

The SAS® Clinical Standards Toolkit

Dave Smith, SAS UK

INTRODUCTION

With the rapid rate of development of CDISC standards, a need has arisen for a tool to create and validate different CDISC structures that can rapidly respond to model changes and be flexible enough to manage individual company variations. This presentation will introduce the audience to the SAS® Clinical Standards Toolkit, which is designed to supersede Proc CDISC and address these changing needs. The metadata needed to generate output will be described and the toolkit will be demonstrated.

CONTEXT FOR THE CLINICAL STANDARDS TOOLKIT

BUSINESS REQUIREMENTS

The clinical standards toolkit is designed to enable companies to manage the demands of implementing clinical standards within their organisations.

The typical task is to create objects that adequately meet the relevant standard (for example a define.xml output). The toolkit needs to be able to cope with changing standards and different interpretations of those standards. Therefore the toolkit has been designed to allow different standards or variations of standards to be defined and then used by the toolkit.

Perhaps the most significant reason for employing the toolkit is to support submissions to the FDA. This requires the creation of valid SDTM structured datasets and a define.xml. It also requires that the data passes the JANUS checks which indicate that the data will successfully load into the FDA's JANUS warehouse. For this reason the SAS Clinical Standards Toolkit includes these checks. Organisations are also able to add their own checks to enhance their own processes.

CONTEXT WITH OTHER SAS® SOLUTIONS

The SAS Clinical Standards Toolkit is an enabling technology which exists as a set of macros and metadata. The macros are called through program code and metadata is passed to it through SAS datasets. These macros are also designed to be called by SAS® Clinical Data Integration, which will automatically populate metadata datasets and perform macro calls. Solutions such as SAS® Drug Development or other advanced analytical applications may also make use of the toolkit as an enabling technology.

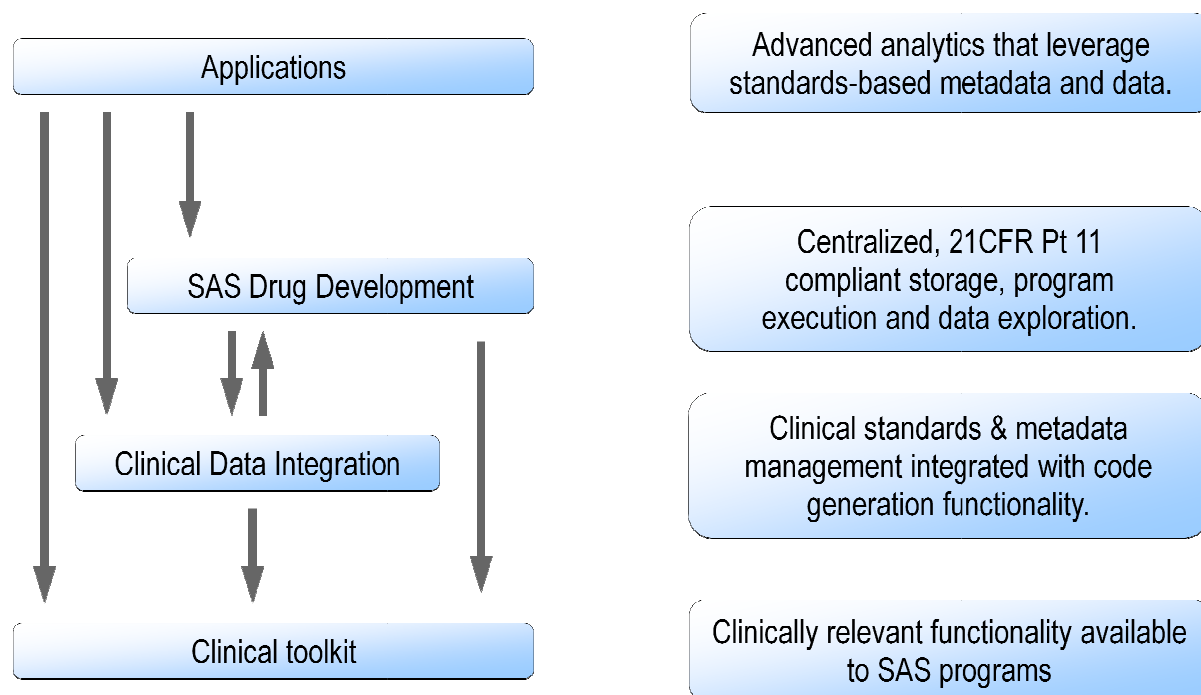


Figure 1 – The context of the SAS Clinical Standards Toolkit with other SAS solutions

CDISC TECHNICAL CHALLENGES

The main technical challenge associated with creating a system to manage standards such as CDISC is the speed at which standards are released compared to the length of clinical trial programs. It is quite common for standards to be updated during the life of a clinical project, and there may be compelling reasons why the new standard might be adopted for some if not all studies within the clinical project. Variations in the implementation of CDISC standards might cause difficulties, such as SDTM +/- where study or therapeutic area specific items are included.

Another technical challenge is that converting data to SDTM can involve complex restructuring, particularly where the data is from legacy studies. This is hard to document and so a flexible framework for creating define.xml is needed.

Finally the increasing adoption of CDISC means that many organisations need to scale up their capabilities in this area very rapidly. This particularly affects CROs, who will by definition be working from many different starting structures.

THE SAS® CLINICAL STANDARDS TOOLKIT

DESCRIPTION

The SAS Clinical Standards Toolkit consists of a set of macros and associated metadata designed to perform the tasks associated with managing clinical standards. The initial focus is around the validation of standards adherence and the creation of CRT-DDS (define.xml). The toolkit is designed to be an extensible framework, with the ability to add support for new models or controlled terminology without changing the core framework.

To perform each major task (such as validating SDTM structure) with the SAS Clinical Standards Toolkit requires the definition of inputs, controls and outputs.

Inputs:	Study data Study metadata
Controls:	Definitions of the standard being applied Physical location of standards metadata Physical location for diagnostic messages Specific metadata for that task (for example a list of validation checks to apply)
Outputs:	Results of the task Metrics about the task execution

PRACTICAL GUIDE TO USING THE MACROS

SDTM VALIDATION

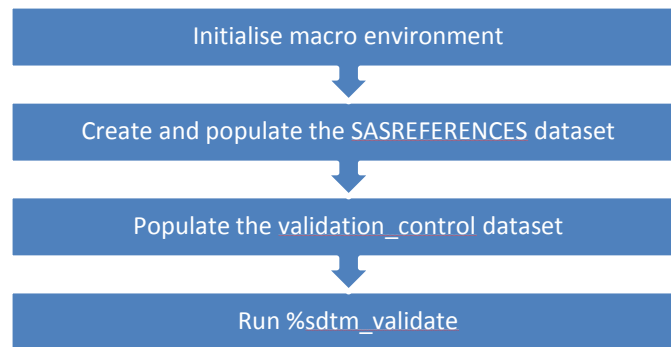


Figure 2 - Typical Calling Program – SDTM validation

To perform any task using the SAS Clinical Standards Toolkit there will be a calling program to manage assignment of options. The first step is always to initialise the macro environment, both for the toolkit framework and the individual standard. The next step will be to either create or refer to the SASREFERENCES dataset. This is the dataset that contains all of the metadata references that the toolkit macro needs to perform the task.

Contents of the SASREFERENCES dataset – SDTM validation

- Location of reference metadata
- Location of study data
- Location of study metadata (Source_columns, Source_tables)
- Locations of formats, codelists, dictionaries
- Location of validation_control dataset
- Output destination datasets

The third step is to populate the validation_control dataset. This is the dataset that defines which validation checks are to be applied during this run of the validation macro, and is typically created by sub-setting the master list of available checks.

In SAS® 9.2 this is found at C:\Program Files\SAS\SASClinicalStandardsToolkitSDTM311\9.2\standards\cdisc-sdtm-3.1.1\validation\control\ validation_master.sas7bdat.

CRT-DDS CREATION

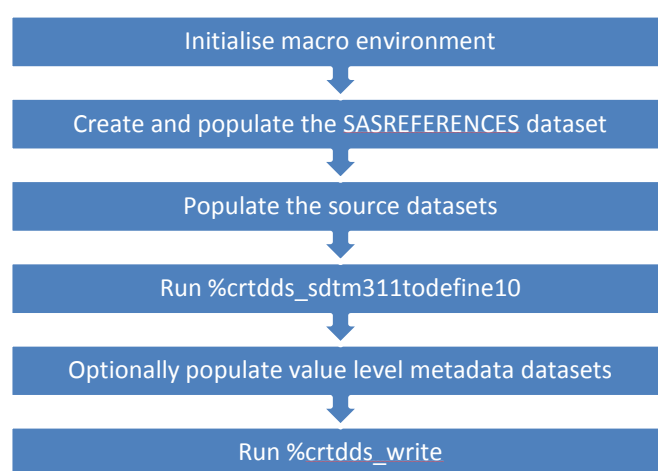


Figure 3 - Typical Calling Program – CRT-DDS Creation

As with the validation of SDTM there will be a calling program to manage assignment of options. The first step is always to initialise the macro environment, both for the toolkit framework and the individual standard. The next step will be to either create or refer to the SASREFERENCES dataset. This is the dataset that contains all of the metadata references that the toolkit macro needs to perform the task.

SASREFERENCES dataset – CRT-DDS

- Where do the messages go?
- Where are the autocall macros?
- Where is the source data? – for CRT-DDS these are metadata datasets
- Where is the CRT-DDS file to go?
- Where is the style sheet?

The next stage is to populate the source datasets that define the metadata about the study, tables and columns.

Examples of these datasets are given below.

Obs	SASREF	StudyName	StudyDescription	ProtocolName	DefineDocumentname
1	1	H2Q-MC-LZZT	H2Q-MC-LZZT	H2Q-MC-LZZT	Define.xml

Figure 4 - source_study dataset

Obs	SAS references source data libref	Table Name	Table Label	Observation Class within Standard	(Relative) path to xpt file	Title for xpt file	Table Structure	Purpose	Table Keys
1	1	CM	Concomitant Medications	Interventions			One record per medication intervention episode per subject	Tabulation	STUDYID USUBJID CMTRT CMSTDTC
2	1	DM	Demographics	Special Purpose			One record per subject	Tabulation	STUDYID USUBJID
3	1	QS	Questionnaires	Findings			One record per question per time point per visit per subject	Tabulation	STUDYID USUBJID QSSEQ QSTESTCD VISITNUM QSTPTNUM
Obs	Date Set State (Final, Draft)	Release Date	Name of Standard		Version of Standard		Associated reference(s) in Standard		Comment
1			CDISC-SDTM		3.1.1				
2			CDISC-SDTM		3.1.1				
3			CDISC-SDTM		3.1.1				

Figure 5 - source_tables dataset

Obs	SAS references source data libref	Table Name	Column Name	Column Description	Column Order	Column Type	Column Length	Display Format	XML Data Type	SAS Format/XML Codelist
1	1	CM	STUDYID	Study Identifier	1	C	40		text	
2	1	CM	DOMAIN	Domain Abbreviation	2	C	8		text	
3	1	CM	USUBJID	Unique Subject Identifier	3	C	40		text	
4	1	CM	CMSEQ	Sequence Number	4	N	8		integer	
5	1	CM	CMGRPID	Group ID	5	C	40		text	
Obs	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	Associated reference(s) in Standard	
1	Req	CRF	Identifier			UPPERCASE	CDISC-SDTM	3.1.1		
2	Req	Derived	Identifier	4*CM		UPPERCASE	CDISC-SDTM	3.1.1		
3	Req	Sponsor Defined	Identifier			UPPERCASE	CDISC-SDTM	3.1.1		
4	Req	CRF or Derived	Identifier				CDISC-SDTM	3.1.1		
5	Perm	Sponsor Defined	Identifier			UPPERCASE	CDISC-SDTM	3.1.1		

Figure 6 - source_columns dataset

These datasets contain a large amount of CDISC standards metadata which will be used in the CRT-DDS output.

Once the source datasets have been prepared the %crtdds_sd311todef10 macro is run to convert them into the more detailed metadata datasets required by the %crtdds_write macro. If formats are applied to the input datasets then the formats catalog will be read to produce the codelists datasets.

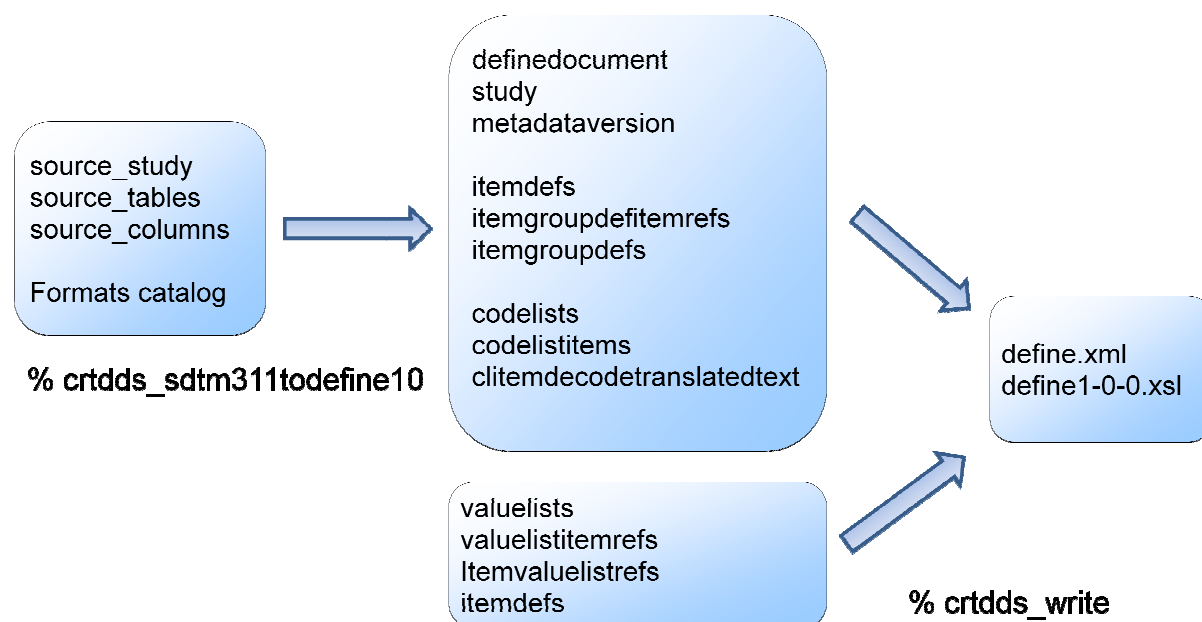


Figure 7 - Datasets feeding the CRT-DDS creation process

At this stage additional metadata datasets can be created to define other components of the CRT-DDS such as value-level metadata, Forms, Events, Annotated CRFs, Computation methods, Range Checks and Measurement Units. In the example above value level metadata is being added.

Finally the %crtdds_write macro is called, which will create the CRT-DDS file and optionally create a copy of the stylesheet define1-0-0.xsl.

AUTOMATION THROUGH SAS® CLINICAL DATA INTEGRATION

The SAS Clinical Standards Toolkit macros are a key enabling technology within SAS Clinical Data Integration. This is a separate product built upon SAS® Data Integration Server and is designed to enable the additional processes needed to create mappings between source formats and standards such as CDISC SDTM. The metadata needed to perform tasks such as CRT-DDS creation is available within SAS Clinical Data Integration, so this can be performed through a wizard interface.

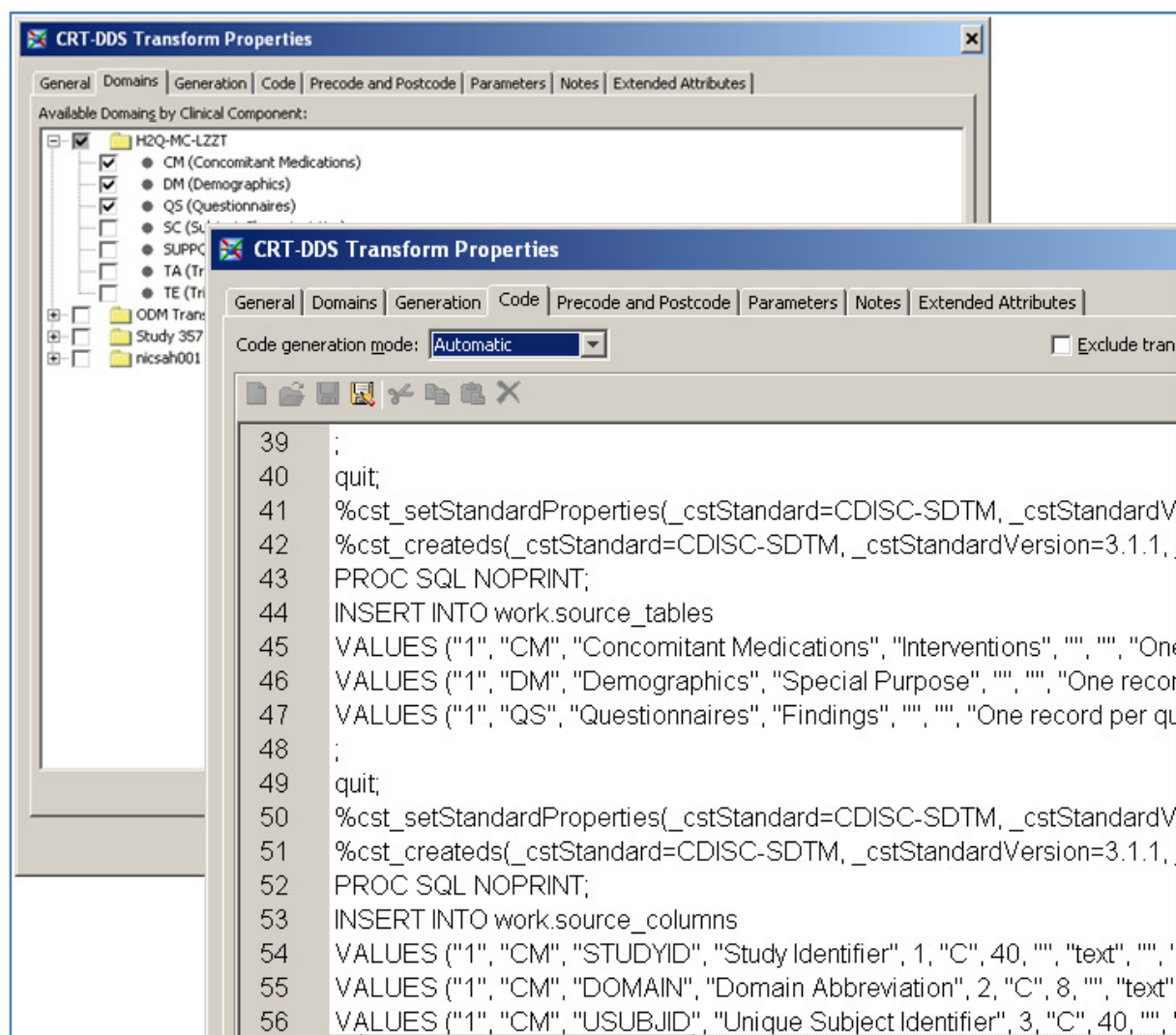


Figure 8 - Clinical Data Integration CRT-DDS transform

In the figure above the CRT-DDS transform is being used to create a CRT-DDS file. The first screen shows selection of the domains to be included in the CRT-DDS and the second screen shows the code that has been generated. This code includes the creation of the metadata datasets and Clinical Standards Toolkit macro calls needed to create the CRT-DDS file.

CONCLUSIONS

The SAS Clinical Standards Toolkit provides a flexible, extensible framework for the creation and validation of clinical standards. Non-SAS files such as XML formats can be created and validated. Customised data standards and validation checks to support those data standards can also be created. The toolkit should therefore provide the tools needed by the pharmaceutical industry to cope with the ever changing demands of CDISC and other standards.

SAS is committed to supporting existing and emerging industry standards through the provision of technical solutions and by contributing expertise and sponsorship to organisations such as CDISC.

ACKNOWLEDGEMENTS

This paper would not have been possible without the support of Bill J Gibson, Pete Villiers, Julie Maddox and Steve Smith in SAS Global Headquarters, Cary, N Carolina, USA.

CONTACT INFORMATION

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

Dave Smith
SAS UK
Wittington House
Henley Road
Medmenham
Marlow
Bucks
SL7 2EB

Work Phone: 01628 404379
Fax: 01628 490550
Email: david.smith@suk.sas.com
Web: <http://www.sas.com>

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 trademarks of their respective companies. Copyright ©2009, SAS Institute Inc. All rights reserved.