SDTM Cartography - Learn To Create SDTM Mapping Specifications

Donna Sattler, Eli Lilly and Company, Indianapolis, IN
Michael Lozano, Eli Lilly and Company, Indianapolis, IN

ABSTRACT

You may have a very good idea of what you need to deliver your clinical trial to regulatory agencies. However, the journey you take to collect, clean, combine, and migrate from multiple data sources can be overwhelming. This paper will explore the Data Integration or Standardization mapping process. It will demonstrate that knowing your destination is harder than it seems when multiple paths are presented. Do you take the path less traveled because you think it better suits your travel plans or should you stick with what the tried and true path proves?

INTRODUCTION

It is helpful but not required to have known how to map (transform, integrate, standardize) data before reading this paper. It will be helpful to understand how Clinical Trial data is collected in relational databases and how it is ultimately used in a Regulatory Submissions.

After this training you will have a High Level understanding of the concepts behind Mapping Data or Standardizing Data Collection. You will understand the questions to ask, the tools to utilize and the common pain points to hopefully avoid, or define and fix along the way.

DEFINING MAPPING CLINICAL DATA AS RELATED TO CARTOGRAPHY

MAPPING IN CLINICAL TRIALS refers to the transformation of one structure of data (metadata and content) to another structure. Simply put, it sometime means, putting a square peg in a round hole. Standardized data is not only required now by regulatory reviewers but also from academia and for the sharing and analysis of data to further a research effort. Consider one company buying another company and all of their data. Wouldn't it be nice to be able to utilize your investment with very little rework and cost associated with making the data useful? In order for Clinical Trial data to be analyzed and understood by the regulator reviewer CDISC standards are now required.

In Clinical Trials, data is collected in several formatted structures. Some are relational (key variables are the same in all datasets to be used together), for example:

In relational databases a unique identifier is used to link the study participant in each dataset. And some not as much, (no real cohesive relationships). Data is collected but no instructions on what the rows and columns represent and how to make connections with other data.

Sometimes keys are the same in the standard structures and sometime you need to retrofit the data into a standards structure. The more standard the data is upfront the more predictive it is when fitting it into a standard structure like CDISC SDTM.

By the end of creating a mapping specification, you will understand where and how the data came from (eDC and CRF data and external data sources: Labs, Biomarker, ECGs…) and where it eventually belongs.

SUBJECT	DOSE	UNIT	TREAT	DATE
123	2.95	mg/kg	good	10JAN2017

Table 1: One type of treatment data collected on CRF

patient	result	un	med	date
000123	14.75	ml	drugx	1/5/2017

Table 2: Another type of data from a hospital for similar medication

USUBJID	xxDOSE	XXDOSU	xxTRT	XXSTDTC
ABC-XYZ-000123	2.95	mg/kg	GOODDRUG	2017-01-10
ABC-XYZ-000123	14.75	ml	DRUGX	2017-01-05

Table 3: Standardized data from previous 2 records

In understanding the data mapping process, it is helpful to understand the data flow. Who touches the data? Who is going to use the data and how is it going to be used?

In map making, the process can be generalized into three groups:

1. **Explorers** who collect information as they go about the land or sea, capturing detailed specifics that are vital in helping others understand what they witnessed (cities, roads, bodies of water, changes in the terrain, etc.)

2. **Cartographers** who compile the information they receive from the explorers and compile it into a diagrammatic representation that follows specific and accepted formats and techniques.

3. **Navigators** who follow the diagrammatic representations to understand and follow (figuratively and/or literally) where the explorers have already gone.

Clinical Trial Mapping has several Roles and Responsibilities:

1. In Clinical Trials, the **physicians, statisticians, trial managers, data managers, etc**. are collecting information as they go.   This information is in the form of clinical trial data and contains the stories of each subject in the trial. Think of these individuals as the Explorers.

2. The **Data Standards Analyst (or SDTM Analyst)** compile the types of information being collected and creates a map to SDTM domains so that each part of these stories will be able to be viewed in a uniformed and understandable way. Think of these individuals as the Cartographers

3. The **SDTM programmers** follow the maps created by the Data Standards Analyst. The programmers ensure that the clinical trial data are delivered to their designated destinations on the map so that the subjects' stories are available and ready to be told. Think of these individuals as the Navigators

Roles and Responsibilities	Explorers	Cartographers	Navigators
Physicians, Statisticians, Trial Managers, Data Managers etc…	x		
Standards or SDTM Analyst		x	
Data Standards/ SDTM Programmers			x

Table 4: Clinical Trial Mapping Roles and Responsibilities

HOW DO WE MAKE A MAP?

So how does the map get created? What steps and considerations need to be taken into account in order to place the data into domains that are not only in compliance with CDISC rules, but also will help those who follow the maps understand the subject's journey through the clinical trial.

Following CDISC standards is the starting point. Understanding Implementation Guides (IGs), Therapeutic Area User Guides (TAUGs), Controlled Terminology, and other CDISC processes and principles must be the primary objective of any mapping exercise.

But CDISC standards only provide so much guidance when difficult data situations arise. Sometimes the old joke "you can't get there from here" really does seem like reality. And just as maps need to account for indirect routes and rough terrains so those who follow them stay on route, the mapping of data must be done in a way that is CDISC compliant, but understandable and useful to those who need it.

As previously mentioned the cartographer must follow specific and accepted formats and techniques in order to create a map that is going to be understood by those trying to use it or follow it. The result is that each map is expected to contain six parts that are necessary for the navigators to understand what the map represents. When the cartographer strays from this practice, either by leaving one or more parts out of the map, or by altering the accepted formats and techniques, it creates uncertainty in the minds of the navigators. The more uncertainly exists, the less useful or effective the map.

1. The first of the six parts of the map is the **Title**. The Title describes what the map is going to show us. What the map is all about. Without a title, it would be difficult for the navigators to understand what they were viewing.
    a. In SDTM, the equivalent to the Title is the **domain names and descriptions**. A domain contains a collection of related observations with a topic-specific data points about the subjects in the trial.
2. The second of the six parts are **Symbols**. Symbols are the points, lines, and patterns that are listed in the map key. Symbols represent real objects on the map. Whenever possible, Symbols and their descriptors should have consistency (if Saint Louis is spelled without abbreviations, then St. Paul with an abbreviation would be discouraged).
    a. **Variables** are the symbols of data. They represent real observations about the subject. The rules for defining variables are the same as defining a symbol on the map; it must have both the symbolic representation and the description to be useful. And just as not all symbols will be used in all states, not all variables in the domain standard will be used in the output datasets.
3. The third part of the map is the **Scale**. The Scale provides perspective to the map. What are the distances between two points? The scale gives a verbal representation (1 inch = 100 miles) that can easily be understood as a visual interpretation.
    a. To the SDTM Cartographer, the scale is the **metadata; Item Oids, Order, Mandatory, Role, Name, DataType, Length, Significant Digits, Origin, etc.** The Metadata not only

gives structure and additional meaning to each variable, but it also provides consistency to like variables across domains.

4. The fourth part of the map is the **Compass**. With the Compass comes direction which provides orientation so the reviews know which side of the map is up.
    a. **Domain classifications** (Events, Findings, Interventions, and Special Purpose), unique identifiers, Topic Variables or Parameters, Timing Variables, and Qualifiers set the direction for the data

5. The fifth part of the map is **Location**. Latitude and Longitude are used to pin point location. To provide coordinates for a single point. To simplify: Latitude is a horizontal line designating the north or south angle from the equator. Longitude is the vertical line designating the east or west angle from the Prime Meridian.
    a. SDTM has to put much of the data into **vertical data structure** when the raw data was horizontal. While this can be straight forward, often more difficult data mappings produce a domain that would be better classified as "vertizontal". Location of the data is not only important to adhere to the rules of the domain's classification, but also to clearly relate **value-level metadata** to the proper topic variables.

6. The sixth part of the map is the **Date**. In May of 1817, the Union only had 19 states. Interstates didn't show up until 1956. Knowing when the map was created adds additional understanding to what is being viewed.
    a. While the actual date might not be that important in mapping SDTM domains, knowing the **SDTM IG version** that is being used is vital. With information available from **SDTM IG v3.1.1 to the expected SDTM IG v3.3**, it is important for the navigators to know what course they are exploring. If and when the mappings are "borrowing" from later versions of SDTM, care must be taken to ensure that the domains and variable borrowed are treated as such – through sponsor-defined domains, supplemental qualifier variables, etc. in order to avoid confusion.

Parts of the Map and how they relate to A Mapping Spec	Title	Symbols	Scale	Compass	Location	Date
Domains	x					
Variables		x				
Metadata			x			
Domain Classification				x		
Horizontal or Vertical Collection					x	
Standard Version(s)						x

Table 5: Six Parts of the Map making process

HOW DO WE FOLLOW IT?

What steps can you take to ensure you do not get lost along the way when programming? Who doesn't like the idea of using a Map to get to a new destination, whether Google Maps is your method or old school Rand McNally, you need to depend on a few things to know where you're going….

1. You can get lost when parts of a map are not provided (like no legend) and when you do not understand how to read the map (example Requirement Flags). If you understand the map and how it was assembled, you should not get lost. But if you do get lost, find your landmarks like MAPPING RULES to help find your true North. These industry guidances can be your **breadcrumbs:**

   a. Protocol

   b. SAP

   c. Data Standards/ eDC/ CRF

   d. Sponsor-Specific/ CDISC Data Standards Catalogs

   e. FDA, Technical Conformance Guidelines

   f. Industry Tool for Compliance Checks

   g. CDISC.org for implementation User Guides

DATASET	VARIABLE	LABEL	MAPPING RULE	ORIGIN	CTLIST	LENGTH	DATATYPE
CM	STUDYID	Study Identifier	"ABC"	Assigned		12	text
CM	DOMAIN	Domain Abbreviation	"CM"	Assigned		4	text
CM	USUBJID	Unique Subject Identifier	Set to (STUDYID||'-'||demog.site||'-'||SUBJID)	Derived	|	14	text
CM	SUBJID	Subject Identifier for the Study	Set to conmed.subject	CRF		6	text
CM	CMTRT	Reported Name of Drug, Med, or Therapy	Set to conmed.treat	CRF	|	200	text
CM	CMCAT	Category for Medication	Set to conmed.cat	CRF		100	text
CM	CMDOSE	Dose per Administration	Set to conmed.dose	CRF		8	float
CM	CMDOSU	Dose Units	Set to conmed.unit	CRF	(UNIT)	20	text
CM	CMSTDTC	Start Date/Time of Medication	Set to conmed.startdt, and format to ISO 8601	CRF		20	datetime
CM	CMENDTC	End Date/Time of Medication	Set to conmed.enddt, and format to ISO 8601	CRF		20	datetime

Table 6: Examples of Concomitant Medications (CM) Metadata Mapping Specification

The Columns of your mapping spec can contain Input and your Target Metadata parameters. You would read your spec left to right. For example, In the CM domain, the USUBJID variable is the Unique Subject Identifier. This variable is **Derived** from concatenating the STUDYID||'-'||demog.site||'-||SUBJID. Example = ABC-XYZ-000123. The length will be a value of 14.

A Direct Mapping example would be, in the CM domain, the CMTRT variable is the Reported Name of the Drug, Med, or Therapy. This variable is **a direct mapping** from the input data = conmed.treat Example = CMTRT = TYLENOL. The length will be a value of 200.

Acronym	Description
CDISC	Clinical Data Interchange Standards Consortium
SDTM	Standard Data Tabulation Model
IG	Implementation Guide
TAUG	Therapeutic Area User Guide
SAP	Statistical Analysis Plan

Table 6: Examples of Concomitant Medications (CM) Metadata Mapping Specification

CONCLUSION

You should have a High Level understanding of the concepts behind Mapping Data or Standardizing Data Collection. Create your mapping spec with knowing where you're going, what is your target keep in mind the 6 parts of your map. You should understand the questions to ask, such as what version of the Standards am I mapping to? Avoid pain points such as CRF collection not aligning to expectation, which will lead to programming rework. Finally, don't forget about the breadcrumbs if you get lost along the way.

RECOMMENDED READING

What is Data Mapping? Copyright © 2017. "Industry Research." Accessed March, 1, 2017

http://www.bridging-the-gap.com/what-is-data-mapping/

What is Data Mapping? 08/08/2014, "Industry Research." Accessed March, 1, 2017

https://adeptia.com/blog/what-data-mapping

Adam J. Sicard and Susan H.M. Boquist. 2016. "It is a standard, so it is simple, right?": Misconceptions and Organizational Challenges of Implementing CDISC at a CRO." PharmaSUG 2016. PAREXEL, Billerica, MA

http://www.lexjansen.com/pharmasug/2016/DS/PharmaSUG-2016-DS06.pdf

Annie Guo, 2011 "Ensuring Consistent Data Mapping Across SDTM-based Studies – a Data Warehouse Approach" PharmaSUG 2011, ICON Clinical Research, North Wales, PA

http://www.lexjansen.com/pharmasug/2011/CD/PharmaSUG-2011-CD21.pdf

REFERENCES

CDISC Study Data Tabulation Model (SDTM) v1.4 and Study Data Tabulation Model Implementation Guide (SDTMIG) v3.2. http://www.cdisc.org/sdtm

CONTACT INFORMATION

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

Donna Sattler
Eli Lilly and Company
484-744-0323
sattler_donna@lilly.com

Mike Lozano
Eli Lilly and Company
(317) 750-5760
lozano_michael_e@lilly.com