Metadata Submission Guidelines

Appendix to the Study Data Tabulation Model Implementation Guide

Prepared by the CDISC SDS Metadata Team

Notes to Readers

This is Version 0.9 of the Metadata Submissions Guidelines, posted for comment by the CDISC Submissions Data Standards Metadata sub team.

Revision History

Date	Version	Summary of Changes
2007-07-25	0.9	• Original draft for comment
		•
		•
		•

1 Introduction

1.1 PURPOSE

The purpose of the Metadata Submission Guidelines, an appendix to the SDTM Implementation Guide (SDTMIG), is to provide an example for sponsors when compiling an electronic submission using the Study Data Tabulation Model (SDTM). This Appendix, which is comprised of this document in conjunction with study-related files, identifies and describes the components highly recommended for an electronic submission of SDTM data. The study-related files associated with this appendix include an annotated CRF, representative sample submission datasets compliant with version 3.1.1 of the SDS Implementation Guide (IG), and the metadata describing the format and content of the submitted datasets.

The SDTM appendix will enable users to become familiar with not only SDTM domains, but also the accompanying submission components. This entire package is meant to illustrate acceptable practices and formats that sponsors may incorporate into their own submissions, but is not intended to dictate the only acceptable practices. The scope of this document and associated study-related files is the SDTM portion of a submission.

In order to view the electronic component of this appendix, download the sample study from the CDISC website (http://www.cdisc.org/) and extract all the components to the same folder. Storing all the files together is essential to preserve the inter-document linking that has been added to the submission components.

1.2 REFERENCES AND ABBREVIATIONS

The following are abbreviations for the documents referenced within this document and the links to the current versions:

- **SDTMIG** — **CDISC SDTM Implementation Guide Version 3.1.1**
  http://www.cdisc.org/models/sdtm/v1.1/index.html
- **SDTM 1.1** — **Study Data Tabulation Model (SDTM) Final Version 1.1**
  http://www.cdisc.org/models/sdtm/v1.1/index.html
- **CRT-DDS (Define.xml)** — **Case report Tabulation Data Definition Specification Version** 1.0
  http://www.cdisc.org/models/def/v1.0/index.html
- **FDA Guidance** — **Electronic Common Technical Document (eCTD)**
  http://www.fda.gov/cder/regulatory/ersr/ectd.htm
  Study Data Specifications Version 1.3
  http://www.fda.gov/cder/regulatory/ersr/Studydata-v1.3.pdf
- **CDISC website** — http://www.cdisc.org/

1.3 ORGANIZATION OF THIS DOCUMENT

This document has been organized into the following sections to facilitate review and understanding of the submission components.

- **Section 1, INTRODUCTION**, delivers an introduction and outlines the organization of this document.
- **Section 2, GENERAL SPECIFICATIONS FOR SUBMITTING DATA**, describes the components that are part of an SDTM submission.

- **Section 3, DATA DEFINITION FILE (DEFINE.XML)**, explains the definition portion of the submission, the define.xml. Descriptions of the various components including the tables, table content, and links are discussed. This section is intended to describe the organization of the define.xml and is not a technical guide to the XML portion of the define.xml. The technical information can be found in the Define.xml document on the CDISC website.
- **Section 4, GUIDELINES FOR ANNOTATING CRFs**, provides guidelines to be used when annotating CRFs according to the SDTM specifications. Recommendations on both content and format are included.

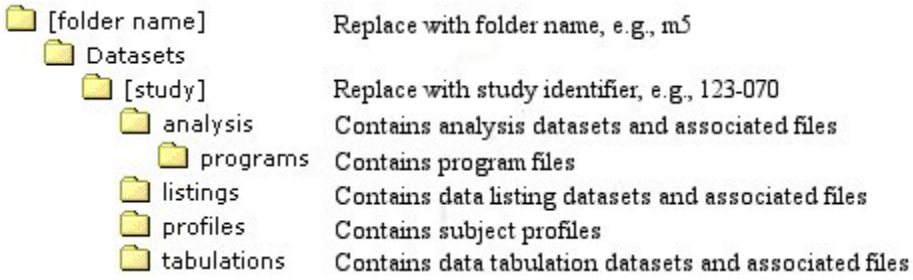
Starting in Section 5 and continuing through Section 11, the SDTM domains contained in this mock submission are explained in detail. The order in which the domains are covered follows the order of the domains as they appear in the define.xml document. The domain definitions are compliant with CDISC SDTM Implementation Guide Version 3.1.1.

- **Section 5, TRIAL DESIGN DOMAINS**, gives examples of both trial- and subject-level domains as defined in the IG Version 3.1.1
- **Section 6, ALL SUBJECT-RELATED DOMAINS**, includes information related to an electronic submission that is applicable to the subject-related domains (excludes Trial Domains). Section 6 specifications are applicable across all subject-related domains whereas the specifications in Sections 7-11 apply to the domain class described in the section and the individual domains within the class
- **Section 7, SPECIAL PURPOSE,** includes information on the Demographics (DM) and Comments (CO) domains.
- **Section 8, INTERVENTIONS**, includes examples of domains in the Interventions class. The Intervention domains capture investigational treatments, therapeutic treatments, and surgical procedures that are intentionally administered to the subject (with some actual or expected physiological effect) either as specified by the study protocol (e.g., “exposure”), coincident with the study assessment period (e.g., “concomitant medications”), or other substances self-administered by the subject (such as alcohol, tobacco, or caffeine).
- **Section 9, EVENTS**, includes examples of domains in the Events class. The Events domains capture occurrences or incidents independent of planned study evaluations occurring during the trial (e.g., 'adverse events' or 'disposition') or prior to the trial (e.g., 'medical history').
- **Section 10, FINDINGS**, includes examples of domains in the Findings class. The Findings domains capture the observations resulting from planned evaluations to address specific questions such as observations made during a physical examination, laboratory tests, ECG testing, and sets of individual questions listed on questionnaires.
- **Section 11, RELATIONSHIP DATASETS**, includes examples of domains that represent relationships among datasets or records, such as relationships between separate domains, or between domains and other special-purpose datasets.

2 General Specifications for Submitting SDTM

2.1 SUBMISSION STRUCTURE

The latest FDA Guidance recommends the following folder structure for the data portion of the eCTD. The Study Data Tabulation Model (SDTM) datasets, definitions, and associated documentation are located in the Tabulations folder.



2.2 DATA TABULATION DATASETS

The Tabulations folder is reserved for datasets conforming to the SDTM standard as defined by CDISC and defined in the SDTMIG. The SDTM datasets provide a framework for organizing and submitting observations collected about the subjects participating in a clinical trial. The SDTM tabulations contain collected data as well as several derived data elements.

2.3 ANNOTATED CRF

The complete CRF from the study should be saved in a pdf document called blankcrf.pdf. Within the pdf all unique CRF pages or forms, visits, and items should be annotated to match the SDTM domains and variables. A page number or similar reference and link to the associated annotated pages should be included on repeated, non-unique, non-annotated CRF pages. The blankcrf.pdf file is stored in the tabulations folder along with the tabulation datasets.

3 Data Definition File (Metadata)

3.1 INTRODUCTION

The data definition file is the metadata describing the format and content of the submitted datasets. It is divided into 2 parts: the metadata definition of the datasets or Table of Contents (TOC) and the metadata definition of the variables within the datasets or Data Definition Table (DDT).

3.2 METADATA FOR DOMAIN DATASETS

The purpose of the domain level metadata is to identify and provide basic information about each of the datasets included in the Tabulation folder. Essentially the domain level metadata is the Table of Contents (TOC) for the tabulation datasets. The format of the metadata is predefined and can be found in the CRT-DDS (http://www.cdisc.org/models/def/v1.0/index.html).

In the event that no records are present in a dataset (e.g., a small PK study where no subjects took concomitant medications), the empty dataset should not be submitted or included in the Define .xml. The annotated CRF will show the data that would have been submitted had data been received; it need not be re-annotated to indicate that no records exist.

3.2.1 Organization

All tabulation datasets must be included in the domain level metadata data definition. The metadata team recommends that the entries within the TOC be organized by the SDTM general class as defined in the **SDTMIG.** It is possible to have both CDISC defined standard domains and sponsor-defined domains within a submission. CDISC defined SDTM domains would appear alphabetically within each class. If sponsor-defined domain(s) are present in the submission, it is recommended that they appear within the defined class following the CDISC defined domains. If there are multiple sponsor-defined domains within a class they should be stored alphabetically.

The recommended order for the classes within the metadata is:

- Trial Design Domains
- Special Purpose Domains
- Intervention Domains
- Event Domains
- Finding Domains
- Relationship Domains

3.2.2 Content

The content of the TOC is predefined in the Define.xml schema and the presentation format is defined through the associated style sheet.

Table 3.1.2.1: Domain Level Data Definition (TOC)

Field	Description
Dataset	The file name of the dataset or data domain name (e.g., "DM", "AE").
Description	A short description of the type of information contained within the dataset (e.g. "Demographics", "Adverse Events"). A hyperlink is defined from the description field to the metadata for the dataset.
Class	The general class, as defined in the SDTM model, of the observations within the domain (e.g. “Events”, “Special Purpose”).
Structure	The level of detail represented by individual records in the dataset (e.g., "One record per subject", "One record per subject per visit", "One record per subject per event").
Purpose	The purpose for the dataset. All SDTM domains have purpose equal to "Tabulation”.
Keys	The domain level keys are a series of variables whose primary purpose is to uniquely identify a data record.
Location	This contains the Folder and filename where the dataset can be found and a hyperlink to the data.

Structure

The description of structure should match the dataset, and, for domains included in the IG, may provide more detail than the description of structure in the IG. The AE entry in Table 3.1.2.2 is an example of a structure description that is more detailed than that in the IG. This example is not intended to recommend a structure description different from that in the IG, but to illustrate the fact that a different description is possible.

Class

The metadata team recommends that the field, “Class”, be present in the submission stylesheet. Although Class is present within the XML schema it is not always specified as part of the stylesheet. The domain Class implies a predefined set of variables and should be helpful in the review of the data.

Keys

As stated in Table 3.1.2.1 the primary purpose of the keys specified in the domain-level table is to provide uniqueness within the domain. Additionally the key variables are sort keys for the domain however they do not necessarily reflect the sort order of the submission dataset. The keys serve as an aide to reviewers in understanding the domain structure and potentially defining a sort order,

The keys shown in Table 3.1.2.2 are only examples and sponsors should organize their datasets according to their preferences and data requirements. For subject related data, the first 2 keys are always STUDYID and USUBJID. These keys are often followed by timing variables, which may be a date for the Events and Interventions domain classes, and a visit number and/or time point for Findings class domains.

In the following table the SDTM defined datasets are listed alphabetically within class. This example shows 3 SDTM defined domains within the Events class in alphabetical order: Adverse Events, Disposition, and Medical History. If a sponsor-defined Events domain was included in this submission, it would be inserted after the Medical History domain and before the first domain in the Findings Class, ECG.

Table 3.1.2.2: Domain Level Data Definition (TOC) Sample

Dataset	Description	Class	Structure	Purpose	Keys	Location
TV	Trial Visits	TRIAL DESIGN	One record per planned visit per arm	Tabulation	STUDYID, VISITNUM, ARMCD	tv.xpt
AE	Adverse Events	EVENTS	One record per event per subject An event is defined as the occurrence of a new term or the increased severity of an existing term.	Tabulation	STUDYID, USUBJID, AESTDTC AETERM	ae.xpt
DS	Disposition	EVENTS	One record per disposition status or protocol milestone per subject	Tabulation	STUDYID, USUBJID, DSSTDTC, DSCAT, DSTERM	ds.xpt
MH	Medical History	EVENTS	One record per medical history event per subject	Tabulation	STUDYID, USUBJID, MHCAT, MHTERM, MHSTDTC, MHENDTC	mh.xpt
EG	ECG	FINDINGS	One record per ECG observation per visit per subject	Tabulation	STUDYID, USUBJID, EGDTC, VISITNUM, EGCAT, EGSPID, EGTESTCD	eg.xpt
VS	Vital Signs	FINDINGS	One record per vital sign measurement per visit per subject	Tabulation	STUDYID, USUBJID, VSDTC, VISITNUM, VSSPID, VSTESTCD, VSPOS	vs.xpt

Note For each of the datasets within the TOC the Description column links to the data definition table for the dataset. The location column links directly to the xport dataset and enables review of the data.

3.3 METADATA FOR DOMAIN CONTENT

The metadata describing the variables within the domains is called the Data Definition Table. This table contains the variable level attributes, descriptions, and usage rules for each variable within each dataset.

3.3.1 Organization

Every entry in the TOC will have an associated Data Definition table. The TOC dataset (Table 3.1.2.2) Description column has hyperlinks to the corresponding variable-level metadata tables. The variable-level definition tables will be in the same order as the domain-level metadata.

The variables within each define table must be ordered according to the sequence specified in the current version of the IG. In addition, the order of the variables in the define.xml data definition tables and the order of the variables within the transport file records must be identical.

3.3.2 Content

The content of the Data Definition Tables is predefined in the Define.xml schema and the format of the display is controlled by the associated style sheet.

The data definition table contains the following items:

Table 3.2.2.1: Data Definition Table

Field	Description
Variable	The name of the variable.
Label	A brief description of the variable
Type	The variable type (e.g., "text", "float", ”date”) as defined in the CRT-DDS section 2.1.2.4.
Controlled Terms Format	The set of controlled terms or variable display information. If controlled terms exist for an item, a hyperlink to the terms is present.
Origin	Indicator of the origin of the variable. Examples could include Case Report Form (CRF) page numbers or form reference, “Derived”, "eDT", "Assigned", or "Protocol."
Role	Information on how a variable is used within the dataset (e.g., "Identifier", "Topic", "Timing", "Qualifier").
Comments	Other information regarding the variable definition, usage, etc.

Variable

In all cases, this column contains the name of the variable within the SDTM dataset. In some domains a sponsor may include value-level metadata within the define.xml to facilitate review and analysis. When value-level metadata is included, the link to the value-level table is from the variable name. The value level metadata table contains all possible values and corresponding data attributes.

Type

Type is the data type of the submitted variable. Eventually data will be submitted to the FDA through ODM and Type will reflect the ODM data type, however currently data are contained in SAS transport files. Submitting data using SAS transport files does restrict the choices within Type. When dates are in ISO-8601 format and submitted using SAS transport files, the Define.xml schema must define the Type for date variables as “text” rather than “date”.

Controlled Terms Format

Controlled terms and formats must be specified in the define.xml. The link from the controlled terms column in the style sheet displays a table of terms. Currently, whether using CDISC controlled terminology, or sponsor-defined terms, the metadata team recommends that all possible values be included within the define.xml. If the controlled terminology reference is extremely large (i.e., MedDRA), then it is recommended to use an external link within the

define.xml. However, the sample submission in this appendix does not include external references to proprietary items.

.

Origin

If the origin is designated as “Derived”, then the derivation definition must be specified within Comments column. If the origin is a page number or form on the annotated CRF then a link to the corresponding page or form within blankcrf.pdf will be provided in Origin.

Comments

Comments are primarily used for defining data derivations. There are multiple ways to specify derivations within the Comments column. The traditional way is to describe the derivation logic or formula for a variable in the Comments column and this is still an accepted method of providing derivation documentation.

Another way to provide a derivation is to specify a link to a “computational method” in the Comments column. Within the define.xml the computational method, or formula for the derivation, is stored separately from the variable and may be linked to multiple times in the SDTM submission. One or more variables within the same domain, or one or more variables across domains, can be linked through the Comments column to the same computational method.

If the derivations for certain variables within a domain are extremely large and involve complex logic or diagrams, the comments column could be used to provide a link to the derivation description. The actual derivation could be stored either within the define.xml or in a separate pdf file. .

Core

For domains described in the IG, the dataset must comply with the requirements described by the values in the "Core" column of the domain table. All "Required" and "Expected" variables must be present, and for required variables, every record must be populated

3.4 CONTROLLED TERMINOLOGY

Sponsor defined, CDISC defined, and external source are all types of controlled terminology that can be referenced within the STDM metadata. Sponsor defined terminology and CDISC defined terminology can overlap.

3.4.1 CDISC Controlled Terminology

Currently there is a phased approach for implementing controlled terminology within the SDTM model. For additional information on codelists to support CDISC standards see the CDISC website. Whenever possible it is recommended that the sponsor use the controlled terminology defined by CDISC. Obviously this has to be a planned transition for the sponsor but the metadata team recommends that this be done as soon as feasible.

3.4.2 External Controlled Terminology

External dictionary controlled terminology will be specified in the comments column of the metadata. A reference to the external dictionary will be embedded within the define.xml schema.

3.5 DEFINE.XML

3.5.1 ItemRef Attribute “Mandatory”

The ItemRef attribute, Mandatory, and the defined Controlled terminology for an item must be consistent within the define.xml. For example, if null is an acceptable entry for a field, and controlled terminology is defined for the item, the Mandatory attribute must be equal to “No”. If mandatory equal to “Yes” is specified for a variable then all occurrences of the variable must contain a non-null value

3.5.2 XML Style Sheets

A style sheet is a tool used to display an XML document, in this case the define.xml, in a meaningful way within a web-browser. While the XML is designed to house the data, the style sheet is designed to display the data. Style sheets are required when submitting define.xml files to the FDA.

3.5.3 Style Sheet Implemented in this Example

For this submission example a custom style sheet, developed and enhanced by several of the Metadata Team members, was used. This style sheet adds functionality, particularly the bookmarks, that the Metadata Team felt would be useful to a reviewer.

3.5.4 What Style Sheets are Available

There is a basic style sheet available for download at the CDISC website, and sponsors can also feel free to use the style sheet included in this example. However, CDISC will not be developing or maintaining style sheets in the future. Sponsors also have the option of developing their own style sheet.

4 Guidelines for Annotating CRFs

4.1 OVERVIEW AND ASSUMPTIONS

When adding annotations, it should be kept in mind that the annotations are meant to help the reviewer at the FDA find the origin of data variables found in the datasets (domains). Annotations are presented to relate the data submitted to its origin on the CRF or other collection instrument. To this end, the annotations should be text based and should be searchable using standard PDF viewers such as Adobe Acrobat Reader, Acrobat Standard, or Acrobat Professional. Acrobat Professional is the tool currently used by the FDA for reviewing PDF documents.

4.2 STANDARDS AND TOOLS

This document assumes, as strongly suggested by the FDA's e-submission guidance and specifications (since 1999), that the annotated CRF is to be provided to the FDA as a PDF file, "blankcrf.pdf", which is used to supplement the define.xml and the datasets. While the metadata team annotated the CRF with Adobe Acrobat 7 Professional, using the comments functionality, sponsors may choose another tool to create the annotation depending upon their business practice. The annotation should be searchable and meet the FDA's guidance and specifications, The annotations should not be handwritten on scanned pages, as such annotations are often illegible and can lead to confusion for the reviewer.

4.3 BASIC PRINCIPLES FOR ANNOTATIONS

Annotations in the blankcrf should appear as simple and as clean to the eye as possible, so that reviewers can find and read them easily. As mentioned earlier, they should be text-based and should be searchable.

The recommended annotation/comment to be used is the "free text" annotation. This may be created in Adobe software using the Text Box tool, found in version 5 or 6 on the Advanced Commenting Toolbar or in version 7 on the Drawing Markups Toolbar.

Since the blankcrf.pdf is supporting the review process, the annotations should reflect the data that is being submitted within the SDTM. Annotations of the operational data, while normally needed by the sponsor for data management, should not be included with the submission. The annotator should avoid adding these additional annotations that were used in the operational database for other internal information sharing (such as domain designations for panels).

4.3.1 Annotating unique CRF pages

The metadata team recommends that each unique occurrence of submitted data be annotated on the CRF. Repetitive pages should refer to the original page that was annotated. In the define.xml Origin column there will be a link to the page where the annotated variables appears. Although the origin will reference all the pages where the variable occurs only the first page will be annotated. This is illustrated in the Metadata sample submission.

The metadata team recommends that when data is recorded on the CRF but is not submitted the CRF be annotated with the text "NOT SUBMITTED". For example, data which were used for operational purposes only, such as trigger questions, will be annotated "[NOT SUBMITTED]" but the data will not be included as part of the domains. There are brackets around "[NOT SUBMIITED]" in the submission sample. The brackets are not required but do help to distinguish the submitted versus not submitted items.

4.3.2 Appearance of Annotation

Since the annotations are created to assist the reviewer in understanding the relationship between the data origin and the variable used for submission, the annotations should not obstruct features of the original collection instrument (CRF/eCRF), such as printed text or entry fields which display printed text. As mentioned before, extraneous annotations should be omitted from the blankcrf. Remember that a clean appearance is preferred to a cluttered one. One reason the metadata team recommends using "free text" annotations is that the text box that is used to bound or contain these comments may be sized to display as much or as little text from the underlying comment as is deemed desirable. A "free text" comment can also be formatted using specific fonts and colors, and can include special characters.

The blankcrf should be prepared in compliance with FDA recommendations. In keeping with the FDA's 2005 "PDF Specifications" document and other PDF recommendations, the team recommends that the comments be created using an Arial Bold Italic font. This font should stand out from other text printed on the collection instrument. The text size should be equivalent in size to 12 point Times Roman, if possible. The font should not be smaller than 9 point Times Roman or equivalent. All text in the annotations that represent variable names should be capitalized.

It is also recommended that the color of the annotation font be a color not used for text on the page (i.e. dark red, dark blue, or purple). To further differentiate the comments from the CRF text select a contrasting color as the "Fill Color" for the comment box. (When selecting colors and contrasting colors, keep in mind that the most common form of color blindness is "Red-Green" which causes those with this color blindness to have difficulty distinguishing between red and green. If there is red or green content on the CRF, these colors should be avoided in the annotations.) Gray fill color should be avoided, as many printed CRFs have gray boxes printed on the page and any gray-filled comments would not be readily apparent as annotations. In the metadata team sample there are borders on the comment boxes although this is not necessary. It may be worth noting that, by using the CTL key, more than one comment may be selected which making setting attributes such as fill color easier by setting multiple comments' formats simultaneously.

4.3.3 Annotating each variable

For the most part, there should be a one-to-one relationship between the data fields on the CRF and the annotations. In some cases it may be necessary to annotate one data collection field using two annotations, such as for a question and a response. When a question and response must be annotated, the metadata team recommends annotating the topic variable, --TESTCD, followed by the result variable. The variable name, by itself, may not express enough information to make the annotation meaningful. In these cases the variable, plus some description about the variables value may be necessary. An example would be in the case of the Vital Signs test for Pulse or Heart Rate (as differentiated from other vital signs tests such as Temperature, Systolic BP, or Diastolic BP, etc.) where the topic variable of VSTESTCD may be annotated as "VSTESTCD when VSTESTCD="PULSE"". The associated result variable of VSORRES may be annotated as "VSORRES when VSTESTCD="PULSE" and VSORRESU may also be annotated as "VSORRESU when VSTESTCD="PULSE". Note that each of these examples uses the topic variable's value to differentiate the current instance of the annotated variable from other tests. In some cases, if the CRF page is extremely full, you may decide that creating a table for annotating the test codes will save space.

Note that it is not necessary to annotate both the test code as well as the test code description. The test code description may be found in the DEFINE file and is redundant to the Topic variable, so the test code should be the variable annotated when describing the topic.

4.3.4 Annotating using wildcard characters for the domain(s)

Some data collection fields, such as dates, may relate to variables in more than one domain, which could result in multiple similar annotations for a single field. An example of this is when a date is filled in at the top of a CRF, and this date relates to more than one domain of data collected on the page. In this case, the metadata team recommends that the comment "free text" start with the text "--", as used in the SDTM Implementation Guide, as a "wildcard", to generically represent the 2-letter domain prefix, followed by the common or shared variable designator (such as "DTC"). The comment string should continue with a comma-delimited string of specific variable enclosed in square brackets. For example, data from the SC domain, the DS domain and the DM domain might all be collected on one page with a single date field. The date field would be annotated with "--DTC [SCDTC, DSDTC, DMDTC]". Providing this information in this manner will allow the reviewer to search for each of the domain variables (using Adobe Acrobat and including the search for comments), and to be able to find any of the distinct domain variables. To provide a "clean" or simple presentation to the reviewer, it is recommended that the comment

box be re-sized such that only the first variable or "common" variable of "*--DTC*" be initially displayed. A reviewer may see the entire comment by opening the Adobe Acrobat "comments" navigation tab, or by double-clicking on the comment.

4.3.5 Annotating Supplemental Qualifier variables

The metadata team recommends that supplemental qualifier variables be annotated using a two part name, with the two parts separated by a dot (.). The first part of the name should be "SUPP--" where "--" indicates the 2-letter domain designation for the domain supplemented. The second part of the name should be the exact value of QNAM used in the supplemental qualifier dataset. An example from the sample submission is the randomization flag, “RAND”, which is supplemental to the DM domain and annotated as "SUPPDM.QVAL when QNAM="RAND"”. While this annotation is different from the other, non-Supplemental-Qualifier annotations, the two part name is important to indicate that the variable is not part of a standard domain. If using Adobe version 7 or higher, in addition to the proscribed comment text, the variables should have the "subject" field populated with the "SUPP--" domain name as well as the 2-character domain which is being supplemented. (These would be "SUPPDM and DM" for the example given above.) This will provide a reviewer who searches for the 2-character "parent" domain a search result which includes the variables in this supplemental file that are related to the parent domain.

4.3.6 Annotating RELREC variables

The metadata team recognizes that some variables may be identified as belonging to a particular domain, but may have a relationship identified within the RELREC domain. Since the RELREC domain typically contains a minimal amount of information about the parent record, it may be helpful for the reviewer to annotate, when possible, to show that some data collection sources are indicative of the existence of RELREC records and subsequent relationships.

The metadata team recommends that when a variable is identified as probably or always qualifying it as inclusion in RELREC, the variable should be annotated with the variable name, any qualifying statements as needed (such as "when" statements as described in the section above titled "Annotating each variable) followed by the statement "consequently -- records exist in RELREC" where "--" represent the 2-letter domain prefix.

An example of this would be a subject disposition page where a field for "Death" is marked and the Adverse Event number associated with that death is reported. The DSTERM annotation for the "Death" would read "DSTERM="DEATH" consequently DS records exist in RELREC" and the Adverse Event number annotation would read "AESPID when DSTERM="DEATH" consequently AE records exist in RELREC". These two annotations would provide the reviewer some assistance in finding the source for the RELREC records as well as pointing to a RELREC relationship existing and documented in the RELREC domain.

4.3.7 Using Adobe Acrobat for annotations and annotation attribute

The FDA reviewers use Adobe Acrobat to review PDF files such as the BLANKCRF.PDF. Annotations created as PDF comments may be searched, filtered by attribute or printed. These comments may also be exported as a separate file such as on an XML file. While the current tools and review processes do not have any identified use for this extra functionality, the presence of these review features may provide additional functionality in the future, as their use becomes more broadly accepted.

5 Trial Design

The domains associated with the trials describe the overall characteristics of the study. Presently there are five domains defined within trial design.

5.1 TRIAL ARMS (TA)

5.1.1 General Considerations

- After study (STUDYID) the keys guaranteeing uniqueness for the Trial Arms (TA) domain are arm code (ARMCD) and TAETORD. Each ARMCD and the associated arm description (ARM) present in the study are defined in the TA domain. All randomized subjects or subjects receiving treatment in an open label study will be assigned an ARMCD in the Demographics (DM) domain. Each study arm defined in TA is comprised of one or more elements. All possible elements are defined within the Trial Element (TE) domain and referenced in the TA. The order that the elements occur within each arm is based on the ARMCD assigned and is defined in the Trial Arms domain using the required variable TAETORD.

- Within this sample study there are three trial arms: low-dose study drug, high-dose study drug, and placebo. This study is a straightforward double-blind design with each arm consisting of two elements: screening and randomized treatment

5.2 TRIAL ELEMENTS (TE)

5.2.1 General Considerations

The Trial Elements (TE) domain defines the “building blocks” of the trial. Each element has one record in the Trial element dataset, even if it is used in multiple arms of the study, or multiple times within one or more arms. For example, the element Screening occurs within each of the three arms of the study, as defined in the TA domain; however the element Screening is defined once within the Trial Element domain. The order that the elements are performed within each study arm is defined in the TA table.

5.3 TRIAL VISITS (TV)

5.3.1 General Considerations

Each planned visit within the study is defined in the Trial Visit (TV) domain. Each visit must be assigned a unique number (VISITNUM) and normally will have an associated visit description (VISIT). Usually the visit numbers are assigned in sequential order based on the chronological visit sequence. .

5.3.2 Define.xml

In studies that have a different visit structure based on the assigned arm code, the ARMCD and ARM will be specified within the TV domain. In the sample submission, the visit structure is the same for all arm codes, and the ARMCD and ARM variables will be null within TV. Since ARMCD is an expected variable in TV, the variable is included however all the occurrences are null. ARM is a permissible variable and has been removed from the domain.

5.4 Trial Inclusion/Exclusion (TI)

5.4.1 General Considerations

The Trial Inclusion (TI) domain contains the actual inclusion/exclusion text for the study. The only time this may become an issue is when the text exceeds 200 characters (see recommended solution under Define.xml).

5.4.2 Define.xml

Presently submission data is transported to the FDA using SAS transport files. As the reader may be aware, there is a 200-character limit on individual fields within the SAS XPORT files. In this example, criteria longer than 200 characters did not occur; however the metadata team recommends one of the following solutions if a study contained data fields greater than 200 characters.

- If the annotated CRF contains the full inclusion/exclusion criteria, insert either the first 200 characters or text criteria abbreviated to 200 characters in TI. In the Define.xml Origin column for STITEXT provide a link to the annotated CRF containing the full inclusion/exclusion text.
- If the annotated CRF does not specify the full text, then it is that a recommended a pdf be created to store the inclusion/exclusion criteria. Within the TI domain, insert either the first 200 characters or text criteria abbreviated to 200 characters. In the Define.xml Comments column for STITEXT, insert a link to the inclusion/exclusion criteria within the pdf.

5.5 Trial Summary (TS)

5.5.1 General Considerations

Trial-specific information is defined in the Trial Summary (TS) domain. Each study will have a unique combination of parameters and values.

5.5.2 Define.xml

The value list for trial summary parameters (TSPARM/TSPARMCD) contains all the parameters applicable to the study and, when controlled terminology is available for the parameter, it can be viewed within the define.xml. Currently the TS domain will consist of both CDISC defined parameters available through controlled terminology and sponsor-defined parameters. Before defining your trial summary study specific parameters you should check the CDISC website for additional information on available parameters and parameter values.

6 All Subject-Related Domains

This section addresses the conventions that apply to all subject-related domains that are included in an electronic submission.

6.1 DATA

The rules that the Metadata Team recommends that sponsors follow when creating SDTM datasets:

- Within SDTM-defined domain datasets, the order of the variables within each record must match the order specified in the IG.
- Within Sponsor-defined domains, the order of the variables should be consistent with the order of the domain model most similar to the domain.
- SDTM-defined domain variables with Core equal to “Req” must be included in the dataset, and null values are not permitted.
- SDTM-defined domain variables with Core equal to “Exp” must be included in the dataset, although some values may be null.
- SDTM-defined domain variables with Core equal to "Perm" are included in the dataset if defined for the study (e.g., appears on an annotated CRF or external file specification). If all values are null for a study, the variable may be removed from the dataset at the sponsor's discretion. SDTM-defined domain variables should have a label identical to the label presented in the IG.

Sponsor defined domain variables should have labels similar to labels presented in the IG. However, sponsors can adjust the labels to properly convey meaning in the context of the domain. Labels should be title case.

6.2 METADATA

6.2.1 Metadata: Domain Level

The rules that the Metadata Team recommends that sponsors follow when developing domain-level metadata:

- All domains must be included in the domain-level metadata table.
- CDISC SDTM-defined domains should be ordered within the domain-level metadata alphabetically with class.
- Sponsor-defined domains, following the SDTM conventions, should appear within the CDISC-defined observation class after the last CDISC SDTM-defined domain. If multiple sponsor-defined domains exist within the same class, they should be arranged alphabetically.

- The "Description" column in the domain level metadata will contain, for each domain, a link to the variable level metadata.
- The "Location" column in the domain level metadata will contain, for each domain, a link to the transport file.

6.2.2 Metadata: Variable Level

The rules that the Metadata Team recommends sponsors follow when developing variable-level metadata:

- Variable-level metadata tables should appear in the same order as listed in the domain-level metadata.
- All variables within the submitted domains must be included in the variable-level metadata.
- Within each domain, the order of the variables within the metadata must match the order of the variables in the dataset.
- All variables must have a variable description in the Variable Label column that matches the label for that variable in the dataset.
- Variables with an origin of CRF must have a link in the Origin column to first occurrence of the variable within the Annotated CRF which is stored as (blankcrf.pdf).
- Variables with controlled terminology must have a link to a list of values.
    - For variables whose values are limited to a discrete number of values within the SDTM data, the allowable values should appear in a controlled terminology list supplied by the sponsor.
    - For variables whose values are dependent on an external third-party dictionary, refer to an external code list identifying the dictionary by name and version. This scenario is not included within the sample submission due to the proprietary nature of external dictionaries.
- Variables with an origin of “Derived” must have an entry in the Comments column that defines the derivation. For derivations based on a logical condition or calculation, it is recommended that Computational Method be used to define the derivation. In this case, the entry in the Comments column will be a link to the computational method.

6.3 ANNOTATED CRF

All variables defined in the value level Metadata tables with an origin of CRF should be annotated on a copy of the CRF and stored as blankcrf.pdf**.**

7 Special-Purpose Domains

7.1 DEMOGRAPHICS (DM)

The CDISC SDTM defined Demographics (DM) domain has one record for each subject either randomized or screened (optional) in the trial.

7.1.1 General Considerations

- The keys for the Demographics guaranteeing uniqueness and defining a potential sort order are STUDYID, USUBJID. By definition, the DM has one record per subject and the USUBJID is both a key and a unique identifier.
- Reference start date (RFSTDTC) and reference end date (RFENDTC) are sponsor-defined items. The Comments column should define the logic used to populate the values. Since these are expected values, null values are allowed and will exist for screening failures (if included); however, for subjects who received treatment, the values should be populated.
- Screening failures may be included in DM. If screening failures are included verify that the subjects are identified according to the recommendation in the SDTMIG (ARM = “Screening Failure” and ARMCD=”SCRNFAIL”).

7.1.2 Define.xml Considerations

- In the example, DM has supplemental qualifiers associated with CRF items and derived variables not included in the SDTM defined DM domain. To locate the supplemental variables select the SUPPDM domain from the domain level table. There are two ways to view the Supplemental Qualifiers using the define.xml style sheet. One way is to select the SUPPDM domain from the domain-level table. Within the submission a link was created in the style sheet” to show an alternative method of accessing the SUPPDM. There are links in the define.xml style sheet at the end of the DM domain to the Supplemental Qualifiers for DM and at the end of the supplemental DM domain back to the parent domain.
- The variable QNAM within SUPPDM has an associated value list. The link on QNAM displays the value list and associated attributes of the supplemental items. If the QNAM variable has an origin of CRF, the link to the annotation will be in the Origin column and, if Origin is specified as derived, the derivation definition will be located by viewing the Comments column.

7.1.3 Additional or Related Data, Supplemental Qualifiers

- The population flags used within the SDTM submission are stored as supplemental qualifiers and defined in SUPPDM. In this example a population flag (SAFETY) and a randomization flag (RAND) are stored in SUPPDM. Within the sample submission a link was created in the style sheet to access the supplemental qualifiers and demonstrate an alternate method to access SUPPDM.

7.2 SUBJECT VISITS (SV)

7.2.1 General Considerations

The Subject Visits (SV) domain relates to the Trial Visits domain. The SV domain records the visits, including the visit start dates, end dates, and planned visit day in relation to the reference start date. The visit number, description and planned visit day for planned visits are identical to the corresponding items in the TV domain.

Unscheduled or unplanned visits are included within the Subject Visits domain. For all unscheduled visits, the reason for the visit appears in the reason for unplanned visit variable (SVUPDES). Sponsors often assign visit numbers to visits that were not part of the study plan (i.e. unscheduled visits) and, if possible, these should also follow within the chronological order.

7.2.2 Additional or Related Data, Supplemental Qualifiers

The team decided to include the actual study day on which the subject visit occurred. The study day (--STDY) is a standard timing variable within most subject domains; however SVSTDY is not defined as a standard variable in SV according to the STDMIG 3.1.1. Adding variables, including standard timing variables, to trial domains is not allowed therefore SVSTDY was defined as a supplemental qualifier.. Supplemental qualifiers are not recommended with the SV domain and the team is implementing SUPPSV.SVSDTY only as a temporary solution. In STDMIG 3.1.2 SVSTDY will be a standard permissible variable in SV and the supplemental qualifier solution will not be necessary.

8 Interventions

8.1 CONCOMITANT MEDICATIONS (CM)

The CDISC SDTM-defined Concomitant Medications (CM) domain has one record for each medication intervention episode per subject. The sponsor defines the intervention episode.

8.1.1 General Considerations

- In this example, the sponsor-defined keys guaranteeing uniqueness and defining a potential sort order for the CM are STUDYID, USUBJID, CMCAT, CMTRT, CMSTDTC, CMENDTC, CMDOSTXT, CMDOSU, CMINDC, and CMDOSFRQ. This is the order defined by the sponsor for efficient review. The SDTM keys that guarantee uniqueness are STUDYID, USUBJID, and CMSEQ.
- The sample submission combines Concomitant Medication and Psychotropic Drug Treatment History within the CM domain. The Concomitant Medication Category (CMCAT) variable is used to identify the type of medication, and contains values of either “Concomitant Medication” or “Psychotropic Drug Treatment History”. The Sponsor ID (CMSPID) can also be used to distinguish the data using the dataset name of origin. The values of CMSPID, which relate to the sponsor’s operational data, in the CM submission dataset are CONMED and PSYCHDR .
- The sample study collected the medication dose as character and dose on the CRF is annotated as CMDOSTXT. Since the data were collected as character and it cannot be guaranteed that the data would convert correctly or consistently from free form character to numeric, the permissible numeric variable CMDOSE is not included in the CM domain. However, sponsors may use both the character and numeric dose representations based on how the data were collected or derived.
- As stated in the Comments column of the define.xml the WHODRUG dictionary was used to code concomitant medications in this study. Although the WHODRUG dictionary allows links to multiple ATC codes, the medications were coded to a single ATC code and class based on the indication. Therefore, for each coded medication, CMCLAS and CMCLASCD were included in the domain and populated with the appropriate ATC level three class description and code based on the indication.

8.1.2 Define.xml Considerations

- CM has Supplemental Qualifiers associated with the CRF items from Psychotropic drug data and derivations applicable to both types of CM data. To view the Supplemental Qualifiers using the define.xml style sheet, select the SUPPCM domain. There are two ways to view the Supplemental Qualifiers using the define.xml style sheet. One way is to select the SUPPCM domain from the domain level table. A link was created within the style sheet following the end of the CM domain and pointing to the supplemental qualifiers for CM. Providing a link at the end of the main domain to the supplemental domain and back from the supplemental domain to the parent domain facilitates the maneuverability within the define.xml.
- The variable QNAM within SUPPCM has an associated value list. The link on QNAM displays the value list and associated attributes of the supplemental items. If the QNAM variable has an origin of CRF the link to the annotation will be in the Origin column and, if Origin is specified as derived, the derivation definition will be located by viewing the Comments column.

8.1.3 Additional or Related Data, Supplemental Qualifiers

- The Psychotropic Medication CRF contains two additional data items not found in Concomitant Medication CRF. These variables, PDRESP and PDDREAS, have been annotated and defined as supplemental qualifiers for CM and are located in the SUPPCM metadata table and transport file. Each Psychotropic Drug record within CM has two associated SUPPCM records. The study and subject

identifiers, along with the CMSEQ from CM and QNAM, uniquely define the associated SUPPCM variables.

8.2 EXPOSURE (EX)

The SDTM IG defined Exposure domain (EX) domain has one record per constant dosing interval per subject.

8.2.1 General Considerations

- In this example, the sponsor-defined keys guaranteeing uniqueness and defining a potential sort order for the EX domain are STUDYID, USUBJID, EXSTDTC and EXENDTC. The SDTM keys that guarantee uniqueness are STUDYID, USUBJID and EXSEQ.
- The value in EXDOSE is the amount of each administration which was assigned based on the ARMCD. The dose per day for this study is recorded on the CRF as number of tablets taken per day. The total daily dose, EXDOSTOT, must be derived by multiplying the number of tablets taken per day, supplemental qualifiers domain (SUPPEX.QNAM=SMNO), by the dose contained in each tablet, EXDOSE.
- EXSTDY and EXENDY are derived variables calculated using the reference dates in DM.

8.2.2 Define.xml

- EX has a supplemental qualifier for “Number of tablets per day” which originates from the CRF. To navigate to the value-level metadata for the values of QNAM, simply click on the link provided at the bottom of the EX table. For easy navigation, there is also a link at the bottom of the SUPPEX value level metadata which leads back to the EX table.

9 Events

9.1 ADVERSE EVENTS (AE)

The CDISC SDTM-defined Adverse Events (AE) domain has one record per adverse event per subject.

9.1.1 General Considerations

- The sponsor-defined keys guaranteeing uniqueness and defining a potential sort order for the AE domain are STUDYID, USUBJID, AESTDTC, AEENDTC, AEDECOD, and AETERM. The SDTM keys that guarantee uniqueness are STUDYID, USUBJID, and AESEQ.
- In this example, the sponsor's instructions for the "Ongoing" checkbox on the CRF corresponded to the IG definition of "AFTER" for the AEENRF variable. Depending on how the data are collected, sponsors might need to derive values for AEENRF and/or populate the variable with a different value.
- Sponsors often include an indicator on their CRF to indicate that there were no adverse events for a subject. Records are included in the AE dataset only when an event actually occurred, so CRF indicators for no events should be annotated as “NOT SUBMITTED.”

9.1.2 Define.xml Considerations

- It is the sponsor’s responsibility to define an adverse event. This definition may vary based on the sponsor’s requirements for characterizing and reporting product safety and is usually defined in the protocol. The sponsor may provide additional detail describing an event in the structure column of the domain-level metadata. See Section 3.2.2.
- Provide the dictionary name and version used to map the terms in the metadata. This information should be displayed in the Comments column for AEDECOD.
- If toxicity grades are assigned, specify the name of the scale and version used in the Comments column for AETOXGR.
- The derivations of AESTDY, AEENDY, and AEENRF are described in the Comments column.

9.2 DISPOSITION (DS)

The CDISC SDTM-defined Disposition (DS) domain has one record per disposition status or protocol milestone per subject.

9.2.1 General Considerations

- The sponsor-defined keys guaranteeing uniqueness and defining a potential sort order for the DS domain are STUDYID, USUBJID, DSSTDTC, DSCAT, and DSTERM. The SDTM keys that guarantee uniqueness are STUDYID, USUBJID, and DSSEQ.
- The DS dataset may include data from a variety of origins. In this example, the data for DS is obtained from the Demography (informed consent date), Randomization, and Termination CRFs.
- It may be necessary to rename, reformat and/or transform the data that was collected on the CRF in order to create appropriately formatted DS records. In this example, a "Yes" response to the question "Did the subject complete the study?" results in a DS record with DSTERM="COMPLETED."

9.2.2 Define.xml Considerations

- CDISC Controlled terminology was supplied for DSCAT. Sponsor defined controlled terminology was defined for both for DSDECOD.

9.3 MEDICAL (HISTORY) MH

The CDISC SDTM-defined Medical History (MH) domain has one record per medical history event per subject.

9.3.1 General Considerations

- The sponsor-defined keys guaranteeing uniqueness and defining a potential sort order for the MH domain are STUDYID, USUBJID, MHCAT, MHTERM, and MHSTDTC. The SDTM keys that guarantee uniqueness are STUDYID, USUBJID, and MHSEQ.
- The data are captured on two CRFs - Medical & Surgical History and Psychiatric History. MHCAT is used to categorize the 2 event types.
- It may be necessary to rename, reformat and/or transform the data that was collected on the CRF in order to create appropriately formatted MH records. In this example, the "Resolved" checkbox on the Medical & Surgical History CRF is transformed to MHENRF="BEFORE" and the "Ongoing" checkbox is transformed to MHENRF="DURING/AFTER."

9.3.2 Define.xml Considerations

- Medical History body system (MHBODSYS) is sponsor-defined controlled terminology and the possible terms are specified within the Define.xml. The reviewer will know to look for sponsor-defined terminology both from the comments associated with MHBODSYS in the Define.xml.

- If medical history terms are coded using an external dictionary, specify the dictionary name and version used in the Comments column for MHDECOD.

10 Findings

FINDINGS domains capture the observations resulting from planned evaluations to address specific tests or questions such as laboratory tests, histopathology, ECG testing, and questions listed on questionnaires. The topic variable for a Findings record would be the test code (--TESTCD) and it is required for all Findings domains and records. In the ECG and Vital Signs domains the test codes (--TESTCD) and the test codes descriptions (--TEST) are Controlled Terms based on CDISC Controlled terminology. All other test codes are purely examples and should not be interpreted as CDISC approved controlled terms.

10.1 ECG (EG)

The CDISC SDTM defined ECG (EG) domain has one record for each ECG observation per visit per subject.

10.1.1 General Considerations

- The sponsor-defined keys guaranteeing uniqueness and defining a potential sort order for the EG domain are STUDYID, USUBJID, EGDTC, VISITNUM, EGCAT, EGSPID, and EGTESTCD. The SDTM keys that guarantee uniqueness are STUDYID, USUBJID, and EGSEQ.
- There are 2 derived test codes (EGTESTCD) defined in the sample EG domain, QTC Fridericia (QTCF) and QTC Bazett (QTCB). Since the test codes are created and the values derived every time there is a value for the test code of QT, a derived record flag must be set. Within the dataset the derived flag (EGDRVFL) is set to Y for each occurrence of an EGTESTCD equal to QTCB or QTCF.
- For the derived records, QTCB and QTCF, the original results variables are null. In this example QTC was defined in the operational database however QTCB and QTCF were derived explicitly for the SDTM and require the creation of derived records. If QTCB and QTCF were included, along with QTC, in the operational database a derived record would not be necessary and QTCB and QTCF would not be defined in derived records or as derived variables.

10.1.2 Define.xml Considerations

- The variable EGTESTCD links to a value level metadata table for the ECG parameters.
- Origin for the EGTESTCD value is CRF for all but two of the ECG parameters. QTC Fridericia and QTC Bazett are derived variables with an origin of “Derived” within derived records. The derivation is defined in the Comments column by using the Computational Algorithms Section. Within the computation method the derived parameter is specified under Reference Name column and the derivation formula is in the Computational Method column.
- At the end of the EG domain, there is a link within the style sheet to the supplemental qualifiers for EG. Providing a link at the end of the main domain to the supplemental domain and back from the supplemental domain to the parent domain facilitated the maneuverability of the define.xml.
- Since the sponsor required clinically significant out-of-range indicators for EG observations, and the EG domain includes only an out-of-range indicator, a Supplemental Qualifier dataset (SUPPEG) was created to hold the clinically significant flag.

10.2 INCLUSION/EXCLUSION (IE)

The CDISC SDTM defined Inclusion/Exclusion (IE) domain has one record per each inclusion/criteria that was an exception for each subject.

10.2.1 General Considerations

- The sponsor defined keys guaranteeing uniqueness and defining a potential sort order for the IE domain. are STUDYID, USUBJID, IECAT, IETESTCD, and VISITNUM. The SDTMkeys that guarantee uniqueness are STUDYID, USUBJID, and IESEQ. Only exceptions to the inclusion or exclusion criteria are included in the domain data.

10.2.2 Define.xml Considerations

- IECAT in the IE domain is required and categorizes the criteria into either "INCLUSION" or "EXCLUSION". In this sample all values of IECAT are the "INCLUSION" since the only exceptions found in the data were for Inclusion Criteria missed.
- The variable IETESTCD links to a value level metadata table for the Inclusion/Exclusion criteria.

10.2.3 Additional or Related Data, Supplemental Qualifiers

- There are only a few records within the transport dataset. Two subjects had answers of "No" for three inclusion criteria. The text of the criteria (IETEXT) is inserted in the IE domain from the Trial Inclusion (TI) domain. In this sample the textual information was less than 200 characters. If the text was greater than 200 characters only the first 200 characters, or if preferred, an abbreviated text is included in IETEXT. This is explained further in Section 5.4 (Trial Inclusion/Exclusion Criteria).

10.3 LABORATORY (LB)

The CDISC SDTM defined Labs (LB) included in the domain has one record per lab test per visit per subject.

10.3.1 General Considerations

The sponsor-defined keys guaranteeing uniqueness and defining a potential sort order for the LB domain are STUDYID, USUBJID, LBDTC, VISITNUM, LBCAT, and LBTESTCD. The SDTM keys that guarantee uniqueness are STUDYID, USUBJID, and LBSEQ.

Currently LBTESTCD and LBTEST can be sponsor test codes and descriptions however controlled terminology will eventually be available. For additional information on terminology to support CDISC standards see the CDISC website.

10.3.2 Define.xml Considerations

- The origin for the LBTESTCD values may be a CRF, a Lab test dictionary, or an EDT. In this example, the Central Lab vendor had sent the LBTESTCD values for all of the records.
- The variable LBTESTCD links to a value-level metadata table for the lab tests.

10.3.3 Additional or Related Data, Supplemental Qualifiers

- The LB records that have a PCS (potentially clinically significant) Indicator will have a corresponding record in the SUPPLB dataset. The normal range indicators are defined within the LB domain but PCS indications are not part of the standard LB domain and must be defined and created in SUPPLB.
- There are two ways to view the supplemental qualifiers using the define.xml style sheet. One way is to select the SUPPLB domain from the domain level table. A link was created within the style sheet following the LB domain and pointing to the supplemental qualifiers for LB. Providing a link at the end of the main domain to the supplemental domain and back from the supplemental domain to the parent domain facilitated the maneuverability of the define.xml.
- The variable QNAM within SUPPLB has an associated value list. The link on QNAM displays the value list and associated attributes of the supplemental items. In this example, the QNAM variable has an origin of derived. The derivation definition will be located by viewing the Comments column.

.

10.4 PE

The CDISC SDTM defined Physical Examinations (PE) domain has one record per body system or abnormality per visit.

10.4.1 General Considerations

- The sponsor-defined keys guaranteeing uniqueness and defining a potential sort order for the PE domain are STUDYID, USUBJID, VISITNUM, PESPID, PETESTCD, and PEDTC. The SDTM keys that guarantee uniqueness are STUDYID, USUBJID, and PESEQ.
- There is one derived variable within the PE domain, PEBLFL. The definition of PEBLFL will be located in the variable level metadata Comments column.

10.4.2 Define.xml Considerations

- The variable PETESTCD links to a value level metadata table for the PE Body Systems.
- Origin for the PETESTCD is CRF for all the records.

10.4.3 Additional or Related Data, Supplemental Qualifiers

- For this example, no additional data have been collected.

10.5 QUESTIONNAIRE (QS)

The Questionnaire (QS) domain contains the data for all questionnaires used the study.

10.5.1 General Considerations

The QS domain can become quite large since it has data for all questionnaires for all subjects at all visits. However, the metadata team recommends that the QS domain be submitted as one domain, as opposed to splitting the data into multiple domains by questionnaire type. The spnsor-defined keys guaranteeing uniqueness and defining a potential sort order for the QS domain are STUDYID, USUBJID, QSCAT, VISITNUM, QSSPID, QSTESTCD, and QSDTC. The SDTM keys that guarantee uniqueness are STUDYID, USUBJID, and QSSEQ.

In order to make the QS domain more usable, QSCAT was designated to contain the questionnaire name. To facilitate review, the sample questionnaire data is organized first by subject and then by questionnaire for each subject. Within each questionnaire and visit, QSSPID stores the logical sequence of the questions. QSSPID often refers to the sequence numbers on the CRF and makes the association between the data and the annotation easier to follow. It will be noticed that QSSEQ guarantees uniqueness across all questionnaires for a subject and the sponsor-defined QSSPID sequences the data logically within each questionnaire for each visit.

10.5.2 Define.xml Considerations

- The define.xml has been organized to facilitate the review of the questionnaire data. The style sheet supports a value list for various questionnaires within the domain through the use of QSCAT. After choosing the QS domain, select the questionnaires category (QSCAT) and an associated value list containing Questionnaire descriptions can be seen. If an individual QSCAT value is chosen, a value list of the test codes (QSTESTCD) or questions specific to the selected questionnaire can be seen. The metadata team recommends that a style sheet similar to this for the questionnaires be used in order to simplify the review.

- In the value-level metadata for QSCAT, the questionnaire version used may be specified in the Comments column.

10.6 SUBJECT CHARACTERISTICS (SC)

10.6.1 General Considerations

The sponsor-defined keys guaranteeing uniqueness and defining a potential sort order for the Subject Characteristic (SC) domain are SCTESTCD and SCSEQ within USUBJID. This is a unique domain key definition since the SDTM domain key, SCSEQ, is used as one of the keys guaranteeing uniqueness. The SDTM keys that guarantee uniqueness are STUDYID, USUBJID, and SCSEQ.

10.6.2 Define.xml Considerations

Since the SC domain varies by sponsor, the value list for the subject characteristics test codes is quite helpful during review. After linking to the SC domain and selecting the SCTESTCD variable, the value list for the variables contained within the domain can be reviewed.

10.7 VITAL SIGN (VS)

The CDISC SDTM defined Vital Signs (VS) domain has one record per vital sign measurement per visit per subject.

10.7.1 General Considerations

- The sponsor-defined keys guaranteeing uniqueness and defining a potential sort order for the VS domain are STUDYID, USUBJID, VSDTC, VISITNUM, VSSPID, VSTESTCD, and VSPOS. The SDTM keys that guarantee uniqueness are STUDYID, USUBJID, and VSSEQ.

10.7.2 Define.xml Considerations

- The variable VSTESTCD links to a value level metadata table for the Vital signs measurements.
- Origin for the VSTESTCD values is a CRF for all of the records.

10.7.3 Additional or Related Data, Supplemental Qualifiers

- The VS records that have a 'Reference Range Indicator –PCS FLAG' are included in the SUPPVS dataset.
- To view the supplemental qualifiers using the define.xml style sheet first select the SUPPVS domain. To locate the supplemental variables, select the SUPPVS domain from the domain-level table. The variable QNAM within SUPPVS has an associated value list. The link on QNAM displays the value list and associated attributes of the supplemental items. In this example, the QNAM variable has an origin of derived. The derivation definition will be located by viewing the Comments column.

10.8 DRUG ACCOUNTABILITY (DA)

The CDISC SDTM defined Drug Accountability (DA) domain has one record per drug accountability finding per visit per subject.

10.8.1 General Considerations

- The sponsor-defined keys guaranteeing uniqueness and defining a potential sort order for the DA domain are STUDYID, USUBJID, DADTC, DATESTCD. The SDTM keys that guarantee uniqueness are STUDYID, USUBJID, and DASEQ.

10.8.2 Define.xml Considerations

- The variable DATESTCD links to a value level metadata table for the drug accountability.
- Origin for the DATESTCD values is CRF for all of the records.

11 RELATIONSHIP DATASETS

11.1 RELREC

The purpose of the RELREC domain is to define a relationship between independent records in separate domains. The REREC records identify the related domains, define the variables that identify the related records, specify the relationship type, and give each relationship a unique identifier. An example of a relationship that could be defined would be when a subject terminates a study due to an adverse event. It is possible, by using a RELREC domain, to link the adverse event in the AE domain that caused the termination to the termination record within the Disposition (DS) domain.

A physical example using RELREC has not been included within the sample submission however; if the data were to be created it would look like the following example.

DS Domain subset

DSDECOD indicates that subject terminated due to an Adverse Event. Identifier for DS record is DSSEQ=3

STUDYID	DOMAIN	USUBJID	DSSEQ	DSDECOD
CDISC01	DS	CDISC01.200001	1	INFORMED CONSENT
CDISC01	DS	CDISC01.200001	2	RANDOMIZED
CDISC01	DS	CDISC01.200001	3	ADVERSE EVENT

AE Domain subset

AEACN indicates that subject discontinued study drug. Identifier for AE record is AESEQ=3

STUDYID	DOMAIN	USUBJID	AESEQ	AEDECOD	AEACN
CDISC01	AE	CDISC01.200001	1	Anxiety	None
CDISC01	AE	CDISC01.200001	2	Nausea	None
CDISC01	AE	CDISC01.200001	3	Constipation	Study Drug Discontinued
CDISC01	AE	CDISC01.200001	4	Fatigue	None
CDISC01	AE	CDISC01.200001	5	Arthralgia	None

RELREC Domain

RELTYPE indicates that the relationship between DS & AE is one **-->** one

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
CDISC01	DS	CDISC01.200001	DSSEQ	3	ONE	1
CDISC01	AE	CDISC01.200001	AESEQ	3	ONE	1

11.2 SUPPQUAL

The CDISC SDTM-defined Supplemental Qualifiers domain is a special-purpose domain that serves as a storage area for the non-standard qualifiers variables sponsors collect.

11.2.1 General Considerations

- The sponsor-defined keys for each SUPPxx dataset guaranteeing uniqueness and defining a potential sort order are STUDYID, RDOMAIN, USUBJID, IDVAR, INDVARVAL and QNAM. There should be one record for each qualifier value.
- The Metadata Team chose to create individual SUPP-- datasets specific to each domain, where needed (e.g., SUPPAE, SUPPCM, SUPPDM, SUPPEX, SUPPLB, SUPPSV, and SUPPVS). Sponsors may opt to store all of the supplemental variables in one large SUPPQUAL dataset. The amount of supplemental data and ease of use of the datasets should be considered before the decision is made.

11.2.2 Define.xml Considerations

- The SUPPxx domains have been placed after all the subject-related domains in the define domain-level table. They appear alphabetically by domain name.
- Each QNAM link displays the value list and associated attributes of the supplemental items. If the QNAM variable has an origin of CRF the link to the annotation will be in the Origin column and if Origin is specified as derived, the derivation definition will be located by viewing the comments column.
- The metadata team recommends, through use of the style sheet, providing a link between the main domain and the associated Supplemental Qualifier domain. It will be noted in the sample submission that at the end of AE domain that there is a link to SUPPAE and at the bottom of the SUPPAE domain there is a return link to the parent AE domain. These links are available for all the domains with an associated supplemental qualifiers domain.

12 Appendices

12.1 CDISC SDS METADATA TEAM

Role	Name	Company
Team Lead	Carolyn Wilson	Forest Research Institute
Team Co-Lead	Richard Lewis	Octagon Research
	Gail Stoner	Centocor
	Madhavi Vemuri	Johnson and Johnson PRD
	Gary Walker	Quintiles
	Trisha Simpson	Schwarz BioSciences, Inc.
	William Friggle	Sanofi-Aventis
	Carol Vaughn	Sanofi-Aventis

12.2 SAMPLE SUBMISSION SOFTWARE ISSUES

12.2.1 Adobe 7 blankcrf.pdf

There is a known problem with Adobe 7 linking when trying to link to a specific page in a PDF from an HTML document, in this case an XML file rendered in a web browser. In the submission example the links within the define.xml are defined to point to a specific page number within the blankcrf.pdf. If you are using other than Adobe 7 the links should work perfectly. However, if you are using Adobe 7 the link will always point to the first page within the blankcrf.pdf.

The link to the Adobe note.

http://www.adobe.com/cfusion/knowledgebase/index.cfm?id=326332

12.2.2 Back Arrows Within the Define.xml

In some cases the back arrows do not work when maneuvering within the define.xml. We believe the functionality is associated with the version of Internet Explorer that is being used. As more information is obtained we will update this information.