

Creating a Centralized Controlled Terminology Mapping Repository

Danny Hsu, Seagen Inc., Bothell, WA

Shreya Chakraborty, Seagen Inc., Bothell, WA

ABSTRACT

Controlled Terminology (CT) is the set of code lists and valid values used with data items within CDISC-defined datasets for regulatory submission. The use of CT helps harmonize the data for all submitted studies and improves the efficiency of data review. It also opens the door to powerful internal automation for heightened quality and efficiency. The question how to most effectively map the collected data, especially any free-text values from CRFs, into CT plays a very important role in each study and product team's development of CDISC datasets that are not only consistent and compliant, but also support analyses and summaries efficiently.

This article shares a concept of creating a centralized repository (such as a spreadsheet) to help streamline the process of CT mapping. A centralized CT mapping sheet is generated by collecting all mapped terms from previous studies and is used to identify the new terms which need attention. Following this process, study team could save time by focusing only on the newly added terms, therefore improving the efficiency of the data-mapping process and increase efficiency and quality of CDISC dataset generation and review. A centralized CT mapping sheet could not only support the creation of code and value list of define.xml, but also provide the format mapping used in SAS® programs.

INTRODUCTION

CDISC (cdisc.org) regularly updates and releases the industry standard CT which helps bring consistency across the industry and helps FDA reviewers relate to the study data more meaningfully, as being an industry standard, it can help reviewers build long-term familiarity and knowledge with these terms.

In a clinical trial, there could be hundreds of pages of case report forms (CRFs) which collect many pre-defined and free-text values. To map all those CRF collected values into CDISC CT could be a challenging and time-consuming process. Given this situation, we propose to create a centralized CT sheet which could help streamline the CT mapping especially during initial setup of a new study.

STEP 1: CREATE A CENTRAL CT LIBRARY

A centralized CT mapping sheet is generated by collecting all mapped terms from previous studies (Step 1 in [Figure 1](#)). By gradually curating the CT mapping terms from multiple studies, we build up a central CT library for a given product or even across multiple products at the department level. All future new studies can then build upon this CT library to further improve the efficiency of CT mapping.

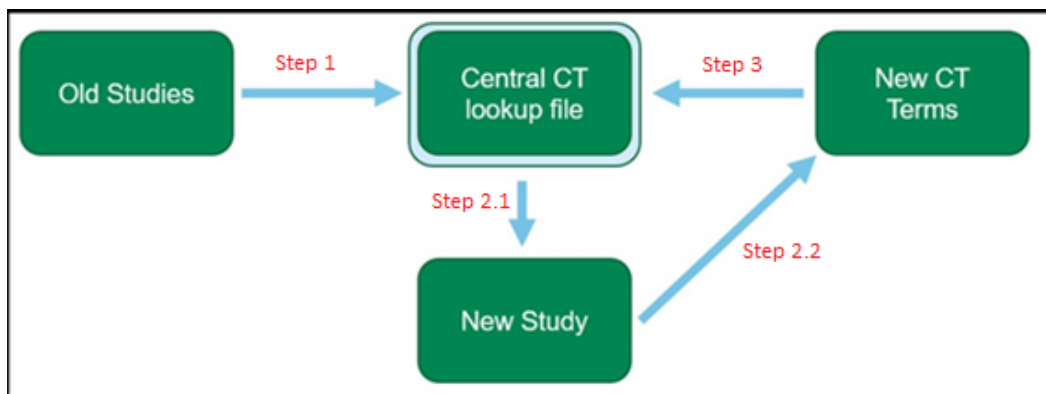


Figure 1

STEP 2: APPLY THE CENTRAL CT SHEET TO NEW STUDY AND IDENTIFY NEW TERMS

This central CT sheet is then used as a lookup table to map the CT for a new study and would also help in identifying new terms which are not in this existing CT library (Step 2 in [Figure 1](#)).

For example, applying the centralized CT sheet in a brand-new study could generate a spreadsheet as shown in [Table 1](#). Most of the raw data set values in the CRFValue and FreeValue columns would have the mapping information automatically populated from the centralized CT sheet. The new terms (e.g., free-text from CRF collection) would be flagged as per the red-highlighted fields. Afterwards, the red text could be filled in either programmatically or manually.

CTName	CTLabel	CRFValue	FreeValue	Code	Display	Extensible	ValueID	ListID	New_Term
UNIT	Unit	ms		msec	Millisecond	Yes	C41140	C71620	
UNIT	Unit		ms	msec	Millisecond	Yes	C41140	C71620	
UNIT	Unit	beats/min		beats/min	Beats per Minute	Yes	C49673	C71620	
UNIT	Unit		bmp	beats/min	Beats per Minute	Yes	C49673	C71620	
UNIT	Unit		beats/minute	beats/min	Beats per Minute	Yes	C49673	C71620	Y
UNIT	Unit	cm2		cm2	Square Centimeter	Yes	C48460	C71620	
UNIT	Unit		cm^2	cm2	Square Centimeter	Yes	C48460	C71620	

TABLE 1

STEP 3: ADD NEW TERMS BACK INTO CT LIBRARY

After identifying and mapping the new CT terms, these are added back into the existing CT library (Step 3 in [Figure 1](#)). For example, in [Figure 2](#), the new term identified in **[New Study]** is added back to **[Existing Central CT file]**, yielding a **[New Central CT file]** that includes this new term. In this way, newly identified CT terms from new studies are added to the central CT mapping file to get a more robust central CT repository. Once a comprehensive CT library is built, it opens the door to programmatic approaches to make the mapping process more automated.

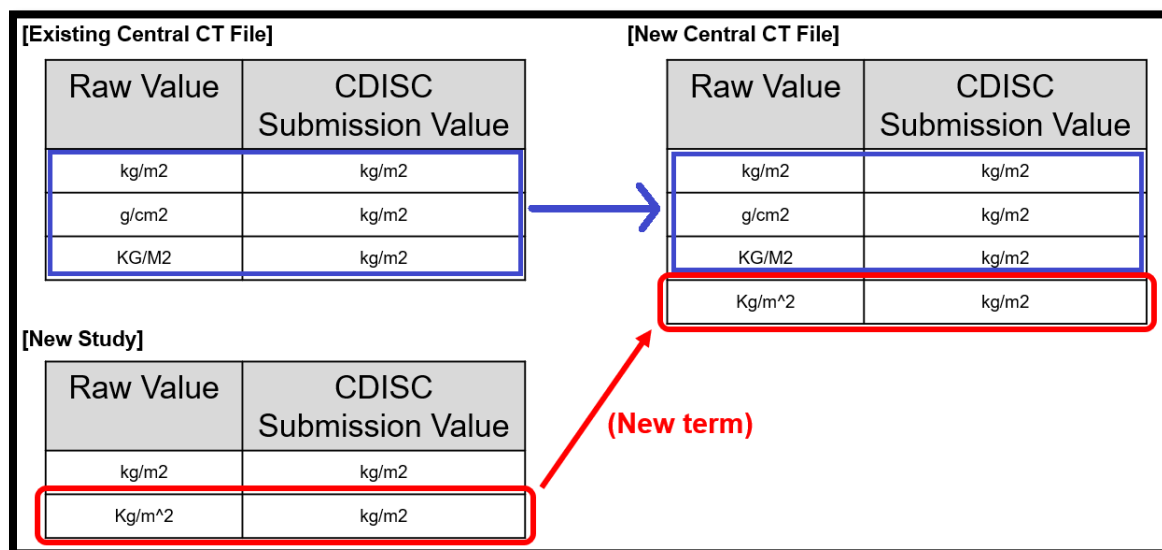


FIGURE 2

CONCLUSION

The mapping of Controlled Terminology is a very important section in the submission preparation, and a centralized CT mapping sheet could help streamline the process. Using the central CT sheet in the CT mapping of a new clinical study saves time for a study team to focus only on the newly added terms.

Furthermore, a centralized CT mapping sheet can not only support the creation of code and value lists in define.xml but also provide the format mapping ([Appendix 1](#)) used within the SAS programs. Therefore, creating a centralized CT mapping library helps improve the efficiency of the data-mapping process and increase efficiency and quality of CDISC dataset generation and review.

REFERENCES

CDISC Standards and Guidelines: CDISC.org

ACKNOWLEDGMENTS

We would like to acknowledge Wei Qian for his valuable feedback and constant support and guidance.

CONTACT INFORMATION

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

Danny Hsu
Seagen Inc.
21823 - 30th Drive S.E.
Bothell, WA 98021
shhsu@seagen.com

Shreya Chakraborty
Seagen Inc.
21823 - 30th Drive S.E.
Bothell, WA 98021
schakraborty@seagen.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.

APPENDIX 1

[Table 1](#) could be converted into an individual SAS program to have the proc format information as shown below and all other SAS programs could then read in this program to obtain format which help in the process of CT mapping.

```
proc format;  
  value $unit  
    'ms'='msec'  
    'beats/min','bmp','beats/minute'='beats/min'  
    'cm2','cm^2'='cm2';  
run;
```