PhUSE 2012

Paper IS04

An Introduction to SDTM – 298 pages in 20 minutes?!

Jennie Mc Guirk, ICON, Dublin, Ireland

ABSTRACT

The Study Data Tabulation Model (SDTM) and its Implementation Guide are very detailed documents, holding vast amounts of information and practical examples. These can seem a daunting set of document's to new comers to the industry, particularly if training consists of the 'read it and get started' approach. This paper will attempt to introduce the Study Data Tabulation Model, by summarizing the guidelines, providing some real life examples, and also some tips on how to confirm the compliance and validity of your mappings by using tools such as OpenCDISC.

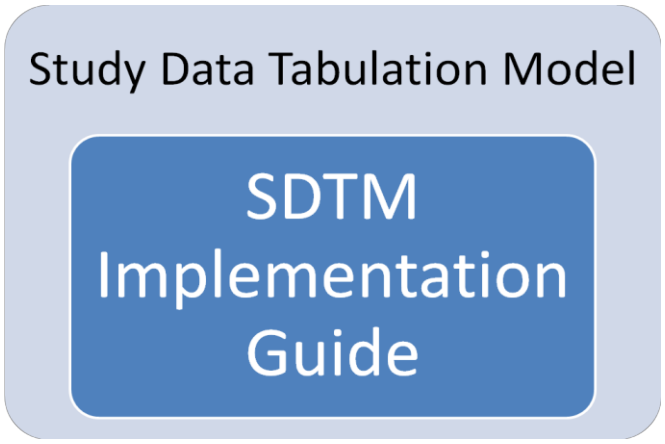
INTRODUCTION

CDISC is a global, open, multidisciplinary, non-profit organization that has established standards to support the acquisition, exchange, submission and archive of clinical research data and metadata. CDISC has a number of Foundation Standards. This paper shall focus on one of those standards - The Study Data Tabulation Model (SDTM)., and provide a high level summary of how to use the SDTM Implementation Guide.

SDTM FUNDAMENTALS

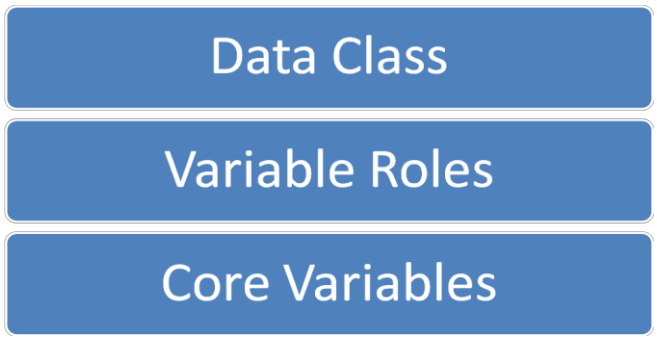
CDISC SDTM is defined within 2 key documents – The Study Data Tabulation Model, and the SDTM Implementation Guide, see Figure 1.

FIGURE 1



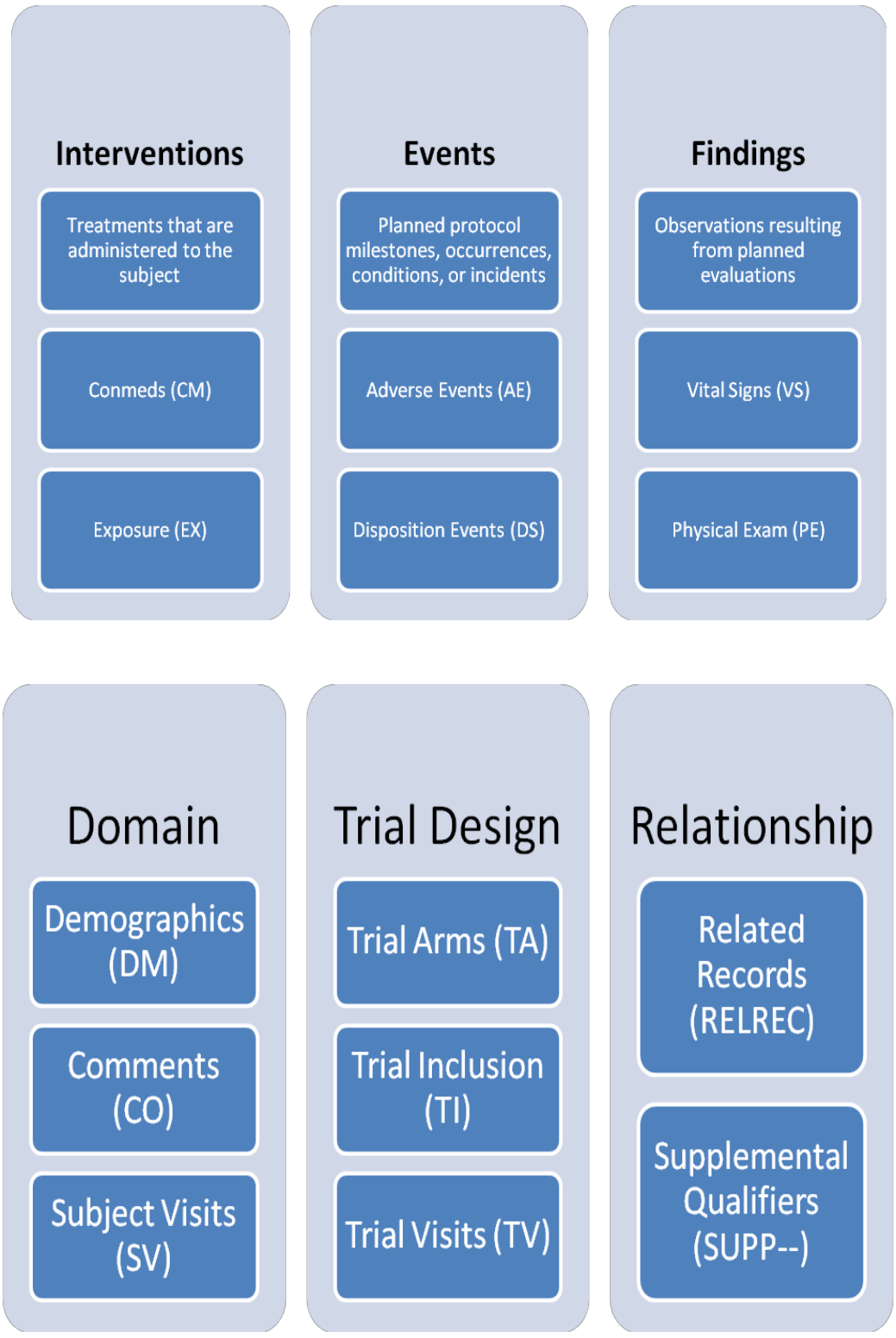
There are 3 key building blocks to how data is structured within the SDTM, see Figure 2.

FIGURE 2



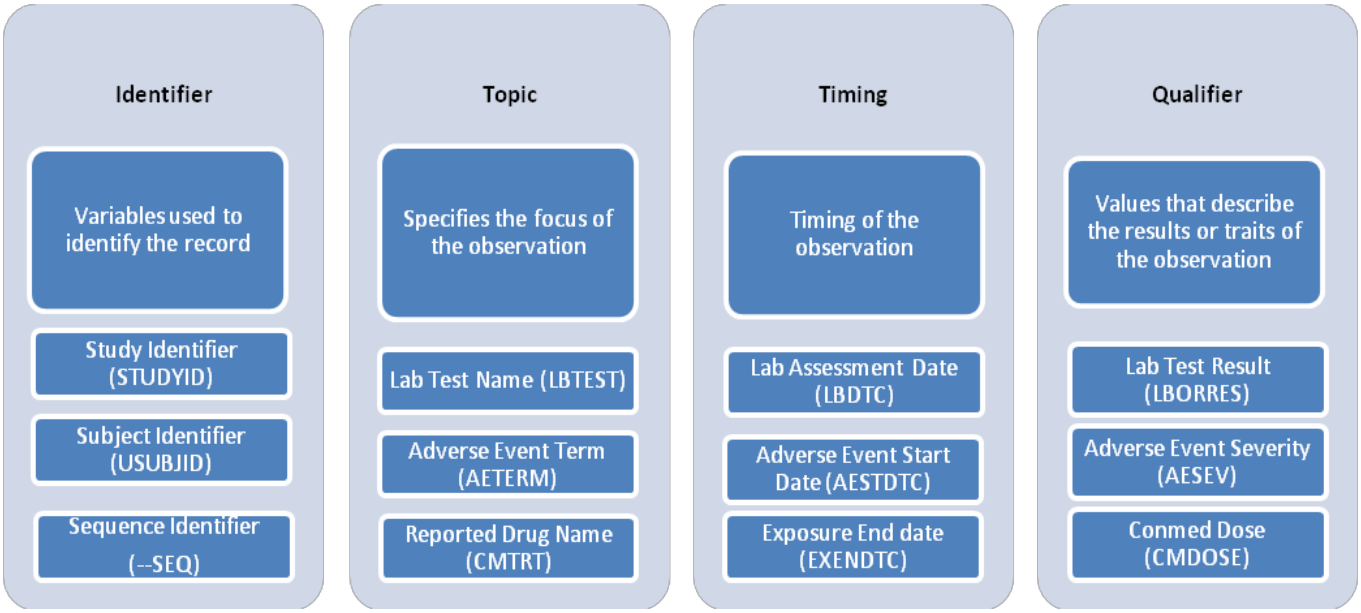
The first of the blocks, the **Data Class**, describes the datasets or domains within the SDTM. These are categorized into 6 classes; see Figure 3, which gives a description of the class, along with some examples.

FIGURE 3 – DATA CLASSES



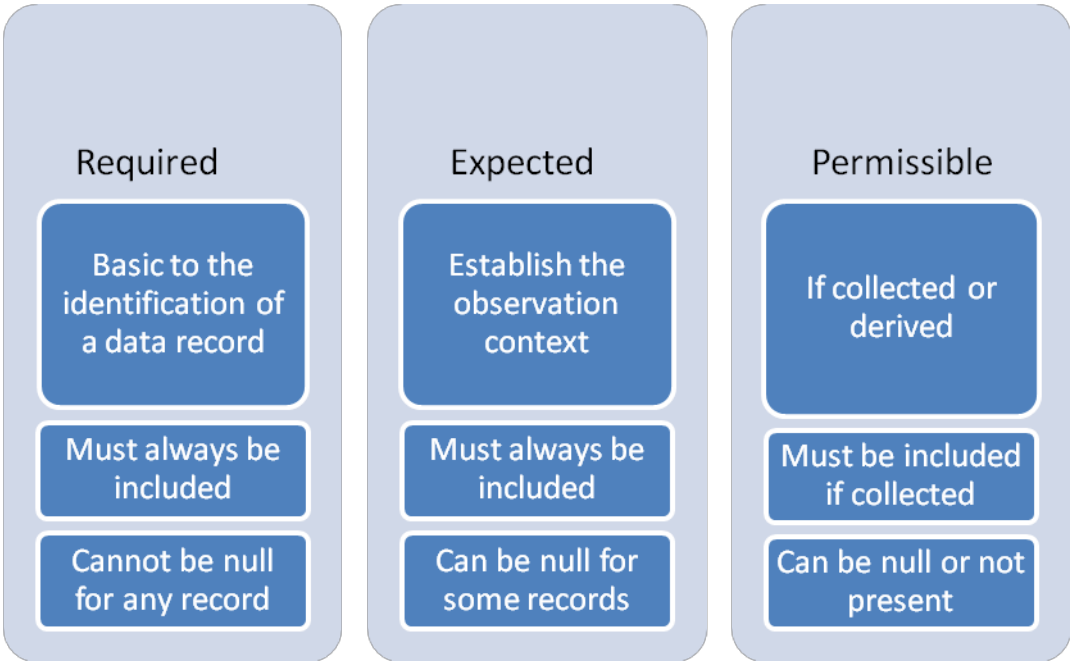
The next block, describes the **Variable Roles**. Variables have 4 main roles within the SDTM, see Figure 4, which gives a description of the role, along with some examples.

FIGURE 4 – VARIABLE ROLES



The final block, describes the **Core Variables**. Variables are divided into 3 core categories within the SDTM, see Figure 5, which gives a description of the core variables, along with the associated rules.

FIGURE 5 – CORE VARIABLES

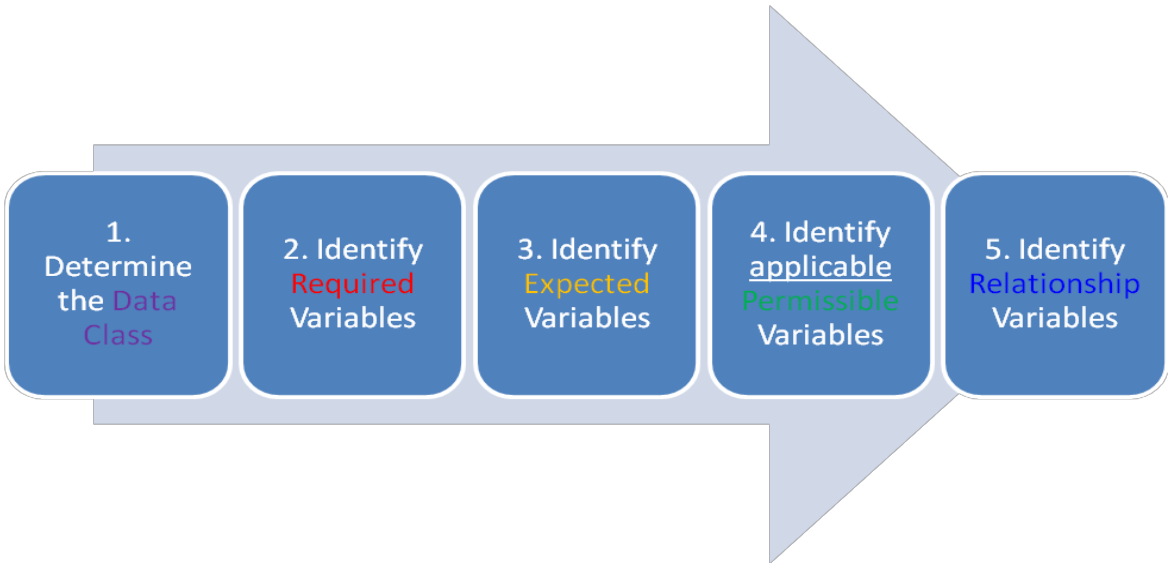


This is the foundation of the SDTM. However it is very theoretical. We now need to interpret the theory and guidelines into some practical examples!

CDISC SDTM IMPLEMENTATION

As this paper is aimed at those starting out with SDTM, the 3 examples selected are simple in complexity. When dealing with such simple examples, there are 5 critical steps to transforming source data to the SDTM, described in Figure 6.

FIGURE 6 – 5 STEPS TO SDTM



Step1: Determine the Data Class.

Although this sounds like a relatively easy step, it can often require some consideration. A simple approach to determining the Data Class is to select keys words from the CRF page, and search the Implementation Guide. For example, a simple search on the text 'Adverse Events' or 'Signs and Symptoms' will indicate immediately that they are within the Events Class. However, if it is not immediately obvious from a key word search, you need to consider the content of the data, alongside the descriptions within the Implementation guide for '**GUIDELINES FOR DETERMINING THE GENERAL OBSERVATION CLASS'.**

Step2: Identify the Required Variables

Once the Data Class has been determined, identifying the required variables is a straightforward look-up of the implementation guide. To start with, STUDYID, DOMAIN, USUBJID, and –SEQ are all required within the General Observation Classes. Then using the Domain model definition select other required variables (for example within EVENTS these are --TERM, --DECOD, within INTERVENTIONS --TRT, and within FINDINGS --TEST –TESTCD.

Step3: Identify the Expected Variables

Follow same process as step 2, except now looking for expected variables. You should now have the minimum set of variables for your domain.

Step4: Identify the Permissible Variables

For permissible variables, as opposed to identify what all the permissible variables are, instead look at your source data to identify what has been captured, but not yet mapped to a required/expected variable in Step2 & 3 above. This enables you to focus on the permissible variables applicable to your study data.

Step5: Identify the Relationship Variables

There may then be variables that are captured, however do not fit into the required, expected or permissible core set of variables. These remaining variables would be mapped to a related domain (such as CO or SUPP--). See below figure 11 on linking a SUPP—and parent domain.

EXAMPLES

Figure 7, 8 and 9 show examples of an INTERVENTION, EVENT and FINDING respectively.
The annotations are color coded based on the 5 step process.

Step 1: Data Class
Step 2: Required
Step 3: Expected
Step 4: Permissible
Step 5: Relationship

FIGURE 7 – INTERVENTIONS EXAMPLE

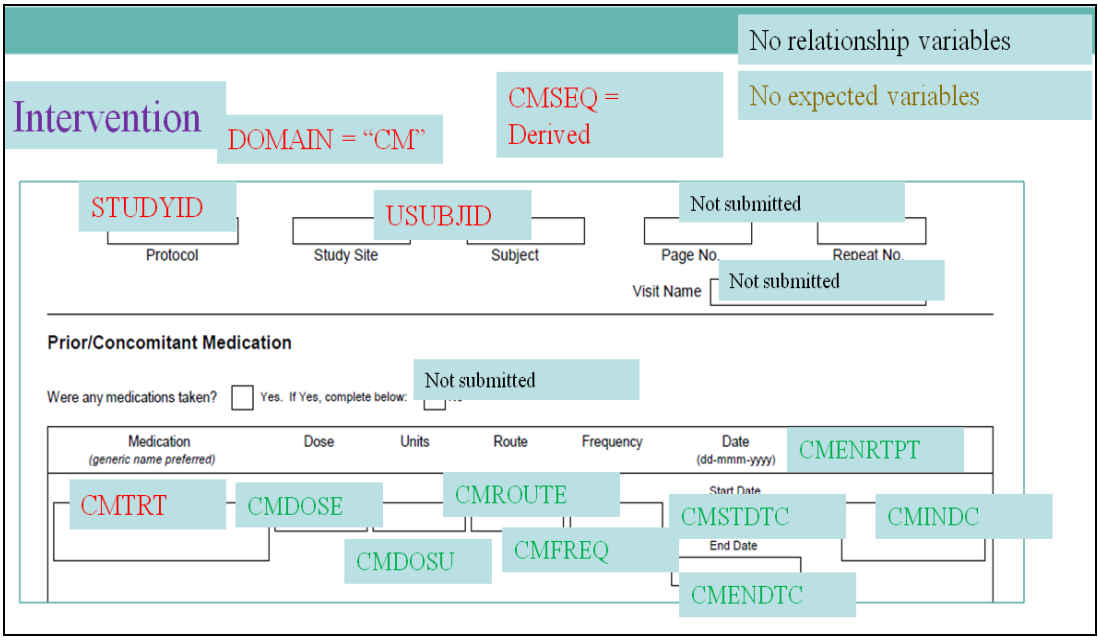


FIGURE 8 – EVENTS EXAMPLE

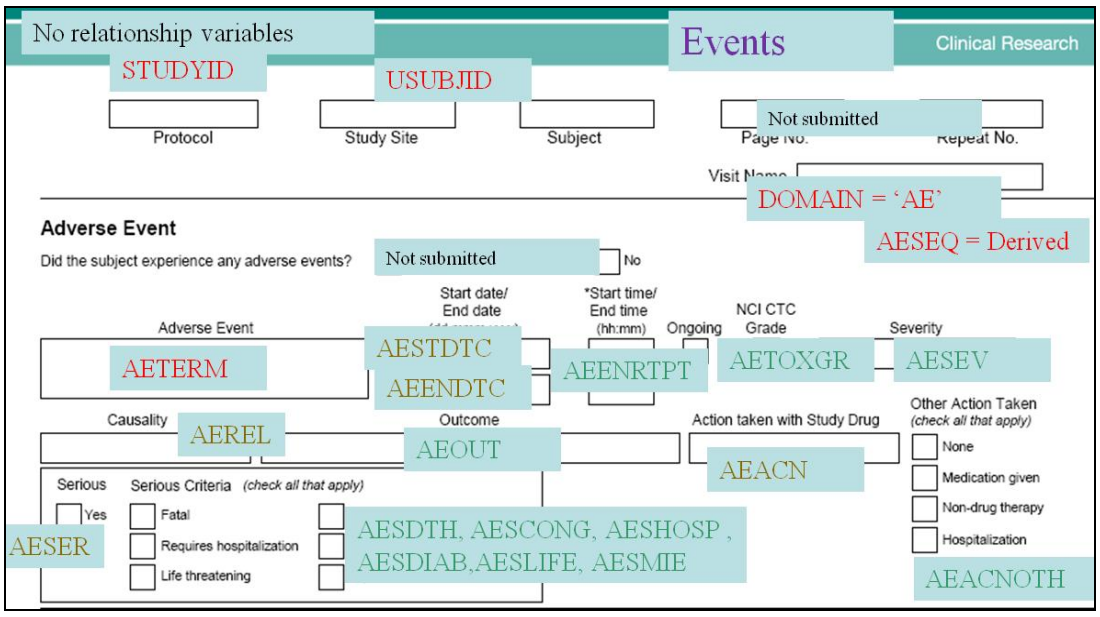
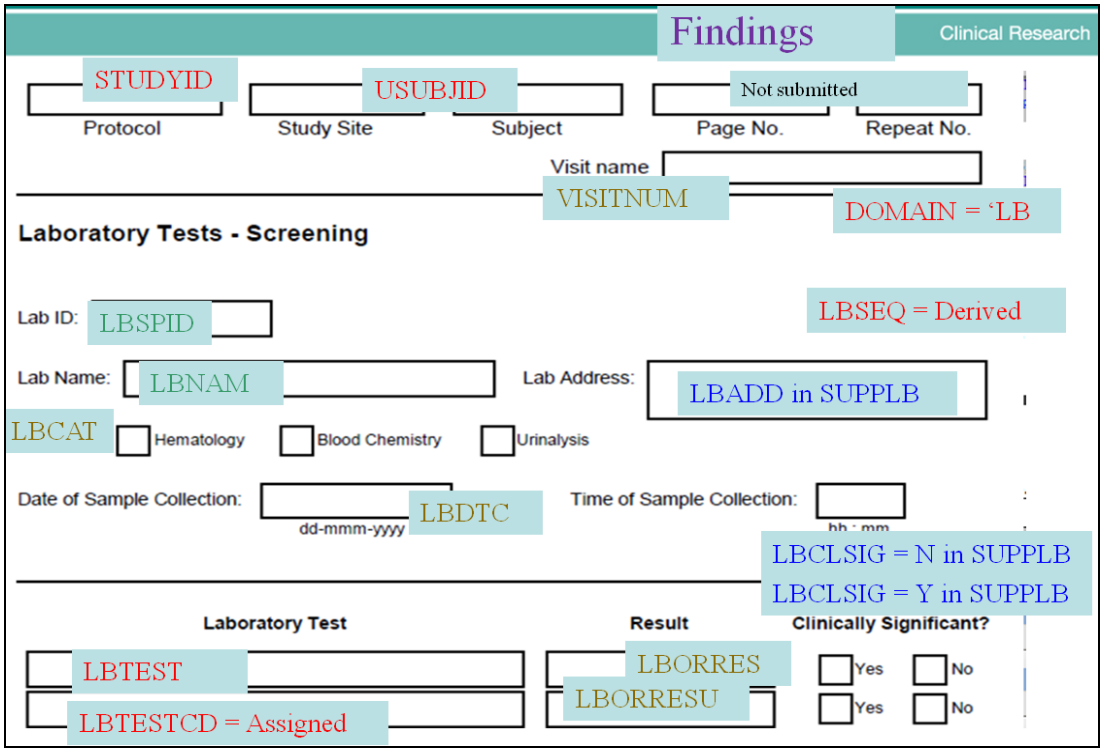


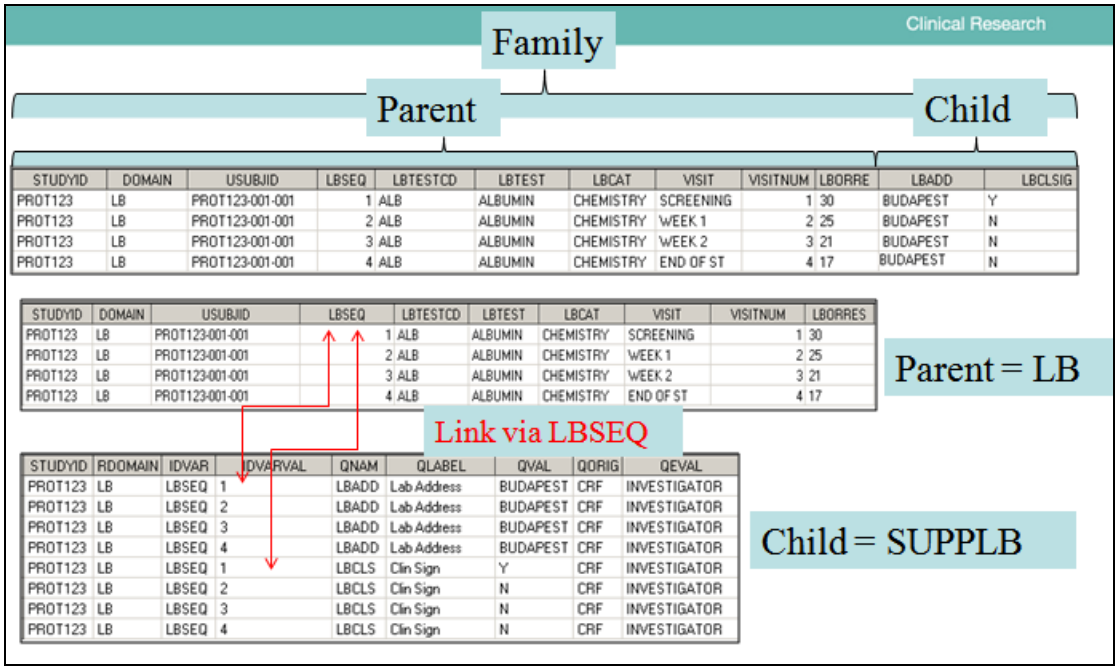
FIGURE 9 – FINDINGS EXAMPLE



LINKING A SUPP—TO A PARENT DOMAIN

Figure 10 is an example of linking a SUPP—and parent domain. This is based on the example provided in Figure 9 above.

FIGURE 10 – LINKING SUPP- TO A PARENT DOMAIN



STUDYID	DOMAIN	USUBJID	LBSEQ	LBTESTCD	LBTEST	LBCAT	VISIT	VISITNUM	LBORRE	LBADD	LBCLSIG
PROT123	LB	PROT123-001-001	1	ALB	ALBUMIN	CHEMISTRY	SCREENING	1	30	BUDAPEST	Y
PROT123	LB	PROT123-001-001	2	ALB	ALBUMIN	CHEMISTRY	WEEK 1	2	25	BUDAPEST	N
PROT123	LB	PROT123-001-001	3	ALB	ALBUMIN	CHEMISTRY	WEEK 2	3	21	BUDAPEST	N
PROT123	LB	PROT123-001-001	4	ALB	ALBUMIN	CHEMISTRY	END OF ST	4	17	BUDAPEST	N

STUDYID	DOMAIN	USUBJID	LBSEQ	LBTESTCD	LBTEST	LBCAT	VISIT	VISITNUM	LBORRES
PROT123	LB	PROT123-001-001	1	ALB	ALBUMIN	CHEMISTRY	SCREENING	1	30
PROT123	LB	PROT123-001-001	2	ALB	ALBUMIN	CHEMISTRY	WEEK 1	2	25
PROT123	LB	PROT123-001-001	3	ALB	ALBUMIN	CHEMISTRY	WEEK 2	3	21
PROT123	LB	PROT123-001-001	4	ALB	ALBUMIN	CHEMISTRY	END OF ST	4	17

STUDYID	RDOMAIN	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG	QEVAL
PROT123	LB	LBSEQ	1	LBADD	Lab Address	BUDAPEST	CRF	INVESTIGATOR
PROT123	LB	LBSEQ	2	LBADD	Lab Address	BUDAPEST	CRF	INVESTIGATOR
PROT123	LB	LBSEQ	3	LBADD	Lab Address	BUDAPEST	CRF	INVESTIGATOR
PROT123	LB	LBSEQ	4	LBADD	Lab Address	BUDAPEST	CRF	INVESTIGATOR
PROT123	LB	LBSEQ	1	LBCLS	Clin Sign	Y	CRF	INVESTIGATOR
PROT123	LB	LBSEQ	2	LBCLS	Clin Sign	N	CRF	INVESTIGATOR
PROT123	LB	LBSEQ	3	LBCLS	Clin Sign	N	CRF	INVESTIGATOR
PROT123	LB	LBSEQ	4	LBCLS	Clin Sign	N	CRF	INVESTIGATOR

CDISC SDTM COMPLIANCE

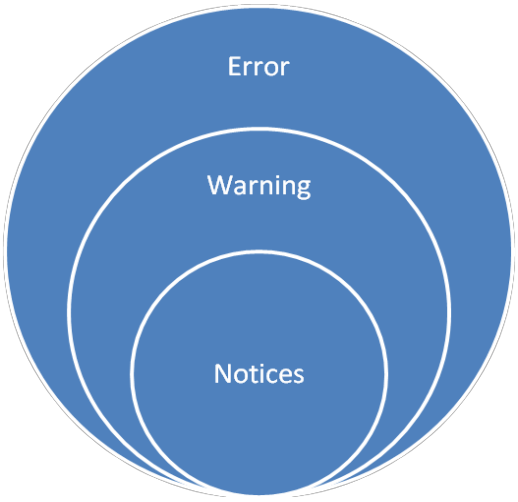
Once you have gone followed the 5 step thinking process, you are now ready to start physically mapping to SDTM.

However, once you have completed your mapping, how will you confirm that the SDTM datasets you have produced actually conform to the SDTM, and the SDTM Implementation Guide? Well, there is a number of ways to validate SDTM mappings, but one of the tools most frequently used is the OpenCDISC Validator.

The OpenCDISC validator provides a method for checking conformance and compliance of mappings against the SDTM Implementation Guide.

OpenCDISC defines issues within 3 severities (see figure 11), and 9 categories (see figure 12).

FIGURE 11 – OPENCDISC SEVERITIES

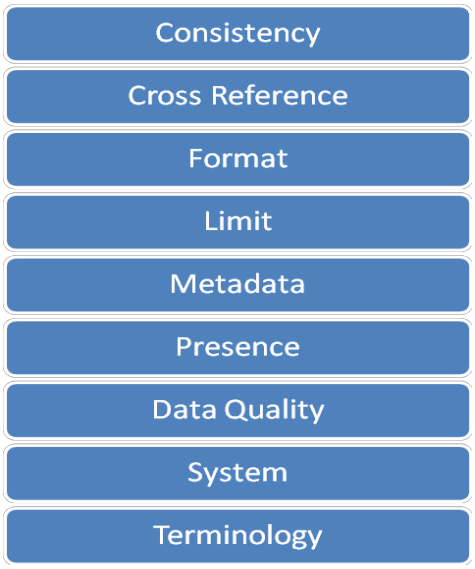


As a rule, errors should be resolved, and all warnings and notices should be at least reviewed and verified.

Sometimes errors are justifiable, for example they are due to underlying data issues (e.g. the study is ongoing, and the database is not yet clean). You can refer to a poster created for the PhUSE FDA conference in March 2012 for some further examples of these (see reference below).

FIGURE 12 – OPENCDISC CATEGORIES

:



Examples of the 9 categories are shown below in Figure 13, also displaying the Severity of each check.

FIGURE 13

Rule ID	Message	Description	Category	Severity
CT0034	Value for SEX not found in SEX controlled terminology codelist	Variable values should be populated with terms found in 'Sex' (C66731) CDISC controlled terminology codelist	**Terminology**	**Error**
SD0002	NULL value in variable marked as Required	Required variables (where Core attribute is 'Req') cannot be NULL for any records	**Presence**	**Error**
SD0003	Invalid ISO 8601 value	Dates and times of day must conform to the ISO 8601 international standard	**Format**	**Error**
SD0005	Duplicate SEQ	The value of Sequence Number (--SEQ) variable must be unique for each record within a subject	**Consistency**	**Error**
SD0012	Day of Start is after the Day of End	Study Day of Start of Event, Exposure or Observation (--STDY) must be less or equal to Study Day of End of Event, Exposure or Observation (--ENDY)	**Limit**	**Error**
SD0056	SDTM Required variable not found	Variables described in SDTM as Required must be included in the dataset	**Metadata**	**Error**
SD0062	Incompatible data source	Domain table must have a valid format (e.g., SAS transport (XPORT) v.5 or text-delimited)	**System**	**Error**
SD0064	Invalid subject	All Subjects (USUBJID) must be present in Demograpics (DM) domain	**Cross-reference**	**Error**
SD1041	Values of --CAT and --SCAT are identical	Values of Category (--CAT) and Subcategory (--SCAT) variables should not be identical	**Data Quality**	**Warning**

There are ~250 checks currently reported from the OpenCDISC validator.

You can learn about the OpenCDISC validator, and download the tool (for free!) by visiting their website (see reference below).

CONCLUSION

The purpose of this paper is to give a high level introduction to the SDTM Fundamentals, and using simple examples provide a foundation to a new comer. This paper does not replace reading - understanding CDISC documentation, or any individual company training courses.

SDTM v1.2 & SDTM I.G. v3.1.2 was used as a basis for this paper, always refer to the CDISC website for most recent versions.

REFERENCES

CDISC Website: http://www.cdisc.org/
OpenCDISC: http://www.opencdisc.org/
PHUSE FDA Poster March 2012: http://www.phuse.eu/download.aspx?type=cms&docID=3965

CONTACT INFORMATION

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

Jennie Mc guirk
ICON
Email: jennie.mcguirk@iconplc.com