

SAS and Open Source Tools for CDISC SDTM Compliance Checks for Regulatory Submissions

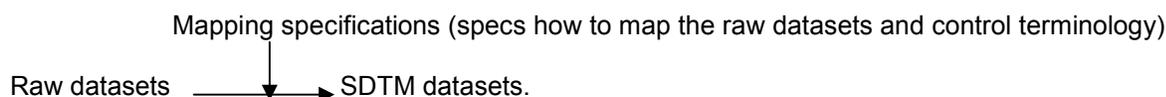
Peter Loonan, Statistical/SDTM Programmer consultant, Eisai Pharmaceuticals, NJ.
 Sandeep Kottam, Sr.Statistical/SDTM Programmer consultant, Eisai Pharmaceuticals, NJ.
 Sree K Tripuraneni, Sr.Statistical Programmer consultant, Purdue Pharma, Stamford, CT.

Abstract:

Implementing CDISC SDTM standards to convert raw data for FDA submissions has become one of the main goals for pharmaceutical companies. Compliance checks of the CDISC SDTM data is also an important step in validating the conversion. Compliance checks are done using lines of SAS code and Metadata files to support this process. Compliance checks for custom domains can become a very tedious process to implement, requiring more information for metadata files used for compliance checks. This paper will discuss the benefits of using a third party tool and how it can be customized to both check the compliance of custom SDTM domains created using SAS and set up specific study or company based checks.

Introduction:

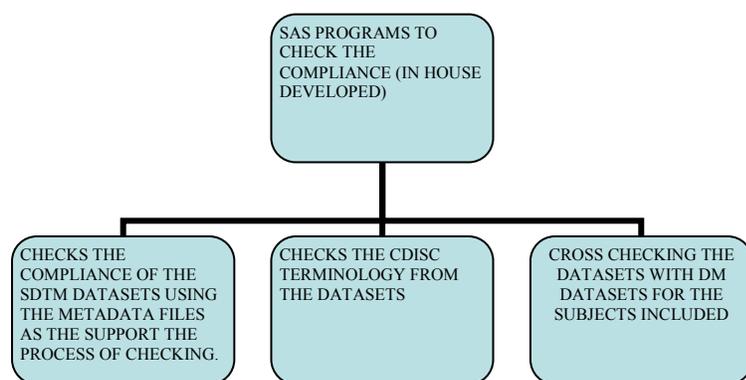
CDISC SDTM defines a standard structure for study data tabulations which are to be submitted for the regulatory. The following illustration shows how the raw data are converted to SDTM standard datasets.



The specifications are developed how to convert raw collected data in to which domain of SDTM dataset. There will be other domains in each study other than described in the SDTM IG 3.1. 1 or 3.1.2 , these customized domains need lot of attention and the specs should be clearly describe how each domain should be classified in to Interventions, Findings or Events Class according to SDTM standards. SAS is the tool used to convert raw datasets to SDTM datasets.

Validation:

After the conversion compliance is the crucial part. The following illustration shows how the validation is done using SAS.



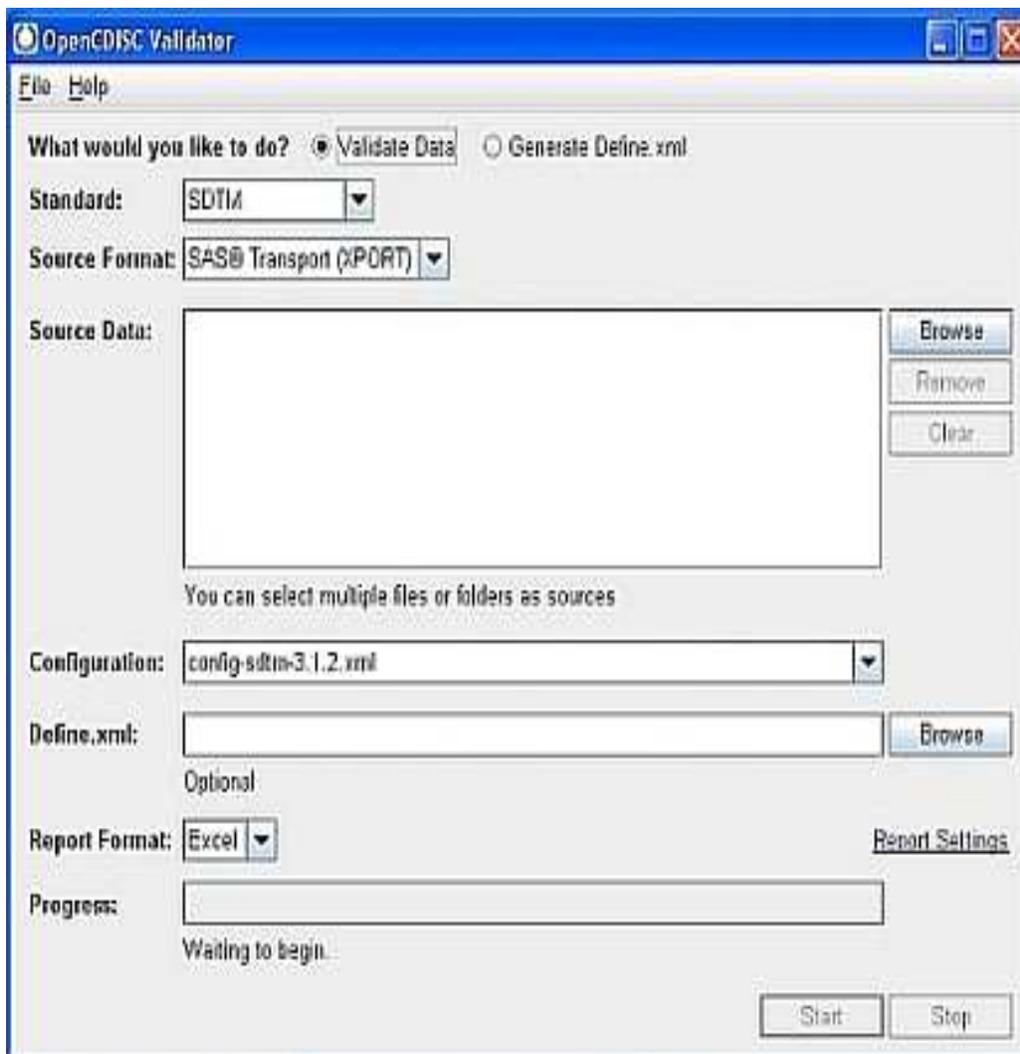
SAS programs are developed in house according to the company specifications to support this process the SDTM metadata files which can be downloaded from CDISC website is used.

Process:

The process consists of checking the data structure and properties, which includes the variable labels, type, names. Checking the CDISC SDTM Terminology like the –testcd, outcome, --stat etc. present in the data. Cross checking the demog data to the other datasets.

Open source tool:

One tool that offers both the ease of use and flexibility of custom user modification is OpenCDISC Validator. OpenCDISC is an open source tool that checks the compliance of SAS transport or delimited files against either SDTM 3.1.1 or SDTM 3.1.2 standards. It also currently offers both a check for validation of and standards for the creation of define.xml documents, and although those are certainly extremely beneficial tools they will not be covered in this paper. The tool has a simple user interface to specify the source files, choose the standard and configuration, choose the output file type, and run the process. The basic validator screen looks like the following picture without modifications.



The OpenCDISC Validator can be run on SAS data in .xpt format as is, with no modification, and it will check the for the same compliance and consistency as the WebSDM tool does using rules that contain the same logic, as well as check community designed rules of the same nature for all domains.

A major advantage to this tool, though, is its ability to allow for user modifications to configuration files and rules being used by the system. Using this advantage, configurations could be set to account for SDTM-, SDTM+, non-conventional, and other data with some simple modifications.

In the OpenCDISC root directory there is a folder named config, where the three default xml configurations are stored. Either one of these documents can be edited to modify a rule or rules that already exist or add/delete rules, or a new configuration xml can be created using the same structure as one of these three files. Once you open one of the configuration files in wordpad or any xml editor, you will see quite a few lines of xml code similar to the following picture.

```

<ODM xmlns="http://www.odisc.org/ns/odm/v1.2" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:df="ht
<Study OID="org.opendisc.validator">
  <GlobalVariables>
    <StudyName>OpenCDISC Validator</StudyName>
    <StudyDescription>OpenCDISC Validator Configuration</StudyDescription>
    <ProtocolName>N/A</ProtocolName>
  </GlobalVariables>
  <MetaDataVersion OID="CDISC.SDTM.3.1.1" Name="OpenCDISC Validator Configuration for CDISC SDTM v3.1.1" Description="OpenCDISC Validator Configuration fo
  <!-- ***** -->
  <!-- ItemGroupDef (domain) information Section -->
  <!-- ***** -->
  <ItemGroupDef OID="GLOBAL" Name="GLOBAL" Repeating="No" IsReferenceData="No" Purpose="Tabulation" def:Label="Global Rules Section" def:Structure="One
  <!-- ***** -->
  <!-- Contains references to submission level rules -->
  <!-- ***** -->
  <ItemRef ItemOID="STUDYID" OrderNumber="1" Mandatory="Yes" Role="Identifier" val:Core="Required"/>
  <!-- ***** -->
  <!-- def:leaf details for hypertext linking the dataset -->
  <!-- ***** -->
  <def:leaf ID="Location.GLOBAL" xlink:href="global.xpt">
    <def:title>global.xpt</def:title>
  </def:leaf>
  <!-- ***** -->
  <!-- Validation rule references for this ItemGroupDef (domain) -->
  <!-- ***** -->
  <!-- Global Rules -->
  <val:ValidationRuleRef RuleID="SD0061" Active="Yes"/>
  <val:ValidationRuleRef RuleID="SD0095" Active="Yes"/>
  <val:ValidationRuleRef RuleID="SD1020" Active="Yes"/>
</ItemGroupDef>
<ItemGroupDef OID="DM" Name="DM" Repeating="No" IsReferenceData="No" Purpose="Tabulation" def:Label="Demographics" def:Structure="One record per subj
  <!-- ***** -->
  <!-- Each variable is listed here for this ItemGroupDef (domain) -->
  <!-- ***** -->
  <ItemRef ItemOID="STUDYID" OrderNumber="1" Mandatory="Yes" Role="Identifier" val:Core="Required"/>
  <ItemRef ItemOID="DOMAIN" OrderNumber="2" Mandatory="Yes" Role="Identifier" val:Core="Required"/>
  <ItemRef ItemOID="SUBJID" OrderNumber="3" Mandatory="Yes" Role="Identifier" val:Core="Required"/>
  <ItemRef ItemOID="SUSBJD" OrderNumber="4" Mandatory="Yes" Role="Topic" val:Core="Required"/>
  <ItemRef ItemOID="DM.RSEIDFC" OrderNumber="5" Mandatory="No" Role="Timing" val:Core="Expected"/>
  <ItemRef ItemOID="DM.RSEIDFC" OrderNumber="6" Mandatory="No" Role="Timing" val:Core="Expected"/>
  <ItemRef ItemOID="DM.STEID" OrderNumber="7" Mandatory="Yes" Role="Record Qualifier" val:Core="Required"/>
  <ItemRef ItemOID="DM.INVID" OrderNumber="8" Mandatory="No" Role="Record Qualifier" val:Core="Permissible"/>
  <ItemRef ItemOID="DM.INVNAM" OrderNumber="9" Mandatory="No" Role="Synonym Qualifier" val:Core="Permissible"/>
  <ItemRef ItemOID="DM.BRTHDTC" OrderNumber="10" Mandatory="No" Role="Result Qualifier" val:Core="Permissible"/>
  <ItemRef ItemOID="DM.AGE" OrderNumber="11" Mandatory="No" Role="Result Qualifier" val:Core="Expected"/>
  <ItemRef ItemOID="DM.AGEU" OrderNumber="12" Mandatory="No" Role="Variable Qualifier" val:Core="Expected"/>
  <ItemRef ItemOID="DM.SEX" OrderNumber="13" Mandatory="Yes" Role="Result Qualifier" val:Core="Required"/>
  <ItemRef ItemOID="DM.RACE" OrderNumber="14" Mandatory="No" Role="Result Qualifier" val:Core="Expected"/>
  <ItemRef ItemOID="DM.ETHNIC" OrderNumber="15" Mandatory="No" Role="Result Qualifier" val:Core="Permissible"/>
  <ItemRef ItemOID="DM.ARMCD" OrderNumber="16" Mandatory="Yes" Role="Result Qualifier" val:Core="Required"/>
  <ItemRef ItemOID="DM.ARM" OrderNumber="17" Mandatory="Yes" Role="Synonym Qualifier" val:Core="Required"/>
  <ItemRef ItemOID="DM.COUNTRY" OrderNumber="18" Mandatory="Yes" Role="Result Qualifier" val:Core="Required"/>
  <ItemRef ItemOID="DM.DMDC" OrderNumber="19" Mandatory="No" Role="Timing" val:Core="Permissible"/>
  <ItemRef ItemOID="DM.DMDY" OrderNumber="20" Mandatory="No" Role="Timing" val:Core="Permissible"/>
  <!-- ***** -->

```

This current section shows just a small glimpse of the beginning of the DM or demographics domain configuration. It would be possible to change the variable order, name, mandatory status, role, and SDTM necessity with some simple changes to the text. Below the definition of variables is the section detailing which rules will be executed for the domain, also allowing the user the ability to add, delete, or modify which rules are actually executed on a domain by domain basis. Further variable level details are stored at the end of the configuration file after all domain details and prior to Janus and OpenCDISC rules.

Modification to the rules is also fairly simple and does not require a vast knowledge of the xml language, as most checks are simple enough to create or modify based on the context of other checks. One example of a simple modification that was made, was when I noticed the SD0007 check for consistency of standard units was only checking the domain --TESTCD variable only.

It is generally agreed that there are lab tests of the same name in different categories that have different units. To make this modification it would simply require a change to the rule in adding a %Domain%CAT call to the GroupBy section of the check. See below for the following check post modification: <val:Unique ID="SD0007" Variable="%Domain%STRESU" GroupBy="%Domain%TESTCD,%Domain%CAT" Matching="Yes" When="%Domain%STRESU != "" Message="Inconsistent value for Standard Units" Description="Standard Units (--STRESU) must be consistent for all records with same Short Name of Measurement, Test or Examination (--TESTCD) and Category (--CAT)." Category="Consistency" Type="Error" Severity="High"/>

There are three other sections, past the Janus rules, containing specific rule syntax as you can see here:

```

<!-- ***** -->
<!-- Additional OpenCDISC Validation Rules -->
<!-- ***** -->
<val:Unique ID="SD1001" Variable="SUBJID" When="SUBJID != ''" Message="Duplicate SUBJID" Description="The value of Subject Identifier for t
<val:Condition ID="SD1002" Test="RFSTDTC @ltcq RFENDTC" When="RFSTDTC != '' @and RFENDTC != ''" Message="Start date must be before end date
<val:Required ID="SD1003" Variable="AGE" When="AGEU != ''" Message="Missing value although units provided" Description="When Age Units (AGE
<val:Lookup ID="SD1005" Variable="STUDYID == STUDYID" From="DM" When="STUDYID != ''" Message="STUDYID value does not match STUDYID in DM dc
<val:Metadata ID="SD1006" Variable="[IDVAR]" From="[RDOMAIN]" When="RDOMAIN != '' @and IDVAR != ''" Message="Referenced key variable not fc
<val:Lookup ID="SD1007" Variable="USUBJID == USUBJID, [IDVAR] == IDVARVAL" From="[RDOMAIN]" When="RDOMAIN != '' @and IDVAR != '' @and USUBJ
<val:Condition ID="SD1008" Test="CODTC == ''" When="RDOMAIN != '' @and IDVAR != '' @and IDVARVAL != ''" Message="CODTC should be null if cc
<val:Condition ID="SD1010" Test="ELEMENT == ''" When="ETCD == 'UNPLAN'" Message="ELEMENT should be null for unplanned element" Description=
<val:Regex ID="SD1011" Variable="%Variables[*DUR,*ELIM,*EVLINT]*" Test="-?P(((0-9)+Y)?((0-9)+M)?((0-9)+D)?(T((0-9)+H)?((0-9)+M)?((0-9)+S)?
<val:Lookup ID="SD1012" Variable="ETCD == ETCD, ELEMENT == ELEMENT" From="TE" When="ETCD @neqic 'UNPLAN' @and ETCD != '' @and ELEMENT != ''
<val:Lookup ID="SD1014" Variable="TAETORD == TAETORD" From="TA" When="TAETORD != ''" Message="Invalid TAETORD" Description="Order of Elemen
<val:Lookup ID="SD1015" Variable="EPOCH == EPOCH" From="TA" When="EPOCH != ''" Message="Invalid EPOCH" Description="Epoch (EPOCH) values mu
<val:Lookup ID="SD1016" Variable="IETESTCD == IETESTCD, IETEST == IETEST, IECAT == IECAT" When="IETESTCD != '' @and IETEST != '' @and IECAT
<val:Lookup ID="SD1017" Variable="VISITNUM == VISITNUM" From="TV" When="VISITNUM != '' @and SVUFDES == ''" Message="Invalid VISITNUM" Descr
<val:Lookup ID="SD1018" Variable="VISITNUM == VISITNUM, VISIT == VISIT, VISITDY == VISITDY" From="TV" When="VISITNUM != '' @and VISIT != ''
<val:Condition ID="SD1019" Test="VISITDY == ''" When="SVUFDES != ''" Message="VISITDY should be null for unplanned visits" Description="For
<val:Find ID="SD1020" Variable="DOMAIN" Terms="DM" Message="Missing DM dataset" Description="Demographics (DM) dataset should be included i
<!-- ***** -->
<!-- Internal Validation Rules -->
<!-- ***** -->
<val:Metadata ID="SD0001" Message="No records in data source" Description="Identifies domain table that has zero rows and therefore contain
<val:Metadata ID="SD0062" Message="Incompatible data source" Description="Identifies a domain whose source data failed to load" Category="S
<!-- ***** -->
<!-- CDISC Controlled Terminology Rules -->
<!-- ***** -->
<val:Lookup ID="CT0004" Variable="CDISCSubmissionValue == AGEU" From="FILE:TAB:%System.ConfigDirectory%/data/SDTM Terminology.txt" Where="C
<val:Lookup ID="CT0006" Variable="CDISCSubmissionValue == COUNTRY" From="FILE:TAB:%System.ConfigDirectory%/data/SDTM Terminology.txt" Where
<val:Lookup ID="CT0020" Variable="CDISCSubmissionValue == IECAT" From="FILE:TAB:%System.ConfigDirectory%/data/SDTM Terminology.txt" Where="

```

Although it is not necessary to separate the rules in the configuration documents, it does allow for an easier time of recognizing and custom checks or added checks. Due to the open source platform of the software, it allows for complete customization, giving the user the ability to create study specific configurations and eliminate the impact on checks for other studies/data. Considering the ever changing set of rules controlling the design of standards such as CDISC SDTM, this benefit allows data to be checked for accuracy using the latest information and possibly more information than is required.

CONCLUSIONS:

By using both SAS and OpenCDISC tool any company can generate SDTM compliant datasets with ease and efficiency.

REFERENCES:

Opencdisc.org

ACKNOWLEDGMENTS:

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

CONTACT INFORMATION:

Your comments and questions are valued and encouraged. Contact the authors at:

Sandeep Kottam

Sr.Statistical programmer consultant,

Eisai Pharmaceuticals,

Woodcliff Lake, NJ -07677.

605-691-3274

kottamsandeep@gmail.com

Peter Loonan

Statistical/SDTM programmer consultant,

Eisai Pharmaceuticals,

Woodcliff Lake, NJ -07677.

ploonan@gmail.com

Sree Lakshmi K Tripuraneni

Sr.Statistical programmer,

Stamford, CT – 06905.

605-691-3312

sreektripuraneni@gmail.com