\\fs20 {\cf3\\b ...Blue ... }  
[ \par\\f1\\fs20 {\cf1 ...Red ... }  
[ \par\\f1\\fs20 {\cf2\\b ...Green ... }  
[ \par\\f1\\fs20 ...Other ...
```

Example

```
[ echo "-R> Each .sas must have a .log (...)" >> $report  
[ echo "-B> Program.sas is more recent than its log" >> $report  
[ echo "-G> Program.sas without any .log associated" >> $report
```



Report file (3/3)

STEP 1

Each program .sas in PGM_RPT and PGM_ANA must have a .log file subsequently created in LOG.

Program.sas is more recent than its log

Program.sas without any .log file associated

PGM_ANA/test2_step1.sas is more recent than its log :

-rwxrwx---+ 1 fr09614 biostat 2281 Jun 25 11:13 PGM_ANA/test2_step1.sas

-rwxrwx---+ 1 fr09614 biostat 30686 Mar 9 2007 LOG/test2_step1.log

PGM_RPT/zzz_test1_step1_s_t.sas has no log associated



Agenda

Scope & Perimeter

File Consistency

Script Operations

Programming & Implementation

Script Details

Report File

Example

Improvements



Example

Remote Submit execution

● %sysexec checkfiles PROJ STUDY ANALYSIS DEV;

Report file generated



Microsoft Word
Document



Agenda

Scope & Perimeter

File Consistency

Script Operations

Programming & Implementation

Script Details

Report File

Example

Improvements



Improvements

Other checks

- For a dataset, datestamp can be different from SAS® internal date
 - 【 Modification through daily operations

Display & Report

- Unix does not display datestamp hour for files older that 6 months
 - 【 -rw-rw-r-- 1 u g 160 Mar 2 2007 test.txt
 - 【 -rw-r--r-- 1 u g 147456 Aug 7 18:22 plot3.sas7bdat



Contact Information

Contact the author at:

● Sanofi Aventis

【 Arnaud.Dauchy@sanofi-aventis.com