PhUSE US Connect 2018

Paper SI04

SDTM: It is not all Black and White

Swapna Pothula, SGS - Agriculture, Food and Life, Mechelen, Belgium

ABSTRACT

CDISC SDTM is no longer recommended, but is rather a required standard specification for all clinical trial data submissions. SDTM comes with SDTMIG, a guide to help us create SDTM compliant databases.

SDTM is a vast standard in theory, but a guide in practice with gaps and gray areas left for interpretation. The more the variety of therapeutic areas, (e)Source systems, providers and sponsors we work with the more will be the differences in how SDTM is interpreted and databases are created. The possibility of SDTMIG being interpreted differently by different people while still being SDTM compliant can jeopardize the very concept of standardization.

This paper will give a programmers' perspective of the conflicts that arise while creating SDTM databases given the requirements, recommendations, expectations, limitations and gaps. Examples along with the approach used at SGS to navigate these gaps to optimize compliance and enhance consistency will be discussed in detail.

INTRODUCTION

CDISC SDTM is no longer recommended or desired but is rather a required standard specification for all clinical trial data submissions as of October 2016 for the PDMA and December 2016 for the FDA. SDTM enhances quality, eases exchange and speeds up the overall review process. SDTM is intended to standardize the way the data is tabulated and comes with SDTMIG a guide to help us create CDISC SDTM compliant databases.

THE GRAY AREAS

SDTM is a vast standard in theory but a guide in practice with gaps and gray areas left for interpretation. The possibility of the implementation guide being interpreted differently by different people while still being SDTM compliant can jeopardize the very concept of standardization.

It is not just the domains but also variables especially those that are derived are open for various interpretations. The more the variety of therapeutic areas, (e)Source systems, providers and sponsors we work with the more will be the differences in how SDTM is interpreted and databases are created.

A few scenarios that lead to conflicts, gaps, gray areas and limitations of SDTM as a standard will be discussed below. The discrepancies between the regulatory authorities such as the FDA, PMDA and CDISC SDTM standards leaves us with conflicting scenarios where the choice of compliance to either the regulatory authorities or the CDISC SDTM has to be chosen.

Screen Failures

One such scenario is the way the ARM/ARMCD are required to be populated by FDA v/s CDISC SDTM for screening failures. The CDISC requirement in case of screening failures is to complete the ARM / ARMCD with Screen Failure / SCRNFAIL (Fig.1, Fig.2). The FDA guidance for screening failures is quite the opposite requiring the ACTARM/ACTARMCD to be left empty (Fig.3)

- Data for screen failure subjects, if submitted, should be included in the Demographics dataset, with ARMCD = "SCRNFAIL" and ARM = "Screen Failure". Sponsors may include a record in the Disposition dataset indicating when the screen failure event occurred. DM/SE Example 6 shows an example of data submitted for a screen failure subject.

Fig.1

Row	SITEID	INVNAM	BRTHDTC	AGE	AGEU	SEX	RACE	ETHNIC
1 (cont)	01	JOHNSON, M	1948-12-13	57	YEARS	M	WHITE	HISPANIC OR LATINO
2 (cont)	01	JOHNSON, M	1955-03-22	50	YEARS	M	WHITE	NOT HISPANIC OR LATINO
3 (cont)	01	JOHNSON, M	1938-01-19	68	YEARS	F	BLACK OR AFRICAN AMERICAN	NOT HISPANIC OR LATINO
4 (cont)	01	JOHNSON, M	1941-07-02			M	ASIAN	NOT HISPANIC OR LATINO
5 (cont)	02	GONZALEZ, E	1950-06-23	55	YEARS	F	AMERICAN INDIAN OR ALASKA NATIVE	NOT HISPANIC OR LATINO
6 (cont)	02	GONZALEZ, E	1956-05-05	49	YEARS	F	NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDERS	NOT HISPANIC OR LATINO

Row	ARMCD	ARM	ACTARMCD	ACTARM	COUNTRY
1 (cont)	A	Drug A	A	Drug A	USA
2 (cont)	P	Placebo	P	Placebo	USA
3 (cont)	P	Placebo	P	Placebo	USA
4 (cont)	SCRNFAIL	Screen Failure	SCRNFAIL	Screen Failure	USA
5 (cont)	P	Placebo	P	Placebo	USA
6 (cont)	A	Drug A	A	Drug A	USA

Fig.2

DM Domain (Demographics)
In the DM domain, each subject should have only one single record per study.

Screen failures, when provided, should be included as a record in DM with the ARM field left blank. For subjects who are randomized in treatment group but not treated, the planned arm variables (ARM and ARMCD) should be populated, but actual treatment arm variables (ACTARM and ACTARMCD) should be left blank.[24]

Fig.3

Given the scenario above, we are forced to either complete the ARM / ARMCD or leave them blank implying the database cannot be compliant to both CDISC and FDA in the case of screening failures. Also, choosing either of the options will generate output on Pinnacle 21.

At SGS, the approach is to leave the ARM / ARMCD blank (Fig 4), thereby being FDA compliant and documenting the explanation for CDISC non conformance on the Study Data Reviewers' Guide. Since the clinical databases are submitted to the FDA we have chosen FDA compliance over CDISC compliance in this particular case.

S..	U..	I..	DOMAIN	S...	RFSTDTC	RFENDTC	RFXSTDTC	RFXENDTC	RFICDTC	RFPENDTC	DTHDTC	DTHFL	SITEID	BRTHDTC	AGE	AGEU	SEX	RACE	ARMCD	ARM	ACTARMCD	ACTARM	COUNTRY
G..	G..	G..	DM	S022					2017-04-27	2017-05-05T08:07			01	1968-09-25	48	YEARS	M	WHITE					BEL
G..	G..	G..	DM	S021					2017-04-27	2017-05-05T08:06			01	1973-04-16	44	YEARS	M	WHITE					BEL
G..	G..	G..	DM	S020					2017-04-18	2017-04-18			01	1970-04-04	47	YEARS	M	WHITE					BEL
G..	G..	G..	DM	S019					2017-04-18	2017-04-18			01	1989-09-30	27	YEARS	M	WHITE					BEL
G..	G..	G..	DM	S018					2017-04-18	2017-04-18			01	1971-10-20	45	YEARS	M	WHITE					BEL
G..	G..	G..	DM	S014					2017-04-18	2017-04-18			01	1972-05-06	44	YEARS	M	WHITE					BEL
G..	G..	G..	DM	S009					2017-04-14	2017-04-14			01	1967-04-14	50	YEARS	M	WHITE					BEL
G..	G..	G..	DM	S008					2017-04-14	2017-04-14			01	1974-12-12	42	YEARS	M	WHITE					BEL
G..	G..	G..	DM	S006					2017-04-14	2017-04-14			01	1976-12-16	40	YEARS	M	WHITE					BEL
G..	G..	G..	DM	S004					2017-04-14	2017-04-14			01	1968-05-06	48	YEARS	M	WHITE					BEL

Fig.4

Conflict of Interest

SDTM serves the purpose of standardization at the core. All the entities that are part of the clinical trial life cycle benefit of SDTM standardization at varying degrees while performing their day to day tasks. SDTM could serve other purposes apart from standardization depending on the entity the database is used by. For example, compliance to SDTM standard could be a major concern for a programmer who creates the databases, while a statistician might be interested in the extent SDTM supports his analysis and a reviewer on the other hand might prioritise traceability over standardization. This is another major factor to be considered while creating SDTM compliant databases which

can lead to different interpretations of the standards due the difference in the purpose of SDTM for different entities.

A classic example of conflict of interest is the way the ACTARM/ACTARMCD are populated for subjects following an unplanned ARM. The CDISC requirement in this case is to complete the ACTARM/ACTARMCD with Unplanned Treatment / UNPLAN (Fig 5).

CDISC Requirement

Variable Name	Variable Label	Type	Controlled Terms, Codelist or Format	Role	CDISC Notes	Core
ACTARMCD	Actual Arm Code	Char	*	Record Qualifier	Code of actual Arm. When an Arm is not planned (not in Trial Arms), ACTARMCD will be UNPLAN. Randomized subjects who were not treated will be given a value of NOTTRT. Values should be "SCRNFAIL" for screen failures and "NOTASSGN" for subjects not assigned to treatment. Restricted to values in Trial Arms in all other cases. ACTARMCD is limited to 20 characters and does not have special character restrictions. The maximum length of ACTARMCD is longer than for other short variables to accommodate the kind of values that are likely to be needed for crossover trials.	Req
ACTARM	Description of Actua	Char	*	Synonym Qualifier	Description of actual Arm. When an Arm is not planned (not in Trial Arms), ACTARM will be "Unplanned Treatment". Randomized subjects who were not treated will be given a value of "Not Treated". Values should be "Screen Failure" for screen failures and "Not Assigned" for subjects not assigned to treatment. Restricted to values in Trial Arms in all other cases.	Req

Fig.5

At SGS we follow the CDISC requirement to complete the ACTARM/ ACTARMCD with Unplanned Treatment / UNPLAN for subjects following a different ARM than the assigned ARM (Fig 6). When a subject does not complete all the foreseen elements, for example (Fig 6) when a subject is assigned DCAB as ARMCD but then dropped out after completing the elements DCA, the subject is assigned UNPLAN ARMCD in the Demographics dataset.

However, lately we have been receiving comments from our clients requesting us to populate the actual treatment received in the ACTARMCD as having the UNPLAN ARMCD value in the DM dataset is not very convenient for analysis of the data.

S..	U..	I..	DOMAIN	SUBJID	RFSTDTC	RFENDTC	RFXSTDTC	RFXENDTC	RFICDTC	RFPENDTC	DTHDTC	DTHFL	SITEID	BRTHDTC	AGE	AGEU	SEX	RACE	ARMCD	ARM	ACTARMCD	ACTARM	COUNTRY
G..	G..	G..	DM	101	2017-05-05T08:35	2017-06-06	2017-05-05T08:35	2017-05-29T08:30	2017-04-14	2017-06-06T09:01			01	1976-12-26	40	YEARS	M	WHITE	DCAB	SCREENING-300 mg ...	DCAB	SCREENING-300 mg ...	BEL
G..	G..	G..	DM	102	2017-05-05T08:40	2017-06-06	2017-05-05T08:40	2017-05-29T08:35	2017-04-14	2017-06-06T08:57			01	1979-02-08	38	YEARS	M	WHITE	BACD	SCREENING-300 mg ...	BACD	SCREENING-300 mg ...	BEL
G..	G..	G..	DM	103	2017 05 05T08:45	2017 06 06	2017 05 05T08:45	2017 05 29T08:40	2017 04 14	2017 06 06T09:20			01	1970 08 21	46	YEARS	M	WHITE	ADBC	SCREENING 300 mg ...	ADBC	SCREENING 300 mg ...	BEL
G..	G..	G..	DM	104	2017-05-05T08:50	2017-06-06	2017-05-05T08:50	2017-05-29T08:45	2017-04-14	2017-06-06T12:11			01	1995-07-13	21	YEARS	M	WHITE	CBDA	SCREENING-300 mg ...	CBDA	SCREENING-300 mg ...	BEL
G..	G..	G..	DM	105	2017-05-05T08:55	2017-06-06	2017-05-05T08:55	2017-05-29T08:50	2017-04-14	2017-06-06T11:00			01	1994-01-07	23	YEARS	M	WHITE	DCAB	SCREENING-300 mg ...	DCAB	SCREENING-300 mg ...	BEL
G..	G..	G..	DM	106	2017-05-05T09:00	2017-06-06	2017-05-05T09:00	2017-05-29T08:55	2017-04-14	2017-06-06T10:06			01	1991-03-08	26	YEARS	M	WHITE	BACD	SCREENING-300 mg ...	BACD	SCREENING-300 mg ...	BEL
G..	G..	G..	DM	107	2017-05-05T09:05	2017-06-06	2017-05-05T09:05	2017-05-29T09:00	2017-04-18	2017-06-06T09:49			01	1970-01-10	47	YEARS	M	WHITE	ADBC	SCREENING-300 mg ...	ADBC	SCREENING-300 mg ...	BEL
G..	G..	G..	DM	108	2017-05-05T09:10	2017-06-06	2017-05-05T09:10	2017-05-29T09:05	2017-04-14	2017-06-06T10:22			01	1992-06-11	24	YEARS	M	WHITE	CBDA	SCREENING-300 mg ...	CBDA	SCREENING-300 mg ...	BEL
G..	G..	G..	DM	109	2017-05-05T09:15	2017-06-06	2017-05-05T09:15	2017-05-29T09:10	2017-04-14	2017-06-06T10:34			01	1967-08-17	49	YEARS	M	WHITE	ADBC	SCREENING-300 mg ...	ADBC	SCREENING-300 mg ...	BEL
G..	G..	G..	DM	110	2017-05-05T09:20	2017-06-06	2017-05-05T09:20	2017-05-29T09:15	2017-04-18	2017-06-06T10:14			01	1966-05-31	50	YEARS	M	WHITE	BACD	SCREENING-300 mg ...	BACD	SCREENING-300 mg ...	BEL
G..	G..	G..	DM	111	2017-05-05T09:25	2017-05-31	2017-05-05T09:25	2017-05-29T09:20	2017-04-18	2017-05-31T10:20			01	1966-06-16	50	YEARS	M	WHITE	DCAB	SCREENING-300 mg ...	UNPLAN	Unplanned Treatment	BEL
G..	G..	G..	DM	112	2017 05 05T09:30	2017 06 06	2017 05 05T09:30	2017 05 29T09:20	2017 04 14	2017 06 06T10:45			01	1994 02 03	23	YEARS	M	WHITE	CBDA	SCREENING 300 mg ...	CBDA	SCREENING 300 mg ...	BEL

Fig.6

In the example above, it has been requested to complete the value DCA in the ACTARMCD of the Demographics dataset since the obvious variables to look for the actual treatment received would be the DM.ACTARMCD or DM.ACTARM from the analysis point of view although this information can easily be obtained from the Exposure dataset. This is pure conflict of interest between that of a programmer mapping the SDTM database and the statistician performing the analysis.

It should also be noted that having the actual treatment received captured in the ACTARM / ACTARMCD variables of the DM dataset will generated output on Pinnacle 21. The check verifying if all the ACTARM / ACTARMCD values used in the Demographics dataset are listed in the Trial Arms, TA dataset, which describes each of the planned ARMs of the clinical trial. Doesn't this imply that all the unforeseen ARM possibilities should be foreseen in the TA dataset?

We have clients willing to have the Pinnacle 21 output documented in the Study Data Reviewers' Guide while having the ACTARM / ACTARMCD completed with the actual treatment received. Another option to avoid output on Pinnacle 21 would be to have the actual treatment information captured in the SUPPDM dataset. But, this would lead to data duplication given the actual treatment information being available in the Exposure domain.

We at SGS are very much in favour having the ACTARM / ACTARMCD completed with Unplanned Treatment /

UNPLAN as expected and in compliance with CDISC SDTM standards.

When capturing data in more than one domain is better

While we agree with the idea of not capturing data twice to avoid data duplication, there are certain kinds of data that need to capture in more than one domain to achieve completeness of the clinical database. One such example situation is illustrated below (Fig. 7), which is a page extract of an annotated CRF.

The page depicted below (Fig.7) is set up for capturing the data from the Audiometric and Vestibular assessments that were performed during the trial. At the first look it is very clear that the data from these assessments will be captured in the Findings About domain, which is the appropriate domain given the type of the data.

However, it has to be noted that Findings about domain data is always about some event that has occurred or was performed. The Findings about domain data in most of the cases is obtained as a result of a procedure, event or intervention that has occurred or was initiated during the trial conduct. So, a better and cleaner approach would be to have the data captured in two domains. In the example below (Fig. 7), the data therefore is captured in both the Procedures (PR) and Findings About (FA) domain.

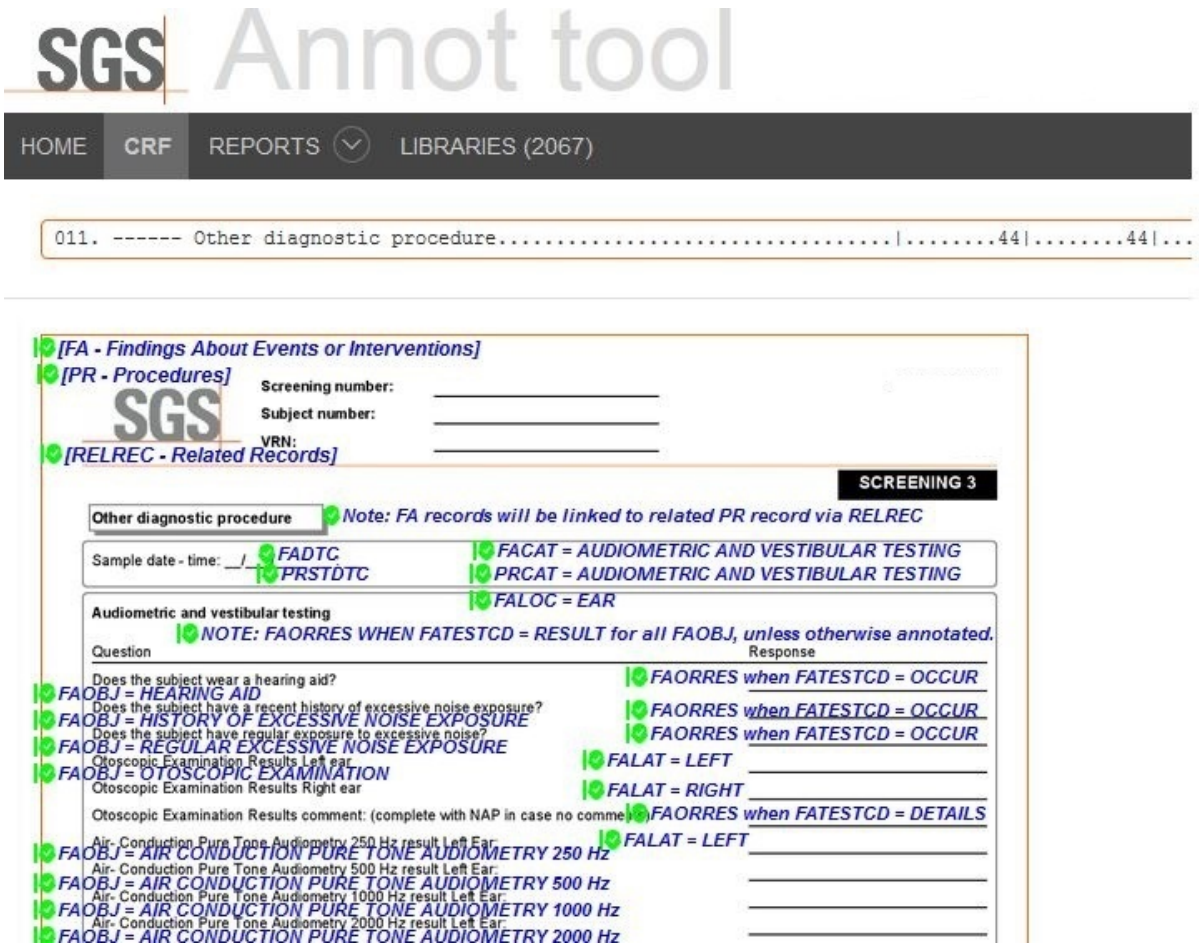

Fig.7

Going a step further in this direction to be able to give more meaning and completeness to the database, these two domains Findings About and Procedures should be linked by creating records in the Related Records (RELREC) domain. CDISC SDTM standards nowhere mention the linking of Findings About domain to the corresponding event, procedure or intervention that led to the Findings About data. And not linking the domains FA and PR in the example below (Fig. 7) would not generate any output corresponding to the missing Related Records on Pinnacle 21.

At SGS our approach to mapping and capturing the Findings About domain data is to identify the event, intervention or procedure that leads to the Findings About data and capture data in both the FA and the corresponding event, intervention or procedure in the applicable events domain and then link both the domain with records in the RELREC

domain.

USUBJIDs for Multiple Enrollments

The FDA requirement for subjects participating in multiple studies or roll over studies that are to pooled for a single submission is to maintain the same USUBJID across all the studies (Fig.8). This requirement, however has many cons when implemented at the operational data management level.

For example the review, reporting, reconciliation and querying become very difficult and confusing. Consider cleaning, reconciliation and querying of provider data. Given the nature of the data, having a same USUBJID across all the studies that are to be pooled becomes an extremely tedious and time consuming process due to lack of the study the issue or query belongs to.

The SUBJID variable available in Demographics dataset can come in handy in such situations, but only to an extent due to limited availability of this variable, in the Demographics domain only.

Unique Subject Identifier (USUBJID)
Each individual subject should be assigned a single unique identifier across the entire application. This is in addition to the subject ID (SUBJID) used to identify subjects in each study and its corresponding study report. An individual subject should have the exact same unique identifier across all datasets, including between SDTM and ADaM datasets. Subjects that participate in more than one study should maintain the same USUBJID across all studies. It is important to follow this convention to enable pooling

U.S. Food & Drug Administration
10903 New Hampshire Avenue
Silver Spring, MD 20903
www.fda.gov

Page 9 of 38

March 2017

Contains Nonbinding Recommendations

FDA U.S. FOOD & DRUG ADMINISTRATION

of a single subject's data across studies (e.g., a randomized control trial and an extension study).

Fig.8

Given the requirement, the cons of its implementation and the limitation of the SDTM model we at SGS have two approaches to work our way around. One approach is to use a unique USUBJID for the same subject participating in each of the studies that are to be pooled together during the setup, cleaning, reconciliation and querying processes and assign the same USUBJID at later stages when the database is clean and ready to be pooled or submitted thus being FDA compliant and efficient with the database setup and cleaning activities.

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG	QEVAL
	DM	001-101-007	SUBJID	101-024	PREVUSUB	Subject ID from Preceding Study	101007	CRF	
	DM	001-101-009	SUBJID	101-017	PREVUSUB	Subject ID from Preceding Study	101-009	CRF	
	DM	001-400-003	SUBJID	400-006	PREVUSUB	Subject ID from Preceding Study	400-003	CRF	
	DM	001-102-001	SUBJID	102-002	PREVUSUB	Subject ID from Preceding Study	102-001	CRF	
	DM	001-300-002	SUBJID	300-003	PREVUSUB	Subject ID from Preceding Study	300-002	CRF	

Fig.9

The other approach is assigning a same USUBJID for a single subject for all the studies he has participated in while capturing each of the corresponding SUBJIDs in the Supplemental Qualifiers, in this case SUPPDM domain (Fig.9). This way the SUBJID from the each of the preceding and subsequent studies can be identified from the SUPPDM dataset. For this approach to work, the SUBJID information from the previous studies has to be captured on the CRF, lack of this information leaves us with the former approach mentioned above.

It is therefore important to collect the information regarding the previous studies in case of studies that are to be pooled at least in cases where we are aware that the studies are going to be pooled. In other cases, we are to find work around that work the best given the (e)Sources systems, client standard specifications and the people performing the various tasks on the clinical databases.

HOW DO WE NAVIGATE THESE GAPS?

The task of creating SDTM compliant databases can be daunting for a programmer given the requirements, recommendations, purpose, expectations, limitations and gaps as these lead to duplicate data, incorrect mapping of data and inconsistencies. At SGS, we have devised an approach to interpret the CDISC SDTM standards as part of the SDTM implementation to enable us navigate the gaps efficiently.

SGS has developed a guide to interpret the CDISC SDTM standards and FDA requirements to be able to optimize the process of creating clinical databases. This guide is named: SGS Interpretation Guide, SGS IG that has been created at sponsor standard level given the variety of client base we have.

The SGS IG is a living document that complements the CDISC SDTM, the SDTM IG (Interpretation Guide) and sponsor standards. The interpretation guide clarifies the variables that are vaguely defined with examples indicating the domains the data can be mapped to. The document is a collection of business rules that define the data to be captured and domains where the data has to be captured in. The document also lists a set of documentation rules describing what should be documented for example in the SDRG, with a standard set of explanations for known output on Pinnacle 21.

Apart from the SGS IG, we have a team of SMEs who curate and version the controlled terminology. This has led to optimization and efficiency especially in case of extensible codelists. The new codelist values are subject to approval of both the sponsor and SGS SMEs before being used to capture data. These process improvements have enabled us to navigate the gray areas, gaps and limitations consistently and efficiently.

CONCLUSIONS

As we have seen from the real time examples, confusions and conflicts arise on a day to day basis while creating SDTM compliant databases. Multiple valid interpretations are made by various people who are interested in the extent SDTM supports their tasks. This possibility of the standard being interpreted lead to the aforementioned data duplication, incorrect mapping and inefficient programming.

The checks that check the structure and content of clinical databases for compliance can only check what has been mapped. They cannot check for what has been incorrectly mapped. Data duplication and incorrect mapping are consequences of gaps, gray areas and limitations in the existing standards.

We at SGS have created an Interpretation guide, SGS IG which has played a major role in fine tuning the process of SDTM mapping while optimizing compliance and enhancing consistency. To add to that there is need for change in the perspective while creating the clinical databases. The databases should not be mapped and created with data perspective or standard perspective but rather with Data-Standard perspective. This change in perspective can bring in consistency and efficiency despite the gaps and limitations.

REFERENCES

Study data technical conformance guide (FDA) – March 2017 release
CDISC Study Data Tabulation Model, 1.4
CDISC Study Data Tabulation Implementation Guide, Version 3.2

CONTACT INFORMATION

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

Swapna Pothula
SGS – Agriculture, Food and Life
Generaal de Wittelaan 19A Bus 5
B-2800 - Mechelen
Email: Swapna.Pothula@sgs.com
Web: www.sgs.com/cro