Study Data and Electronic Submission Formats

Soma Sekhar Koramatla
Senior Statistical Programmer

PhUSE 2017 Single Day Event
22nd April 2017

Disclaimer

ALL OPINIONS EXPRESSED IN THIS PRESENTATION ARE THE PRESENTER'S PERSONAL VIEWS, AND DO NOT REFLECT THE VIEWS OR OPINIONS OF QUARTESIAN

Agenda

- Introduction
- Study Data Submission Format
- Electronic Submission Format
- Q & A

Study Data Technical Conformance Guide

FDA document that Provides specifications, recommendations, and general considerations on how to submit standardized study data

- Planning and providing standardized study data
- Exchange format
- **Study data submission format**
- Therapeutic area standards
- Terminology
- **Electronic submission format**
- Study data validation and traceability

https://www.fda.gov/downloads/forindustry/datastandards/studydatastandards/ucm384744.pdf

Study Data Submission Format (SDTM)

Annotated Case Report Form (aCRF) for SDTM

- Regardless of whether the clinical database is legacy or SDTM compliant, an aCRF should be submitted.
- SDTM Metadata Submission Guidelines should be followed to annotate CRF.
- aCRF should be provided as a PDF with the file name "**acrf.pdf**".
- When data are recorded on the CRF but are not submitted, then CRF should be annotated with the text "**NOT SUBMITTED**" and description should be provided in the cSDRG stating why data was not submitted.

SDTM Domain Specifications

DM (Demographics)

- Each subject should have only one record per study.
- ARM should be left blank if a subject is Screen Failure.
- If subjects randomized to treatment group but not treated then planned arm variables (ARM and ARMCD) should be populated and actual treatment arm variables (ACTARM and ACTARMCD) should be left blank.
- Although this convention is inconsistent with the SDTMIG, FDA recommends its use so that “Screen Failure” is not specified as a treatment arm

DM (Demographics) (contd.)

Subject Identifier (SUBJID)

- If same subject is screened more than once in a trial, then the subject's SUBJID should be different.

Unique Subject Identifier (USUBJID)

- Subjects should have unique Identifier across all the datasets including 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 of a single subject's data across studies.
- USUBJID should not have leading or trailing spaces in any dataset.

DS (Disposition)

- If there is more than one disposition event, then EPOCH or DSSCAT variable should be used to distinguish between event. This will allow identification of the EPOCH in which each event occurred or DSSCAT to differentiate if the disposition is for treatment or study.

Row	STUDYID	DOMAIN	USUBJID	DSSEQ	DSTERM	DSDECOD	DSCAT	EPOCH	DSDTC	DSSTDTC
1	ABC123	DS	123101	1	INFORMED CONSENT OBTAINED	INFORMED CONSENT OBTAINED	PROTOCOL MILESTONE		2003-09-21	2003-09-21
2	ABC123	DS	123101	2	COMPLETED	COMPLETED	DISPOSITION EVENT	SCREENING	2003-09-29	2003-09-29
3	ABC123	DS	123101	3	RANDOMIZED	RANDOMIZED	PROTOCOL MILESTONE		2003-09-30	2003-09-30
4	ABC123	DS	123101	4	COMPLETED	COMPLETED	DISPOSITION EVENT	TREATMENT	2003-10-31	2003-10-31
5	ABC123	DS	123101	5	COMPLETED	COMPLETED	DISPOSITION EVENT	FOLLOW-UP	2003-11-15	2003-11-15

- If a death occurs, it should be the last record and should include its associated EPOCH.

AE (Adverse Events)

- AE domain should include all adverse events that were recorded in the subjects' CRF, regardless of whether the sponsor determined that particular events were or were not treatment-emergent.
- If Subject has SAE it should have the assessment indicated, (e.g., as a death, hospitalization, or disability/permanent damage).

LB (Laboratory Test Results)

- The size of the LB domain dataset is often too large to process, so by splitting a large LB dataset into smaller datasets according to LBCAT and LBSCAT, using LBCAT for initial splitting.

  Example : lb1.xpt for chemistry, lb2.xpt for hematology, and lb3.xpt for urinalysis.

- If the size is still too large, then LBSCAT should be used for further splitting.

- Splitting the dataset in other ways (e.g., by subject or file size) makes the data less useable.
- Sponsors should submit these smaller files in addition to the larger non-split standard LB domain file.

LB (Laboratory Test Results) (contd.)

- Sponsors should submit the split files in a separate sub-directory/split that is clearly documented in addition to the non-split standard LB domain file in the SDTM datasets directory.

SE(Subject Elements)

- SE dataset should be included to aid in the association of subject data (e.g., findings, events, and interventions) with the study element in which they occurred.

Trial Design Model (TDM)

- All TD datasets should be included, as appropriate for the specific clinical trial, in SDTM submissions as a way to describe the planned conduct of a clinical trial.
- Specifically, the Trial Summary dataset will be used to determine the time of study start.
- Sponsors submitting legacy data should provide a TS dataset (ts.xpt) which includes the study start date in the form of SSTDTC (TSPARMCD = SSTDTC) and TSVAL= “yyyy-mm-dd”.

Custom Domains

- SDTMIG permits the creation of custom domains if the data do not fit into an existing domain.
- Prior to creating a custom domain, sponsors should confirm that the data do not fit into an existing domain
- To create custom domains, sponsors should follow the recommendations in the SDTMIG
- Sponsors should present their implementation approach in the cSDRG.

SUPPQUAL (Supplemental Qualifiers)

- It is a special SDTM dataset that contains non-standard variables which cannot be represented in the existing SDTM domains.
- SUPPQUAL should be used only when key data cannot be represented in SDTM domains.
- In general, variables used to support key analyses should not be represented in SUPPQUAL.
- If the sponsor intends to include important variables (e.g., that support key analyses) in SUPPQUAL datasets then explanation should be provided in cSDRG.

SDTM General Considerations

- SDTM Versions, When submitting clinical or nonclinical data, sponsors should not mix versions within a study.
- SDTM datasets should not contain imputed data.
- Naming conventions (variable name and label) and variable formats should be followed as specified in the SDTMIG.
- Dates in SDTM domains should conform to the ISO 8601 format.
- SDTM variables are categorize as Required, Expected, and Permissible.

SDTM General Considerations (contd.)

- For SDTM submissions, all Required, Expected, and Permissible variables that were collected, plus any variables that are used to compute derivations, should be submitted.
- FDA recognizes that SDTM contains certain operationally derived variables that have standard derivations across all studies (e.g., --BLFL,--STDY, EPOCH).
  - Baseline flags for finding domains should be submitted if the data were collected or can be derived.
  - EPOCH should be included for clinical subject-level observation. This will allow the reviewer to easily determine during which phase of the trial the observation occurred (e.g., screening, treatment, follow-up).

SDTM General Considerations (contd.)

- If Timing variables –DTC, --STDTC or –ENDTC are included, then matching Study Day variables (--DY, --STDY, or -- ENDY, respectively) should be included. For example, in most Findings domains, --DTC is Expected, which means that --DY should also be included.

Data Definition Files for SDTM

- The data definition file is the most important part of the electronic dataset submission for regulatory review.
- Separate data definition files should be included for each type of electronic dataset submission, Example SDTM, SEND and ADaM .
- The data definition file should be submitted in XML format, i.e., a properly functioning define.xml.
- In addition to the define.xml, a printable define.pdf should be provided if the define.xml cannot be printed.

Data Definition Files for SDTM (contd.)

- Define.xml version 2.0 is the preferred version.
- Sponsors should include a reference to the style sheet as defined in the specification and place the corresponding style sheet in the same submission folder as the define.xml file.

Electronic Submission Format

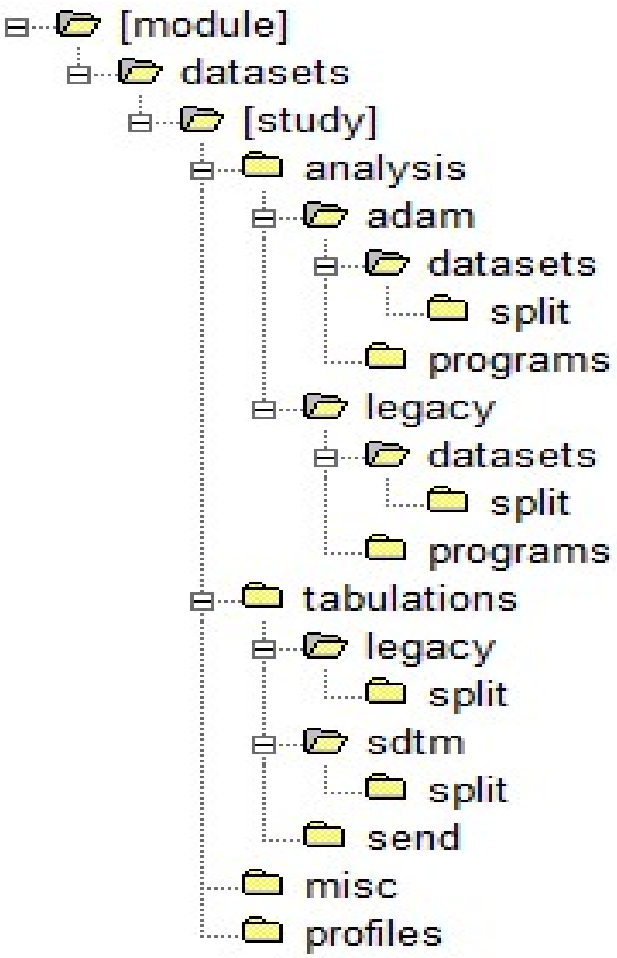
eCTD File Directory Structure

- eCTD - Electronic Common Technical Document
- Study datasets and their supportive files should be organized into a specific file directory structure when submitted in the eCTD format.
- Submission of files within the appropriate folders allows automated systems to detect and prepare datasets for review, and minimizes the need for manual processing.
- Do not submit empty file folders and additional subfolders.
- If additional folders are needed, consult with the appropriate center in advance for guidance.

eCTD File Directory Structure (contd.)

- If you need to split a file that exceeds file size limits, you should submit the smaller split files in the “split” sub-folder in addition to the larger non-split file in the original data folder.
- There is no need for a second define.xml file to be submitted within the split subfolder.

Folder Structure for Study Datasets



Study Dataset and File Folder Structure and Description

Folder Name	Folder Level	Description/Contents
[module]	1	Refers to the eCTD module in which study data is being submitted. Name this folder m4 for nonclinical data and m5 for clinical data. Do not place files at this level.
datasets	2	Resides within the module folder as the top-level folder for study data (nonclinical or clinical) being submitted for the specified module (m4 or m5). Do not place files at this level.
[study]	3	Name this folder with the study identifier or analysis type performed (e.g., study123, iss, ise). Do not place files at this level.
analysis	4	Contains folders for analysis datasets and software programs; arrange in designated level 6 subfolders. Do not place files at this level.
adam	5	Contains subfolders for ADaM datasets and corresponding software programs. Do not place files at this level.
datasets	6	Place ADaM datasets in this subfolder.
split	7	Place any split ADaM datasets in this subfolder.
programs	6	Place software programs for ADaM datasets, tables and figures in this subfolder.
legacy	5	Contains legacy formatted analysis datasets and corresponding software programs. Do not place files at this level.
datasets	6	Place legacy analysis datasets in this subfolder.
split	7	Place split legacy analysis datasets in this subfolder.
programs	6	Place software programs for legacy analysis datasets, tables and figures in this subfolder.
misc	4	Place miscellaneous datasets that don't qualify as analysis, profile, or tabulation datasets in this subfolder. This subfolder was formerly named "listings".
profiles	4	Place patient profiles in this subfolder.

Study Dataset and File Folder Structure and Description

Folder Name	Folder Level	Description/Contents
tabulations	4	Contains subfolders for tabulation datasets. Do not place files at this level.
legacy	5	Place legacy (non-standardized) tabulation datasets in this folder.
split	6	Place any split legacy tabulations datasets in this subfolder.
sdtm	5	Place SDTM tabulation datasets in this subfolder. Should only be used in m5 for clinical data.
split	6	Place any split SDTM files in this subfolder.
send	5	Place SEND tabulation datasets in this subfolder. Should only be used in m4 for animal data.

Quartesian
The Data Company

Q & A

Thank You

Soma Sekhar Koramatla

Senior Statistical Programmer.

Quartesian Clinical Research Pvt. Ltd.

E-mail : somasekhar.k@quartesian.com

Ph : +91-9036568901