



Traceability confirmation at your fingertips

By: Shafi Chowdhury

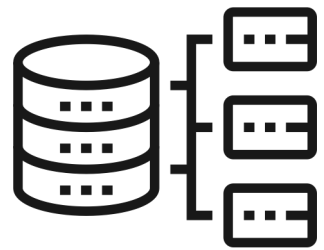
SM12 - PHUSE EU 2023



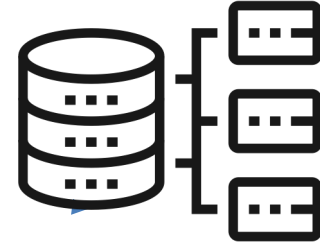
Delivering Efficiency, Standards, Quality and Oversight

Traceability normally refers to data flow

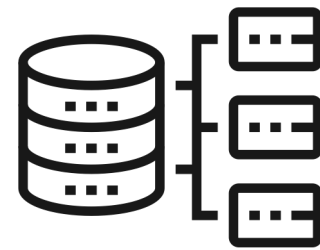
RAW data



SDTM data



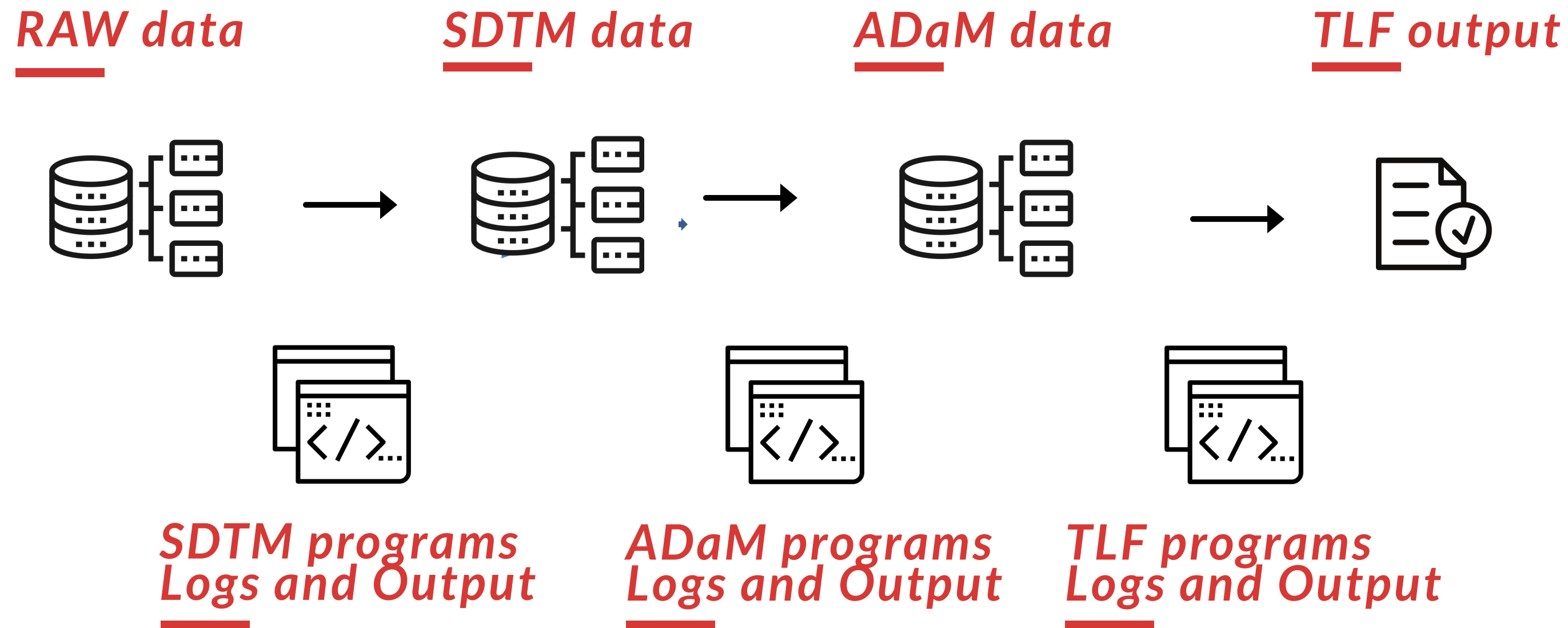
ADaM data



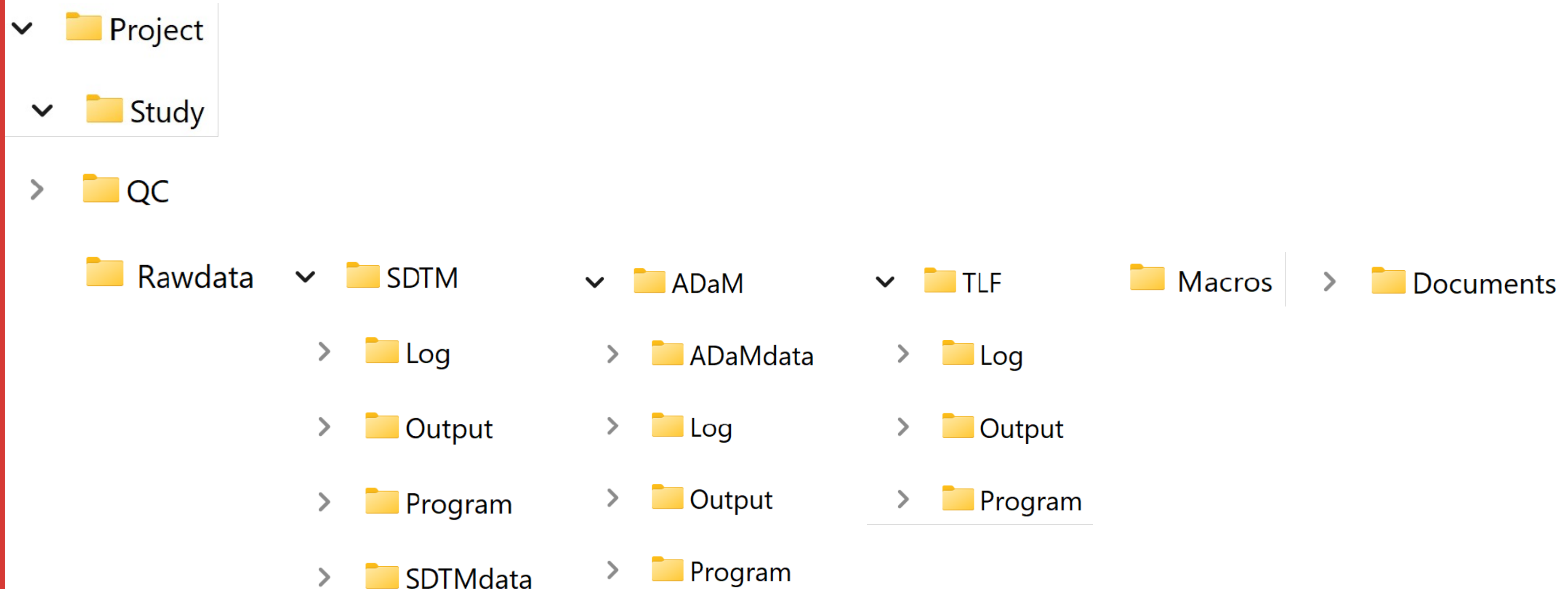
TLF output



Traceability also refers to association of programs, logs and output files



Common study folder structure



What is the traceability problem?

- Have everything been run in the correct order?
- Are all the files from the latest submit of the programs?

▼  SDTM

>  Log

>  Output

>  Program

>  SDTMdata

▼  ADaM

>  ADaMdata

>  Log

>  Output

>  Program

- We can list all files and their dates and times

What is the traceability problem?

How do we check the associated files in multiple subfolders?

How do we identify the associated files?

 SDTM  Program  Log  Output  SDTMdata

QC

 Program  Log  Output  SDTMdata

Traceability problems with associated files

- A program can have multiple output files:
 - sdtm.<domain> and SUPP<domain> datasets
 - Multiple TLFs
- Output file names may not reflect the program name
- Associated validation programs and validation datasets may have different names

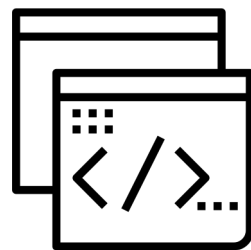
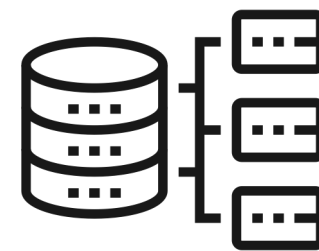
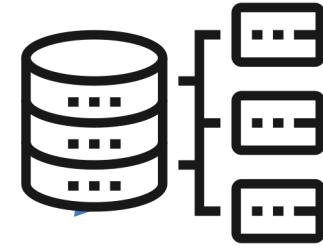
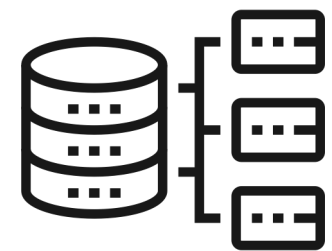
Where do we want to be?

RAW data

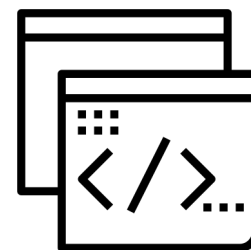
SDTM data

ADaM data

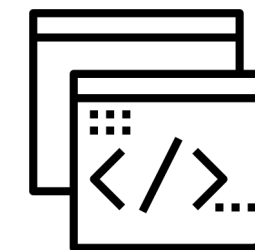
TLF output



SDTM programs
Logs and Output



ADaM programs
Logs and Output



TLF programs
Logs and Output

Program < Log/Output < validation output
Minimum time difference between Log and output

Standard checks – how do we get there?

```
%traceability(folder      = raw  
                ,folder_num = 1  
                );
```

```
%traceability(folder      = sdtm  
                ,folder_num = 2  
                ,qc_log     = qc_  
                ,qc_data    =  
                ,log_dttm_fmt = 2  
                ,log_sas_diff = 10  
                );
```

```
%traceability(folder      = t1f  
                ,folder_num = 4  
                ,qc_log     = qc_  
                ,qc_data    =  
                ,use_rtf_ftnt = Y  
                ,rtf_ftnt_fmt = 4  
                ,log_dttm_fmt = 2  
                ,log_rtf_diff = 10  
                ,log_sas_diff = 10  
                ,rtf_sas_diff = 10  
                ,create_xlsx  = Y  
                );
```

What tasks are needed to perform a traceability check?

- List all files (including subfolders) and their dates/times
- Identify associated files
- Merge associated files
- Check consistency with regards to dates and times
- Perform additional checks
 - Check log files for errors and warning messages
 - Check if all expected outputs are present

Most challenging task – Identify associated files

- If there is a spreadsheet or metadata containing program names and outputs that each program should produce, then this should be referenced. (Best option)
- Otherwise – it is complicated
 - Output with the same name as the program
 - Initial part is the same with the exception of a number or so
 - Output file is mentioned in the program log file

Traceability result – Summary number of files

Folder	Path	Total	CSV	XLS	XLSX	SAS7BDAT	XPT	SAS	DOC	DOCX	PDF	LOG	LST
Documents	Project\Study\Documents	15		1	7				1	1	5		
Macros	Project\Study\Macros	20	1		11				2	1	5		
SDTM	Project\Study\SDTM\SDTMdata	92				46	46						
SDTM	Project\Study\SDTM\SDTMdata	92				46	46						
SDTM	Project\Study\SDTM\log	29										29	
SDTM	Project\Study\SDTM\output	29											29
SDTM	Project\Study\SDTM\program	29						29					
SDTM	Project\Study\QC\macros	3						3					
SDTM	Project\Study\QC\SDTM\SDTMdata	46				46							
SDTM	Project\Study\QC\SDTM\log	29										29	
SDTM	Project\Study\QC\SDTM\output	29											29
SDTM	Project\Study\QC\SDTM\program	29						29					

Traceability result – Summary of dates and times

Folder	Path	Total	SAS7BDAT	XPT	LOG	LST	Earliest	Latest	Number of files with issues	Number of issues	Types of issues (N)
Documents	Project\Study\Documents	15									
Macros	Project\Study\Macros	20									
SDTM	Project\Study\SDTM\SDTMdata	92	46	46			12JAN22:03:24:28	12JAN22:03:35:39			
SDTM	Project\Study\SDTM\SDTMdata	92	46	46			12JAN22:03:24:28	12JAN22:03:35:39			
SDTM	Project\Study\SDTM\log	29			29		12JAN22:03:24:15	12JAN22:03:35:13	5	18	INPUT DATA SET IS EMPTY(18)
SDTM	Project\Study\SDTM\output	29				29					
SDTM	Project\Study\SDTM\program	29									
SDTM	Project\Study\QC\macros	3									
SDTM	Project\Study\QC\SDTM\SDTMdata	46	46				12JAN22:03:41:55	12JAN22:03:57:43			
SDTM	Project\Study\QC\SDTM\log	29			29		12JAN22:03:41:12	12JAN22:03:57:22			
SDTM	Project\Study\QC\SDTM\output	29				29					
SDTM	Project\Study\QC\SDTM\program	29									

Traceability result

SDTM - Summary of Programs, Logs, Output and Datasets

Yellow rows highlight rows with date/association issues. Red rows highlight LOG files with issues.

Validation details					Check associations	No. of issues in LOG	Check associations	No. of issues in LOG
LOG filename	LOG date and time	SAS dataset name	SAS Dataset date and time	No. of obs. in SAS Dataset	Issues between associated files	INPUT DATA SET IS EMPTY	Issues between associated files	INPUT DATA SET IS EMPTY
		TE	12JAN22:03:35:28	10	The SAS dataset has no program associated with it		The SAS dataset has no program associated with it	
qc_ex	12JAN2022:03:35:00	EX	12JAN22:03:35:46	264				1
qc_ex	12JAN2022:03:35:00	SUPPEX	12JAN22:03:35:47	264				1
qc_dm	12JAN2022:03:36:00	DM	12JAN22:03:36:36	84		1		
qc_dm	12JAN2022:03:36:00	SUPPDM	12JAN22:03:36:37	92		1		
qc_ec	12JAN2022:03:37:00	EC	12JAN22:03:37:08	528				

Summary

Traceability

Essential? No, but really helpful for everyone

When to produce it

Best if done once a delivery is ready by those involved.
Best to do it on the server where everything is produced.
If not provided, and you receive files from a vendor then it is also very useful to produce it, see what was actually sent.

Finally

One simple program that calls the macro to produce the spreadsheet. Does not need to be changed between studies.
Once you have it, you always want it.....

Thank you

Website

www.shaficonsultancy.com

Email Address

shafi@shaficonsultancy.com

