Evolution of SDTMIG 3.1.1 to 3.1.2: A mapping specialist must reference on these changes

Rachit Desai, eClinical Solutions, New London, CT
Anirudh Gautam, MaxisIT, Metuchen, NJ
Vikash Jain, eClinical Solutions, New London, CT

ABSTRACT

With the evolution of changes to the SDTM data standards model which incorporated in minor and major changes to the Classes, Domains, Variables, Controlled Terminology and its respective assumptions in SDTMIG 3.1.2 version has played a great deal of efforts for a mapping specialist to understand these evolving changes and their respective principals been implementation in compliance to these data standards and keep themselves updated on this knowledgebase with their ongoing work with in the industry to keep them in par with the competitive global market. This paper will further dive into the details of the new domains incorporated in the current version and also give an overview and details on their structure and its respective implementation with real case scenarios. The purpose of this presentation is to benefit the user by providing a clear comparison of changes incorporated to these standards and its impact to implemented standards in previous versions and also discusses further on the new guidelines outlines on which observation class the data belongs to was not part of the previous versions of implementation guide.

INTRODUCTION

The SDTM submission data standards model has been evolving since its introduction in the year 2004. The current SDTM v3.1.2 is the latest standard available and also accepted by FDA since 6 November 2009 for electronic submission of data. The current version has come a long way since its introduction, but here we would be comparing it with the previous approved version (3.1.1)

The key changes to v3.1.2 are for been outline for the following sections:

- Addition of new domains
- NCI governed Controlled Terminology
- Domain level changes to pre-existing domain models
- Changes to Assumptions for Domain Models
- Domain level changes

We will now drive into the details of new additions and changes to the standards for each if the above sections outlined.

- **Addition of new domains in Model for Special Purpose domains**
    - Subject Element (SE) and Subject Visits (SV):

      The Trial Elements, Trial Arms, and Trial Visits datasets in the Trial Design model describe the planned design of the study but it is also necessary to collect

the corresponding actual data. Subject assignment to an Arm is reported in the ARM variable in Demographics. Actual Elements and Visits data for each subject are described in two additional datasets:

- The Subject Elements dataset (SE)
- The Subject Visits dataset (SV)

(SE) and Subject Visits (SV) are moved under special-purpose domain datasets from Trial Design model.

- **Addition of new domains in General Observation Class (Events)**
  - Protocol Deviation (DV):

    The DV domain is an Events model for collected protocol deviations and not for derived protocol deviations that are more likely to be part of analysis. Events typically include what the event was, captured in --TERM (the topic variable), and when it happened (captured in its start and/or end dates). The intent of the domain model is to capture protocol deviations that occurred during the course of the study (see ICH E3, Section 10.2). Usually these are deviations that occur after the subject has been randomized or received the first treatment. This domain should not be used to collect entry criteria information. Violated inclusion/exclusion criteria are stored in IE. The Deviations domain is for more general deviation data. A protocol may indicate that violating an inclusion/exclusion criterion during the course of the study (after first dose) is a protocol violation. In this case, this information would go into DV.

  - Clinical Events (CE):

    The intent of the domain model is to capture clinical events of interest that would not be classified as adverse events. The data may be data about episodes of symptoms of the disease under study (often known as signs and symptoms), or about events that do not constitute adverse events in themselves, though they might lead to the identification of an adverse event. For example, in a study of an investigational treatment for migraine headaches, migraine headaches may not be considered to be adverse events per protocol. The occurrence of migraines or associated signs and symptoms might be reported in CE. Other studies might track the occurrence of specific events as efficacy endpoints. For example, in a study of an investigational treatment for prevention of ischemic stroke, all occurrences of TIA, stroke and death might be captured as clinical events and assessed as to whether they meet endpoint criteria. Note that other information about these events may also be reported in other datasets. For example, the event leading to death would be reported in AE and death would also be a reason for study discontinuation in DS.

- **Addition of new domains in General Observation Class (Findings)**
  - PK Concentrations (PC):

    PK Concentrations domain model contains data collected about tissue (e.g., serum or plasma) concentrations of analytes (usually study drugs and/or their metabolites) as a function of time after dosing the study drug.

Example:

Blood Sampling for Determination of Drug Concentrations in Serum

Sample No.	Sampling Time (Nominal Time)	Real Time (24 hr clocktime)	Comment	Initials
00	0 hr (= just prior to infusion of test medication)	⌊_⌋⌊_⌋ : ⌊_⌋⌊_⌋		

STUDYID	DOMAIN	USUBJID	PCSEQ	PCGRPID	PCREFID	PCTESTCD	PCTEST	PCCAT	PCSPEC	PCORRES	PCORRESU	PCSTRESC	PCSTRESN	PCSTRESU
ABC-123	PC	123-0001	1	Day 1	A554134-10	DRGA_MET	Drug A Metabolite	ANALYTE	PLASMA	<0.1	ng/mL	<0.1		ng/mL
ABC-123	PC	123-0001	2	Day 1	A554134-10	DRGA_PAR	Drug A Parent	ANALYTE	PLASMA	<0.1	ng/mL	<0.1		ng/mL
ABC-123	PC	123-0001	3	Day 1	A554134-11	DRGA_MET	Drug A Metabolite	ANALYTE	URINE	<2	ng/mL	<2		ng/mL
ABC-123	PC	123-0001	4	Day 1	A554134-11	DRGA_PAR	Drug A Parent	ANALYTE	URINE	<2	ng/mL	<2		ng/mL
ABC-123	PC	123-0001	5	Day 1	A554134-11	VOLUME	Volume	SPECIMEN	URINE	3500	mL	100	100	mL
ABC-123	PC	123-0001	6	Day 1	A554134-11	PH	PH	SPECIMEN	URINE	5.5		5.5	5.5	

PCSTAT	PCLLOQ	VISITNUM	VISIT	VISITDY	PCDTC	PCENDTC	PCDY	PCTPT	PCTPTNUM	PCTPTREF	PCRFTDTC	PCELTM	PCEVLINT
	0.10	1	DAY 1	1	2001-02-01T07:45		1	PREDOSE	0	Day 1 Dose	2001-02-01T08:00	-PT15M	
	0.10	1	DAY 1	1	2001-02-01T07:45		1	PREDOSE	0	Day 1 Dose	2001-02-01T08:00	-PT15M	
	2.00	1	DAY 1	1	2001-02-01T07:45	2001-02-01T07:45	1	PREDOSE	0	Day 1 Dose	2001-02-01T08:00	-PT15M	
	2.00	1	DAY 1	1	2001-02-01T07:45	2001-02-01T07:45	1	PREDOSE	0	Day 1 Dose	2001-02-01T08:00	-PT15M	
		1	DAY 1	1	2001-02-01T07:45	2001-02-01T07:45	1	PREDOSE	0	Day 1 Dose	2001-02-01T08:00	-PT15M	
		1	DAY 1	1	2001-02-01T07:45	2001-02-01T07:45	1	PREDOSE	0	Day 1 Dose	2001-02-01T08:00	-PT15M	

- PK Parameters (PP):

  PP Definition: Data describing the parameters of the time-concentration curve for PC data (e.g., area under the curve, Cmax, Tmax). It is recognized that PP is a derived dataset, and may be produced from an analysis dataset that might have a different structure. As a result, some sponsors may need to normalize their analysis dataset in order for it to fit into the SDTM-based PP domain

STUDYID	DOMAIN	USUBJID	PPSEQ	PPGRPID	PPTESTCD	PPTEST	PPCAT	PPORRES	PPORRESU
ABC-123	PP	ABC-123-0001	1	DAY1_PAR	TMAX	Time to Max Effect	DRUG A PARENT	1.87	h
ABC-123	PP	ABC-123-0001	2	DAY1_PAR	CMAX	Max Effect Concentration	DRUG A PARENT	44.5	ug/L
ABC-123	PP	ABC-123-0001	3	DAY1_PAR	AUC	Area Under Curve	DRUG A PARENT	294.7	h*mg/L
ABC-123	PP	ABC-123-0001	4	DAY1_PAR	THALF_1	Half-life of 1st exp phase	DRUG A PARENT	0.75	h
ABC-123	PP	ABC-123-0001	5	DAY1_PAR	THALF_2	Half-life of 2nd exp phase	DRUG A PARENT	4.69	h
ABC-123	PP	ABC-123-0001	6	DAY1_PAR	VD	Vol of Distribution	DRUG A PARENT	10.9	L

PPSTRESC	PPSTRESN	PPSTRESU	PPSPEC	VISITNUM	VISIT	PPDTC	PPRFTDTC
1.87	1.87	h	PLASMA	1	DAY 1	2001-03-01	2001-02-01T08:00
44.5	44.5	ug/L	PLASMA	1	DAY 1	2001-03-01	2001-02-01T08:00
294.7	294.7	h.mg/L	PLASMA	1	DAY 1	2001-03-01	2001-02-01T08:00
0.75	0.75	h	PLASMA	1	DAY 1	2001-03-01	2001-02-01T08:00
4.69	4.69	h	PLASMA	1	DAY 1	2001-03-01	2001-02-01T08:00
10.9	10.9	L	PLASMA	1	DAY 1	2001-03-01	2001-02-01T08:00

- Drug Accountability (DA):

  Drug Accountability is for data regarding the accountability of study drug, such as information on receipt, dispensing, return, and packaging.

- Microbiology Specimen (MB):

  The MB domain is designed to store microbiology findings that include organisms found, grain stain results and organism growth status.

- Microbiology Susceptibility Test (MS):

  The MS domain is designed to store any findings related to the organisms found and submitted in MB. This will usually consist of susceptibility testing results, but can also be other organism-related findings such as extent of growth of an

organism. This domain is intended to be used in conjunction with the MB domain described above.

- Findings About:

  Findings About Events or Interventions is a specialization of the Findings General Observation Class. Findings About domain needs to be used when collected CRF data represent an Event or Intervention which can not be part of Event/ Intervention or Supplemental Qualifier.

Example:

Pre-Specified Adverse Events In this example, three AEs are pre-specified and are scheduled to be asked at each visit. If the occurrence is yes, then a complete AE record is collected on the AE form. Pre-Specified Adverse Events.

The Findings About record and the AE record may be linked via RELREC

Pre-Specified Adverse Events of Clinical Interest	
Date of Assessment	DD-MMM-YYYY
Did the following occur? If Yes, then enter a complete record in the AE CRF	
Headache	○ No ○ Yes ○ Not Done
Respiratory infection	○ No ○ Yes ○ Not Done
Nausea	○ No ○ Yes ○ Not Done

fa.xpt

STUDYID	DOMAIN	USUBJID	FASEQ	FATESTCD	FATEST	FAOBJ	FAORRES	FASTRESC	FASTAT	FADTC	VISITNUM	VISIT
QRS	FA	1234	1	OCCUR	Occurrence	Headache	Y	Y		2005-10-01	2	VISIT 2
QRS	FA	1234	2	OCCUR	Occurrence	Respiratory Infection	N	N		2005-10-01	2	VISIT 2
QRS	FA	1234	3	OCCUR	Occurrence	Nausea			NOT DONE	2005-10-01	2	VISIT 2
QRS	FA	1234	4	OCCUR	Occurrence	Headache	Y	Y		2005-10-10	3	VISIT 3
QRS	FA	1234	5	OCCUR	Occurrence	Respiratory Infection	N	N		2005-10-10	3	VISIT 3
QRS	FA	1234	6	OCCUR	Occurrence	Nausea	Y	Y		2005-10-10	3	VISIT 3

ae.xpt

STUDYID	DOMAIN	USUBJID	AESEQ	AETERM	AEDECOD	AEBODSYS	AESEV	AEACN	AEPRESP	AESTDTC	AEENDTC
QRS	AE	1234	1	Headache	Headache	Nervous system disorders	MILD	NONE	Y	2005-09-30	
QRS	AE	1234	2	Nausea	Nausea	Gastrointestinal disorders	MODERATE	NONE	Y	2005-10-08	2005-10-09

- **NCI governed Controlled Terminology**

  For SDTMIG V3.1.1 the presence of a single asterisk (*) or a double asterisk (**) in the “Controlled Terms or Format” column indicated that a discrete set of values (controlled terminology) was expected for the variable. This set of values was sponsor-defined in cases where standard vocabularies had not yet been defined (represented by *) or from an external published source such as MedDRA (represented by **). For V3.1.2, controlled terminology is now represented one of three ways which is been governed by NCI:

  - A single asterisk when there is no specific CT available at the current time, but the SDS Team expects that sponsors may have their own CT and/or the CDISC Controlled Terminology Team may be developing CT
  - A list of controlled terms for the variable when values are not yet maintained externally
  - The name of an external code lists whose values can be found via the hyperlinks in either the domain or Appendix C form SDTMIG 3.1.2 approved version

  The current list of controlled terminology is located on the CDISC website at http://www.cancer.gov/cancertopics/terminologyresources/CDISC

- **Domain level changes to pre-existing domain models**
    - Addition of new variables for respective domains
        - Domain: CONCOMITANT MEDICATIONS — CM
            - Added new variable CMPRESP, CMSTRTPT, CMSTTPT, CMENRTPT
        - Domain: ADVERSE EVENTS — AE
            - Added new variable AEPRESP, AEENRTPT, AEENTPT
        - Domain: SUBSTANCE USE — SU
            - Added new variable SUPRESP
        - Domain: MEDICAL HISTORY — MH
            - Added new variable MHPRESP, MHENRTPT, MHENTPT
        - Domain: PHYSICAL EXAMINATION — PE
            - Removed expected variable PESTRESN and PESTRESU
        - Domain: EXPOSURE — EX
            - Added permissible variables EXVAMT and EXVAMTU
- **Changes to Assumptions for Domain Models**
    - **Submission of free text**
      Data collected as free text is usually a part of a list of choices and accompanied by "Other, specify". The placement of free-text value would differ as per the below mentioned scenarios.
        - When associated with a non-result qualifier then the free text value should be included in the SUPPQUAL.
        - But when the free text value is associated with a qualifier field then the sponsor may choose to either include it in SUPPQUAL or directly in the qualifier variable field within the domain.
        - If the free text value is associated with a result field then the free text is always included in the Original result (--ORRES) and its up to the sponsor to either include the full/partial free text value in the character result (--STRESC) or put 'OTHER'.
        - For Intervention and Events class, the respective Topic variables can have the free text value collected on CRF and --PRESP could be used to distinguish between pre-specified and free text responses. For Findings class, TEST would have the value of the free text and TESTCD would have to coded.
    - **Splitting Domains**
      This is a completely new assumption that has been introduced in version 3.1.2 which allows the sponsor to split a domain into physically different datasets by a non missing category or sub-category. There are certain rules that need to be followed so as to ensure that the spitted domains can be appended back together.
    - **Grouping and Categorization of variables**
      The sponsor may choose to categorize a topic variable into --CAT and --SCAT so as to uniquely identify a record, however these values should not be redundant within --DECOD and --BODSYS. The implementation guide also suggests on how to group variables and how does --GRPID differ from --CAT and --SCAT.

Hierarchy of Grouping Variables
STUDYID
DOMAIN
--CAT
--SCAT
USUBJID
--GRPID

- **Assignment of Natural Keys**

  Variables contributing to the natural key for a domain should be included in the metadata. These are not static for a domain model and may differ from sponsor to sponsor based on the collection of the data so as to uniquely identify a record within a domain. This provides freedom to the sponsor to include result variables as a part the natural key.

- **Origin of Metadata for Variables**

  The SDTMIG defines certain controlled terms that would populate the Origin column of the define.xml. These controlled terms are as follows:

  CRF – For data collected on the CRF

  eDT – For data received via electronic Data Transfer. Applicable for lab, ECG or IVRS

  Derived – For variables derived by the Sponsor based on an algorithm from other data values

  Assigned – For variables that are either coded as a part of coding process (--DECOD) or values set independently so as to fill SDTM fields (DOMAIN etc.)

  Protocol – For variables having values coming from the protocol as a part of the trial (EXDOSE, VSPOS etc.)

- **Submitting multiple values for a variable**

  A Topic variable, for an Intervention (--TRT) or Event (--TERM) class, may have multiple values reported. For such cases the sponsor usually either resolves the multiplicity or splits the values in to independent records.
  A Result variable for a Finding class (--ORRES) having multiple values reported for a test must be submitted in multiple records for the same test (--TESTCD). But in order to preserve uniqueness of the data, the sponsor may choose to include result variable as a part of the natural key for that domain.
  Multiple values for a Qualifier variable must be stored in the SUPPQUAL of the domain and the value for the Qualifier variable should be “MULTIPLE”. The values of QNAM and QLABEL should be more specific in order to clearly link the multiple values with its Domain.

- **Supplemental Reason variables**

  The sponsor sometimes collects the reason for something was done or not done. Then variable --REASND, a part of the general observation class, handles the reason for not done. Similarly --INDC and --ADJ variables are present in Intervention class to handle the reason for an intervention. The Findings class

has no predefined variable that could handle such scenario; therefore any reasoning value must go to SUPPQUAL.

- **Relative Timing variables**

  The latest version of SDTMIG (v3.1.2) adds four new timing variables, --STRTPT, --STTPT, --ENRTPT, and –ENTPT, so as to express relative timing assignments such as “Prior” or “Ongoing” that are collected at other times of the study and is not handled by --STRF and --ENRF variables. Although these new variables contain similar values as --STRF and --ENRF but is not limited to the study reference period and can be linked to any timing description.

  For a reference time point corresponding to the date of collection or assessment, an observation can start BEFORE that time point, can start COINCIDENT with that time point, or it is unknown (U) when it started. An observation can end BEFORE that time point, can end COINCIDENT with that time point, can be known that it didn‘t end but was ONGOING, or it is unknown (U) at all when it ended or if it was ongoing. In this case AFTER is not a valid value because it would represent an event after the date of collection.

  If the reference time point is prior to the date of collection or assessment then an observation can start BEFORE the reference point, can start COINCIDENT with the reference point, can start AFTER the reference point, or it may not be known (U) when it started. An observation can end BEFORE the reference point, can end COINCIDENT with the reference point, can end AFTER the reference point, can be known that it didn‘t end but was ONGOING, or it is unknown (U) when it ended or if it was ongoing.

- **Domain level changes**

  - DM
    - Role of BRTHDTC changed from “Result Qualifier” to “Record Qualifier”
    - Changed variable label for AGE from “Age in AGEU at RFSTDTC” to “Age” to remove names of other variables in variable labels. AGE does not have to be derived from RFSTDTC
    - ARMCD is restricted to 20 characters and not 8 characters

  - CO
    - Added VISITNUM, VISITDY and VISIT variables
    - IDVAR and IDVARVAL role changed from “Identifier” to “Record Qualifier” and COVAL role changed from “Result Qualifier” to “Topic”

  - CM
    - Changed variable label for CMDOSTOT from ” Total Daily Dose Using DOSU” to “Total Daily Dose” to remove names of other variables in variable labels

  - EX
    - Added assumption that Exposure data is required
    - Changed variable label for EXDOSTOT from “Total Daily Dose Using DOSU” to “Total Daily Dose” to remove names of other variables in variable labels

  - EG
    - VISITNUM changed from Required to Expected
    - EGEVAL changed from Expected to Permissible

- LB
  - VISITNUM changed from Required to Expected
- PE
  - Removed expected variables PESTRESN and PESTRESU and also removed variable PEBLFL
- SC
  - SCDTC changed from Expected to Permissible
- VS
  - VISITNUM changed from Required to Expected

CONCLUSION

The outlined changes and additions described in this paper are the once described as major changes in SDTMIG 3.1.2 version. But there are many detail changes at micro level which is a task for a mapping specialist to keep a close eye on. These changes to standards are ongoing and evolving which will be updated in future versions, and always opens up a scope of reference for the mapping specialist to refer to such changes, new addition to standards either at class, domain or variable levels and many more aspects of the functionality in implementing these standards to streamline the data reporting process.

REFERENCES

CDSIC SDTM Implementation Guide 3.1.1 and 3.1.2

ACKNOWLEDGMENTS

I would like to acknowledge eClinical Solution for providing the opportunity to work on this paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Please one of the authors at:

Rachit Desai
Email: rdesai@eclinicalsol.com

Anirudh Gautam
Email: agautam@egistar.com

Vikash Jain
Email: vjain@eclinicalsol.com or jainvikash77@yahoo.com