

Validating CDISC Compliant data and Creating Define.xml made easy

Sandeep Purwar, Bodhi Global Solutions

Abstract:

The goal of this paper is to list possible techniques available in the industry to validate the SDTM metadata and to generate Define.xml. This paper concentrates how to use SAS Clinical Toolkit to achieve above mentioned tasks.

Introduction:

With CDISC SDTM and ADAM being standards used industry wide, validating the standards make it important for smooth eSubmission and timely approval of the studies by the FDA.

The different techniques used to validate the SDTM data are double programming (In this case both the programmers follow the same specification document and use Proc Compare to check if both the datasets have same metadata), In house macros, open CDISC, Proc CDISC and SAS Clinical Toolkit (we will talk in detail on how to use toolkit to validate the SDTM data).

As per the FDA eSubmission preferred way of submitting the data is SDTM/ADAM standard and define.xml. We will discuss how to create define.xml using SAS Clinical Toolkit also.

What is SAS Clinical toolkit?

Its utility tool developed by SAS to enable you to manage the registration of standards, and provides the metadata and API infrastructure to interact with standards. Toolkit is capable of validating SDTM data (Includes JANUS, WEBSDM and SAS checks) and also can create Define.xml.

Validation Using SAS Clinical Toolkit:

Toolkit compares the metadata, data values of source data (study data) with reference metadata (Standards) to find compliance with SDTM standards. SAS clinical toolkit provides framework for validation, creating define.xml and also to build customized process according the client's needs.

From the below figure you see that the validation process uses Source (study) Data and Metadata , Process Controls (Reference/Standard Metadata , properties and checks to be run) and Controlled terminology to create process results (results data set , metrics data set). Once the result datasets are created you can also create PDF report by running appropriate macros.

Following figure shows bigger picture of process and framework on which SAS Clinical toolkit is built on.

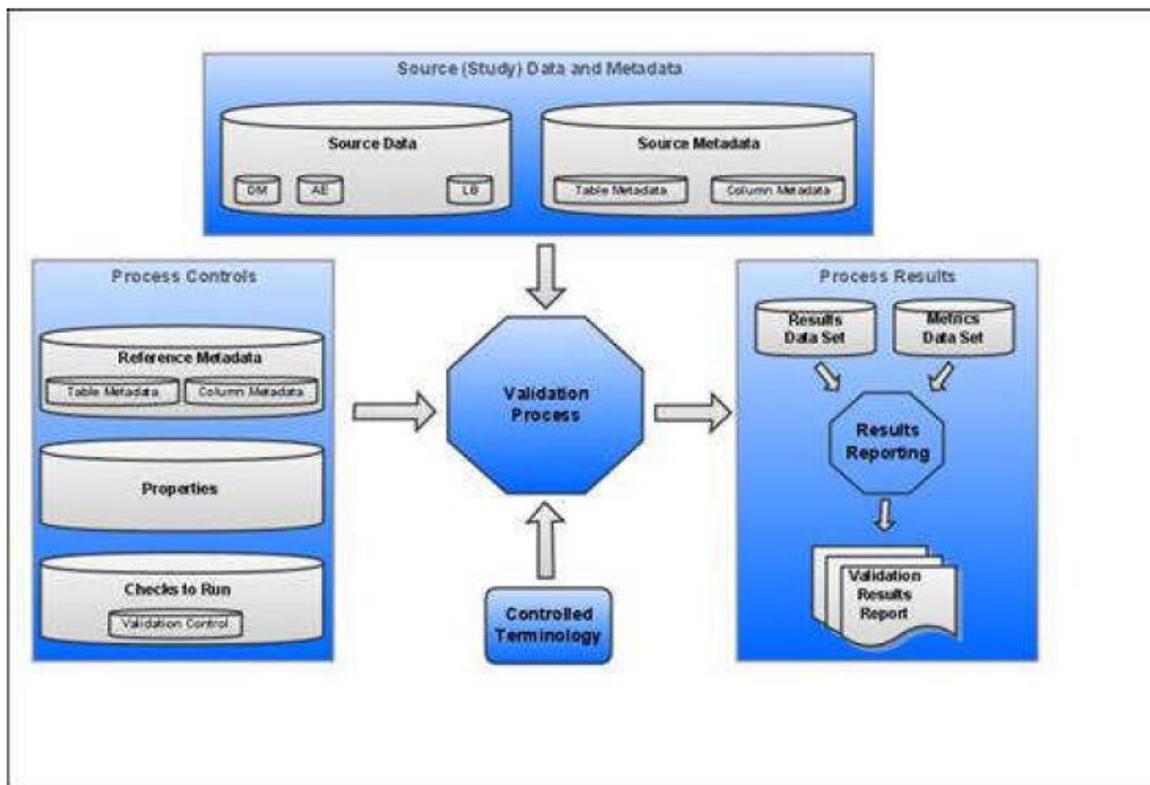


Figure 1: From SAS Clinical Toolkit 1.3 User guide Pg 84

Here are the steps to validate study (Source) data:

1. Create source metadata
 - Create source meta by running create_sourcemetadata.SAS
 - Change _cstsrctype = SDTMLIBRARY instead of CRTDDS (If SDTM data is used)
 - Update the macro : sdtmutil_createsrcmetafromSASlib.SAS to give right path for Supp and custom domains in the source_table (If made changes in this macro , make sure you compile before running the create_sourcemetadata.SAS program)
 - Run create_sourcemetadata.SAS (I/P : SAS Datasets)
2. Validation_control.SAS7bdat (Subset of master control dataset) should be updated depending upon what checks have to be conducted.
3. Source SDTM as Input data

4. Folders required :

- “Control “ with validation_control.SAS7bdat
- “Metadata” with Study_columns ,Source_study , Source_tables SAS datasets
- “Program” with validation.properties
- “Data” with SDTM datasets
- “Results”

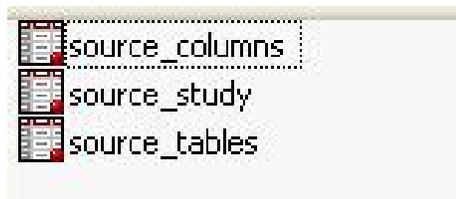
5. Macros Running :

- cstutil_processsetup()
- Sdtm_validate

6. Run Validate_data.SAS

7. Check the folder “Results” for two SAS datasets

- Validation_metrics.SAS7bdat
- Validation_metrics.SAS7bdat



8. Run Cst_report.SAS without closing the previous session as this program requires SASreferences.SAS7bdat which was created by validate_data.SAS

9. Check pdf file in “Results” folder called cstreport.pdf . Following figure shows the part of the cstreport.pdf.

**SAS Clinical Standards Toolkit 1.3
CDISC-SDTM 3.1.2 VALIDATION**

Process Metrics

Summary Metrics		Check Metrics				
Metric	#	Check ID	# Check Invocations	# Recs (if available)	# Errors	# Check Invocations Not Run
# of distinct check invocations	44	SDTM0002	1	3	0	0
# check invocations not run	21	SDTM0003	1	1	0	0
Errors (severity=High) reported	558	SDTM0012	1	6	6	0
Warnings (severity=Medium) reported	0	SDTM0017	1	1	0	0
Notes (severity=Low) reported	0	SDTM0023	1	123	123	0
Structural errors, warnings and notes	217	SDTM0031	1	86	86	0
Content errors, warnings and notes	543	SDTM0036	1	1	0	0
		SDTM0038	1	1	0	0
		SDTM0201	1	23	0	0
		SDTM0209	1	1	0	1
		SDTM0210	1	2	0	2
		SDTM0217	1	1	0	1
		SDTM0220	1	1	0	1
		SDTM0352	1	1	0	0
		SDTM0407	1	1	0	1
		SDTM0410	1	1	0	1
		SDTM0414	1	3	0	1
		SDTM0415	1	1	0	1
		SDTM0422	1	1	0	1
		SDTM0423	1	16	15	1
		SDTM0506	1	1	0	0
		SDTM0521	1	1	0	1
		SDTM0603	1	17	0	5
		SDTM0606	1	23	0	0

10.

Report generated 2011-06-28T13:12:46 on process run 2011-06-28T13:02:12

Creating Define.xml using SAS Clinical Toolkit:

Toolkit creates Define.xml using 39 CRT-DDS (CDISC ODM) i.e. SAS representation of CRT-DDS datasets and 176 columns, which are created using SDTM metadata. Toolkit is also capable of creating metadata from define.xml, topic which is out of the scope of this paper.

Here are the steps to create Define.xml:

Toolkit provides driver programs to run following steps and macros. Driver programs also help to setup the process to perform specific toolkit task.

1. Follow this step if you have not already created source metadata
 - Create source meta by running create_sourcemetadatas.SAS
 - Change _cstsrctype = SDTMLIBRARY
 - XPT datasets should be created from source data & be kept in right folder in accordance with path specified
 - Update the macro : sdtmutil_createsrcmetafromSASlib.SAS to give right path for Supp and custom domains in the source_table (If made changes in this macro , make sure you compile before running the create_sourcemetadatas.SAS program)
 - Run create_sourcemetadatas.SAS (I/P : SAS Datasets)
 - Update Source_table by using program update_source_table.SAS , to add all the missing values for SUPP and Study specific domains unmatched domains in master source_table dataset
2. Run create_crtds10_from_sdtm311.SAS (creates crtds datasets)
 - Input: source metadata from above steps
 - Output : crtds datasets
3. Run create_crtds_define.SAS (creates define.xml)
 - Make changes to crtds SAS datasets: annotatedcrfs . mdvleaf , mdvleaftitles to incorporate blankcrf.pdf in define.xml before running this program
 - Input : crtds datasets (created from above steps)

Conclusion:

The cumbersome process of validating SDTM data and creating Define.xml can be easy done by SAS Clinical Toolkit by understanding the framework and driver programs which access the macros. The flexibility of adding user defined validation checks, capability of creating and validating Define.xml makes SAS Clinical Toolkit one of the good solutions available out in the market.

References:

- Understanding the define.xml File and Converting It to a Relational Database
<http://support.sas.com/resources/papers/proceedings10/157-2010.pdf>
- Power point presentation by SAS
<http://www.lexjansen.com/pharmasug/2010/SAS/SAS-cd-SAS01.pdf>
- SAS®Clinical Standards Toolkit 1.3 User's Guide
<http://support.sas.com/documentation/cdl/en/clinstdtktug/63608/PDF/default/clinstdtktug.pdf>