PhUSE 2010

Paper CD14

Implementing, Managing, and Validating a Clinical Standard Using SAS Clinical Standards Toolkit 1.3®

Gene Lightfoot, SAS Institute, Cary, North Carolina, USA

Abstract

Many pharmaceutical companies are starting to implement the various CDISC models that are now available for download. Unquestionably the most popular standard is the Submission Data Tabulation Model version 3.1.2, most commonly referred to as SDTM 3.1.2. It is the primary model for data submitted to the Federal Drug Administration (FDA). There are many ways to implement, manage, and validate standards and this paper will discuss this process using SDTM 3.1.2 as the example standard.

Introduction

The SAS Clinical Standards Toolkit version 1.3 (Toolkit) is the product used throughout this paper to achieve the desired goal of standards management. The Toolkit is a set of SAS macros that are supplied with SAS version 9.1.3 and 9.2. For 9.1.3 it is available as a download only, for 9.2 it is available with base SAS when purchased from SAS. All examples throughout this paper will be using SAS 9.2 (notes will be included if the process differs from SAS 9.1.3). The Toolkit macros are provided as open source and are accessible to the user. The Toolkit also provides domain metadata for SDTM 3.1.1 and 3.1.2, and in addition, it also provides recognized validation checks from WedSDM, OpenCDISC, and SAS. Sample studies and data are provided for the user along with driver programs that allow the Toolkit to be run from base SAS. The Toolkit allows the user to create their own validation checks or to modify existing checks. This paper will limit the discussion of the Toolkit to a few basic concepts that the new user will encounter in order to use the product effectively. Discussion will be limited to the reference data sets and the study level data sets.

The Toolkit is part of SAS Clinical Data Integration® (CDI) solution and full benefit is derived when used with this product. Interaction with the Toolkit outside of CDI requires moderate to advanced skills in SAS programming language and the SAS macro language.

SDTM 3.1.2 Submission Data Tabulation Model

SDTM 3.1.2 contains at least 32 definable domains as stated in the SDTM Implementation guide available at www.cdisc.org. Using the excel spreadsheet from CDISC for SDTM 3.1.2 the information was imported into SAS data sets. The Toolkit uses this metadata in the SDTM validation process and it is also used within CDI, via the Toolkit, for mapping and validating data. The Toolkit uses two primary data sets to provide this metadata: reference_tables and reference columns. Reference_tables contains domain

level metadata as provided by the SDTM model as well as metadata needed by the Toolkit. Shown below is a contents listing of the reference_tables data set.

Reference_tables.sas7bdat				
Variable Name	**Variable Length**	**Variable Label**	**Variable Type**	**Notes**
class	40	Observation Class within Standard	Char	SDTM 3.1.2 supplied
Comment	200	Comment	Char	SAS/User supplied
Date	20	Release Date	Char	Derived from SDTM
Keys	200	Table Keys	Char	SDTM 3.1.2 supplied
Label	40	Table Label	Char	SDTM 3.1.2 supplied
Purpose	20	Purpose	Char	SDTM 3.1.2 supplied
SASref	8	SASreferences sourcedata libref	Char	SAS/User supplied
Standard	20	Name of Standard	Char	SDTM 3.1.2 supplied
StandardVersion	20	Version of Standard	Char	SDTM 3.1.2 supplied
Standardref	200	Associated reference(s) in Standard	Char	SAS/User derived from any available source
State	20	Data Set State (Final, Draft)	Char	SAS/User supplied
Structure	200	Table Structure	Char	Derived from SDTM
Table	32	Table Name	Char	SDTM 3.1.2 supplied
XmlPath	200	(Relative) path to xpt file	Char	SAS/User supplied
XmlTitle	200	Title for xpt file	Char	SAS/User supplied

The reference_tables data set is the global standard for the Toolkit. All study metadata will be compared to this table to determine if the study domain tables are in compliance with the SDTM model. If this were not an SDTM model but a user standard or model, the user would be responsible for creating, populating, and maintaining the reference_tables data set. Since the references_tables data set represents the global standard, a study level data set containing the same metadata is provided in the Toolkit. It is called the source_tables data set. Depending on the complexity or size of the study, source_tables could be an exact copy of reference_tables. But in some cases, the study may contain fewer domains, and in these cases source_tables would be a subset of the global standard. In addition, source_tables could contain domains not present in the global standard.

Shown below in Figure 1 is an example of the contents for reference_tables from Toolkit 1.3 for the SDTM 3.1.2 standard.

Figure 1 Reference_tables.sas7bdat

SASreferences sourcedata libref	Table Name	Table Label	Observation Class within Standard	(Relative) path to xpt file
REFDATA	AE	Adverse Events	Events	../transport/ae.xpt
REFDATA	CE	Clinical Events	Events	../transport/ce.xpt
REFDATA	CM	Concomitant Medications	Interventions	../transport/cm.xpt
REFDATA	CO	Comments	Special Purpose Domains	../transport/co.xpt

Table Name	Title for xpt file	Table Structure	Purpose	Table Keys	Data Set State (Final, Draft)
AE	Adverse Events SAS transport file	One record per adverse event per subject	Tabulation	STUDYID USUBJID AEDECOD AESTDTC	Final
CE	Clinical Events SAS transport file	One record per event per subject	Tabulation	STUDYID USUBJID CETERM CESTDTC	Final
CM	Concomitant Medications SAS transport file	One record per recorded medication occurrence or constant-dosing interval per subject	Tabulation	STUDYID USUBJID CMTRT CMSTDTC	Final
CO	Comments SAS transport file	One record per comment per subject	Tabulation	STUDYID USUBJID COSEQ	Final

Table Name	Release Date	Name of Standard	Version of Standard	Associated reference(s) in Standard	Comment
AE	November 12, 2008	CDISC-SDTM	3.1.2		
CE	November 12, 2008	CDISC-SDTM	3.1.2		
CM	November 12, 2008	CDISC-SDTM	3.1.2		
CO	November 12, 2008	CDISC-SDTM	3.1.2		

The reference_columns data set contains metadata at the column level for each domain defined in the reference_tables data set. In SDTM 3.1.2, this information is derived from the Implementation Guide provided by CDISC. Much like the reference_tables data set, reference_columns contains standard derived values, SAS Toolkit values, and user defined values. This data set is also a global standard and all column metadata at the study level are compared for compliance during the Toolkit validation process. Below is a listing of the variables currently defined in reference_columns.

Reference_columns.sas7bdat				
Variable Name	**Variable Length**	**Variable Label**	**Variable Type**	**Notes**
SASref	8	SASreferences sourcedata libref	Char	User supplied SAS libname statement (SAS default is REFMETA)
algorithm	1000	Computational Algorithm or Method	Char	User supplied standard algorithm for computed values
column	32	Column Name	Char	SDTM 3.1.2 supplied
comment	1000	Comment	Char	User supplied
core	10	Column Required or Optional	Char	SDTM 3.1.2 supplied
displayformat	32	Display Format	Char	User/Standard supplied
label	200	Column Description	Char	SDTM 3.1.2 supplied
length	8	Column Length	Num	User supplied
order	8	Column Order	Num	SDTM 3.1.2 supplied
origin	40	Column Origin	Char	User supplied (SAS default)
qualifiers	200	Column qualifiers (space delimited)	Char	
role	200	Column Role	Char	SDTM 3.1.2 supplied
standard	20	Name of Standard	Char	SDTM 3.1.2 supplied
standardref	200	Associated reference(s) in	Char	User supplied (SAS default)

		Standard		
standardversion	20	Version of Standard	Char	SDTM 3.1.2 supplied
table	32	Table Name	Char	SDTM 3.1.2 supplied
term	80	Controlled Term of Format in Standard	Char	SDTM 3.1.2 supplied
type	1	Column Type	Char	SDTM 3.1.2 supplied
xmlcodelist	32	SAS Format/XML Codelist	Char	SDTM 3.1.2 supplied
xmldatatype	8	XML Data Type	Char	SDTM 3.1.2 supplied

The source_columns data set stores the column metadata at the study level and is compared to the global reference_columns data set during the Toolkit validation process to verify compliance. At the study level, for every domain listed in source_tables there should be a corresponding group of column records for that domain in source_columns. Some content of reference_columns is shown below in Figure2.

Figure 2 Reference_columns.sas7bdat

SASreferences sourcedata libref	Table Name	Column Name	Column Description	Column Order	Column Type	Column Length	Display Format	XML Data Type	SAS Format/XML Codelist
REFDATA	AE	STUDYID	Study Identifier	1	C	40		text	
REFDATA	AE	DOMAIN	Domain Abbreviation	2	C	8		text	
REFDATA	AE	USUBJID	Unique Subject Identifier	3	C	40		text	
REFDATA	AE	AESEQ	Sequence Number	4	N	8		integer	
REFDATA	AE	AEGRPID	Group ID	5	C	40		text	
REFDATA	AE	AEREFID	Reference ID	6	C	40		text	
REFDATA	AE	AESPID	Sponsor-Defined Identifier	7	C	40		text	

Column Name	Column Required or Optional	Column Origin	Column Role	Controlled Term or Format in Standard	Computational Algorithm or Method	Column qualifiers (space delimited)	Name of Standard	Version of Standard
STUDYID	Req		Identifier			UPPERCASE	CDISC-SDTM	3.1.2
DOMAIN	Req		Identifier	AE		UPPERCASE	CDISC-SDTM	3.1.2
USUBJID	Req		Identifier			UPPERCASE	CDISC-SDTM	3.1.2
AESEQ	Req		Identifier				CDISC-SDTM	3.1.2
AEGRPID	Perm		Identifier			UPPERCASE	CDISC-SDTM	3.1.2
AEREFID	Perm		Identifier			UPPERCASE	CDISC-SDTM	3.1.2
AESPID	Perm		Identifier			UPPERCASE	CDISC-SDTM	3.1.2

Column Name	Associated reference(s) in Standard	Comment
STUDYID	SDTM2.2.4	Unique identifier for a study.
DOMAIN	SDTM2.2.4,SDTMIG4.1.2.2,AppendixC2	Two-character abbreviation for the domain.
USUBJID	SDTM2.2.4,SDTMIG4.1.2.3	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.
AESEQ	SDTM2.2.4	Sequence Number given to ensure uniqueness of subject records within a domain. May be any valid number.
AEGRPID	SDTM2.2.4	Used to tie together a block of related records in a single domain for a subject.
AEREFID	SDTM2.2.4	Internal or external identifier such as a serial number on an SAE reporting form
AESPID	SDTM2.2.4	Sponsor-defined identifier. It may be pre-printed on the CRF as an explicit line identifier or defined in the sponsor's operational database. Example: Line number on an Adverse Events page.

In the SDTM model some of the metadata is not provided and needs to be created by the user. A good example of this is the lengths of each of the variables in the domains. SAS has included lengths for the domain level variables. Users should review these values. SAS ships its interpretation of the SDTM model with the Toolkit.

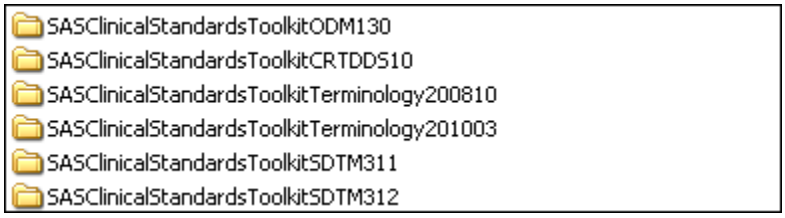
The reference_tables data set currently contains 32 domains (rows) of data. For each domain in reference_table, 15 columns of metadata are collected for each domain. The reference_column data set currently contains 723 rows of column metadata for the 32 domains listed in the reference_tables data set. For each column observation, 20 columns of metadata are collected for each observation. The reference_tables and reference_columns data sets are the foundation for all of the models included with the Toolkit as well as any custom models designed by the user.

Since the reference_tables and reference_columns data sets are global standards, they should change very little after being created and finalized. Additional data can be added as the model grows.

Clinical Toolkit 1.3 Concepts

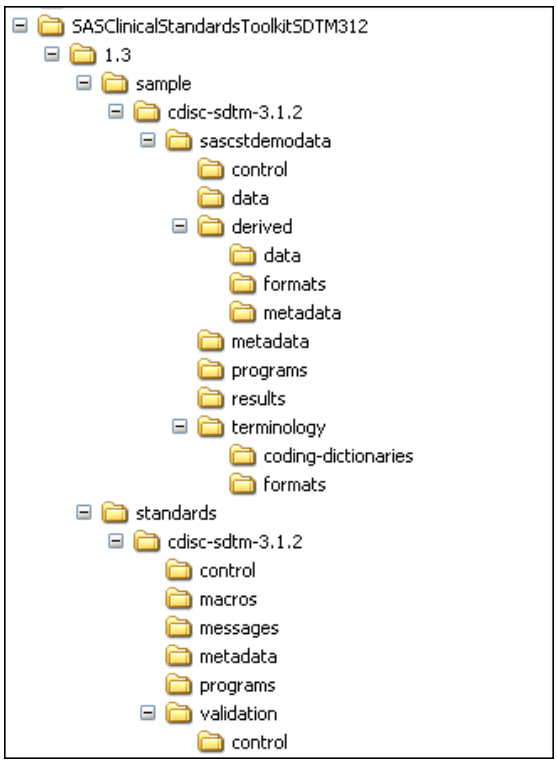
The Toolkit is installed in two areas: ***!SASROOT*** and **C:*/CSTGlobalLibrary***. In ***!SASROOT,*** directories designed specifically for each of the standards represented within Toolkit are created. The following 6 directories in Figure 3 are created.

Figure 3 Directories Created at !SASROOT



Further breakdown of the SASClinicalStandardsToolkitSDTM312 directory reveals the following structure in Figure 4.

Figure 4 Further Breakdown



The ***1.3*** directory represents the version of Toolkit (V1.3). The ***/sample*** directory contains a sample study and metadata that is shipped with the product. The SAS supplied IQ/OQ process uses this directory to test installation. The ***/standards*** directory contains the SAS supplied information regarding the SDTM 3.1.2 model. Nothing should ever be modified in the standards directory, it is used to populate the CSTGlobalLibrary and is overwritten whenever the Toolkit is re-installed or updated.

In the ***!SASROOT/SASFoundation/9.2*** directory a new ***/cstframework/sasmacro*** directory is created and contains the SAS macros that are used within the Toolkit. Below in Figure 5 is a partial listing of these macros.

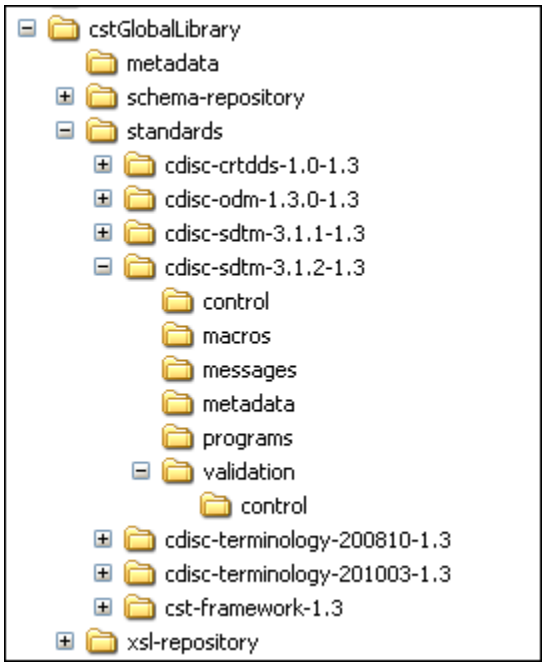
Figure 5 CST Framework Macros

cstcheck_column.sas	23 KB	SAS Program	8/20/2010 9:13 AM
cstcheck_columncompare.sas	26 KB	SAS Program	8/20/2010 9:13 AM
cstcheck_comparedomains.sas	24 KB	SAS Program	8/20/2010 9:13 AM
cstcheck_dsmismatch.sas	13 KB	SAS Program	8/20/2010 9:13 AM
cstcheck_metamismatch.sas	17 KB	SAS Program	8/20/2010 9:13 AM
cstcheck_notconsistent.sas	24 KB	SAS Program	8/20/2010 9:13 AM
cstcheck_notimplemented.sas	2 KB	SAS Program	8/20/2010 9:13 AM
cstcheck_notincodelist.sas	59 KB	SAS Program	8/20/2010 9:13 AM
cstcheck_notsorted.sas	13 KB	SAS Program	8/20/2010 9:13 AM
cstcheck_notunique.sas	76 KB	SAS Program	8/20/2010 9:13 AM

These macros are readily available to the user. SAS discourages changing these files since they can be overwritten or deprecated with the next installation of the Toolkit. It is recommended the user create their own macros or copy these to a separate area to make changes and then access this folder as an additional macro library. Any of these macros can be used as a template if the user needs to create their own macro. Review of the different styles of Toolkit macros will help the user decide which one to use. Styles include self contained macros that require no metadata from the validation_master/control data set, macros that compare only tables or columns against each other, and macros that rely on code logic supplied by the user

The ***C:/CSTGlobalLibrary*** is created during the default installation of the Toolkit and contains copies of the information stored in ***!SASROOT/SASClinicalStandardsToolkitSDTM312/1.3/standards***. Figure 6 is a screenshot of the CSTGlobalLibrary.

Figure 6 cstGlobalLibrary Directory



The cdisc-sdtm-3.1.2-1.3 directory is identical to the structure represented in ***!SASROOT:/SASClinical StandardsToolkitSDTM312/1.3/standards/cdisc-sdtm-3.1.2***. To the Toolkit this is the global standard. Stored in the ***/macros*** directory are standard specific macros. These are different than the macros stored under the SAS Foundation framework directory. The framework directory macros can be used across all standards whereas the standard specific macros are designed to be used within the standard. The ***/messages*** directory contains the validation messages data set and will be discussed later. The ***/metadata*** directory contains the reference_tables and the reference_columns data sets, and the ***/validation/control*** directory contains the validation_master data set. The ***/programs*** directory contains several properties files and cimport code that is used during installation. The property files set up various global macro variables for the Toolkit.

Validating SDTM 3.1.2 in Clinical Toolkit 1.3

Validation is achieved in the Toolkit through a group of specialized macros designed to handle common conditions such as missing data, data inconsistencies, etc. In addition, the user may also create their own macros. The validation rules for SDTM are provided through a collection of checks supplied through several sources. Currently these sources include WebSDM, OpenCDISC , and SAS. As with the SDTM reference metadata data sets mentioned earlier, SAS has developed a group of metadata data sets to handle the validation checks from these various sources. A validation_master data set is created that contains all of the currently available SDTM 3.1.2 checks. This data set contains 247 checks and for each check contains 21 columns of metadata. In addition to the validation_master data set, SAS collects the text or verbiage for each check in a separate messages data set. Just like the reference_tables and reference_columns data set, the validation_master and messages data sets are considered global standards used by multiple studies and should change very little. At the study level there is a validation_control data set that can be an exact copy or a subset of the valudation_master data set. If there are any modifications needed at the study level they are done here. The validation_control data set is the “control center” of the Toolkit validation process. It determines which checks, tables (domains), and columns will be submitted for validation. Below is a listing of the contents of the validation_master/control data set.

Validation_master.sas7ddat (validation_control)				
Column Name	Column Length	Column Label	Column Type	Notes
checkid	8	Validation check identifier	Char	Provided by SAS, and used to access the messages data set
standard	20	Standard model	Char	Provided from the standard
standardversion	20	Standard version	Char	Provided from the standard
checksource	40	Source of check	Char	WebSDM, OpenCDISC, SAS
sourceid	8	Record identifier used by checksource	Char	Provided by WebSDM, OpenCDISC, SAS
checkseverity	40	Severity of check	Char	Warning, Error, Note, etc.
checktype	20	Category of check	Char	Metadata, Date, etc.
codesource	32	SAS macro module name	Char	SAS Toolkit macro that handles the check
usesourcemetadata	1	Check should use source metadata	Char	Checks against the source_tables/columns data set if Y, otherwise use the reference_tables/columns
tablescope	200	Domains/data sets to which check applies	Char	The domains or tables to be checked.
columnscope	200	Columns to which check applies	Char	The columns within the domains or tables to be checked
codelogic	2000	Code logic used within code	Char	SAS code, if needed, that contains the actual check logic. It is submitted in the codesource macro above. Not all check macros allow codelogic.
lookuptype	20	Lookup standard type	Char	
lookupsource	32	SAS format name	Char	
standardref	200	Reference in standard supporting check	Char	Where to find external information about the check
reportingcolumns	200	Column values to be reported	Char	Columns to report in addition to those in columnscope.
checkstatus	8	Current check status	Num	Active vs inactive
reportall	1	Report all possible records in error	Char	Determines is all or single occurrences of a check ar reported.
uniqueid	48	Unique check identifier	Char	Unique number generated by SAS and used by CDI
codetype	8	Code logic type	Num	
comment	200	Check comment	Char	Comments about the validation

Following is a condensed example of the contents of validation_master/control. There are several columns the Toolkit user may find useful when validating data. The SAS macro module name (codesource) identifies which SAS validation macro (stored in **!SASROOT/SASFoundation/ cstframework/macros**) is used for the specified check. The Domains/data sets to which check applies

(tablesource) determines which tables are submitted for validation. In the example below the keyword _ALL_ is used to signify that all of the available domains in the study will be validated. Toolkit knows which tables to use when _ALL_ is specified buy deriving the list of tables/domains from the source_tables data set. The Columns to which check applies (columnsource) contains the variables from the tables that will be subject to validation. In the example below the **DTC is read by the macro as <domain>DTC or AEDTC for the AE domain. In the example below 3 columns are being validated for the AE domain, AEDTC, AESTDTC, and AEEN. These two fields in the validation_control data set can be modified by the user, for example if the user wanted to limit the check to AE only, table scope would be AE instead of _ALL_. Columnscope could remain as is, or be modified as AEDTC, or contain another date value to validate that was not in the original list. When present, the Code logic used within code (codelogic) is a very important key piece of information for the validation process. This field contains the actual logic, or in the example here, another call to a macro that validates ISO8601 date structures.

Figure 7 Validation_master/control Data Set

	Validation check identifier	Standard model	Source of check	Record identifier used by checksource	Severity of check	Category of check	SAS macro module name
28	SDTM0101	CDISC-SDTM	WebSDM	IR5002	Warning	Date	cstcheck_column

	Validation check identifier	Domains/data sets to which check applies	Columns to which check applies	Code logic used within code	Lookup standard type	SAS format r
28	SDTM0101	_ALL_	**DTC+**STDTC+**EN	%sdtmutil_iso8601(_cstString=&_cstColu if not _cstISOisValid then do; _cstError=1; _cstMsgParm1=_cstISOInfo; end;		

Below is a more readable form of the codelogic value which is substituted as &_csCodeLogic in the validation check macros:

```
%sdtmutil_iso8601(_cstString=&_cstColumn.);
 if not _cstISOisValid then
do;
  _cstError=1;
  _cstMsgParm1= _cstISOInfo;
end;
```

This codelogic is executed within the cstcheck_column (codesource) macro. Not all Toolkit macros allow the use of the codelogic column, but those that do allow the user to modify this field to make the check work for them. This is particularly useful when customizing the checks or adding additional check logic to an existing check.

Clinical data is validated at the study level. In the default installation of the Toolkit this is located in ***!SASROOT:/SASClinicalStandardsToolkitSDTM312/1.3/sample/cdisc-sdtm-3.1.2/sascstdemodata***. A user's study location will be different. In order for the Toolkit to work, certain data sets are required in the study location. There can be many studies and all will either be duplicates or subsets of the global standard. The source_tables and source_columns data sets contain the same column names as reference_tables and reference_columns. These source data sets handle study data and allow the user to customize the SDTM model at the study level. For example, a study might not have all of the domains listed in the global standard (reference_tables). In this case source_tables would only contain those domains needed by the study. The same is true for the source_columns data set, it will only contain those domains and columns that are needed for the study. In the end, the user will have one (or more) source_table and source_column data set for EACH study, and one reference_table and reference_column data set as the global standard.

In the Toolkit validation process, the source_tables data set determines which domains will undergo validation checks. The source_columns data set determines which columns are provided to the validation process. Source_columns should match source_tables, for each table in source_tables there should be a corresponding table in source_columns.

To manage all of the metadata relationships within the Toolkit, a special data set is created called sasreferences. This data set contains pointers to the reference library, the format catalogs, controlled terminology, study data, and any other information a Toolkit process may require. The sasreferences data set has its own standard structure that is required by the Toolkit. The sasreferences data set can be created either as a permanent data set or as a temporary work file. Most users prefer the work file approach which can be modified in the code if needed and requires generation of the sasreferences data set each time a validation is submitted. More detailed information for the sasreferences data set is available from the SAS Clinical Standards Toolkit 1.3 User Manual. Figure 8 is a partial screen shot of the creation of a sasreferences data set that is created in the work directory.

Figure 8 Sasreferences.sas7bdat Generation

```
proc sql;
  insert into work.sasreferences
  values ("CST-FRAMEWORK"     "1.2"    "messages"          ""                  "messages"  "libref"
  values ("CDISC-SDTM"        "3.1.2"  "autocall"          ""                  "sdtmauto"  "fileref"
  values ("CDISC-SDTM"        "3.1.2"  "control"           "reference"         "cntl_s"    "libref"
  values ("CDISC-SDTM"        "3.1.2"  "control"           "validation"        "cntl_v"    "libref"
  values ("CDISC-SDTM"        "3.1.2"  "fmtsearch"         ""                  "srcfmt"    "libref"
  values ("CDISC-SDTM"        "3.1.2"  "messages"          ""                  "sdtmmsg"   "libref"
  values ("CDISC-SDTM"        "3.1.2"  "properties"        "initialize"        "inprop"    "fileref"
  values ("CDISC-SDTM"        "3.1.2"  "properties"        "validation"        "valprop"   "fileref"
  values ("CDISC-SDTM"        "3.1.2"  "referencecontrol"  "validation"        "refcntl"   "libref"
```

```
""                                                  1 ""                              "")
""                                                  1 ""                              "")
"&workpath"                                         1 "sasreferences.sas7bdat"        "")
"&studyRootPath/control"                            2 "validation_control.sas7bdat"   "")
"&studyRootPath/terminology/formats"                1 "formats.sas7bcat"              "")
""                                                  2 ""                              "")
""                                                  1 "initialize.properties"         "")
"&studyRootPath/programs"                           2 "validation.properties"         "")
""                                                    ""                              "")
```

Since the Toolkit is designed around the Clinical Data Integration (CDI) solution, there are no user interface programs supplied with base SAS as the interface to the Toolkit is through CDI. In order to access Toolkit macros a driver program is needed to supply the required information. Several driver programs are provided in the sample area under the ***/programs*** directory and can be used as user templates. In addition, users can also create their own user interface to wrapper around the Toolkit macros. To run the validation process, the driver program validate_data.sas for SDTM 3.1.2 located in ***!SASROOT:/SASClinicalStandardsToolkitSDTM312/1.3/sample/cdisc-sdtm-3.1.2/sascstdemodata/programs*** is used. Each standard supplied by SAS in the Toolkit has a set of driver programs located in the sample study areas that can be submitted for processiong and modified by the user.

Reporting in Clinical Toolkit 1.3

After the process is submitted, the Toolkit generates both a validation results data set and a validation metrics data set. These data sets contain information about the validation submission and are available to the user for customized reporting. The validation_results data set contains observations about each failure for the checks, checks that were not run, checks that generated errors due to insufficient metadata, and checks that ran successfully. The validation_metrics data set contains run time information about the checks, how many were run, number of observations, number of data failures, and successes. For version 1.3, the Toolkit provides two types of reports out of the box, a validation report and a validation metadata report. Figure 9 displays the validation report that was run against a sample Toolkit SDTM 3.1.1 study validation submission. The driver program used to create this report is create_report.sas. As can be seen, it lists the checks for SDTM0011 and in this particular example it reported that the SUPPAE data set was not found in the reference standard. This means the reference_tables data set for SDTM 3.1.1 did not contain the domain SUPPAE, but this domain was present at the study level in the source_tables data set.

Figure 9 Example of Validation Report (3.1.1)

SAS Clinical Standards Toolkit 1.3
CDISC-SDTM 3.1.1 VALIDATION

Process Results, CheckID: SDTM0011

Description: Identifies a column that was described in the domain description but not included in the SAS dataset for that domain
Check scope: (Tables) _ALL_, (Columns)
Source: WebSDM (IR4250)
Validation check macro: cstcheck_metamismatch, using source metadata

Check Invocation	Seq #	Source Data	Result Identifier	Message	Severity	Problem Detected?	Actual Value	Keys
1	1	SRCDATA.SUPPAE	CST0025	Data set not found in reference standard - compliance not assessed	Warning: Check incomplete	Yes		
2	1	SRCDATA.SUPPAE	CST0025	Data set not found in reference standard - compliance not assessed	Warning: Check incomplete	Yes		

The validation metadata report contains information surrounding each validation check. It gives the user all of the information about that validation check contained in the Toolkit. Any metadata data set that contains validation checks is used to populate this table. Figure 10 shows a small sample of the metadata that is available from the report. This report is generated by running the create_metadata.sas driver program available in the sample study area.

Figure 10 Example of Validation Metadata Report (3.1.1)

SAS Clinical Standards Toolkit 1.3
CDISC-SDTM 3.1.1 Validation Check Metadata

Check Overview

Validation Check Identifier	Version of Standard	Source of Check	Record Identifier used by Check Source	Rule Description from Checksource	Severity of Check	Domains/Data Sets to which Check Applies	Columns to which Check Applies
SDTM0001	***	WebSDM	IR4000	Identifies domain table that has zero rows and therefore contains no data	Warning	_ALL_	
	***	Janus	IR4000	Identifies domain table that has zero rows and therefore contains no data	Note	_ALL_	
SDTM0002	***	JanusFR	SAS0017	A load of data into JANUS requires that the DM, DS and EX domains be submitted for each study to be loaded.	Error	DM+DS+EX	
SDTM0003	***	WebSDM	SAS0018	WebSDM and the SDTM model require only the DM domain be present.	Error	DM	
SDTM0004	***	SAS	SAS0033	Source metadata includes domain data set not found in reference metadata	Note	_ALL_	

The report programs are available to the users as open SAS source code and can be customized or modified by the user. The reports are generated using PROC REPORT and the user has the options of creating HTML, PDF, or RTF output from the provided reporting macros.

Conclusion

SAS Clinical Toolkit 1.3 is a tool designed to implement, manage, and validate a company's standard. Its use of open SAS code in the form of macros, allows users to customize and model the Toolkit to fit their needs. This paper addresses the most common bits of metadata that may need to be modified or even customized by the user in order to validate their instance of SDTM. Looking forward to ADaM, the ability to write customized checks looks like it will be even more important than for SDTM. The data set structure is more flexible and users may add any number of new columns, and may wish to add any number of custom checks related to those columns.

Further Reading

CDISC SDTM3.1.2 Implementation Guide available at www.cdisc.org
SAS Clinical Standards Toolkit V1.3 User's Guide

Contact Information

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

Gene Lightfoot
SAS Institute Inc.
SAS Campus Drive S2074
Cary, North Carolina 27513 USA
919-677-8000
919-531-0700 (Fax)
gene.lightfoot@sas.com
www.sas.com