

## Submission-Ready Define.xml Files Using SAS® Clinical Data Integration

Melissa R. Martinez, SAS Institute, Cary, NC USA

### ABSTRACT

SAS Clinical Data Integration simplifies the transformation of raw data into submission-ready datasets that conform to CDISC data standards. It also has a built-in transformation that creates a define.xml file from a study's CDISC domains, with just a few simple selections required from the end user. With the appropriate metadata definitions, the built-in transformation will pick up and include computational algorithms and controlled terminology codelists in the resulting define.xml file. In SAS Clinical Data Integration 2.6, a new feature was added that simplifies the process of adding information about supplemental documents and value-level metadata to the define.xml file. This paper will provide examples and instructions for creating a submission-ready define.xml file complete with the appropriate computational algorithms, controlled terminology codelists, value-level metadata, and supplemental documents using SAS Clinical Data Integration.

### INTRODUCTION

This paper uses the CDISC01 example study SDTM domain metadata and define.xml version 2.0 file available from CDISC as a reference for a complete, submission-ready defin.xml file. All screenshots of the define.xml file use the default stylesheet for easy viewing in a web browser. This paper also highlights the available features of SAS Clinical Data Integration (CDI) version 2.6. This paper assumes the reader is familiar with the use of SAS Clinical Data Integration and has an intermediate understanding of define.xml files.

The major sections of a define file are

- Links to external documents (annotated case report form [CRF], reviewer's guide, complex algorithms document, etc.)
- Tabulation Datasets
- Value Level Metadata
- Controlled Terminology
- Computational Algorithms
- Comments

### GAPS BETWEEN THE DEFAULT SAS CLINICAL DATA INTEGRATION GENERATED DEFINE.XML FILE AND A SUBMISSION-READY DEFINE.XML FILE

SAS Clinical Data Integration makes it very easy to generate a define.xml file. The product stores the metadata for the study and its domains. The user runs the CDISC-Define Creation transformation, which uses the stored metadata to generate most of the define.xml file. The following examples assume that a user has imported the CDISC SDTM 3.1.2 data standard into CDI as provided with CDI and has not modified any of the clinical metadata for any of the domains or its columns. In that case, there are a few gaps between what CDI will produce and what a company needs in order to have a submission-ready define.xml file. In the following screen shots comparisons, major features of a submission-ready define.xml file are highlighted alongside the basic define.xml file produced by CDI.

#### SDTM-IG 3.1.2

- [Annotated Case Report Form](#)
- [Reviewers Guide](#)
- [Complex Algorithms](#)
- ▶ [Tabulation Datasets](#)
- ▶ [Value Level Metadata](#)
- ▶ [Controlled Terminology](#)
- ▶ [Computational Algorithms](#)
- ▶ [Comments](#)

#### SDTM-IG 3.1.2

- ▶ [Tabulation Datasets](#)
- ▶ [Controlled Terminology](#)
- ▶ [Comments](#)

FIGURE 1: CONTENTS, SUBMISSION-READY DEFINE.XML

FIGURE 2: CONTENTS, DEFAULT CDI DEFINE.XML

Figure1 shows the table of contents link that appears in the submission-ready define.xml file. It includes all of the major sections of a define file listed above. Figure 2 shows the table of contents link that appears in the define.xml file

generated by SAS Clinical Data Integration when none of the included CDISC metadata is modified. You'll notice there are no links to external documents, no value level metadata, and no computational algorithms.

The summary of the tabulation datasets is displayed at the beginning of the define file. The Description field is a hyperlink to the details for the dataset. The Location field is a hyperlink to the SAS V5 transport file for the dataset. Figure 3 and Figure 4 below show the submission-ready define.xml file and the default CDI define.xml file, respectively. Notice that the Documentation field contains a reference to the appropriate section of the Reviewer's Guide and a hyperlink to the document in the submission-ready define.xml file. The default CDI define.xml file has no documentation comments for any domains.

### Tabulation Datasets for Study CDISC01 (SDTM-IG 3.1.2)

Dataset	Description	Class	Structure	Purpose	Keys	Location	Documentation
TA	<a href="#">Trial Arms</a>	TRIAL DESIGN	One record per planned Element per Arm	Tabulation	STUDYID, ARMCD, TAETORD	<a href="#">ta.xpt</a>	
TE	<a href="#">Trial Elements</a>	TRIAL DESIGN	One record per planned Element	Tabulation	STUDYID, ETCD	<a href="#">te.xpt</a>	
TI	<a href="#">Trial Inclusion/Exclusion Criteria</a>	TRIAL DESIGN	One record per I/E criterion	Tabulation	STUDYID, IETESTCD	<a href="#">ti.xpt</a>	
TS	<a href="#">Trial Summary</a>	TRIAL DESIGN	One record per trial summary parameter value	Tabulation	STUDYID, TSPARMCD, TSSEQ	<a href="#">ts.xpt</a>	
TV	<a href="#">Trial Visits</a>	TRIAL DESIGN	One record per planned Visit per Arm	Tabulation	STUDYID, VISITNUM, ARMCD	<a href="#">tv.xpt</a>	
DM	<a href="#">Demographics</a>	SPECIAL PURPOSE	One record per subject	Tabulation	STUDYID, USUBJID	<a href="#">dm.xpt</a>	See Reviewer's Guide, Section 2.1 Demographics <a href="#">Reviewers Guide</a>

FIGURE 3: SUMMARY OF TABULATION DATASETS, SUBMISSION-READ DEFINE.XML

### Tabulation Datasets for Study CDISC01 (SDTM-IG 3.1.2)

Dataset	Description	Class	Structure	Purpose	Keys	Location	Documentation
AE	<a href="#">Adverse Events</a>	Events	One record per adverse event per subject	Tabulation	STUDYID, USUBJID, AEDECOD, AESTDTC	<a href="#">Adverse Events SAS transport file</a>	
CM	<a href="#">Concomitant Medications</a>	Interventions	One record per recorded medication occurrence or constant-dosing interval per subject	Tabulation	STUDYID, USUBJID, CMTRT, CMSTDTC	<a href="#">Concomitant Medications SAS transport file</a>	
DA	<a href="#">Drug Accountability</a>	Findings	One record per drug accountability finding per subject	Tabulation	STUDYID, USUBJID, DATESTCD, DADTC	<a href="#">Drug Accountability SAS transport file</a>	
DM	<a href="#">Demographics</a>	Special Purpose Domains	One record per subject	Tabulation	STUDYID, USUBJID	<a href="#">Demographics SAS transport file</a>	
DS	<a href="#">Disposition</a>	Events	One record per disposition status or protocol milestone per subject	Tabulation	STUDYID, USUBJID, DSDECOD, DSSTDTC	<a href="#">Disposition SAS transport file</a>	

FIGURE 4: SUMMARY OF TABULATION DATASETS, DEFAULT CDI DEFINE.XML

Below the list of tabulation datasets, each dataset's metadata is listed. This includes the variable names, labels, key indicator, variable type, and length. It also includes links to controlled terminology codelists within the define.xml document, the origin of each variable, and any derivation notes (computational algorithms) or comments. Figure 5 below displays the TI domain metadata from the submission-ready define.xml file. Notice that the origin is populated, and some of the variables hyperlink to specific pages in the annotated CRF. Figure 6 below displays the TI domain metadata

from the default CDI define.xml file. Three things to note are that the codelist for IECAT is not hyperlinked, Origin is not populated for any variables, and there are extensive comments for every variable. The comment field is being populated with the CDISC Notes from the SDTM Implementation Guide. While these notes and comments are helpful in understanding how to implement the SDTM data model, these comments are generally not appropriate to pass on to regulatory agencies as comments for a specific study. The Origin field is blank for all variables because this is not populated in the SDTM 3.1.2 default data standard since these values are study-specific. The IECAT codelist is specified for the IECAT variable by default in the SDTM 3.1.2 data standard included with CDI. However, this study does not use a codelist for IECAT. Therefore it is not hyperlinked because it is not included in the study's Controlled Terminology Package in CDI and hence is not available to be added to the define.xml file. Additionally, the IETESTCD and IECAT variables are using codelists according to the submission-ready define.xml file. They are in the study's Controlled Terminology Package, but since the SDTM 3.1.2 default data standard does not specify those codelists for those variables, it is not part of the define.xml file.

**Trial Inclusion/Exclusion Criteria (TI)** [Location: [ti.xpt](#)]

Variable	Label	Key	Type	Length	Controlled Terms or Format	Origin	Derivation/Comment
STUDYID	Study Identifier	1	text	7		Protocol	
DOMAIN	Domain Abbreviation		text	2	["TI" = "Trial Inclusion/Exclusion Criteria"] <Domain Abbreviation (TI)>	Assigned	
IETESTCD	Incl/Excl Criterion Short Name	2	text	6	<a href="#">IE Test Codes</a>	Assigned	
IETEST	Inclusion/Exclusion Criterion		text	92	<a href="#">IE Tests</a>	CRF Pages <a href="#">4</a> - <a href="#">5</a>	
IECAT	Inclusion/Exclusion Category		text	9		CRF Pages <a href="#">4</a> - <a href="#">5</a>	

**FIGURE 5: TI DOMAIN METADATA, SUBMISSION-READY DEFINE.XML**

**Trial Inclusion/Exclusion Criteria (TI)** [Location: [Trial Inclusion/Exclusion Criteria SAS transport file](#)]

Variable	Label	Key	Type	Length	Controlled Terms or Format	Origin	Derivation/Comment
STUDYID	Study Identifier	1	text	7			Unique identifier for a study.
DOMAIN	Domain Abbreviation		text	2			Two-character abbreviation for the domain.
IETESTCD	Incl/Excl Criterion Short Name	2	text	6			Short name IETEST. It can be used as a column name when converting a dataset from a vertical to a horizontal format. The value in IETESTCD cannot be longer than 8 characters, nor can it start with a number (e.g., "1TEST"). IETESTCD cannot contain characters other than letters, numbers, or underscores. The prefix "IE" is used to ensure consistency with the IE domain.
IETEST	Inclusion/Exclusion Criterion		text	92			Full text of the inclusion or exclusion criterion. The prefix "IE" is used to ensure consistency with the IE domain.
IECAT	Inclusion/Exclusion Category		text	9	CL.IECAT		Used for categorization of the inclusion or exclusion criteria.

**FIGURE 6: TI DOMAIN METADATA, DEFAULT CDI DEFINE.XML**

Figure 7 and Figure 8 below illustrate additional information found in the tabulation dataset listings. Figure 7 shows the submission-ready define.xml file. The RFSTDTC and RFENDTC variables have computational algorithms specified in the Derivation/Comment field. The AGE variable specifies a computational algorithm and directs the reviewer to the Complex Algorithms external document with a hyperlink. Figure 8 shows the default CDI define.xml file, which again contains the CDISC Notes for comments and no values for the Origin field.

**Demographics (DM)** [Location: [dm.xpt](#)]

Variable	Label	Key	Type	Length	Controlled Terms or Format	Origin	Derivation/Comment
STUDYID	Study Identifier	1	text	7		Protocol	
DOMAIN	Domain Abbreviation		text	2	["DM" = "Demographics"] <Domain Abbreviation (DM)>	Assigned	
USUBJID	Unique Subject Identifier	2	text	14		Derived	Concatenation of STUDYID and SUBJID
SUBJID	Subject Identifier for the Study		text	6		CRF Page <a href="#">3</a>	
RFSTDTC	Subject Reference Start Date/Time		date		ISO8601	Derived	RFSTDTC = first date/time of study drug, for safety subject. Null for screen failures.
RFENDTC	Subject Reference End Date/Time		date		ISO8601	Derived	RFENDTC = termination date, for safety subjects. Null for screen failures.
SITEID	Study Site Identifier		text	3		CRF Page <a href="#">3</a>	
BRTHDTC	Date/Time of Birth		date		ISO8601	CRF Page <a href="#">6</a>	
AGE	Age		integer	2		Derived	Age at Screening Date (Screening Date - Birth date). For the complete algorithm see the referenced external document. Complex Algorithms ( <a href="#">complexalgorithms.pdf</a> )
AGEU	Age Units		text	5		Assigned	Defaulted to YEARS

**FIGURE 7: DM DOMAIN METADATA, SUBMISSION-READY DEFINE.XML**

**Demographics (DM)** [Location: [Demographics SAS transport file](#)]

Variable	Label	Key	Type	Length	Controlled Terms or Format	Origin	Derivation/Comment
STUDYID	Study Identifier	1	text	7			Unique identifier for a study.
DOMAIN	Domain Abbreviation		text	2			Two-character abbreviation for the domain.
USUBJID	Unique Subject Identifier	2	text	14			Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product. This must be a unique number, and could be a compound identifier formed by concatenating STUDYID-SITEID-SUBJID.
SUBJID	Subject Identifier for the Study		text	6			Subject identifier, which must be unique within the study. Often the ID of the subject as recorded on a CRF.
RFSTDTC	Subject Reference Start Date/Time		datetime		ISO8601		Reference Start Date/time for the subject in ISO 8601 character format. Usually equivalent to date/time when subject was first exposed to study treatment. Required for all randomized subjects; will be null for all subjects who did not meet the milestone the date requires, such as screen failures or unassigned subjects.
RFENDTC	Subject Reference End Date/Time		datetime		ISO8601		Reference End Date/time for the subject in ISO 8601 character format. Usually equivalent to the date/time when subject was determined to have ended the trial, and often equivalent to date/time of last exposure to study treatment. Required for all randomized subjects; null for screen failures or unassigned subjects.
SITEID	Study Site Identifier		text	3			Unique identifier for a site within a study.
BRTHDTC	Date/Time of Birth		datetime		ISO8601		Date/time of birth of the subject.
AGE	Age		float	8			Age expressed in AGEU. May be derived from RFSTDTC and BRTHDTC, but BRTHDTC may not be available in all cases (due to subject privacy concerns).

**FIGURE 8: DM DOMAIN METADATA, DEFAULT CDI DEFINE.XML**

SAS Clinical Data Integration 2.6 does offer support for value level metadata, but the procedure for adding it to a study is new in this version. This paper will cover how to add the value level metadata in the following section that describes how to turn the default CDI define.xml file into a submission-ready define.xml file. Figure 9 below illustrates the value level metadata for the SC domain in the submission-ready define.xml file.

### Value Level Metadata - SC [SCORRES]

Variable	Where	Type	Length / Display Format	Controlled Terms or Format	Origin	Derivation/Comment
SCORRES	<a href="#">SCTESTCD</a> EQ EDLEVEL (Education Level)	text	24		CRF Page <a href="#">6</a>	
SCORRES	<a href="#">SCTESTCD</a> EQ MARISTAT (Marital Status)	text	17	<a href="#">Marital Status</a>	CRF Page <a href="#">6</a>	
SCORRES	<a href="#">SCTESTCD</a> EQ SUBJINIT (Subject Initials)	text	3		CRF Page <a href="#">3</a>	

FIGURE 9: VALUE LEVEL METADATA, SUBMISSION-READY DEFINE.XML

The next section of the define file is the computational algorithms. Just like the Origin, these algorithms are study-specific and are not included in the SDTM 3.1.2 data standard that is included with CDI. Figure 10 below shows part of the computational algorithm section from the submission-ready define.xml file. Notice that the Age computational algorithm is identical to the Derivation/Comment for Age in the DM domain metadata in Figure 7.

### Computational Algorithms

Method	Type	Description
Algorithm to derive AEENDY	Computation	AEENDY = AEENDTC - RFSTDTC+1 if AEENDTC is on or after RFSTDTC. AEENDTC - RFSTDTC if AEENDTC precedes RFSTDTC
Algorithm to derive AESTDY	Computation	AESTDY = AESTDTC - RFSTDTC+1 if AESTDTC is on or after RFSTDTC. AESTDTC - RFSTDTC if AESTDTC precedes RFSTDTC
Algorithm to derive the AETRTEM flag	Computation	AETRTEM = "Y" if Adverse Event was not present prior to the RFSTDTC, or it was present prior to the RFSTDTC but increased in severity during the treatment period. Null otherwise.
Algorithm to derive AGE	Computation	Age at Screening Date (Screening Date - Birth date). For the complete algorithm see the referenced external document. Complex Algorithms ( <a href="#">complexalgorithms.pdf</a> )

FIGURE 10: COMPUTATIONAL ALGORITHMS, SUBMISSION-READY DEFINE.XML

The final section in the define.xml file is the comments. The comments listed here are a copy of the comments listed in the Derivation/Comments field in each domain's metadata listing. Notice that the comments from the submission-ready define.xml file in Figure 11 are the comments intended for the regulatory agency, while the comments from the default CDI define.xml file in Figure 12 are the CDISC Notes from the SDTM Implementation Guide.

### Comments

CommentOID	Description
COM.LBREFID	Accession number
COM.RELTYPE	All values are null since this is used only when identifying a dataset-level relationship.
COM.ARMCD	Assigned based on Randomization Number. See Note 2.1 Reviewers Guide ( <a href="#">reviewersguide.pdf</a> )
COM.MHBODSYS	Assigned for Medical History but not Psychiatric History

FIGURE 11: COMMENTS, SUBMISSION-READY DEFINE.XML

### Comments

CommentOID	Description
COM.AE.AEACN	Describes changes to the study treatment as a result of the event. AEACN is specifically for the relationship to study treatment. AEACNOTH is for actions unrelated to dose adjustments of study treatment. Examples of AEACN values include ICH E2B values: DRUG WITHDRAWN, DOSE REDUCED, DOSE INCREASED, DOSE NOT CHANGED, UNKNOWN or NOT APPLICABLE
COM.AE.AEBODSYS	Dictionary derived. Body system or organ class used by the sponsor from the coding dictionary (e.g., MedDRA). When using a multi-axial dictionary such as MedDRA, this should contain the SOC used for the sponsor's analyses and summary tables which may not necessarily be the primary SOC.
COM.AE.AEDECOD	Dictionary-derived text description of AETERM or AEMODIFY. Equivalent to the Preferred Term (PT in MedDRA). The sponsor is expected to provide the dictionary name and version used to map the terms utilizing the define.xml external codelist attributes

FIGURE 12: COMMENTS, DEFAULT CDI DEFINE.XML

Clearly, the default CDI define.xml file needs some work. However, many of the issues highlighted in the previous section are easily resolved by making use of study-specific controlled terminology and custom data standards. I have written papers previously on both topics. [A How-To Guide for Extending Controlled Terminology in SAS Clinical Data Integration](#) discusses how to specify the precise terminology needed for a study in a way that CDI can recognize and use it, including placing the codeslists in the define.xml file. [Managing Custom Data Standards in SAS Clinical Data Integration](#) gives step by step instructions on defining a study-specific data standard that would include the correct values for origin, computational method, comments, and codeslists for a study. This would eliminate most of the manual work that will be described in the following section and is far more efficient.

## HOW TO TURN THE DEFAULT SAS CLINICAL DATA INTEGRATION GENERATED DEFINE.XML FILE INTO A SUBMISSION-READY DEFINE.XML FILE

### DOMAIN PROPERTIES: MODIFYING DOMAIN-LEVEL COMMENTS AND LOCATION DESCRIPTION

To view a domain's properties, double click the domain in the Folders tree in the study. The Notes tab of the domain properties has a field called Quick Note. This value is what is used in the Documentation field in the summary of Tabulation Datasets at the very beginning of the define.xml file. Figure 3 showed a comment for the DM domain directing the reviewer to the Reviewer's Guide. Instructions for adding the hyperlink are described below in the Study Properties section. For the remaining text, just type it into the Quick Notes field (Figure 13).

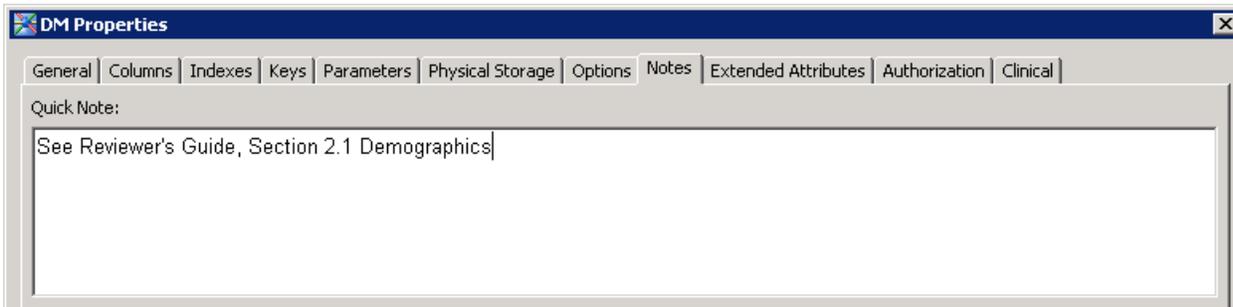


FIGURE 13: QUICK NOTE FIELD IN DOMAIN PROPERTIES

Notice in Figure 3 and Figure 4 the value of the description under Location is different; one document uses the text "dm.xpt" as the hyperlink while the other uses "Demographics SAS Transport File". This value is flexible and may be changed by modifying the Archive Title value on the Clinical tab in the domain properties (Figure 14).

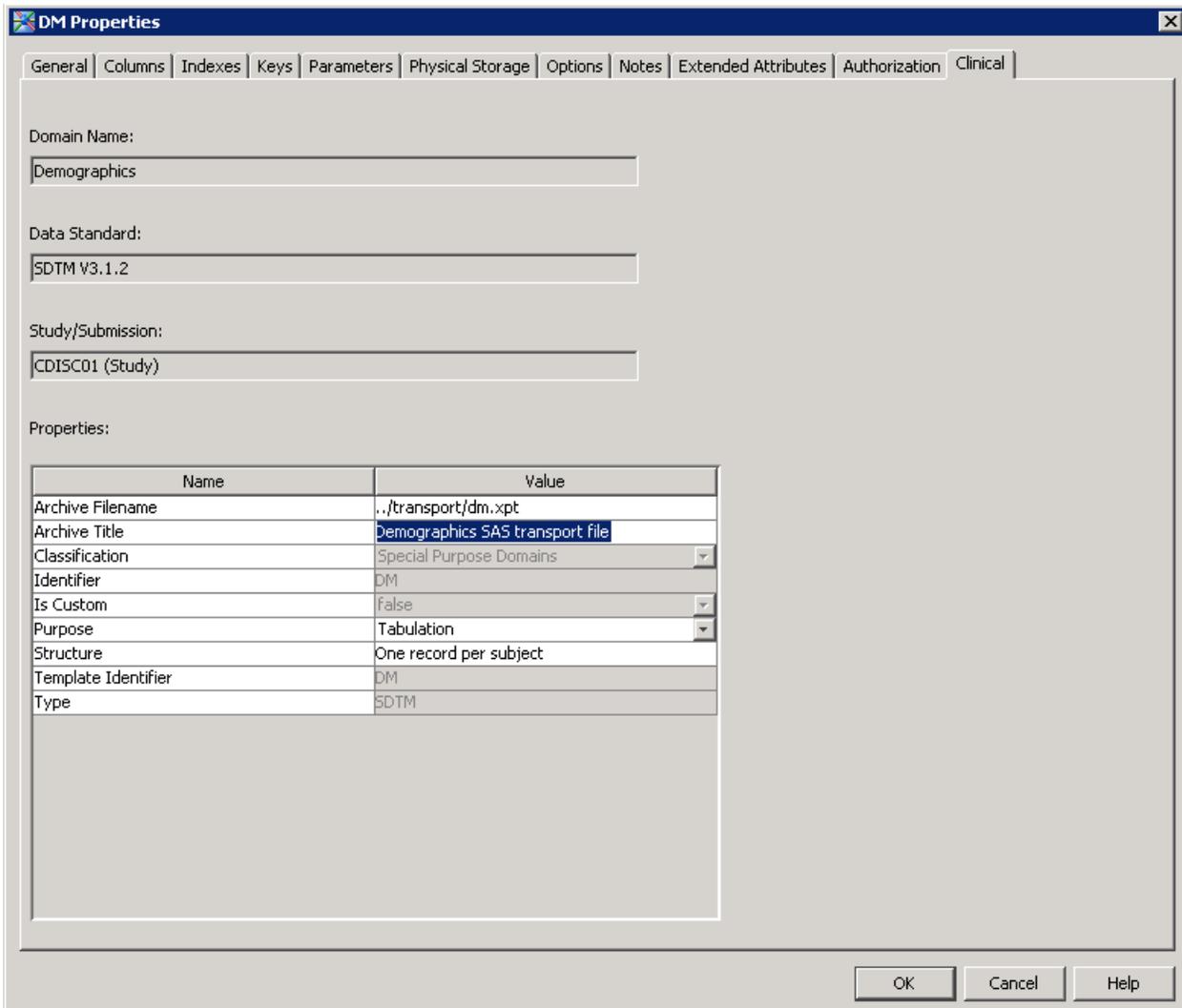


FIGURE 14: ARCHIVE TITLE FIELD IN DOMAIN PROPERTIES

### COLUMN PROPERTIES: MODIFYING VARIABLE-LEVEL COMMENTS, COMPUTATIONAL ALGORITHMS, ORIGIN, AND CODELISTS USED

To view a column's properties, select the column from the Columns tab of the domain properties, right click, and select Properties. Similarly to domain properties, there is a Quick Note field on the Notes tab of the column properties, which is used as the Derivation/Comment value when it is populated. Delete the CDISC Note populated by default. Replace it with a comment for the reviewer, if necessary (Figure 15).

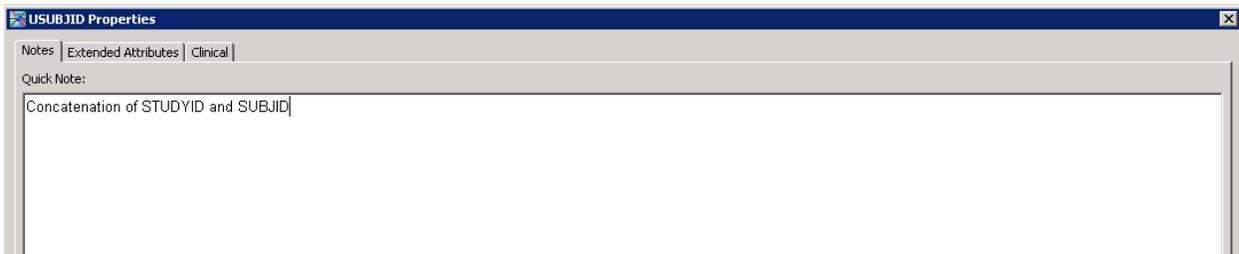


FIGURE 15: QUICK NOTE FIELD IN COLUMN PROPERTIES

On the Clinical tab, several more properties affect the define.xml file. The Method field is empty by default; enter the computational algorithm for a variable in this field. If the field needs to be hyperlinked to an external document for

complex algorithms, that will be described in the Study Properties section below. Just add the text for the algorithm, leaving off the hyperlink's text (Figure 16).

The Origin field is also empty by default; enter the origin for the variable (Figure 16). If the Origin is to be hyperlinked to the annotated CRF, the value entered in the Origin field should be CRF (Figure 17). The following section, Study Properties, describes how to specify the actual pages to be hyperlinked

The screenshot shows the 'AGE Properties' dialog box with the 'Clinical' tab selected. The 'Label' field contains 'Age'. Below it is a 'Properties' table with the following data:

Name	Value
Contributes to Key	false
Core	Exp
Display Format	8.1
Method	Age at Screening Date (Screening Date - Birth
Origin	Derived
Prefix by Identifier	false
Qualifiers	
Role	RecordQualifier
Term	
Type	SDTM
XML Codelist	
XML Type	float

**FIGURE 16: METHOD, ORIGIN FIELD IN DOMAIN PROPERTIES**

The Term and XML Codelist fields specify the controlled terminology codelist the variable uses. Many variables have default values set for these by the SDTM 3.1.2 data standard. However, a study may not use that default. It may not use any codelist at all, or it may use a different codelist. Some variables that do not have a default codelist specified do actually use a codelist. To specify the codelist to use for a study, enter the codelist name in the XML Codelist field and also in the Term field, but in the Term field it should be enclosed in parenthesis (Figure 17). Delete both fields if the variable does not use a codelist, but a default one has been specified.

The screenshot shows the 'IETEST Properties' dialog box with the 'Clinical' tab selected. The 'Label' field contains 'Inclusion/Exclusion Criterion'. Below it is a 'Properties' table with the following data:

Name	Value
Contributes to Key	false
Core	Req
Display Format	
Method	
Origin	CRF
Prefix by Identifier	false
Qualifiers	MIXEDCASE
Role	SynonymQualifier
Term	(IETEST)
Type	SDTM
XML Codelist	IETEST
XML Type	text

**FIGURE 17: ORIGIN, TERM, XML CODELIST FIELDS IN COLUMN PROPERTIES**

## STUDY PROPERTIES: ADDING VALUE LEVEL METADATA AND EXTERNAL DOCUMENTS

Newly available in CDI 2.6 is the ability to add value level metadata and supplemental documents from within the CDI application. If you open a Study's Properties and select the Study tab, you will find fields to add a Value Level Metadata Table and a Supplemental Documents Table. Selecting the Add... button will walk you through a wizard asking you to specify the library and location in the Folder tree for the table. It will then create a zero-observation dataset that must then be populated with the value level metadata or supplemental document information intended for the define.xml file.

The screenshot shows the 'Study Properties' dialog box with the 'Study' tab selected. The 'Root folder' is set to '/ClinicalDI Course/CDISC01'. The 'Properties' table is as follows:

Name	Value
Protocol Phase	Phase II
Protocol Title	Study CDISC01
Short Protocol Title	CDISC01
Study Identifier	795c2618-cf7e-418a-a464-2b82cdc0d073

Below the table, the 'Terminology Package' is 'CDISC01'. There are 'Add...' and 'Remove' buttons for the Terminology Package, Value Level Metadata Table (currently 'SOURCE\_VALUES'), and Supplemental Documents Table (currently 'SOURCE\_DOCUMENTS'). 'OK' and 'Cancel' buttons are at the bottom.

FIGURE 18: STUDY PROPERTIES – VALUE LEVEL METADATA AND SUPPLEMENTAL DOCUMENTS TABLES

## CDISC-DEFINE CREATION PROPERTIES: ORDERING DOMAINS FOR TABULATION DATASETS LISTING ORDER

On the tables tab of the CDISC-Define Creation transformation, on the bottom right, there is a button labeled "Order Tables". Selecting this button allows the user to specify the order in which the domains will appear in the Tabulation Datasets section of the define.xml document. The default is alphabetical order, but domains may be moved using the Move Up and Move Down buttons (Figure 19).

The 'Order Tables' dialog box contains a list of domains: AE (Adverse Events), CM (Concomitant Medications), DA (Drug Accountability), DM (Demographics), DS (Disposition), EG (ECG Test Results), and EX (Exposure). Below the list are 'Move Up' and 'Move Down' buttons. 'OK', 'Cancel', and 'Help' buttons are at the bottom.

FIGURE 19: ORDER TABLES DIALOG BOX

## CONCLUSION

SAS Clinical Data Integration makes the generation of a define.xml file simple. The built-in CDISC-Define Creation transformation does the work of reading the study, domain, and column metadata stored inside CDI, using SAS Clinical Standards Toolkit macros to generate the xml document. When using the CDISC standard included with CDI, some of the pieces of a submission-ready define.xml file are missing, such as the origin, hyperlinks to external documents, computational algorithms, and value level metadata. By making use of study-specific controlled terminology codelists, study-specific data standards, and the new feature that supports value level metadata and supplemental documents, CDI can generate a fully submission-ready define.xml file.

## ACKNOWLEDGMENTS

Thank you to Lex Jansen for the guidance and explanations of SAS Clinical Standards Toolkit and its generation of Define 2.0 files and to Angela Lightfoot for always providing the best guidance and support.

## RECOMMENDED READING

- [A How-To Guide for Extending Controlled Terminology in SAS Clinical Data Integration](#)
- [Managing Custom Data Standards in SAS Clinical Data Integration](#)

## CONTACT INFORMATION

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

Melissa R. Martinez  
SAS Institute Inc.  
720 SAS Campus Dr.  
Cary, NC 27513, USA  
Work Phone: +1 (919) 531-9277  
E-mail: [Melissa.Martinez@sas.com](mailto:Melissa.Martinez@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.