CDISC SDTM Conversion, a Case Study in Perfection
Abstract

This paper discusses the conversion of raw datasets to Clinical Data Interchange Standards Consortium (CDISC), Study Data Tabulation Model (SDTM) format and Define XML including reviewers guide for a global, tier one pharmaceutical company. The conversion strategy, challenges faced, and process based conversion approach which facilitated highly complex studies conversion. This paper also discusses NAVITAS’ success in quality of deliverables with-in a given budget and time frame for successful submission without RTF (Refusal to File.) Quality is paramount, as regulatory agencies will not accept any errors in this area.

Introduction

This ongoing multi-year project involves conversion of raw data from multiple therapeutic areas into CDISC SDTM standards and creating submission ready documents. Different therapeutic areas include Reproductive System, Cardiovascular, Oncology, Imaging, and Gastrointestinal. Each area contains multiple studies of Phases I to III. The following diagram shows therapeutic areas and time of conversion.

Journey
Conversion Process

Within the scope of these projects were (1) Creating SDTM aCRF (2) Trial design domains (3) Mapping specification based on requirement (4) Macro development to create SDTM datasets from source datasets (5) Validation (6) Generating submission ready documents define.xml (7) Preparation of Reviewers Guide (8) Compliance Checks [WebSDM, OpenCDISC Validator] (9) Documentation

There are two conversion methodologies based on sponsor requirement.

- Transforming raw data to SDTM model using SDTM metadata - This involves in creating Mapping specification document which contains source metadata, SDTM metadata and rules to migrate from source to SDTM. Programming is done based on this mapping specifications document.

Highlights of the conversion:
- Successfully delivered over 20 Pivotal studies across 6 submissions in the last five years
- 100% track record of On-Time, Quality Delivery within budget
- Zero Refusal-to-File by Agency
- Capability to deliver highly complex studies
- Successfully validated CDISC conversions of other CRO, closely collaborating with Sponsor and other vendors.
Transforming raw data to SDTM model using Sponsor specific target metadata – Mapping specification document is not created for this method. Target metadata sent by Sponsor contains information whether a variable migrates to SDTM or ADaM. Programming is done based on this metadata and variables are mapped to SDTM/SUPP accordingly. Gap analysis is done prior to start of programming to check if sponsor metadata is SDTM compliant and also if CDISC Controlled terminology is applied correctly in the data as per standards. Issues identified are discussed with sponsor and updates to metadata are done when required. Details of issues identified is discussed in challenges section.

Figure 2: Conversion process using sponsor Target metadata model – Mapping spec is not created

Challenges

Issues in sponsor specific metadata:

- Sponsor metadata has issues with Length i.e., for certain standard variables like TEST/CD, QNAM, QLABEL, DTC length in metadata exceeded that required. For some variables, length in metadata is less than actual length in raw dataset.
- Variables collecting Prompt questions information in CRF which are not needed for submission are noted as SDTM SUPP variables.
- Variables collecting important information needed for submission in CRF are not considered in sponsor metadata.
- Variable values in Trial summary dataset are not SDTM compliant.
- Sponsor requirement was to strictly follow their metadata for SDTM conversion as they created ADaM datasets using this metadata. Any changes in metadata happens during SDTM conversion will impact ADaM conversion. This resulted in firing some Open CDISC checks in SDTM conversion.
**Coding Issues**

- Dictionary coding values are provided at the end of conversion which affected length of dictionary variables.

**Custom domain Issues**

- Naming a custom domain has to be in line with the sponsor specific standards. Hence before proposing a custom domain, needing to get approval from the sponsor.

**Achieving Efficiency**

The project implementation strategy includes the following:

**Gap analysis and Communication Plan:** Gap analysis is done to check if sponsor specific metadata is SDTM compliant. This involves both programmatic checks using SAS based utilities as well as manual checks. Open CDISC checks are also run to check for SDTM compliance. Issues identified during these checks are tracked in the issue log. Issues are classified in different categories to ease sponsor in decision making, e.g., Sponsor metadata required but Metadata update can’t be done for various reasons but document in Reviewers Guide for submission, Update metadata as needed. Issues with Length etc. are documented and proposed length values are suggested and sent for client’s confirmation. If variable values are not SDTM compliant, proposed SDTM controlled terminology values are documented in the issue tracking repository. The Communication Plan created between Navitas and the sponsor is showed in the figure below. Issues are discussed regularly in weekly teleconferences, emails or face to face meetings if needed. All resolutions and decisions taken are documented properly in this repository to make consistent interpretation of the logical mapping and reusability of annotation, sponsor metadata updates and also for auditing purposes. All related documents (source data, conversion data and issue tracking spreadsheet) shared between Navitas and the sponsor are located in a secured SharePoint area.

**ISSUE TRACKING AND RESOLUTION REPOSITORY MANAGEMENT**
Sponsor Involvement with Deliverables

The Sponsor review and approval on the aCRF is the first step. This helped Navitas in creating custom domain based on sponsor requirements. The Sponsor suggested changes based on custom domain naming and were communicated early in the process before creating mapping specification and programming. Updates to be done in metadata are confirmed by sponsor early in the process after issues are sent based on Gap analysis. Several Decisions were made after reviewing first few studies to go back and change the logical mapping which was identified easily thorough impact analysis.

Conclusion

Converting legacy data to SDTM is a complex process that requires a zero tolerance factor for errors. The structure has been defined by the agency (FDA) and is a strict standard that enables the agency to use in house programs and tools for their review. The conversion of legacy data requires an experienced vendor, who possesses in-depth knowledge of the CDISC model and experts in the field who are familiar with CDISC and its overall mission. Process is key in relation to quality control of the final deliverables. Therapeutic area experience only enhances the level of comfort, which Sponsors must have with their vendor.