PharmaSUG 2019 - Paper SI-320

Vendor's Guide to Consistent, Reliable, and Timely CDISC Deliverables

Dharmendra Tirumalasetti, Vita Data Sciences, Waltham, USA;

Santosh Lekkala, Vita Data Sciences, Waltham, USA;
Bhavin Busa, Vita Data Sciences, Waltham, USA

ABSTRACT

When working in a CRO/FSP model, programmers have to work on multiple clinical studies across different sponsor companies and various data collection systems. In addition, even though the submission standards are common across the industry, the requirements and expectations for CDISC deliverables could differ from one sponsor to another. The differences could be based on multiple factors such as therapeutic areas, internal data standards, and study-specific needs. Also, the interpretation and handling of the data may differ between the sponsor companies. If not understood and documented earlier by the vendor, the differences could cause re-work at a later stage which adds up to delayed deliverables, rework time and extra cost. To meet the sponsor's specific needs and expectations, it is highly recommended to have and follow effective processes with-in the organization. This paper will describe processes we follow at our organization beforehand, during the specification development and programming of the CDISC datasets in order to achieve consistent, reliable and timely deliverables thus benefiting both the Sponsor and the vendor. This paper also provides details on the pre-processing steps that can be followed before writing the specification, quality checks and data handling during the development process of the datasets and informed notes during the delivery of datasets to the Sponsor to ensure expected outcome.

INTRODUCTION

The process of generating the SDTM datasets starts with understanding the requirements and continues with annotating the CRF, developing the specifications, programming and validation of the datasets and checking for compliance against the current validation rules. Generation of consistent and reliable datasets needs a two-way co-ordination effort which requires a clear set of expectations from the sponsor, meticulous attention to detail and timely response from the vendor to make sure that expectations are met. If there is no clear set of expectations for various components beforehand, it results in re-work at a later stage where implementation becomes difficult and this increases the overall cost and time.

Although each vendor may have their own SDTM development processes, there are various components and steps that can be implemented to ensure that we have required information right before the start. This paper provides an insight to the steps and a set of pre and well-defined processes to develop effective and consistent SDTM and strongly believe that these processes would help any vendor/CRO to achieve a goal of reliable and submission ready SDTM data.

PROCESSES

THE PRE-REQUISITE STEPS THAT CAN BE FOLLOWED BEFORE WRITING THE SPECIFICATION

The pre-processing steps are those that are to be taken into consideration before proceeding with writing the specifications for the study. The following questions answered beforehand would be beneficial and this would involve a few confirmations from the sponsor. The lead programmer must review all the documents related to study and make sure they have the following information.

Topic	Required Checks
Protocol (All versions)	All the protocol versions are to be requested from the sponsor before proceeding to write the specifications and need to make sure if the latest one being used.

Topic	Required Checks
SDTM Model and IG Version	The general practice is to use the latest version available. But, in some cases, if the sponsor determines to use a previously available version to make it consistent with the other studies - they may request to use a previously available version.
CT, MedDRA and WHO Drug Dictionary versions	Get the information from Sponsor on Control Terminology, MedDRA and WHO Drug versions to specify in and assist during the specifications writing/programming.
eCRF With Raw Annotations	Electronic CRF with RAW annotations would be helpful to navigate and search for required variable names and assist during specification writing.
Source Data Analysis	It would be better to have a list of all the raw datasets that are being used in the study along with the associated SDTM dataset to which they are being mapped to and confirmation on list of SDTM domains from the sponsor before writing the specifications. This would help avoid re-work which when done at a later stage might affect the quality and timeline.
External Data Transfer Specifications	Review and let the sponsor know of any changes that might be required such as the TEST and TESTCD, addition of new variables, format of the transfer, frequency of the transfer, Original and Standard units for the data received, Conversion factor if required. This review helps to make sure all the required variables are present and checks for any data inconsistencies that could affect the specification writing/programming.
Frequency of Transfer	We need to determine the frequency of transfer to plan and execute the delivery process so that delay can be minimized. These include the frequency of transfer for - RAW data, external data from sponsor and SDTM datasets to the sponsor. Appropriate care must be taken to have enough buffer time
P21/ Define Version	The version of P21 used to validate the define and datasets along with the version of the define required by the sponsor needs to be confirmed.
Study Tracker	A study tracker is to be prepared with all the tasks such as the datasets planned, production and QC programmer names, point of contact for questions
Randomization	Randomization, if any, in the study needs to be taken care of. A dummy randomization file needs to be requested with the various fields that would be available when the final randomization file is received so that it can be incorporated during programming and would not require an update at a later stage
Study Milestones	Determining the study milestones is an important part which provides an idea on various deliverable components to both the Sponsor and the Vendor. It determines various important steps during the study and helps in determining the deliverables
CRF Annotation	It is recommended to annotate the CRF, without any bookmarks, for sponsor review before proceeding with specification writing.

Table 1. Pre-requisite Steps

QUALITY CHECKS AND DATA HANDLING DURING THE DEVELOPMENT PROCESS

The quality checks and data handling are done during the specification development and programming part. Once the specifications are ready it is recommended to have the sponsor review the specifications. A checklist is recommended to be included along with the specification with major points of focus which requires critical attention to detail so that errors are minimized at a later stage. The following points would not only help the reviewer get an overall picture but also makes the review process faster and easier. The following checklist can be provided to the sponsor:

Domain	Topic	Values Expected/Derivation	Agree/Disagree
TA/DM	ARM and ARMCD	'XX01', 'XX02'	
TA/DM	Actual vs Planned Arm	'XX01', 'XX02'	
TA/TE/SE	Elements	Screening, Treatment, Follow-	
TA/TE	Epoch	SCREENING, TREATMENT,	
TV/SV	Planned Visits	Visit 1, Visit 2, etc.,	
DM	RFSTDTC vs RFXSTDTC	First day and time of dosing in	
DM	RFENDTC vs RFXENDTC	Last drug dose date in EX	
DM	RFPENDTC	Last available date for a subject in the entire study	
SV	Data sets to be considered for VISITS	Considerations for creation of SV dataset. Check with the Sponsor whether a single master dataset, if available, is to be used or all the datasets are to be used. Also check if any external data such as the LB or PC	
SV	Handling of "UNSCHEDULED VISITS"	Naming of UNSCHEDULED VISITS such as 1.01	
TS	TSPARM/TSVAL	Missing information or information unavailable through open sources for various TSPARM's are to be confirmed from sponsor	
AE	Derivation of AETRTEM	Set to "Y" for all AE's occurring after treatment start date (or) existing AE's which have	
AE	Control terminology used for AEREL, AEACN, AEENRF, etc.	Control terminology used for mapping the collected values are to be confirmed	
DS	Values for DSTERM	Screening, Randomization,	
All Findings Domain	BLFL derivation method	Last non-missing value collected before study drug administration	
All Findings Domain	TEST and TESTCD per CT	TEST and TESTCD's being used for all findings domains are to be confirmed.	
All Domains as required	Determination of EPOCH derivation.	If SE is being used or any other derivation method is being used to determine EPOCH.	
All Domains as required	Mapping of External data	Example: SDTM.LB dataset. Merging process to be followed and how missing data is being represented (EX: LBALL for missing TESTS by each category)	

Domain	Topic	Values Expected/Derivation	Agree/Disagree
XX	Custom datasets	Custom datasets are to be confirmed by the sponsor for the approach and appropriate type into which they are being mapped to like events or interventions.	

Table 2. Checklist to sponsor when Specification Document sent for review

During the programming and validation part, we follow the following processes to make sure that the SDTM domains are generated with high quality:

- **Metadata Difference Check:** A standard macro which performs a check on the RAW data transfers received from sponsor and lists all the changes such as addition of new variables, any changes to the field types etc. Any changes between two RAW data transfers are sent to the sponsor and get the confirmation from sponsor that these are expected and to be mapped.
- **SDTM Metadata Shells:** Create a list of SDTM shell datasets from define.xml with zero records and all the attributes from define.xml so that the programmers need not worry about the lengths and labels while writing the program and just incorporate the shell datasets in their program for all the attributes.
- **Log Checks:** A macro which summarizes any errors or warnings in a log while programming for any SDTM dataset. This would help programmers to fix any errors/warnings as soon as he/she finds the m in the log and no need to search through the log for any errors/warnings manually.
- **Compare Outputs:** A macro which scans all the compare SAS® output files in Validation folder and lists any unequal results, attribute differences between production and validation datasets. This helps reviewer from vendor side/programming lead to understand and confirm that there are no mismatches between production and validation datasets.
- **Compliance Check:** Pass the generated datasets through P21 validator and address all issues as required. It is recommended to generate and pass the define along with the datasets when validating through P21 to make any changes as required to either define, datasets or both.
- **Manual Check:** A manual check of all the datasets is required before sending the delivery to the sponsor (such as running customized edit checks) for those that cannot be determined through P21 and requires a diligent Human Eye. (Ex: P21 does not throw an Error (or) Warning when RFSTDTC is present but ACTARM is provided as “Not Assigned”).

INFORMED NOTES DURING THE DELIVERY OF DATASETS TO THE SPONSOR

The informed notes provided during the delivery of Datasets to the sponsor is an important part which ensures that the basic high-level information regarding the delivery has been provided to the sponsor. This may include information such as:

- The total number of subjects present in the transfer (if possible, a breakdown by number of subjects per ARM, such as 5 Screen failures, 10 treated to drug, 10 to placebo, etc.)
- The total number of datasets in the transfer
- Missing datasets if any.
- Any information regarding specific subjects that the vendor thinks would be helpful for review by the sponsor.
- Data Issues recorded in a log.
- The Data cut for RAW data source such as RAW data cut date, External data cutoff date, etc.,
- Clarification Log for any information that requires sponsor response.
- Information regarding any blank datasets that have not been included in the transfer, due to lack of corresponding information, but are part of the delivery.
- Any errors, warnings from P21 that standout from the normal ones.

CONCLUSION

Achieving consistent and reliable datasets starts right from determining what the expectations are and is an ongoing process until the end of the study. The different components mentioned above make sure the interpretation, handling of data is done correctly and all the pieces such as specification development, programming, compliance and ease of understanding are met. This in turn would benefit both the sponsor and vendor by making sure that all the deliverables are done on time without any delay or re-work. Following the above steps as part of SDTM development process, one can achieve reliable datasets right from the start.

REFERENCES

CONTACT INFORMATION

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

Name: Dharmendra Tirumalasetti

Company: Vita Data Sciences

Address: 281 Winter Street, Suite 100

City / Postcode: Waltham, MA 02451

Phone: 781-833-0257

Email: dtirumalasetti@vitadatasciences.com

Web: www.vitadatasciences.com

Name: Santosh Lekkala

Company: Vita Data Sciences

Address: 281 Winter Street, Suite 100

City / Postcode: Waltham, MA 02451

Phone: 713-851-7559

Email: slekkala@vitadatasciences.com

Web: www.vitadatasciences.com

Name: Bhavin Busa

Company: Vita Data Sciences

Address: 281 Winter Street, Suite 100

City / Postcode: Waltham, MA 02451

Phone: 781-373-8455

Email: bbusa@vitadatasciences.com

Web: www.vitadatasciences.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.
Other brand and product names are trademarks of their respective companies.