PhUSE 2017

Paper SI10

Define.xml tools supporting SEND/SDTM data process - Using SAS® Clinical Standards Toolkit

Kirsten Walther Langendorf, S-cubed, Copenhagen, Denmark

ABSTRACT

Creating SEND (or SDTM) data often involves multiple data stakeholders (CRO, LABs, and Pharma Company) which requires clear and understandable data specifications. This paper demonstrates a set of define.xml tools developed for a customer (Pharma Company). The tools create and compare define.xml and are used to specify what data to receive, description of the actual data received and a way to compare the two. The tools are developed using SAS® 9.4 and SAS Clinical Standards Toolkit (CST) 1.7. The paper will provide an overview of how the define.xml tools are used in a SEND data process. The three define.xml tools will be described and demonstrated. To support the maintenance of the CDISC data standards and the CDISC controlled terminology, two tools have been developed to update the SAS Clinical Toolkit's Global Library of standards. These will also be described and demonstrated.

INTRODUCTION

Define.xml is often a file that is made before data is submitted to the Regulatory Agencies and intended for the data reviewer. However, Define.xml can be used throughout the data collection and reporting process as a tool to:

- Provide specifications to a data provider, e.g. lab or CRO
- Provide ability to compare received data against specified
- Provide data description for internal users of the data, e.g. Data Reviewers and Statisticians
- Data archiving, serving as a 'table of contents' of a company's data

The tools described and demonstrated in this paper has been developed for a customer to support a SEND data process, but the tools are also applicable for SDTM.

There are three (3) define tools:

- **Spec-driven define**
  Based on (Excel) specification for domains, variables, controlled terminology (study/sponsor specific) and value-level metadata, a define.xml is created and a set of empty SEND/SDTM data sets. Only metadata is used and no study (patient/animal) data records are needed.

- **Data-driven define**
  Based on xpt files containing study data a define.xml is created. Reading the content of the data it creates the code lists (and subset code lists) and value-level metadata (VLMD). Additional VLMD can be added using Excel (template).

- **Compare define**
  Reads two define files and compares them on different levels, e.g. domains, variables, code lists, VLMD and provides the differences in a report.

To support the maintenance of the CDISC data standards and the controlled terminology two (2) tools have been developed:

- **Register new controlled terminology from NCI/CDISC**
  Based on the odm.xml file which can be downloaded from the NCI homepage, a new controlled terminology can be registered in the SAS CST Global Library (cstGlobalLibrary). This serves as the 'master' for controlled terminology and, code lists not specified in the study/sponsor specific, is drawn from this master when creating the spec-driven define.xml.

- **Register a new data model**
  Based on the (Excel) specification either from CDISC SHARE or a customized version of it (made by the customer) the model is registered in the SAS CST Global Library. This then serves as the master reference for the model used for a study.

In the following an overall process for using the tools will be described focusing on the SEND process. Then each of the tools will be described and demonstrated (screen shots). The tools have been developed using SAS

9.4 and SAS Clinical Standards Toolkit 1.7. The simple user interface (UI) is made using %WINDOW and %DISPLAY, which are standard functionalities in BASE SAS.

HIGH-LEVEL SEND DEFINE.XML PROCESS

Typically, a pre-clinical study is described in a study plan (protocol) which is sent to the CRO along with a set of data delivery agreements for them to conduct the study. Additional sub-laboratories can be involved in analyzing samples for the SEND data but the CRO will need all the results since they will deliver a study report of the SEND data to the pharmaceutical company.

The pharmaceutical company will send all their studies (including SEND data) to the FDA and the company needs to ensure consistency across the studies and that the studies follow the standards communicated in the Standardization Plan communicated to the FDA. To fulfill this requirement, the company typically defines a SEND master model that each of the studies must adhere to.

To ensure that the CRO follows the master model, the pharmaceutical company will send a (spec-driven) define.xml for the study to the CRO. The CRO delivers SEND data and a define.xml (CRO define.xml) or the pharmaceutical company makes one (data-driven) based on the data received, and compares the spec-driven against the data-driven (or the CRO define.xml) and gets a report of any differences. This process can occur for each data delivery agreed with the CRO and issues can be identified early in the study conduct phase, see Figure 1.

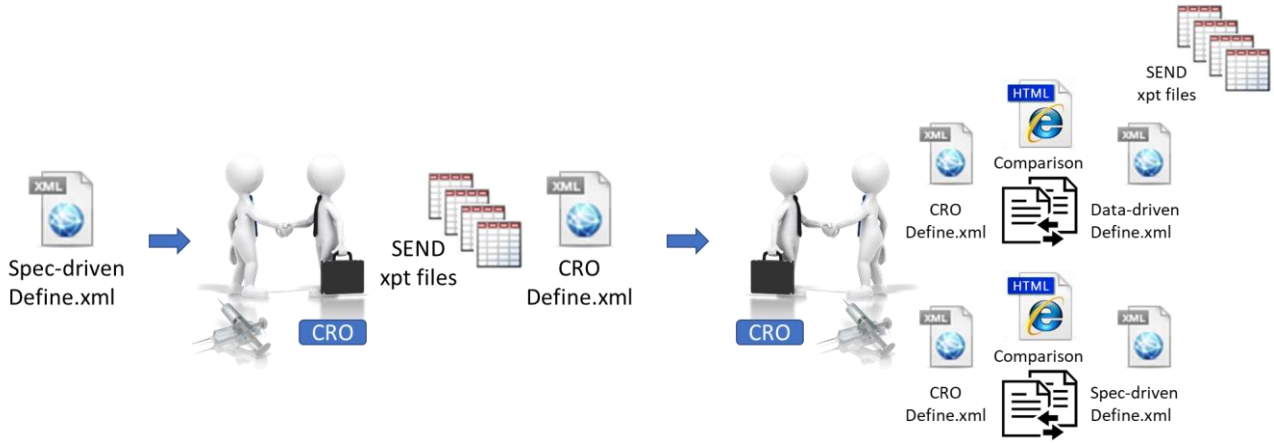


Figure 1 Define.xml high-level process

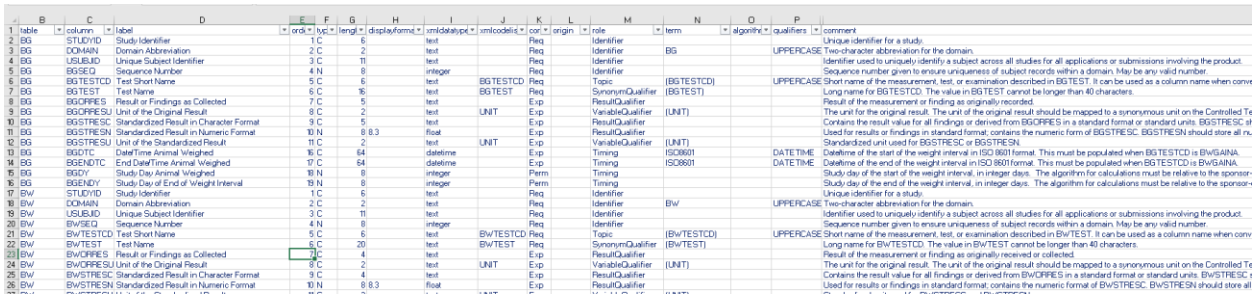
CREATING AND COMPARING DEFINE.XML

The creation of the define files (spec- and data-driven and compare) are based on the sample programs in the SAS Clinical Standards Toolkit sample library. These macros have been wrapped into a main-macro for each of the define file tools.

SPEC-DRIVEN DEFINE FILE

The spec-driven define will be based on an Excel-input file representing the tables(domains) and columns of the model used for the study, see Figure 2.

	B	C	D	E	F	G	H	I	
1	table	label	class	xmlpath	xmltitle	structure	purpose	keys	comment
2	BG	Body Weight Gains	Findings	/transportlbg.xpt	Body Weight Gains SAS transport file	One record per test per interval per subject	Tabulation	STUDYID USUBJID BGTESTCD BGDTC BGENDTC	Body weight gain is the act
3	BW	Body Weights	Findings	/transportlbw.xpt	Body Weights SAS transport file	One record per test per observation time per subject	Tabulation	STUDYID USUBJID BWTESTCD BWDTC	This domain captures body
4	CL	Clinical Observations	Findings	/transportlcl.xpt	Clinical Observations SAS transport file	One record per finding per observation time per subject or pool	Tabulation	STUDYID USUBJID CLTESTCD CLCAT CLORRES CLLOC CLDTC	This domain captures clinic
5	CO	Comments	Special Purpose Domains	/transportlco.xpt	Comments SAS transport file	One record per comment	Tabulation	STUDYID COSEQ	The Comments special-pur
6	DS	Disposition	Events	/transportlds.xpt	Disposition SAS transport file	One record per subject	Tabulation	STUDYID USUBJID	The disposition dataset pro
7	EX	Exposure	Interventions	/transportlex.xpt	Exposure SAS transport file	One record per constant dosing interval per treatment per subject or pool	Tabulation	STUDYID USUBJID EXTRT EXSTDTC	The Exposure Domain Mod
8	LB	Laboratory Tests Results	Findings	/transportllb.xpt	Laboratory Tests Results SAS transport file	One record per test per specimen per observation time per subject or pool	Tabulation	STUDYID USUBJID LBTESTCD LBSPEC LBDTC LBTPTNUM	This domain captures labor
9	MA	Macroscopic Findings	Findings	/transportlma.xpt	Macroscopic Findings SAS transport file	One record per finding per specimen per subject	Tabulation	STUDYID USUBJID MATESTCD MADRES MASPEC MAANTREG MALAT MADIR	The macroscopic findings
10	MI	Microscopic Findings	Findings	/transportlmi.xpt	Microscopic Findings SAS transport file	One record per finding per specimen per subject	Tabulation	STUDYID USUBJID MITESTCD MISTRESC MISPEC MIANTREG MILAT MIDIR	This domain captures the n
11	OM	Organ Measurements	Findings	/transportlom.xpt	Organ Measurements SAS transport file	One record per test per specimen per subject	Tabulation	STUDYID USUBJID OMTESTCD OMSPEC OMANTREG OMLAT OMDIR	The Organ Measurements
12	RELREC	Related Records	Relationship Datasets	/transportlrelrec.xpt	Related Records SAS transport file	One record per related record, related group of records (e.g., --GRPID), or related datas	Tabulation	STUDYID RDOMAIN USUBJID IDVAR IDVARVAL RELID	This domain represents rel
13	SUPPBG	Supplemental Qualifiers	Relationship Datasets	/transportlsuppbg.xpt	Supplemental Qualifiers SAS transport file	One record per IDVAR, IDVARVAL, and QNAM value per subject per related domain	Tabulation	STUDYID RDOMAIN USUBJID IDVAR IDVARVAL QNAM	The Supplemental Qualifie
14	SUPPBW	Supplemental Qualifiers	Relationship Datasets	/transportlsuppbw.xpt	Supplemental Qualifiers SAS transport file	One record per IDVAR, IDVARVAL, and QNAM value per subject per related domain	Tabulation	STUDYID RDOMAIN USUBJID IDVAR IDVARVAL QNAM	The Supplemental Qualifie
15	SUPPCL	Supplemental Qualifiers	Relationship Datasets	/transportlsuppcl.xpt	Supplemental Qualifiers SAS transport file	One record per IDVAR, IDVARVAL, and QNAM value per subject per related domain	Tabulation	STUDYID RDOMAIN USUBJID IDVAR IDVARVAL QNAM	The Supplemental Qualifie
16	SUPPDS	Supplemental Qualifiers	Relationship Datasets	/transportlsuppds.xpt	Supplemental Qualifiers SAS transport file	One record per IDVAR, IDVARVAL, and QNAM value per subject per related domain	Tabulation	STUDYID RDOMAIN USUBJID IDVAR IDVARVAL QNAM	The Supplemental Qualifie
17	SUPPLB	Supplemental Qualifiers	Relationship Datasets	/transportlsupplb.xpt	Supplemental Qualifiers SAS transport file	One record per IDVAR, IDVARVAL, and QNAM value per subject per related domain	Tabulation	STUDYID RDOMAIN USUBJID IDVAR IDVARVAL QNAM	The Supplemental Qualifie
18	SUPPMA	Supplemental Qualifiers	Relationship Datasets	/transportlsuppma.xpt	Supplemental Qualifiers SAS transport file	One record per IDVAR, IDVARVAL, and QNAM value per subject per related domain	Tabulation	STUDYID RDOMAIN USUBJID IDVAR IDVARVAL QNAM	The Supplemental Qualifie
19	SUPPMI	Supplemental Qualifiers	Relationship Datasets	/transportlsuppmi.xpt	Supplemental Qualifiers SAS transport file	One record per IDVAR, IDVARVAL, and QNAM value per subject per related domain	Tabulation	STUDYID RDOMAIN USUBJID IDVAR IDVARVAL QNAM	The Supplemental Qualifie
20	SUPPVS	Supplemental Qualifiers	Relationship Datasets	/transportlsuppvs.xpt	Supplemental Qualifiers SAS transport file	One record per IDVAR, IDVARVAL, and QNAM value per subject per related domain	Tabulation	STUDYID RDOMAIN USUBJID IDVAR IDVARVAL QNAM	The Supplemental Qualifie
21	VS	Vital Signs	Findings	/transportlvs.xpt	Vital Signs SAS transport file	One record per measurement per observation time per subject	Tabulation	STUDYID USUBJID VSTESTCD VSDTC	This domain is intended to



	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	table	column	label	ordi	type	length	displayformat	xmldatatype	xmlcodelist	core	origin	role	term	algorithm	qualifiers	comment
2	BG	STUDYID	Study Identifier	1	C	6		text		Req		Identifier				Unique identifier for a study.
3	BG	DOMAIN	Domain Abbreviation	2	C	2		text		Req		Identifier	BG		UPPERCASE	Two-character abbreviation for the domain.
4	BG	USUBJID	Unique Subject Identifier	3	C	11		text		Req		Identifier				Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.
5	BG	BGSEQ	Sequence Number	4	N	8		integer		Req		Identifier				Sequence number given to ensure uniqueness of subject records within a domain. May be any valid number.
6	BG	BGTESTCD	Test Short Name	5	C	6		text	BGTESTCD	Req		Topic	(BGTESTCD)		UPPERCASE	Short name of the measurement, test, or examination described in BGTEST. It can be used as a column name when conve
7	BG	BGTEST	Test Name	6	C	16		text	BGTEST	Req		SynonymQualifier	(BGTEST)			Long name for BGTESTCD. The value in BGTEST cannot be longer than 40 characters.
8	BG	BGORRES	Result or Findings as Collected	7	C	5		text		Exp		ResultQualifier				Result of the measurement or finding as originally recorded.
9	BG	BGORRESU	Unit of the Original Result	8	C	2		text	UNIT	Exp		VariableQualifier	(UNIT)			The unit for the original result. The unit of the original result should be mapped to a synonymous unit on the Controlled Te
10	BG	BGSTRESC	Standardized Result in Character Format	9	C	5		text		Exp		ResultQualifier				Contains the result value for all findings or derived from BGORRES in a standard format or standard units. BGSTRESC sh
11	BG	BGSTRESN	Standardized Result in Numeric Format	10	N	8	8.3	float		Exp		ResultQualifier				Used for results or findings in standard format; contains the numeric form of BGSTRESC. BGSTRESN should store all nu
12	BG	BGSTRESU	Unit of the Standardized Result	11	C	2		text	UNIT	Exp		VariableQualifier	(UNIT)			Standardized unit used for BGSTRESC or BGSTRESN.
13	BG	BGDTC	DateTime Animal Weighed	16	C	64		datetime		Exp		Timing	ISO8601		DATETIME	Datetime of the start of the weight interval in ISO 8601 format. This must be populated when BGTESTCD is BWGAINA.
14	BG	BGENDTC	End DateTime Animal Weighed	17	C	64		datetime		Exp		Timing	ISO8601		DATETIME	Datetime of the end of the weight interval in ISO 8601 format. This must be populated when BGTESTCD is BWGAINA.
15	BG	BGDY	Study Day Animal Weighed	18	N	8		integer		Perm		Timing				Study day of the start of the weight interval, in integer days. The algorithm for calculations must be relative to the sponsor-
16	BG	BGENDY	Study Day of End of Weight Interval	19	N	8		integer		Perm		Timing				Study day of the end of the weight interval, in integer days. The algorithm for calculations must be relative to the sponsor-
17	BW	STUDYID	Study Identifier	1	C	6		text		Req		Identifier				Unique identifier for a study.
18	BW	DOMAIN	Domain Abbreviation	2	C	2		text		Req		Identifier	BW		UPPERCASE	Two-character abbreviation for the domain.
19	BW	USUBJID	Unique Subject Identifier	3	C	11		text		Req		Identifier				Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.
20	BW	BWSEQ	Sequence Number	4	N	8		integer		Req		Identifier				Sequence number given to ensure uniqueness of subject records within a domain. May be any valid number.
21	BW	BWTESTCD	Test Short Name	5	C	6		text	BWTESTCD	Req		Topic	(BWTESTCD)		UPPERCASE	Short name of the measurement, test, or examination described in BWTEST. It can be used as a column name when conv
22	BW	BWTEST	Test Name	6	C	20		text	BWTEST	Req		SynonymQualifier	(BWTEST)			Long name for BWTESTCD. The value in BWTEST cannot be longer than 40 characters.
23	BW	BWORRES	Result or Findings as Collected	7	C	4		text		Exp		ResultQualifier				Result of the measurement or finding as originally received or collected.
24	BW	BWORRESU	Unit of the Original Result	8	C	2		text	UNIT	Exp		VariableQualifier	(UNIT)			The unit for the original result. The unit of the original result should be mapped to a synonymous unit on the Controlled Te
25	BW	BWSTRESC	Standardized Result in Character Format	9	C	4		text		Exp		ResultQualifier				Contains the result value for all findings or derived from BWORRES in a standard format or standard units. BWSTRESC s
26	BW	BWSTRESN	Standardized Result in Numeric Format	10	N	8	8.3	float		Exp		ResultQualifier				Used for results or findings in standard format; contains the numeric format of BWSTRESC. BWSTRESN should store all

Figure 2 Model specification file for Spec-driven define file

A template has been made from the master model (see page 9 for description of master model). The user then deletes the domains not being used for the study, as well as the columns in the two tabs. Since this can be a tedious task, templates for typical study designs or products could be made. The specification file is stored in a pre-defined place in the study folder hierarchy and must have a specified name. This was made to limit the number of parameters in the main macro.

Controlled terminology is by default read from the registered terminology in the cstGlobalLibrary of the SAS CST Toolkit. If the study is only using a subset of the terminology, e.g. country is limited to only a couple of countries, then the study specific terminology specification will specify this, see Figure 3.

A	C	D	E	F	G	H	I	N	O
sourcesystem	description	codelist	codelist	codelist_nam	codelist_extensible	codelist_synonym	codelist	code	cdisc_submission_value
NCI Thesaurus	CDISC SEND Controlled Terminology, 2014-06-27	COUNTRY	C66786	Country	No	Country	A collectiv	C16496	DNK
NCI Thesaurus	CDISC SEND Controlled Terminology, 2014-06-27	COUNTRY	C66786	Country	No	Country	A collectiv	C16592	FRA
NCI Thesaurus	CDISC SEND Controlled Terminology, 2014-06-27	COUNTRY	C66786	Country	No	Country	A collectiv	C16636	DEU
NCI Thesaurus	CDISC SEND Controlled Terminology, 2014-06-27	COUNTRY	C66786	Country	No	Country	A collectiv	C17180	SWE
NCI Thesaurus	CDISC SEND Controlled Terminology, 2014-06-27	COUNTRY	C66786	Country	No	Country	A collectiv	C17181	CHE
NCI Thesaurus	CDISC SEND Controlled Terminology, 2014-06-27	COUNTRY	C66786	Country	No	Country	A collectiv	C17233	GBR
NCI Thesaurus	CDISC SEND Controlled Terminology, 2014-06-27	COUNTRY	C66786	Country	No	Country	A collectiv	C17234	USA

Figure 3 Controlled terminology specification file for Spec-driven define file

Next step is to set the relevant parameters for the study in a parameter file called studyparms.sas:

```
*This file contains settings of study specific parameters;

*Below are the study specific parameters. These should be changed to fit the
specific study before any programs are executed;

*The name of the study, e.g. %let studyname=117002CRO. To be used in the header
of the define file.;
%let studyname=117002; **<------ user input;

*Folder path to the study. ROOT is set as part of your system parameters;
*%let studyroot=&root\&studyname\10_SAS;
%let studyroot=&root\Study\10_SAS;

*The internal version of the SEND standard (in cstGlobalLibrary).
E.g. %let standardver=3.0 if the ordinary SEND standard is used or
%let standardver=3.0-_1.0 which could be the Company version 1.0 of the SEND 3.0
version.
Naming should be consistent with what is registered in the CST Global library.
;
*%let standard_ver=3.0-1.0; **<------ user input;
%let standard_ver=3.0-1.0-2.0;

*The version used on the header of the define.xml. It must be one of the official
versions to avoid error in define validation.;
%let reportversion=3.0; **<------ user input;

*The version of the Controlled terminology to be used;
%let CTDate=201606;**<------ user input;

*Short description of the study;
%let studydescription=Specification 117002CRO;**<------ user input;

*A short description of the protocol;
%let protocolname=Tox evaluation of drug 117002; **<------ user input;

*The version of the study;
```

```
%let studyversion=1.0;**<------ user input;
```

Finally, the main macro is executed:

```
*includes the standard parameters;
%include "%program_path_v01()\..\..\study_parms.sas";

*executes the create-define_v01_spec;
%create_define_v01_spec;
```

The spec-driven define.xml is created, see Figure 4.

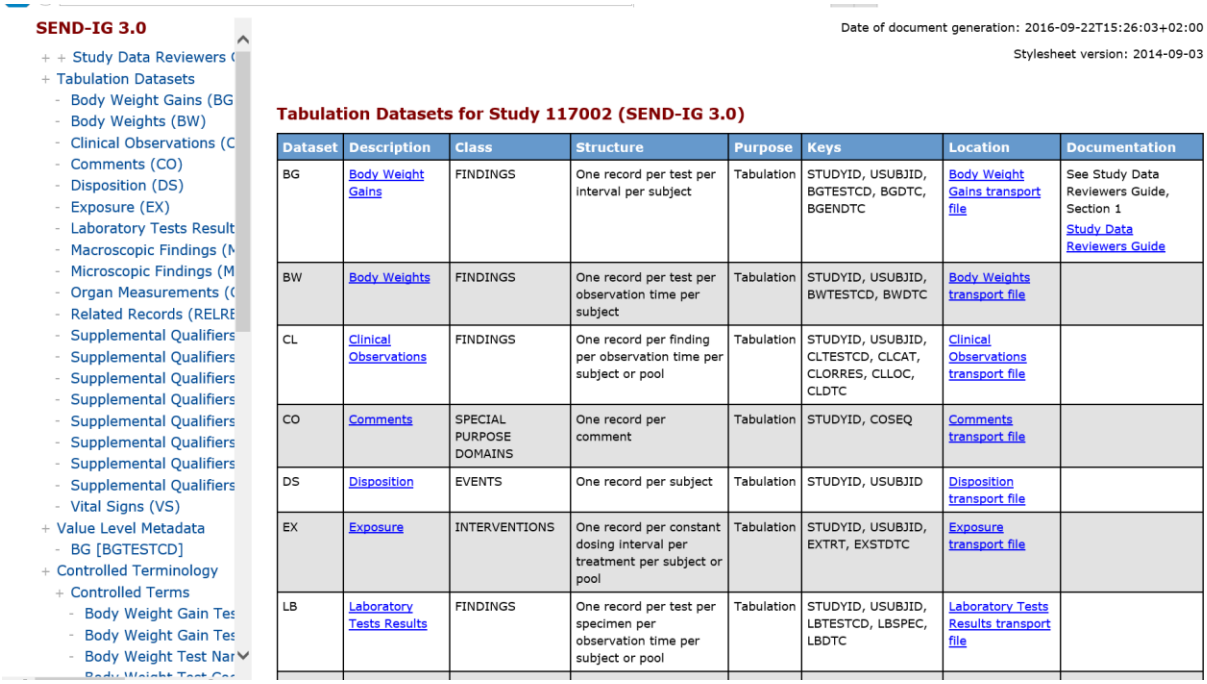


Figure 4 Extract of Spec-driven define.xml

Supporting documents (sdrg.pdf) and specification of additional metadata, including value level metadata, is done in the ValueLevelMetaData.xls sheet, see page 5. Note that for SEND studies no CRF is created and hence this solution does not include a CRF reference and hyper-linking in the define.xml.

Since different CROs can contribute with a subset of the domains, the spec-driven define-xml can be tailored to fit the domains to be delivered by a certain CRO.

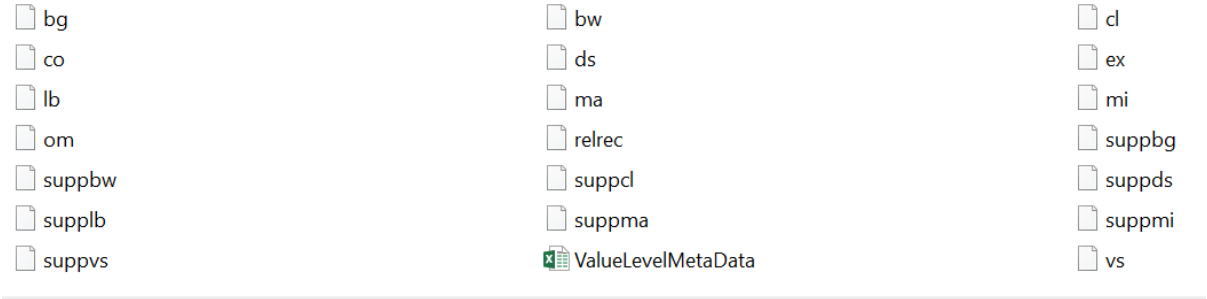
DATA-DRIVEN DEFINE FILE

The approach for the data-driven define.xml is similar, only this only requires the data (the xpt files), the ValueLevelMetaData.xls sheet and the sdrg.pdf.

The study_parms.sas file is the same as for the spec-driven define.xml. The xpt-files are placed in the input folder, the supporting documents are placed in the output folder, and the main macro is called:

```
%include "%program_path_v01()\..\..\study_parms.sas";
%create_define_v01_study;
```

The data-driven define.xml is created based on the data in the input folder, see Figure 5.



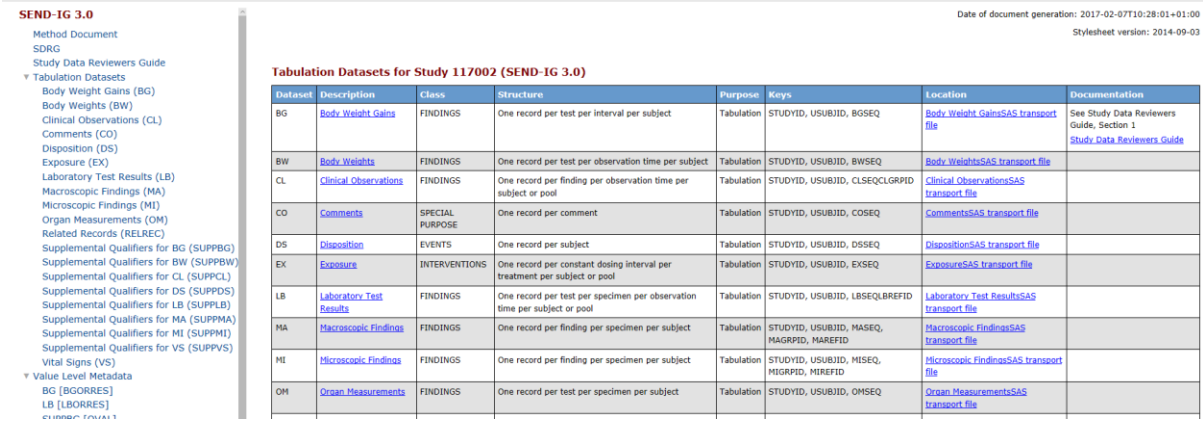


Figure 5 Input and output of data-driven define.xml

The macro also checks terminology used in the data against the CDISC controlled terminology version specified in the study_parms.sas file. If invalid codes are being used, a report is produced, see Figure 6.

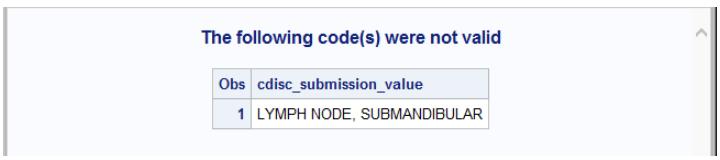


Figure 6 Report of in-valid codes in data

These codes can then be added as sponsor-defined codes in the ValueLevelMetaData.xls sheet, if the code is accepted, see Section below.

VALUE LEVEL METADATA AND SUPPORTING DOCUMENTS

The SAS Clinical Standards Toolkit is creating the define.xml based on a set of SAS datasets, see Figure 7.

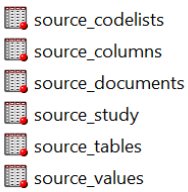


Figure 7 Source Metadata Dataset for define.xml

The ValueLevelMetadata.xls sheet will allow the user to add rows and cell information to these metadata datasets before the define.xml is created, see Figure 8. The source_study dataset is populated based on the parameters set in the study_parms.sas file.

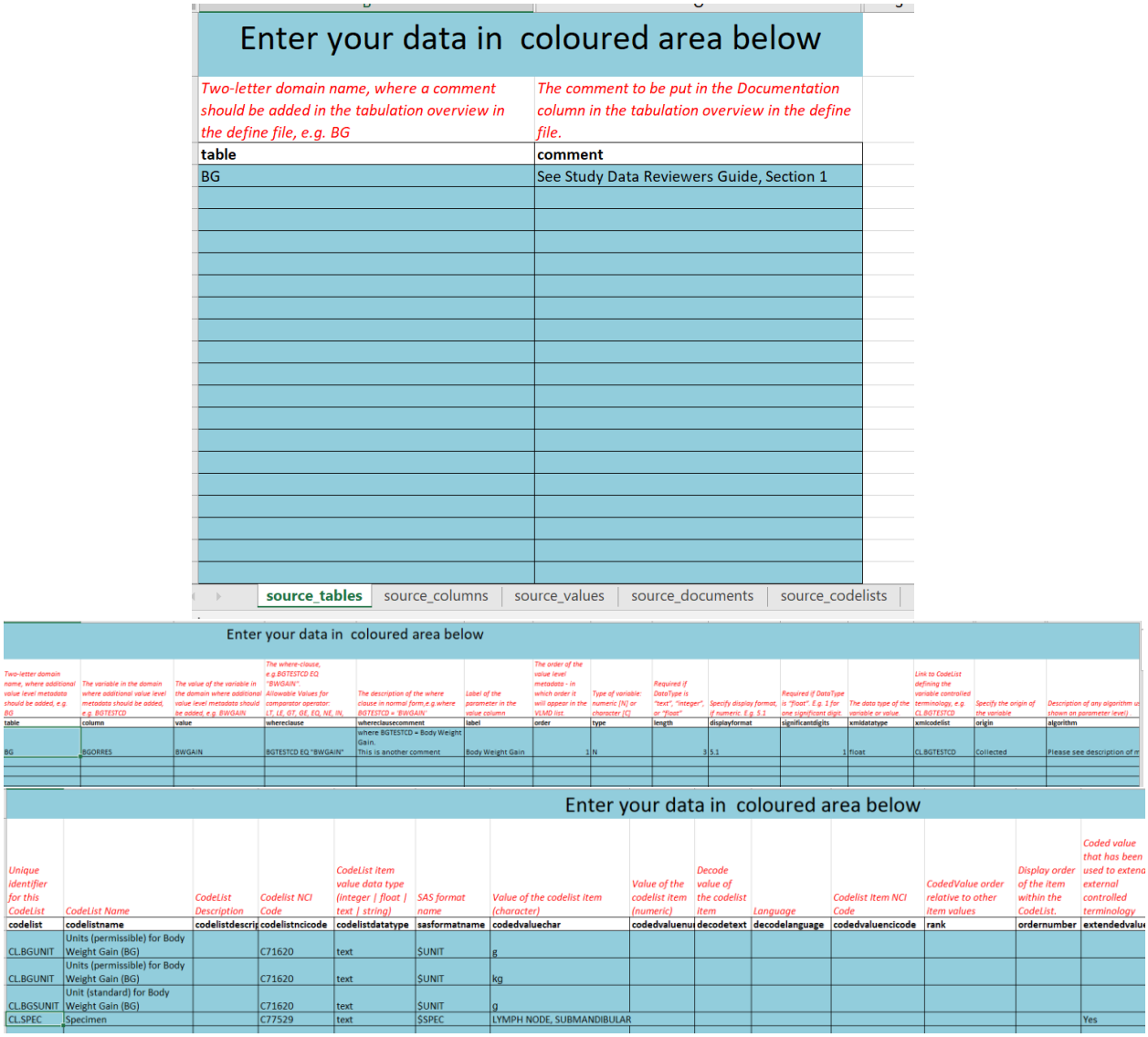


Figure 8 User input in ValueLevelMetadata.xls – source_tables, _values, and _codelists example

The main macro for the data-driven define.xml also adds information automatically based on the data, e.g. sub-setting code lists, based on observed data and providing VLMD for SUPP domains.

COMAPRE DEFINE REPORT

The compare macro reads in two define.xml ('Base' and 'Compare'). It uses SAS xml.map (SAs XML-mapper tool) to convert the files into SAS datasets, see Figure 9.

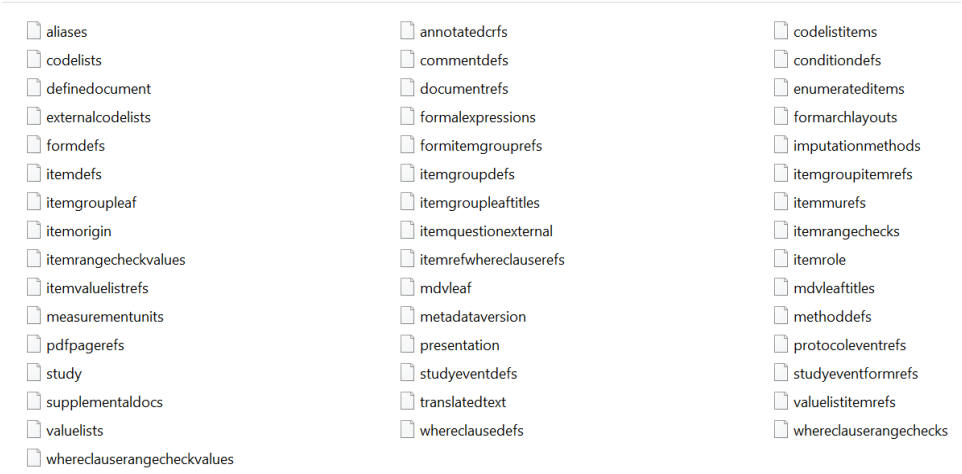


Figure 9 SAS datasets representing define.xml

The macro is executed the following way:

```
*includes the standard parameters;
%include "%program_path_v01()\..\..\study_parms.sas";

/*change specdriven define file name*/
%let BASE_DEFINE1 =
&studyroot\03_Define\01_Spec_driven\02_output\spec_define_2016-05-02T13-54.xml;

/*change data driven define file name*/
%let COMP_DEFINE1 =
&studyroot\03_Define\02_data_driven\02_output\data_define_2016-05-02T14-06.xml;

/*location of the comparison output*/
%let RESULT_LOC1=&studyroot\03_Define\03_Compare;

*executes the create-define_v01_spec;
%compare_define_v01(
      BASE_DEFINE=&BASE_DEFINE1
   ,  COMP_DEFINE=&COMP_DEFINE1
   ,  BASE_std=CDISC_SEND
   ,  stdvers=&standard_ver
   ,  RESULT_LOC=&RESULT_LOC1
);
```

The following is compared using some of the datasets in Figure 9:

Table compared	Variables compared
Study	OID Studyname Protocolname StudyDescription
Metadataversion	OID Name DefineVersion
ItemGroupDefs	OID Name Repeating TranslatedText Class
ValueListItemRefs	OID ItemOID OrderNumber Mandatory
ItemDefs	OID Name DataType Length Comment TranslatedText
Codelists	OID Name DataType
CodelistItems	CodedValue
TranslatedText	*See description below*
Valuelists ValueListItemrefs	OID

The output is provided in an html report. The beginning of the report compares basic information from the two studies, such as file type, study name and protocol name among other things. This information comes from the STUDY and METADATAVERSION tables. The rest of the output html document is divided into 8 additional sections, where each section in general comprises smaller reports on one specific table from the list above. The TranslatedText table, which contains information about multiple tables is divided into the relevant sections where the corresponding table is analyzed. See Figure 10 for examples of the report output.

(B) BASELIB: \\FSDKHQ001\x$\SENDProject\IT system\SAS Implementation\Study\10_SAS\03_Define\01_Spec_driven\02_output\spec_define_new_STANDARD_STANDARD_2016-09-22T15-23.xml
(C) COMPLIB: \\FSDKHQ001\x$\SENDProject\IT system\SAS Implementation\Study\10_SAS\03_Define\02_data_driven\02_output\del1c_data_define_2016-11-25T12-14.xml

COMPARING ODM attributes

COMPARED_VARIABLE	BASELIB_VALUE	COMPLIB_VALUE	status
ODM_FILE_TYPE	Snapshot	Snapshot	OK

COMPARED_VARIABLE	BASELIB_VALUE	COMPLIB_VALUE	status
ODM_ODMVERSIONSPEC	Snapshot	Snapshot	OK

STUDY COMPARISON

COMPARED_VARIABLE	BASELIB_VALUE	COMPLIB_VALUE	status
OID	STUDY1	STUDY1	OK
StudyName	117002	216127	NOT OK
ProtocolName	Tox evaluation of drug [redacted]117002	Tox evaluation of drug [redacted]0765	NOT OK
StudyDescription	Specification 117002CRO	Specification 216127	NOT OK

METADATAVERSION COMPARISON

COMPARED_VARIABLE	BASELIB_VALUE	COMPLIB_VALUE	status
OID	1.0	1.0	OK
NAME	Metadata for 117002, SEND-IG 3.0	Metadata for 216127, SEND-IG 3.0	NOT OK
DefineVersion	2.0.0	2.0.0	OK

ITEMGROUPDEF COMPARISON

BASECNTOID	COMPCNTOID	status
20	5	NOT OK MISSMATCH IN COUNTS

B.OID	STATUS
IG.BG	NOT OK CORRESPONDING OID CANNOT BE FOUND IN COMPLIB
IG.BW	NOT OK CORRESPONDING OID CANNOT BE FOUND IN COMPLIB
IG.CL	NOT OK CORRESPONDING OID CANNOT BE FOUND IN COMPLIB
IG.CO	NOT OK CORRESPONDING OID CANNOT BE FOUND IN COMPLIB
IG.DS	NOT OK CORRESPONDING OID CANNOT BE FOUND IN COMPLIB
IG.EX	NOT OK CORRESPONDING OID CANNOT BE FOUND IN COMPLIB
IG.LB	NOT OK CORRESPONDING OID CANNOT BE FOUND IN COMPLIB
IG.MA	NOT OK CORRESPONDING OID CANNOT BE FOUND IN COMPLIB
IG.MI	NOT OK CORRESPONDING OID CANNOT BE FOUND IN COMPLIB
IG.OM	NOT OK CORRESPONDING OID CANNOT BE FOUND IN COMPLIB
IG.RELREC	NOT OK CORRESPONDING OID CANNOT BE FOUND IN COMPLIB
IG.SUPPBG	NOT OK CORRESPONDING OID CANNOT BE FOUND IN COMPLIB
IG.SUPPBW	NOT OK CORRESPONDING OID CANNOT BE FOUND IN COMPLIB
IG.SUPPCL	NOT OK CORRESPONDING OID CANNOT BE FOUND IN COMPLIB
IG.SUPPDS	NOT OK CORRESPONDING OID CANNOT BE FOUND IN COMPLIB
IG.SUPPLB	NOT OK CORRESPONDING OID CANNOT BE FOUND IN COMPLIB
IG.SUPPMA	NOT OK CORRESPONDING OID CANNOT BE FOUND IN COMPLIB
IG.SUPPMI	NOT OK CORRESPONDING OID CANNOT BE FOUND IN COMPLIB
IG.SUPPVS	NOT OK CORRESPONDING OID CANNOT BE FOUND IN COMPLIB
IG.VS	NOT OK CORRESPONDING OID CANNOT BE FOUND IN COMPLIB

C.OID	STATUS
IG.DM	NOT OK CORRESPONDING OID CANNOT BE FOUND IN BASELIB
IG.TA	NOT OK CORRESPONDING OID CANNOT BE FOUND IN BASELIB
IG.TE	NOT OK CORRESPONDING OID CANNOT BE FOUND IN BASELIB
IG.TS	NOT OK CORRESPONDING OID CANNOT BE FOUND IN BASELIB
IG.TX	NOT OK CORRESPONDING OID CANNOT BE FOUND IN BASELIB

Figure 10 Extract from compare define report

MANAGING THE MODEL AND CONTROLLED TERMINOLOGY

In addition to a library of macros, see ref. [1], the SAS Clinical Standards Toolkit consists of a library of metadata. This library is stored on the C-drive and is called cstGlobalLibrary. It comes with a set of standards already registered, see Figure 11

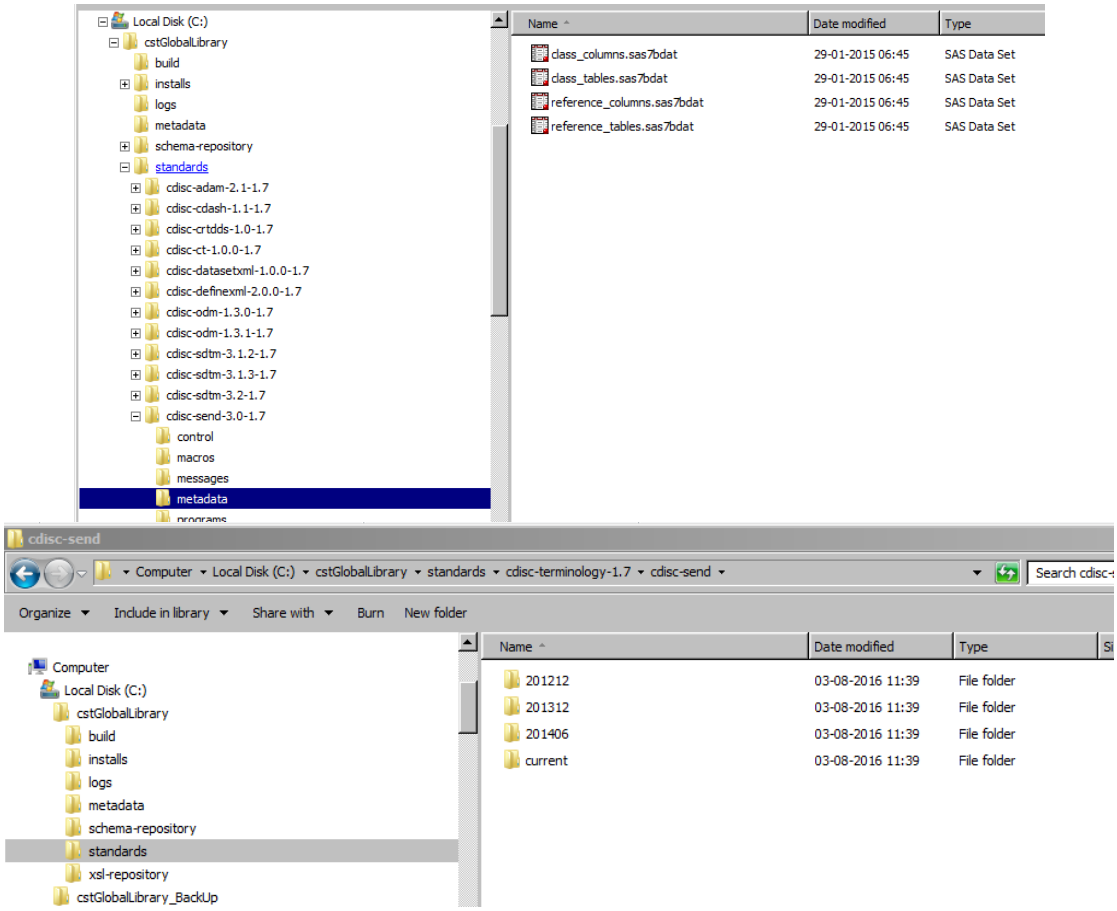


Figure 11 Pre-registered standards in cstGlobalLibrary (models and terminology)

Since standards and terminology evolve over time, the toolkit allows for registering new standards. Based on programs from the sample programs in the SAS Clinical Standards Toolkit sample library, two macros have been developed to enable the user to register new standards (SEND model and controlled terminology).

REGISTER NEW CONTROLLED TERMINOLOGY FROM NCI/CDISC

A new version of CDISC controlled terminology can be downloaded from NC's (National Cancer Institute) website, see Figure 12.

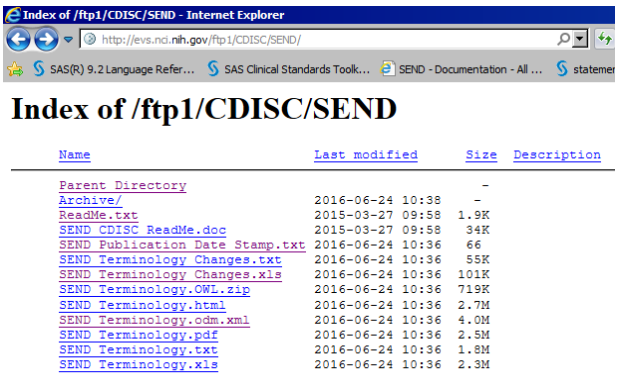


Figure 12 Download from the CDISC SEND directory on an NCI File Transfer Protocol (FTP) site

The tool is using the odm.xml format. This file is stored in a predefined folder for standards administration, see **Figure 13**.

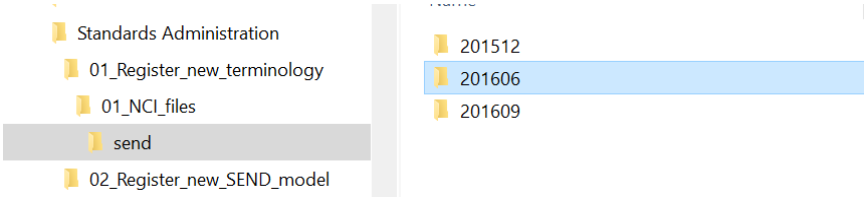


Figure 13 Folder hierarchy for Controlled Terminology Administration

The macro **register_new_version_ct_v01.sas** is then executed in SAS, see Figure 14. The example here is registering the controlled terminology from December 2015.

```
31
32 %include "%program_path_v01()\..\init_parms.sas";
33
34 *The macro register_new_version_ct_v01;
35
36 %register_new_version_ct_v01(CTFile=SEND_Terminology.odm,
37                              CTDate=201512,
38                              CTDescription=CDISC SEND Controlled Terminology released by NCI on 2015-12-18,
39                              DEFAULT_STD=Y);
40
```

Figure 14 Sample program calling register_new_version_ct_v01.sas macro

Executing the program above with the parameters as specified in Figure 14, the following will be available in \Standards Administration \01_Register_new_terminology\01_NCI_files\send\201512:

- A **results** folder containing two toolkit generated datasets. The **read_results_send_201512.sas7bdat** containing information about how the process went reading in the .xml file. The **cterms_results_send_201512.sas7bdat** containing information about how the process went when creating the cterms dataset. If something went wrong, the error and warning explanations can be found here. If everything went fine, only records of type 'Info' (resultsseverity=Info) will be present.
- A **data** folder containing the xml file translated into a set of SAS datasets. These are used in the process when creating the cterms dataset.

Executing the program above with the parameters as specified in Figure 14, the following will change in the cstGlobalLibrary:

- A new folder in the cstGlobalLibrary will be created, see Figure 15, and a SAS dataset with the terminology (cterms.sas7bdat) will be populated with controlled terminology as specified in the CTFile ("..\StandardsAdministration\01_Register_new_terminology\01_NCI files\send\201512\SEND_Terminology.odm.xml")

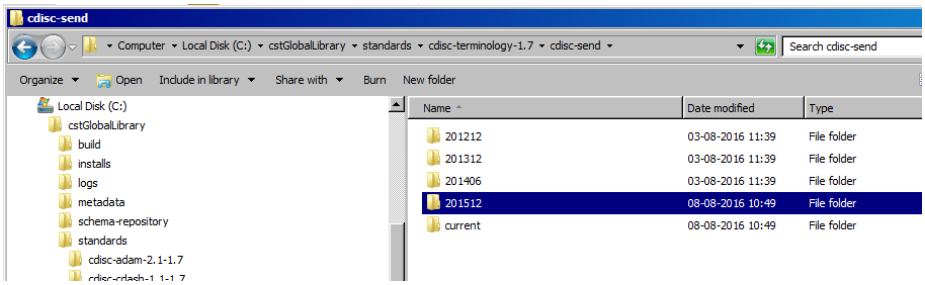


Figure 15 New folder in the cstGlobalLibrary

The new controlled terminology will also be copied to the current folder since the DEFAULT_STD parameter was set to Y. The controlled terminology is being registered in the toolkit's overview datasets adding a record in the **standardsubtypes** dataset, see **Figure 16**.

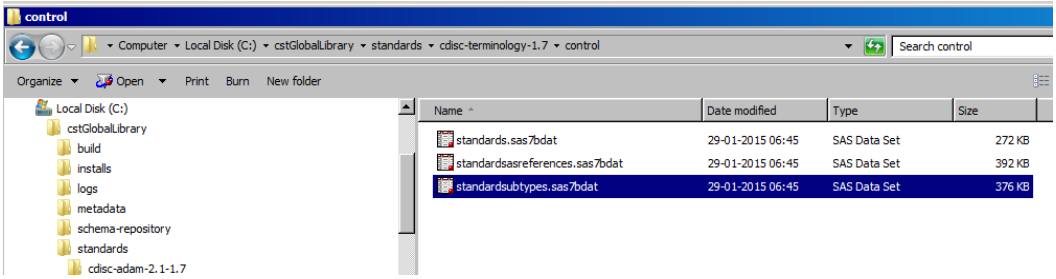


Figure 16 Dataset listing all CTs registered in the cstGlobalLibrary

The macro checks if the version specified in the odm.xml file fits with the CTDate specified in the parameter of the macro. Also, it checks if the terminology already exists in the cstGlobalLibrary and interacts with the user if

it does, see Figure 17 and Figure 18. First an error message is displayed in the log instructing the user what to do if the CT must be re-registered. After deleting the folder in the cstGlobalLibrary and running the macro again, then the macro will launch the window in Figure 18, since the controlled terminology is (still) registered in the **standardsubtypes** dataset. Typing Y will then unregister the terminology and then run the full registration again and re-create the deleted folder.

```
ERROR: The Controlled Terminology from 201512 is already in the CST Global library
If you want to redo the process then delete the folder
(C:\cstGlobalLibrary\standards\cdisc-terminology-1.7\cdisc-send\201512\)
in CST Global Library and then run mk_register_new_version_ct program again
```

Figure 17 Error message - Controlled Terminology already exists

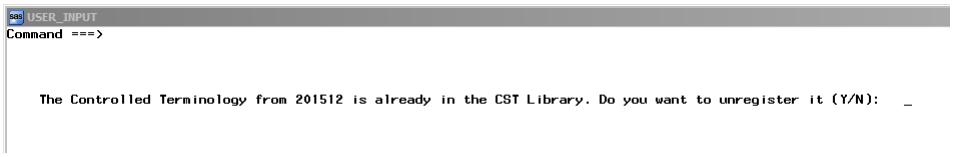


Figure 18 User input - unregister controlled terminology

REGISTER A NEW DATA MODEL

Registering a new version of the SEND model can either be a new version from CDISC or sponsor modified version of a CDISC released version. Making sponsor modifications to a CDISC model is allowed, if it conforms with the overall model principles, as described in the implementation guide.

A model is described by its domains (tables) and variables in each domain (columns). In the cstGlobalLibrary, the model is represented in reference_tables and reference_columns dataset. (The class of domains (FINDING, EVENT, INTERVENTION etc.) and variables (IDENTIFIER, TIMING, etc.) are described in class_tables and class_columns dataset), see Figure 19.

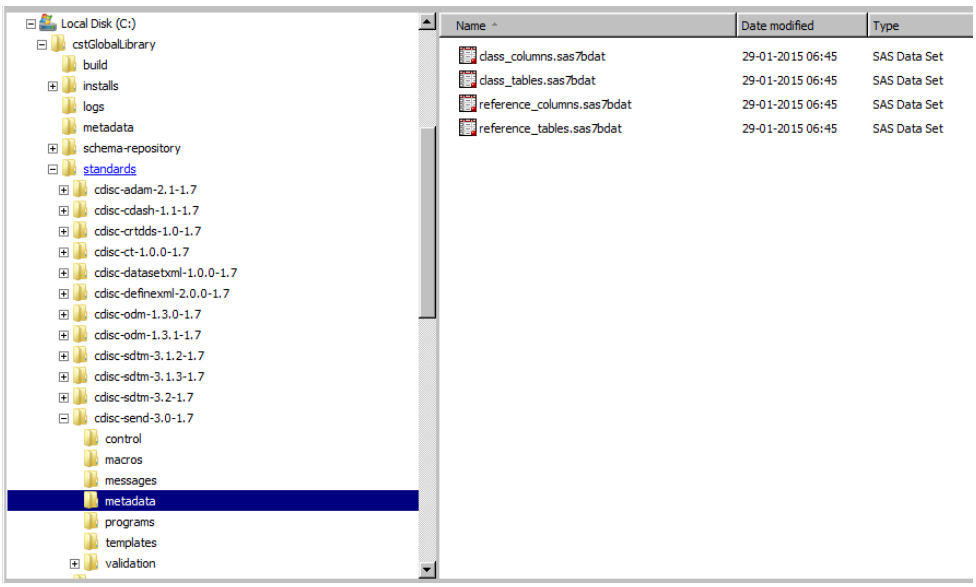


Figure 19 Reference tables for a model in cstGlobalLibrary

To create and register a new model in the cstGlobalLibrary, input will be the tables and column specification for the new model (in Excel).

During the development of the tools it became apparent that the metadata for the domains (tables) also needs to be published (SHARE download). This information describes domains: name, label, structure and keys, see ref. [2] (Table 3.2.1 Dataset Definition Metadata example). After some dialogue with the CDISC organization, some of this information is now provided for SEND.

Since the macro was developed before this was in place, the macro expects two excel sheets as input: one for the columns and one for the tables, see Figure 20 and Figure 21.

Seq. For Order	Observation Class	Domain Prefix	Variable Name (minus domain prefix)	Variable Name	Variable Label	Type	Controlled Terms or Format	Role	CDISC Notes (for domains) Description (for General Classes)	Core
1	Findings	BG	STUDYID	STUDYID	Study Identifier	Char		Identifier	Unique identifier for a study.	Req
2	Findings	BG	DOMAIN	DOMAIN	Domain Abbreviation	Char	(DOMAIN)	Identifier	Two-character abbreviation for the domain.	Req
3	Findings	BG	USUBJID	USUBJID	Unique Subject Identifier	Char		Identifier	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.	Req
4	Findings	BG	BGSEQ	BGSEQ	Sequence Number	Num		Identifier	Sequence number given to ensure uniqueness of subject records within a domain. May be any valid number.	Req
5	Findings	BG	BGTESTCD	BGTESTCD	Test Short Name	Char	(BGTESTCD)	Topic	Short name of the measurement, test, or examination described in BGTEST. It can be used as a column name when converting a dataset from a vertical to a horizontal format. The value in BGTESTCD cannot be longer than 8 characters, nor can it start with a number (e.g., "1TEST" is not valid). BGTESTCD cannot contain characters other than letters, numbers, or underscores.	Req
6	Findings	BG	BGTEST	BGTEST	Test Name	Char	(BGTEST)	Synonym Qualifier	Long name for BGTESTCD. The value in BGTEST cannot be longer than 40 characters.	Req
7	Findings	BG	BGORRES	BGORRES	Result or Findings as Collected	Char		Result Qualifier	Result of the measurement or finding as originally recorded.	Exp
8	Findings	BG	BGORRESU	BGORRESU	Unit of the Original Result	Char	(BGUNIT)	Variable Qualifier	The unit for the original result. The unit of the original result should be mapped to a synonymous unit on the Controlled Terminology list.	Exp
9	Findings	BG	BGSTRESC	BGSTRESC	Standardized Result in Character Format	Char		Result Qualifier	Contains the result value for all findings or derived from BGORRES in a standard format or standard units. BGSTRESC should store all results or findings in character format; if results are numeric, they should also be stored in numeric format in BGSTRESN.	Exp
10	Findings	BG	BGSTRESN	BGSTRESN	Standardized Result in Numeric Format	Num		Result Qualifier	Used for results or findings in standard format; contains the numeric form of BGSTRESC. BGSTRESN should store all numeric test results or findings.	Exp
11	Findings	BG	BGSTRESU	BGSTRESU	Unit of the Standardized Result	Char	(BGSUNIT)	Variable Qualifier	Standardized unit used for BGSTRESC or BGSTRESN.	Exp

Figure 20 Specification for columns (variables)

Domain Prefix	Domain Description	Domain Structure	Domain Label	Keys
BG	Body weight gain is the actual difference between two body weight measurements for any given interval for a subject. This is most commonly shown as the difference between two consecutive body weight measurements.	One record per test per interval per subject	Body Weight Gains	STUDYID USUBJID BGTESTCD BGDTC BGENDTC
BW	This domain captures body weights collected for subjects during the study and at the end of the study (terminal body weights).	One record per test per observation time per subject	Body Weights	STUDYID USUBJID BWTESTCD BWDTC
CL	This domain captures clinical sign information in addition to ophthalmology, physical examination, and dermal examination collected during in-life while executing the study.	One record per finding per observation time per subject or pool	Clinical Observations	STUDYID USUBJID CLTESTCD CLCAT CLORRES CLLOC CLDTC
CO	The Comments special-purpose domain provides a solution for submitting free-text comments related to data in one or more SEND domains. Comments are generally not responses to specific questions; instead, comments usually consist of voluntary, free-text, or unsolicited observations.	One record per comment	Comments	STUDYID COSEQ
DD	This domain captures the diagnosis of the cause of death for a subject.	One record per diagnosis per subject (for unscheduled deaths only)	Death Diagnosis	STUDYID USUBJID
DM	This domain captures subject demographics, with one record for each subject used in the study.	One record per subject	Demographics	STUDYID USUBJID
DS	The disposition dataset provides a record for the disposition of subjects throughout the study. One record will be created for each subject used in the study, whether it ultimately died or was removed from the study.	One record per subject	Disposition	STUDYID USUBJID
EG	EG: This domain captures electrocardiographic (ECG) specific parameters.	One record per test per observation time per subject	ECG Test Results	STUDYID USUBJID EGTESTCD EGDTC
EX	The Exposure Domain Model records the details of a subject's administered dose of protocol-specified study treatment. Study treatment may be any intervention that is prospectively defined as a test material within a study, and is typically, but not always, administered to the subject. Examples include, but are not limited to, placebo, active comparators, and investigational products. Only protocol-specified treatments should be included in this domain.	One record per constant dosing interval per treatment per subject or pool	Exposure	STUDYID USUBJID EXTRT EXSTDTC
FW	This domain captures food/water consumption of animals in the study. The data in this domain is derived data.	One record per test per interval per subject or pool	Food and Water Consumption	STUDYID USUBJID FWTESTCD FWDTC FWENDTC
LB	This domain captures laboratory data collected by the lab executing the study or received from a central provider.	One record per test per specimen per observation time per subject or pool	Laboratory Test Results	STUDYID USUBJID LBTESTCD LBSPEC LBDTC LBTPTNUM

Figure 21 Specification for tables (domains)

The KEYS column has been added compared to specification published by SHARE. The specification files are stored in a Model Specification folder and then the macro is executed:

```
%include "%program_path_v01()\..\init_parms.sas";


%register_new_send_standard_v01(STD_SPEC_NAME=send-3-0-2-0.xls,
                                TABLE_SPEC_NAME=reference_tables_SHARE.xlsx,
                                standardver=3.0-2.0,
                                DESCRIPTION=Company SEND V2.0 based on CDISC SEND V3.0);
```

The init_parms.sas contains a few parameters like the base CDISC SEND standard (v. 3.0).

Executing the program with the parameters as specified above, the following will change in the cstGlobalLibrary:

- A new folder in the cstGlobalLibrary will be created (named cdisc-send-3.0-2.0-1.7) and populated with model metadata based on the specification file ..\Standards Administration\02_Register_new_SEND_model\Model Specification\send-3-0-2-0.xls.
- A new record in the standards.sas7bdat dataset is added, see Figure 22 which then makes it registered in the cstGlobalLibrary.

	Name of standard	Mnemonic for standard	Standard version	Standard group	Standard group version	Comments
1	CDISC-ADAM	ADAM	2.1	ADAM	2.1	CDISC ADAM V2.1
2	CDISC-CDASH	DASH	1.1	CDASH	1.1	CDISC CDASH V1.1
3	CDISC-CRTDDS	CRT	1.0	DEFINE	1.0	CDISC CRT-DDS V1.0
4	CDISC-CT	CTX	1.0.0	CTX	1.0.0	CDISC CT XML V1.0.0
5	CDISC-DATASET-XML	DATA	1.0.0	DATASET	1.0.0	CDISC Dataset-XML V1.0.0
6	CDISC-DEFINE-XML	DEF	2.0.0	DEFINE	2.0.0	CDISC Define-XML V2.0.0
7	CDISC-ODM	ODM	1.3.0	ODM	1.3.0	CDISC ODM V1.3.0
8	CDISC-ODM	ODM	1.3.1	ODM	1.3.1	CDISC ODM V1.3.1
9	CDISC-SDTM	SDTM	3.1.2	SDTM	3.1.2	CDISC SDTM V3.1.2
10	CDISC-SDTM	SDTM	3.1.3	SDTM	3.1.3	CDISC SDTM V3.1.3
11	CDISC-SDTM	SDTM	3.2	SDTM	3.2	CDISC SDTM V3.2
12	CDISC-SEND	SEND	3.0	SEND	3.0	CDISC SEND V3.0
13	CDISC-SEND	SEND	3.0-2.0	SEND	3.0	Company SEND V2.0 based on CDISC SEND V3.0
14	CDISC-TERMINOLOGY	CT	NCI_THESAURUS	TERMINOLOGY	NCI_THESAURUS	CDISC Terminology
15	CST-FRAMEWORK	CST	1.2	FRAMEWORK	1.2	Clinical Standards Toolkit Framework

Figure 22 New record in standards.sas7bdat

As for the controlled terminology, checks have also been built into the macro %register_new_send_standard_v01. If the standard is already registered, an ERROR message is displayed in the log instructing the user. If the user deletes the standards folder in the cstGlobalLibrary and runs the macro again, a UI appears allowing the user to decide to re-register (un-register) the standard. The standard is then un-registered and the standard is being generated again and re-registered.

A standard can also be unregistered separately, using the unregister_standard_v01.sas macro. This macro is based on the un-register-standard.sas macro in the CST sample library.

CONCLUSION

The macros described have been developed to fit customer needs and business process. However, improvements could be considered:

- The VLMD provided in the excel-sheet (Figure 8) could be created automatically, e.g. adding sponsor-defined codes to the code list
- The model specification (Figure 20 and Figure 21) could be extended to include more columns providing VLMD on a master model level
- The define.xml comparison report could be based on the source metadata datasets and not the xml-like datasets. The source metadata datasets are typically more familiar to the user due to the structure of the VLMD input file.
- The UI used is based on the %WINDOWS and %DISPLAY commands in SAS. To make the interface more up to date a web-based interface could be developed using SAS stored processes. Using this approach would also free the user from executing the SAS macros.

When using SAS and being familiar with programing, the SAS Clinical Standards Toolkit provides a flexible and efficient way of managing the SEND (SDTM) model, controlled terminology and generating define.xml. The toolkit provides additional features like dataset.xml and ADaM, see ref. [1].

REFERENCES

[1] Clinical Standard Toolkit v 1.7 System Documentation
[2] SEND IG 3.1

ACKNOWLEDGMENTS

During the development of the tools I was employed at SAS Institute. I would like to thank my colleague at that time, Lex Janssen (SAS Institute), for help and support during the development of the tools. Also, I would like to thank Allan Lyngby Lassen, Morten Badensø and Kim Sebastian Lillandt, colleagues from SAS Institute that were part of the implementation team at the customer site.

CONTACT INFORMATION

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

Kirsten Walther Langendorf, S-CUBED
Lille Strandstraede 20C 5., DK-1254 Copenhagen K
Work Phone: +45 31 65 76 55
Email: kl@s-cubed.dk
Web: www.s-cubed.dk