

Streamline end to end traceability for A&R Standard Programs

Merck & Co., Inc.,
Kenilworth, NJ, USA

- Gurubaran Veeravel
- Guowei Wu
- Suhas Sanjee
- Ram Radhakrishnan

Topics

- Need to streamline standards
- Factors affecting standards
- Solution
- Variable dependency
- Challenges and Strategies
- Summary & Conclusion

Need to Streamline Standards



Robust data standards and reporting standards are essential for the automation and optimization of Analysis & Reporting (A&R).



Once these standards are established and programs developed, maintaining them is challenging due to the dynamic environment of changes with internal business needs and regulatory requirements.



It is vital to assess the impact of these changes to the A&R standard programs. Impact assessment usually relies on the institutional knowledge of the A&R standard programmers and is typically manual and hence error prone.

Factors Affecting Standards

Changes to address internal business needs

- ❖ eCRF updates
- ❖ Bug fixes
- ❖ Enhancements
- ❖ Platform Changes (Additional Features)

Changes to address external factors

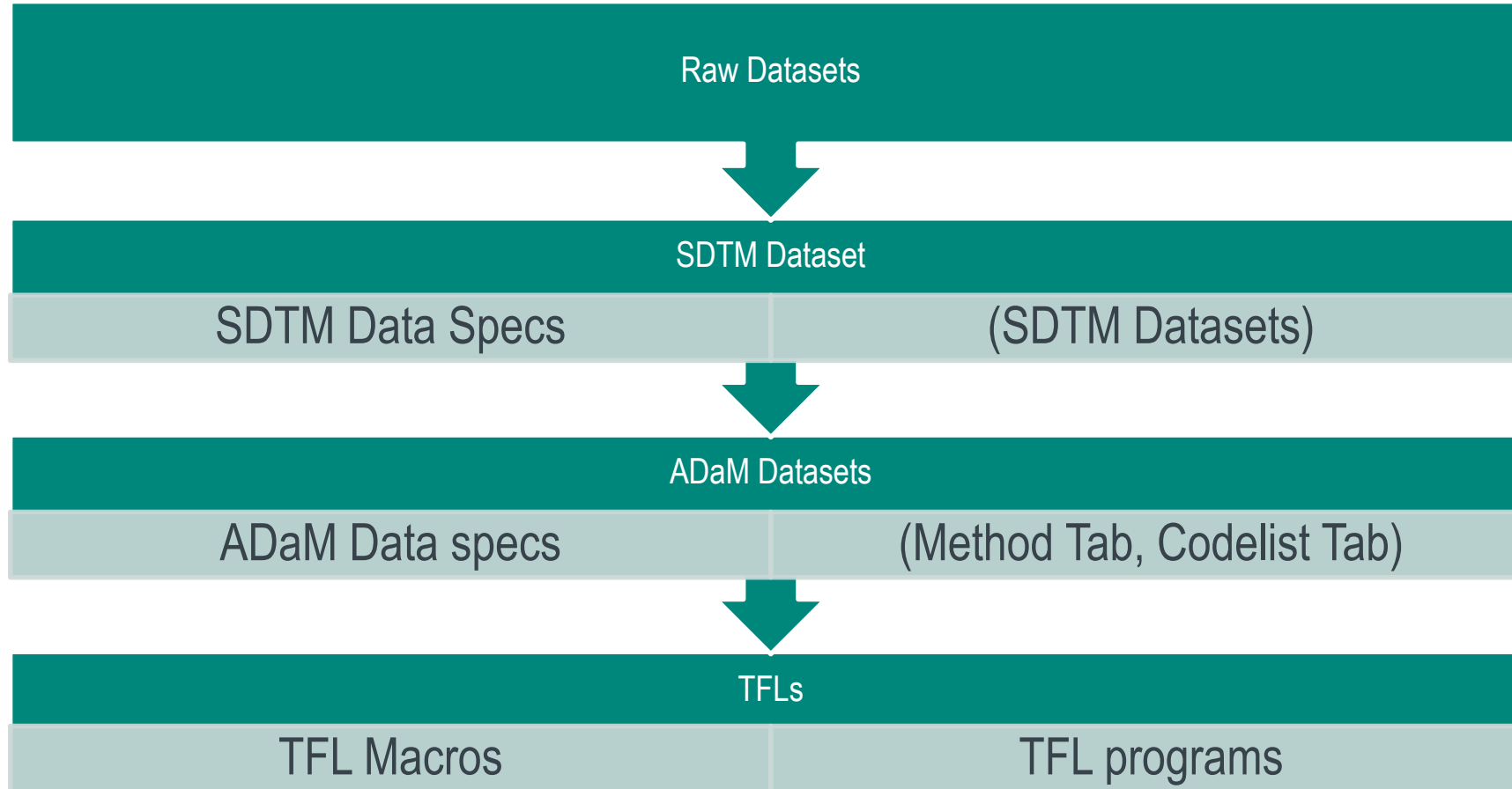
- ❖ SAS version changes (SAS 9.3 to 9.4)
- ❖ Platform changes (Modern Platforms Implementation)
- ❖ SDTM Version Changes (SDTM IG 3.2 to IG 3.3)
- ❖ TCG updates (Technical Conformance Guide – FDA)

Solution

- ✓ Impact analysis is essential while understanding the downstream effect of changes.
- ✓ Perform dependency checking on the standard macros and programs and identify the dependency on datasets and the variables.
- ✓ Develop an automated macro or a system to identify the dependency of all the standard macros, programs and templates.

Variable dependency

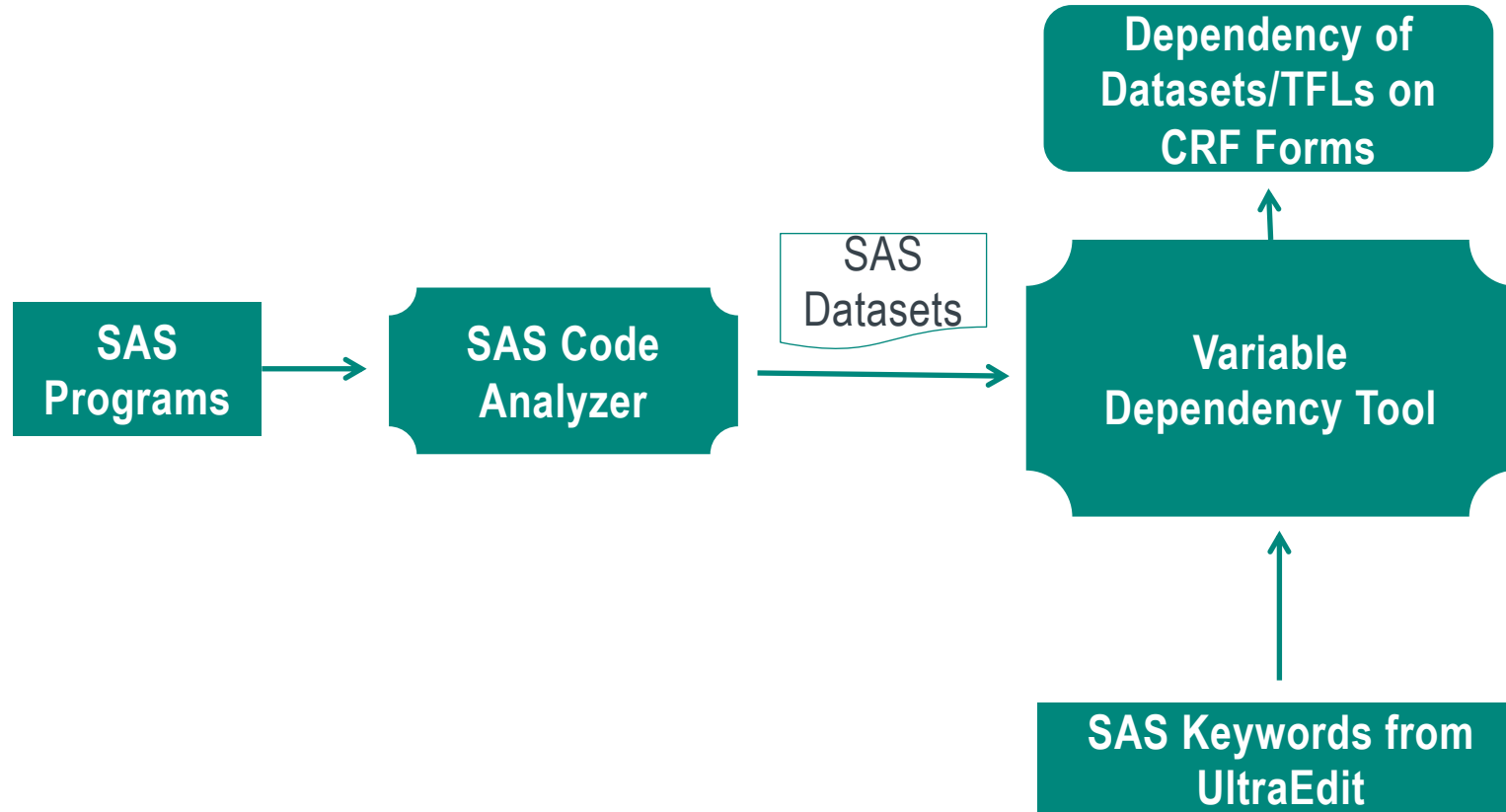
Variable Dependencies - Flow Chart



Dependency Testing

- Dependency testing is checking all the SAS programs or resolved macro code to identify the dataset names and the variable names the program is dependent on.
- Automated dependency check is important while performing impact analysis on the macros.
 - ✓ Impact analysis is essential while understanding the downstream effect of any changes
 - ✓ Accurate analysis on the macros and programs to make sure all the affected programs are captured
- Using SAS® keywords collected from Ultraedit® to parse out all the keywords to identify the datasets and variables

Design: Dependency Tool Input and Output



SAS Code Analyzer

The SCAPROC procedure implements the SAS Code Analyzer

- - Captures information about input, output and the use of macro symbols from a SAS job while it is running.
 - Can write this information and the information that is in the original SAS file to a file.
 - Can also generate a grid-enabled job that can concurrently run independent pieces of the job.
 - Can issue the SCAPROC procedure on your operating system's command line or in SAS code in the SAS Editor window.

The following command runs your SAS job with the SAS Code Analyzer from your operating system's command line:

```
sas yourjob.sas -initstmt "proc scaproc; record 'yourjob.txt' ; run;"
```

sas

is the command used at your site to start SAS.

yourjob.sas

is the name of the SAS job that you want to analyze.

yourjob.txt

is the name of the file that will contain a copy of your SAS code.

The file will also contain the comments that are inserted to show input and output information, macro symbol usage, and other aspects of your job.

Macro: Variable Dependency

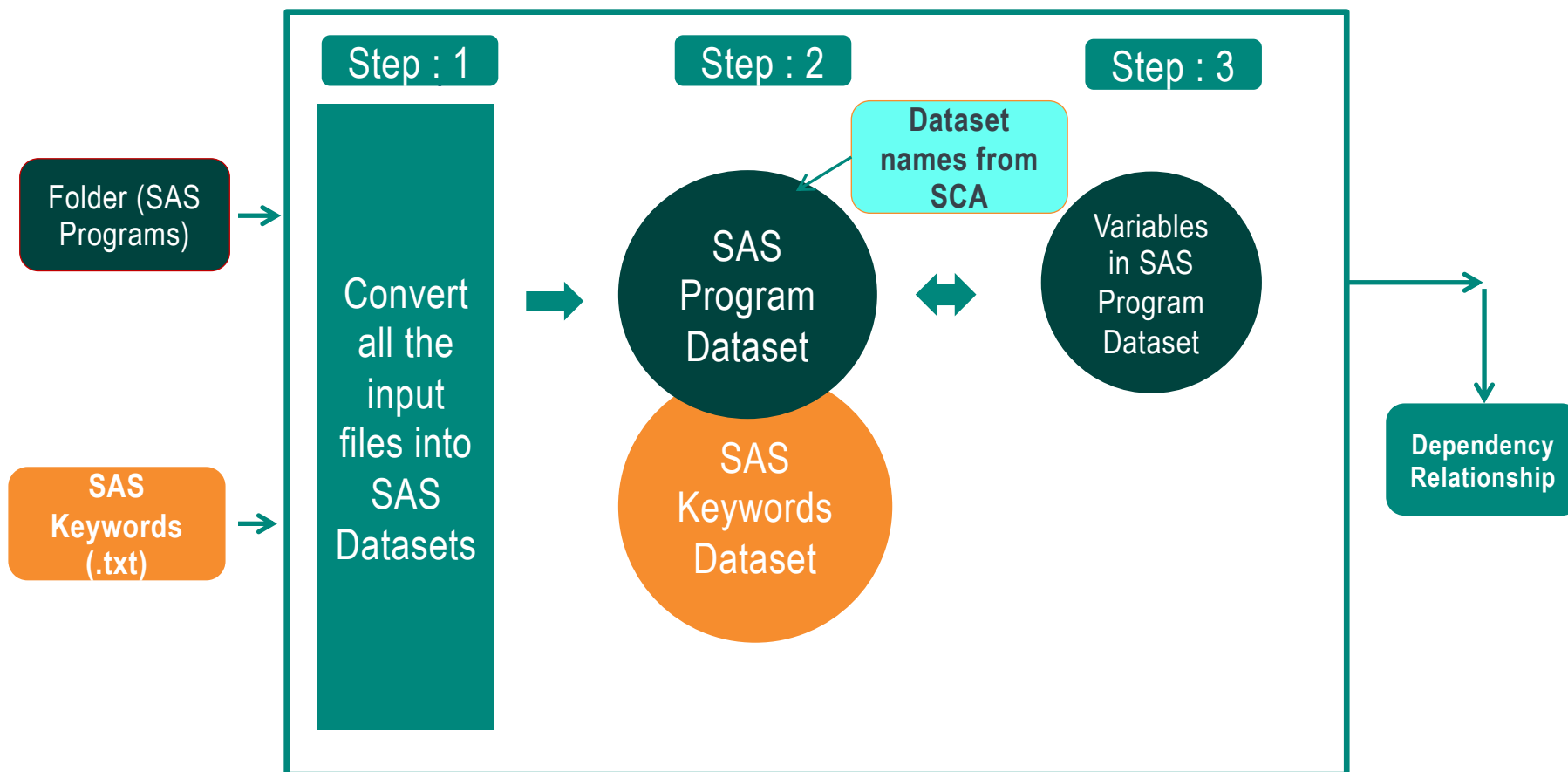
Name of the macro: %var0dependency

Function: Run the call programs for any SAS® program. The macro uses SAS keywords as input to identify the dependent datasets and variables.

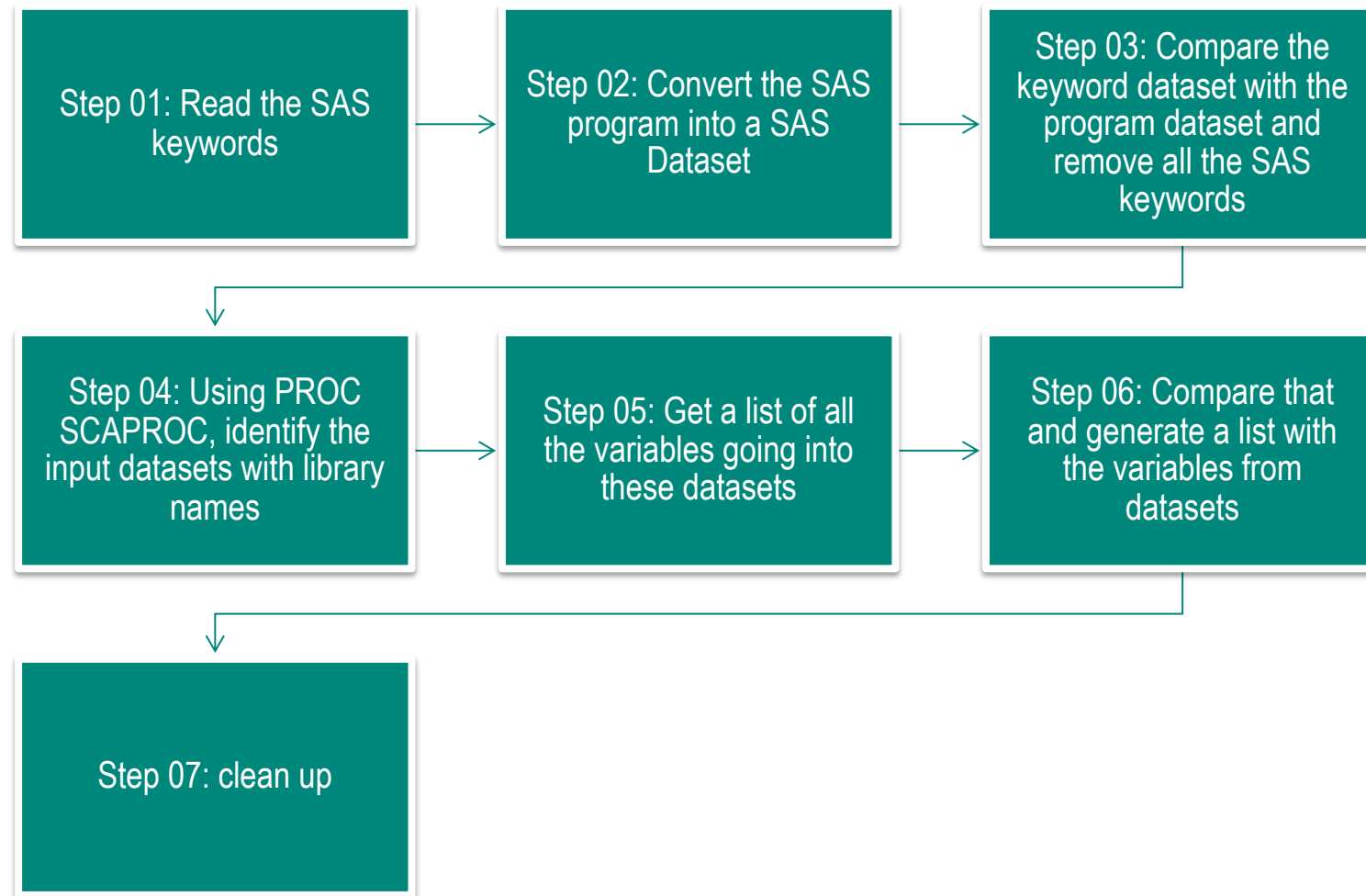
Parameter Description:

Parameter	Description
Input_libname	Input library
Output_libname	Output library
Input_excel	Input excel full path and filename
Excel_col_name	The column name in the excel file to use
Excel_where	Statement to subset excel sheets observations
sheet	Sheet name in the excel file
debug	Debug parameter. Valid values are Y.

Implementation: Internal Mechanics of the Dependency Tool



Macro: Step-by-Step Processing



SAS Keywords

SAS - [VIEWTABLE: TMP1.keywords]

File Edit View Tools Data Solutions Window Help

	F1	F2	F3	F4	F5	F6	F7
1	Nocase	Line	Comment	=	*	Line	Comment
2	String	Chars	=	"" DisableMLS File Extensions = SAS LOG ..# % &'()*+,-./:;<=>@[\\]^ }~ ..%macro data do proc { ..%mend end; run; quit; } ..{([..])} ..%"macro"	data	proc sql	
3	%mend	run;	quit;				
4	ACCESS	ACECLUS	ALLELE	ANOM	ANOVA	APPEND	APPSRV
5	BMDP	BOM	BOXPLOT	BUILD			
6	CALENDAR	CALIS	CANCORR	CANDISC	CAPABILITY	CASECONTROL	CATALOG
7	DATASETS	DATASOURCE	DB2EXT	DB2UTIL	DBCSTAB	DBF	DBLOAD
8	ENTROPY	EXPAND	EXPLODE	EXPORT			
9	FACTEX	FACTOR	FAMILY	FASTCLUS	FCMP	FONTREG	FORECAST
10	G3D	G3GRID	G4GRID	GA	GAM	GANNO	GANTT
11	HAPLOTYPE	HTSNP					
12	IML	IMPORT	INBREED	INFOMAPS	INTPOINT	IRP	ISHIKAWA
13	KDE	KRIGE2D					
14	LATTICE	LIFEREG	LIFETEST	LOAN	LOESS	LOGISTIC	LP
15	MACONTROL	MAPIMPORT	MDC	MDDDB	MDS	MEANS	METADATA
16	NESTED	NETDRAW	NETFLOW	NLIN	NLMIXED	NLP	NPARIWAY
17	OLAP	OPERATE	OPTEX	OPTIONS	OPTLOAD	OPTLP	OPTMODEL
18	PARETO	PDLREG	PDS	PDSCOPY	PHREG	PLAN	PLOT
19	QLIM	QUEST					
20	RANK	REG	REGISTRY	RELEASE	RELIABILITY	REPORT	ROBUSTREG
21	SCORE	SERVER	SHEWHART	SIM2D	SIMLIN	SORT	SOURCE
22	TABULATE	TAPECOPY	TAPELABEL	TEMPLATE:	TIMEPLOT	TIMESERIES	TPHREG
23	UCM	UNIVARIATE	UPLOAD	USERPROC			
24	VARCLUS	VARCOMP	VARIOGRAM	VARMAX			
25	WEBMDDDB						
26	X11	X12					

Sample: Impact Assessment on ADaM Macros

- 1) For a Technical Conformance Guide (FDA Guidance) update, a handful of variables are affected by the update.
- 2) Perform impact analysis on all the ADaM Macros/Templates to identify which datasets/variables are affected.
- 3) Input is in the form of Excel file, which contains the variable name affected.
- 4) Output will be in the form of a dataset which will contain all the datasets and variables affected by the TCG update.

Sample: Impact Assessment on ADaM Macros

ADaM programs/Macros

Name	Date modified
adae.sas	2/11/2019 12:16 PM
adae0sur.sas	7/25/2018 11:10 AM
adae0wrsndt.sas	3/27/2019 10:58 AM
ADaMIG1_1 Dataset Specification Templa...	3/27/2019 1:39 PM
adbase.sas	7/30/2018 8:32 PM
adcm.sas	12/18/2018 12:03 ...
adcm0whodd.sas	2/6/2019 11:45 AM
addili.sas	12/17/2018 3:08 PM
adeff.sas	7/15/2018 10:44 AM
adeg.sas	5/23/2018 9:05 AM
adex.sas	7/23/2018 2:02 PM
adexsum.sas	7/6/2018 3:29 PM
adlb.sas	5/24/2018 2:08 PM
adlb0con.sas	5/31/2018 8:56 AM
admh.sas	6/6/2018 2:04 PM
adpc.sas	9/3/2018 10:03 PM
adpm.sas	7/25/2018 2:40 PM
adpp.sas	10/24/2018 11:18 ...
adsite.sas	6/8/2018 11:56 AM
adsl.sas	10/25/2018 12:42 ...

Variables affected by TCG Update

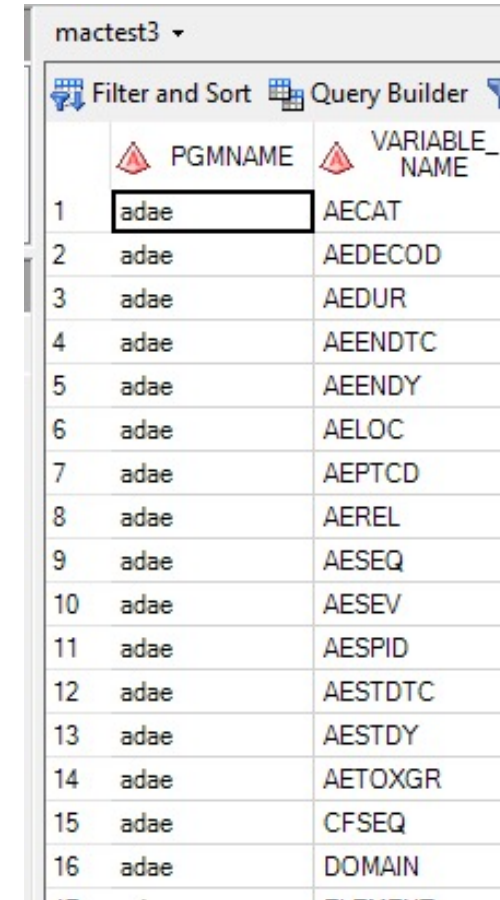
	D	E
Prefix	Variable Name (minus domain prefix)	Variable Name
	STUDYID	UNALLOC
	DOMAIN	ARMCD
	USUBJID	ARMCD
	SUBJID	ACTARM
	RFSTDTC	ACTARMCD
	RFENDTC	RFENDTC
	RFXSTDTC	RFXSTDTC
	RFXENDTC	RFXENDTC
	RFICDTC	RFICDTC
	RFPENDTC	RFPENDTC
	DTHDTC	DTHDTC
	DTHFL	DTHFL
	SITEID	SITEID
	INVID	INVID
	INVNAM	INVNAM
	BRTHDTC	BRTHDTC
	AGE	AGE
	AGEU	AGEU
	SEX	SEX
	RACE	RACE
	ETHNIC	ETHNIC
	ARMCD	ARMCD
	ARM	ARM
	ACTARMCD	ACTARMCD
	ACTARM	ACTARM

Dependency Macro Call Program

Call Program

```
%var0dependency(  
  input_libname=mactest3  
  ,output_libname=lpod  
  ,input_excel= %str(/root/test/tcgimpact.xls)  
  ,sheet=SDTM32  
  ,excel_col_name=variable_name  
);
```

Output – SAS Dataset



	PGMNAME	VARIABLE_NAME
1	adae	AECAT
2	adae	AEDECOD
3	adae	AEDUR
4	adae	AEENDTC
5	adae	AEENDY
6	adae	AELOC
7	adae	AEPTCD
8	adae	AEREL
9	adae	AESEQ
10	adae	AESEV
11	adae	AESPID
12	adae	AESTDTC
13	adae	AESTDY
14	adae	AETOXGR
15	adae	CFSEQ
16	adae	DOMAIN

Challenges & Strategies

Challenges	Strategies
Identifying the dataset names	Use SAS Code Analyzer – PROC SCAPROC
Performance/Scalability	Performs well in production environment
Distinguish SAS keywords and Variable names	Use PROC CONTENTS and link variables to the data specs to identify the right variable
Ability to use on Standard level macros	Use the Resolved SAS code, using options MPRINT
Dealing with legacy programs	Develop the Tool to read the keywords intelligently

Summary and Conclusion

- It is important that every organization that develops and maintains standards have an automated method to identify dependencies.
- Identifying the dependencies of datasets and variables for a SAS program is the first step in identifying the end-to-end traceability.
- Automated approach helps identifying the dependencies in all the macros for a particular update in one macro run.
- In on our implementation experience, we can say that variable dependency macro identifies all the dependencies for all the macros which will have an impact on any major updates/changes.

THANK YOU