PharmaSUG 2016 - Paper DS16

Codelists Here, Versions There, Controlled Terminology Everywhere

Shelley Dunn, Regulus Therapeutics, San Diego, California

ABSTRACT

Programming SDTM and ADaM data sets for a single study based on a one quarterly version of NCI Controlled Terminology (CT) can give a false sense that implementing CT correctly is a straight forward process. The most complex issues for a single study may rely on generating sufficient extensible codelists and ensuring all possible values on a CRF are accounted for within the study metadata. However, when looking beyond a single study towards developing processes to support end to end management, the maintenance of controlled terminology involves more complex processes requiring ongoing upkeep, compliance checking, archiving, updating, and governance.

To ensure FDA compliance, it behooves sponsors to develop processes for maintenance and organization of CT. It is not enough to hire an external vendor and hope they apply CT correctly and/or to assume the CT used is correct based on a clean vendor compliance report. This is primarily due to the iterative changes to industry terminology over time. Additionally, sponsor-defined terms must constantly be re-evaluated and compared to regularly published CDISC controlled terminology. Some best practices include:

- maintaining a CT repository/library
- running checks to compare current sponsor CT to new NCI CT, codelist codes, and other metadata
- adopting strategies for version control and developing processes for up-versioning

This paper will provide an in depth look at requirements and governance needed to ensure consistent and compliant use of controlled terminology across an entire company. Example issues and workable solutions will also be provided to illustrate a number of the challenges requiring this recommended rigor.

INTRODUCTION

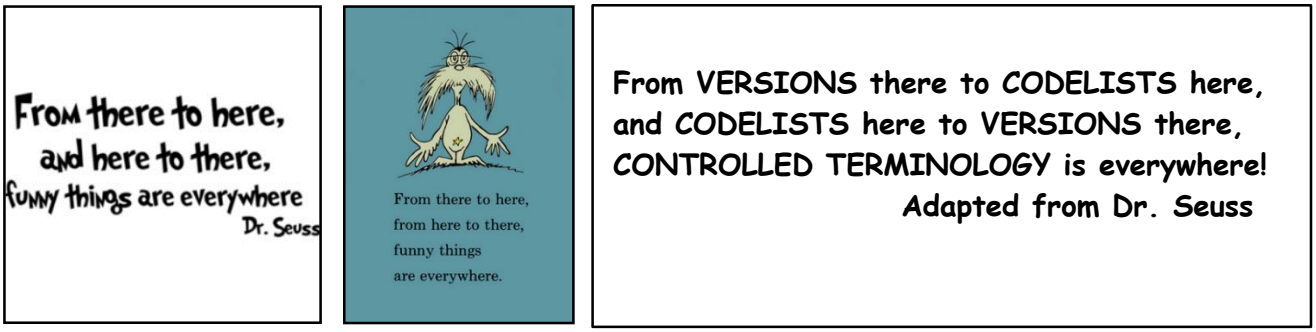
The knowledge base regarding what is Controlled Terminology (CT) and how best to implement CT varies greatly across sponsors, CROs, study sites, consultants, and consulting companies. This paper will begin by defining and providing examples of controlled terminology and codelists used in clinical studies. Industry resources, e.g., NCI CT, CDISC SDTMIG, etc., will be introduced to illustrate the general concepts surrounding how to apply CT correctly. After providing a basic overview of CT, a big picture view of how to implement different versions of CT will be provided.

With the goal of submitting standard compliant data to the FDA, this paper will propose a process for how to organize CT within a company to include both NCI CT and extensible codelists. As standard CT changes over time and new versions of NCI CT are published quarterly, the need to create a standards governance around the upkeep and versioning of Sponsor CT Library becomes a large task.

This paper will provide a clear step by step process for how to ensure CT compliance, archive different versions of a Sponsor CT Library, and provide a pathway for compliance and maintenance.

NOTE: Throughout this paper, National Cancer Institute Controlled Terminology (NCI CT) and CDISC CT may be used interchangeably.

Display 1



Display 1. Dr. Seuss quote and related paper quote

THE BASICS – CONTROLLED TERMINOLOGY (CT)

Let's start with an example. In order to analyze adverse events, medical coders have, for years, used a variety of coding software packages, e.g., MedDRA, to convert verbatim terms to System Organ Class, Preferred Terms, and other hierarchical coding terms. If one subject reported experiencing a "myocardial infarction" and another subject reported experiencing a "heart attack," medical coders use a coding dictionary to code both of these adverse events. In this example, "myocardial infarction" and "heart attack" are synonyms and should be mapped to the same thing. In order to analyze these two events together, a medical coder would need to code both of these to the same hierarchy of adverse event (AE) coding. That is, when a programmer gets the data for analysis, these two different AEs can be combined when summarizing AE incidence tables. Medical coders also implement a similar practice for coding concomitant medications (CM) using a dictionary such as WHO DRUG to code CMs to generic and/or trade names.

WHAT IS CT?

In much the same way medical coders have been coding AEs and CMs for years, with the FDA mandate to submit standardized data based on the standards outlined in the FDA Data Standards Catalog, additional data is now subject to a similar process that requires like terms to be standardized.

For example, when collecting gender on a case report form (CRF) the possibly values might be "Male" and "Female." Before standards, sponsors could submit this data in a number of ways. That is, the name of the variable might be "GENDER," "SEX," or any other name a sponsor chose. Likewise, the values of this variable could take on values such as "Male/Female," "M/F," "1/2" etc. By standardizing both the name of the variable and the possible variable values, the industry can increase efficiencies, simplify the review of data within an organization, and/or simplify the review of data at the FDA. Per the current standard, CDISC SDTMIG (Study Data Tabulation Model Implementation Guide), this variable will always be called "SEX" and per the current NCI CT standard, the only possible values for this variable are "F," "M," "U," or "UNDIFFERENTIATED." The possible values that specific variables can have is called Controlled Terminology.

Display 2

Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Synonym(s)	CDISC Definition	NCI Preferred Term
C66731		No	Sex	SEX	Sex	The assemblage of physical properties or qualities by which male is distinguished from	CDISC SDTM Sex of Individual Terminology
C16576	C66731		Sex	F	Female	A person who belongs to the sex that normally produces ova. The term is used to indicate	Female
C20197	C66731		Sex	M	Male	A person who belongs to the sex that normally produces sperm. The term is used to	Male
C17998	C66731		Sex	U	U; Unknown	Not known, not observed, not	Unknown
C45908	C66731		Sex	UNDIFFERENTIATED		A person (one of unisexual specimens) who is born with genitalia and/or secondary	Intersex

Display 2. NCI CT for the Codelist Name "Sex"

HOW TO APPLY TO A SINGLE STUDY?

Within a single study, the CT metadata is based on all of the possible collected values for a variable. Given the sample CRF below in Table 1, and the possible values for ETHNIC in Table 2, consider what values to include as part of the CT metadata if the data only includes the value of "Hispanic or Latino." In this example, the CT would need to include all three values on the CRF as all three values are possible, i.e., "Hispanic or Latino," "Not Hispanic or Latino," and "Unknown." The value "NOT REPORTED," although in the list of possible CDISC Submission Values in Table 2, would not be submitted since it was neither collected on the CRF nor is it included as a value in the data.

There are two main points to keep in mind when preparing study CT:

1. The data alone does not determine the CT for a given variable. All possible CRF values must be included.
2. All possible values, according to NCI CT, do not determine the CT for a given variable. In many cases the study CT may be a subset of the NCI CT. In other cases, where a codelist is extensible, the study CT may include additional terms not included in the NCI CT

Consider what happens to CT when trying to integrate several studies and it is easy to see how a simple concept can easily become more complex. Ensuring consistency of all controlled terminology across studies and throughout a

company are essential. Add to this the idea that NCI CT changes over time, the maintenance of the NCI CT versions, and the sponsor-defined CT for extensible codelists, and one can see the increased challenges in maintaining accurate and up the date CT.

Table 1

Ethnicity	
Hispanic or Latino	
Not Hispanic or Latino	
Unknown	

Table 1. Sample DM CRF for Ethnicity

Table 2

Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value
C66790		No	Ethnic Group	ETHNIC
C17459	C66790		Ethnic Group	HISPANIC OR LATINO
C41222	C66790		Ethnic Group	NOT HISPANIC OR LATINO
C43234	C66790		Ethnic Group	NOT REPORTED
C17998	C66790		Ethnic Group	UNKNOWN

Table 2. NCI CT for Ethnic Group

WHY DO WE NEED CT?

"Providing Regulatory Submissions in an Electronic Format – Standardized Study Data" was published on 17Dec2014 with the key message that submissions must be electronic for studies starting on or after December 2016. Additionally, submissions must be in a supported standard as documented in the FDA Data Standards Catalog (also referred to as "the Catalog"). "The Study Data Technical Conformance Guide" provides strong recommendations for submitting standard data and controlled terminology. One recommendation is to use common dictionaries across all clinical studies and throughout a submission. Below if a screen shot showing the "Terminology Standards" tab of the FDA Data Standards Catalog v4.4 (08-17-2015). According to this version of the Catalog, NCI CT versions of 06/10/2011 or later will be required beginning on 12/17/2016. Prior studies may use all previous NCI CT versions.

Currently the Catalog has fairly loose requirements around Terminology Standards. However, this may change moving forward and it is in a sponsor's best interest to be prepared to comply with new versions of Terminology Standards as the Catalog is updated.

Display 3

FDA Data Standards Catalog v4.4 (08-17-2015)

This table contains a listing of the standard terminology code sets. When the Catalog expresses support for more than one terminology for a given type of regulatory information, the submitter may choose which one to use. Submissions using any terminology not listed should be discussed with the Agency in advance.
The listing of the data exchange standards developed at FDA are listed in a separate tab. Please look at the "Data Exchange Standards" tab to find data exchange standards information supported by FDA. The data exchange standards listed have established processes and technology infrastructure to support the process, review, and archive of the data. The submission of standardized data using any standard not listed, or to an FDA component not listed, should be discussed with the Agency in advance.

Terminology Standard	Terminology Type	Terminology Standards Development and/or Maintenance Organization	Version(s)	FDA Centers That Use This Terminology	Date Support Begins (MM/DD/YYYY)	Date Support Ends	Date Requirement Begins (MM/DD/YYYY)	Date Requirement Ends	Examples of Use	Regulatory References and Information Sources
Clinical Data Interchange Standards Consortium (CDISC) Terminology	General Clinical Data	CDISC	2011-06-10 or later	CBER, CDER	06/13/2011		12/17/2016 [1] 12/17/2017 [2]		Use CDISC Submission values	Index of CDISC SDTM Terminology
Clinical Data Interchange Standards Consortium (CDISC) Terminology	General Clinical Data	CDISC	All Previous Version	CBER, CDER	Ongoing				Use CDISC Submission Values. Do not use for studies initiated after 2011-06-13.	Index of CDISC SDTM Terminology

Instructions | Data Exchange Standards | **Terminology Standards** | Change History

Display 3. Catalog Terminology Standards Tab

CODELISTS HERE

WHAT ARE CODELISTS?

The name given to a list of controlled terminology is called a "codelist." That sounds easy enough until you realize that a codelist is not actually a list. It is sort of like Mike Myers doing a Saturday Night Live skit of Coffee Talk where he says, "codelist is neither a code nor a list…discuss!" It might be helpful if "codelist" was called something else such as "codename" since it is actually the name associated with a given list.

The industry has standards related to which "variables" are subject to controlled terminology and the allowable "values" for these variables. However, the industry does not have standards around codelist naming conventions. That is, a sponsor may create whatever name they choose to identify a list of CT as long as they provide all possible values for the codelist, e.g., submission values, decodes, and other corresponding metadata.

To illustrate this, assume there are variables that can only take on the values of "Y" of "N" ("Y" when a value is "Yes" and "N" when a value is "No"). In the first example below, the codelist is named YESNO and in the second example below the codelist is named YN. Notice that the actual lists are exactly the same; only the name of the list is different. This is acceptable and codelist naming conventions can be sponsor specific.

Table 3

Codelist (Name)	Submission Value	Decode
YESNO	Y	Yes
	N	No

Table 3. Codelist YESNO (Code List Example #1)

Table 4

Codelist (Name)	Submission Value	Decode
YN	Y	Yes
	N	No

Table 4. Codelist YN (Code List Example #2)

RESOURCES

NCI Controlled Terminology

The National Cancer Institute (NCI) developed Controlled Terminology prior to partnering with CDISC. CDISC Controlled Terminology is now the same as NCI Controlled Terminology through a partnership between the two organizations to develop and support CT for all CDISC Standards (e.g., SDTM, ADaM, etc.). For all intents and purposes, CDISC CT and NCI CT are now considered the same. New versions of CT are published on a quarterly basis and can be found on the NCI web site or hyperlinked from the CDISC web site. Files can be downloaded in the following formats: Excel, text, odm.cml, pdf, html, and OWL/RDF.

NCI/CDISC Controlled Terminology

Below, in Display 4, is an example of the codelist named "Race" in the NCI CT file. This file contains information used to accurately assign CT values to the RACE variable in the SDTM domain DM. Additional columns provide metadata associated with the CT that will be necessary to submit as part of the Define.xml.

Here is a brief overview of each column:

1. **Code:** For the row highlighted in light blue, the Code (e.g., C74457) is the Code assigned to the entire codelist named "Race." In all of the non-blue columns, the Code is the unique number assigned to a given submission value (e.g., C412260 is the Code for the CDISC Submission Value "ASIAN")
2. **Codelist Code:** This is the same as the Code in the row highlighted in light blue and is the Codelist Code assigned to the entire codelist named "Race"
3. **Codelist Extensible (Yes/No):** A value of "No" in this column indicate no additional values can be added to the given codelist. A value of "Yes" in this column (see Display 5 for codelist "LOC") indicates a sponsor may add new values if an equivalent value doesn't already exist. When adding new CT, care should be taken to ensure it is formatted consistently with the other values.

4. **CDISC Submission Value:** For the row highlighted in light blue, this is the name of the codelist. For all other rows, these are the possible values a variable with this codelist can assume. For example, in SDTM.DM the variable DM.RACE may take on the values of “AMERICAN INDIAN OR ALASKA NATIVE, “ASIAN,” “BLACK OR AFRICAN AMERICAN,” “NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER,” or “WHITE.” The variable DM.RACE cannot have any other value.
5. **CDISC Synonym(s):** If available, this column may contain synonyms for the CDISC Submission Value, e.g., see Display 5, in the codelist named “LOC” the CDISC Submission Value “ABDOMINAL CAVITY” has the CDISC Synonym(s) of “Abdomen” or “Abdominal Cavity.” This column can help in determining is a codelist contains a given equivalent Submission Value. That is, there is no need to add a new Submission Value for “ABDOMEN” as “ABDOMINAL CAVITY” is meant to cover both CDISC Synonyms
6. **CDISC Definition:** This column, although truncated here, provides information about assigning CT. For a full description, view the NCI CT via one of the web sites provided
7. **NCI Preferred Term:** Sponsors may use this column as the “Decode” of the “CDISC Submission Value”

Display 4

Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Synonym(s)	CDISC Definition	NCI Preferred Term
C74457		No	Race	RACE	Race	Terminology codelist used to identify the	CDISC SDTM Race Terminology
C41259	C74457		Race	AMERICAN INDIAN OR ALASKA NATIVE		A person having origins in any of the	American Indian or Alaska Native
C41260	C74457		Race	ASIAN		A person having origins in any of the	Asian
C16352	C74457		Race	BLACK OR AFRICAN AMERICAN		A person having origins in any of the	African American
C41219	C74457		Race	NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER		Denotes a person having origins in any	Native Hawaiian or Other Pacific Islander
C41261	C74457		Race	WHITE		Denotes a person with European, Middle	White

Display 4. NCI CT for Codelist RACE

Display 5

Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Synonym(s)	CDISC Definition	NCI Preferred Term
C74456		Yes	Anatomical Location	LOC	Anatomical Location	Terminology codelist used for	CDISC SDTM Anatomical Location Terminology
C116163	C74456		Anatomical Location	5TH LUMBAR SPINOUS		The spinous	Fifth Lumbar Spinous Process
C32038	C74456		Anatomical Location	ABDOMINAL AORTA		The portion of the descending	Abdominal Aorta
C12664	C74456		Anatomical Location	ABDOMINAL CAVITY	Abdomen; Abdominal Cavity	The portion of the body that lies	Abdomen
C12360	C74456		Anatomical Location	ABDOMINAL LYMPH NODE		Any lymph node within the	Intra-Abdominal Lymph Node
C52758	C74456		Anatomical Location	ABDOMINAL SKIN	Abdominal Skin	The skin or	Abdominal Skin

Display 5. NCI CT for Codelist LOC

SPONSOR DEFINED CODELISTS

When a codelist is “Extensible,” sponsors may add new CT that is currently not provided in the NCI CT. Lab tests are a good example where most sponsors will need to supplement the codelist provided by NCI. The SDTM CT is much more extensive than the ADaM CT. In most cases, for analysis, sponsors will be required to add to extensible codelists to ensure all parameters are accounted for in their CT.

Implementation Guides (SDTMIG and ADaMIG)

The Implementation Guides, SDTMIG and ADaMIG, provide guidance on which variables have CT. The column entitled “Controlled Terms, Codelists or Formats,” provides guidance on the recommended CDISC codelist for a given variable. The values of (AGEU), (SEX), etc. can be referenced in the NCI CT and implemented accordingly.

Table 5

Variable Name	Variable Label	Type	Controlled Terms, Codelists or Format
AGEU	Age Units	Char	(AGE)
SEX	Sex	Char	(SEX)
RACE	Race	Char	(RACE)

Table 5. SDTMIG example for some DM variables

VERSIONS THERE

NCI VERSIONING

CDISC publishes new versions of controlled terminology every quarter. Codelists can change in a number of different ways. Sometimes values are added or removed. Other times, the "codes" associated with given values may change. To keep up with the quarterly published changes it is recommended that sponsors run compliance checks on the CT used for their studies.

Once the Sponsor CT Library has been updated to match a given NCI CT version, it is recommended that the sponsor archive a version of their Sponsor CT Library. This process can be repeated quarterly as new NCI CT versions are published.

The following sections will describe a process for updating versions.

CONTROLLED TERMINOLOGY EVERYWHERE

The goal is to maintain and organize controlled terminology while ensuring compliance and consistency with each new NCI CT version.

MAINTAINING A SPONSOR CT LIBRARY/REPOSITORY

It is recommended that sponsors maintain a "Sponsor CT Library" for every CT version they are using. The Sponsor CT Library will contain the following for a given version:

1. All applicable NCI CT controlled terminology to be used by sponsor
2. All extensible CT added by sponsor
3. All new codelists added by sponsor

SPECIFICATIONS AND COMPLIANCE CHECKING

In preparation of creating CT compliance specifications, there are some "one-time" tasks that need to be performed:

One Time Tasks

1. Determine the checks needed to compare Sponsor CT to NCI CT

   This paper provides example checks based on prior experience with NCI CT and a sponsor CT library. To generate the appropriate checks for your organization, a thorough review of the Sponsor CT Library and requirements for maintenance are necessary.

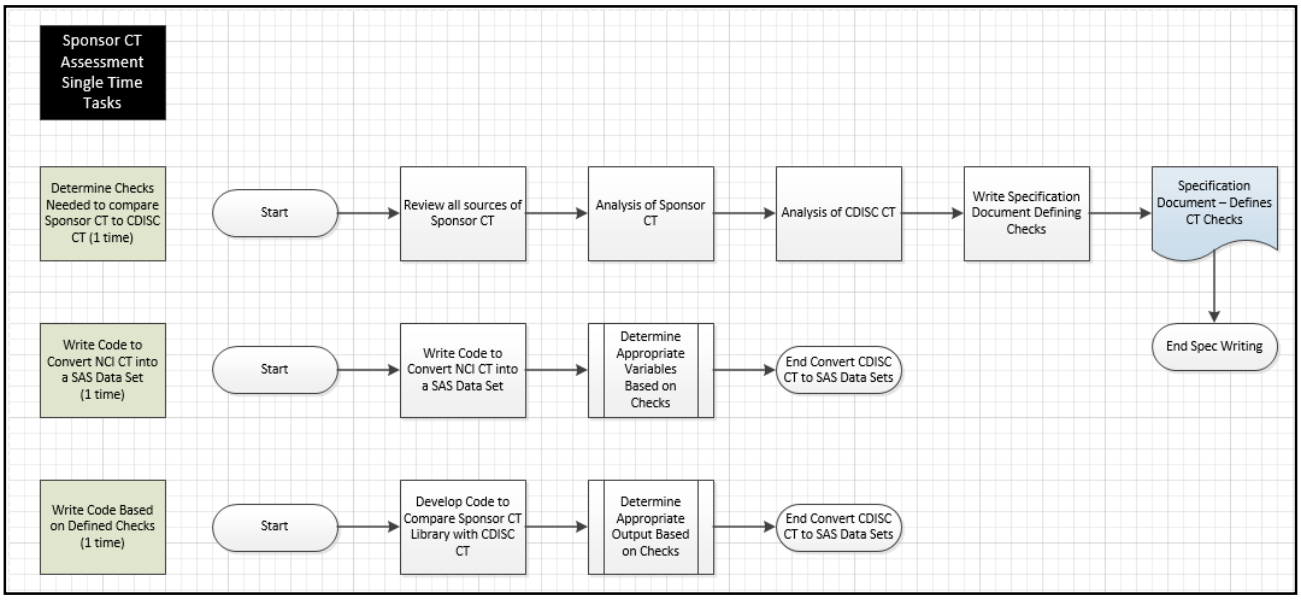
2. Write code to convert NCI CT into a (SAS) data set (Similarly, write code to convert Sponsor CT Library into a (SAS) data set). This example assumes Sponsor CT Library is already available as a SAS data set

   As previously discussed, there are many file formats available for the NCI CT. Ensure you are using the same file format each time to create a consistent data set for each new version. Note: SAS data sets were used for this example. Sponsors are welcome to use the software and/or file format of their choice.

3. Write code based on defined checks

   Program checks based on programming specifications. The program(s) are meant to access the NCI CT data and the Sponsor CT Library once they are both converted to a consistent format where they can be compared programmatically.

Display 6



Display 6. One-Time Steps to Prepare Specifications for CT Checks

EXAMPLE CT CHECKS

Before defining the CT Check specifications, it is best to familiarize yourself with content of both the NCI CT files and the Sponsor CT Library. In the event a Sponsor CT Library has not yet been established, using the basic format of the NCI CT is a good place to start. Additional metadata will most likely need to be added if the contents of this file will be used to generate a Define.xml. However, Define.xml metadata is beyond the scope of this paper.

Comparing Metadata: NCI CT to Example Sponsor CT Library

Display 7

SAMPLE from NCI CT					
Code	**Codelist Code**	**Codelist Extensible (Yes/No)**	**Codelist Name**	**CDISC Submission Value**	**NCI Preferred Term**
C66767		No	Action Taken with Study Treatment	ACN	CDISC SDTM Action Taken with Study Treatment Terminology
C49503	C66767		Action Taken with Study Treatment	DOSE INCREASED	Dose Increased
C49504	C66767		Action Taken with Study Treatment	DOSE NOT CHANGED	Dose Not Changed
C49505	C66767		Action Taken with Study Treatment	DOSE REDUCED	Dose Reduced
C49501	C66767		Action Taken with Study Treatment	DRUG INTERRUPTED	Drug Interrupted
C49502	C66767		Action Taken with Study Treatment	DRUG WITHDRAWN	Drug Withdrawn
C48660	C66767		Action Taken with Study Treatment	NOT APPLICABLE	Not Applicable
C17998	C66767		Action Taken with Study Treatment	UNKNOWN	Unknown
SAMPLE from Sponsor CT Library					
C_CODE	**CODELIST_CODE**	**EXTENSIBLE**	**CODELIST_DESCRIPTION**	**SUBMISSION_VALUE**	**DECODE**
C49503	C66767		Action Taken with Study Treatment	DOSE INCREASED	Dose Increased
C49504	C66767		Action Taken with Study Treatment	DOSE NOT CHANGED	Dose Not Changed
C49505	C66767		Action Taken with Study Treatment	DOSE REDUCED	Dose Reduced
C49501	C66767		Action Taken with Study Treatment	DRUG INTERRUPTED	Drug Interrupted
C49502	C66767		Action Taken with Study Treatment	DRUG WITHDRAWN	Drug Withdrawn
C48660	C66767		Action Taken with Study Treatment	NOT APPLICABLE	Not Applicable
C17998	C66767		Action Taken with Study Treatment	UNKNOWN	Unknown

Display 7. Example of NCI CT and Sponsor CT Library (column headers do not need to be the same)

Determine Common Terminology for Specifications

It is possible that NCI CT and a Sponsor CT Library use different labels or descriptions to define the same content. Setting up a chart to identify naming conventions between the two sources may be helpful when communicating about the different sources and can aid in programming the CT Compliance Checks.

Table 6

Columns *(variables)* in NCI CT	**Columns *(variables)* in Sponsor Internal CT Library**	**Common Terminology Used for Checks**
Code	C_CODE	C_CODE
Codelist Code	CODELIST_CODE	CODELIST_CODE
Codelist Extensible (Yes/No)	EXTENSIBLE	EXTENSIBLE
Codelist Name	CODELIST_DESCRIPTION	CODELIST_DESCRIPTION
CDISC Submission Value	SUBMISSION_VALUE	SUBMISSION_VALUE
NCI Preferred Term	DECODE	DECODE

Table 6. Comparison of NCI CT to Sponsor CT Library to Determine Common Terminology for Compliance

Example CT Compliance Check Specifications Document

Table 7

Check #	**Sponsor CT Library Variable(s) Used in Check**	**NCI Variable(s) Used in Check**	**English Version of Check**
CT001	CODELIST_ CODE	Codelist Code	CODELIST_CODE is in NCI CT and is not in Sponsor CT Library.
CT002	CODELIST_ CODE	Codelist Code	CODELIST_CODE is in Sponsor CT Library and is not in NCI CT.
CT003	SUBMISSION_ VALUE_CODE	Code: where Codelist Code is not missing	C_CODE is in NCI CT and is not in Sponsor CT Library.
CT004	SUBMISSION_VALUE_CODE	Code: where Codelist Code is not missing	C_CODE is in Sponsor CT Library and is not in NCI CT.
CT005	CODELIST_DESCRIPTION	Codelist Name	CODELIST_DESCRIPTION is in NCI CT and is not in Sponsor CT Library.
CT006	CODELIST_DESCRIPTION	Codelist Name	CODELIST_DESCRIPTION is in Sponsor CT Library and is not in NCI CT.
CT007	SUBMISSION_VALUE	CDISC Submission Value	SUBMISSION_VALUE is in NCI CT and is not in Sponsor CT Library.
CT008	SUBMISSION_VALUE	CDISC Submission Value	SUBMISSION_VALUE is in Sponsor CT Library and is not in NCI CT.
CT009	DECODE	NCI Preferred Term	DECODE is in NCI CT and is not in Sponsor CT Library.
CT010	DECODE	NCI Preferred Term	DECODE is in Sponsor CT Library and is not in NCI CT.
CT011	CODELIST_ CODE SUBMISSION_VALUE_CODE	Codelist Code Code	CODELIST_CODE is in NCI CT and in Sponsor CT Library; however, C_CODE is in one and not the other.
CT012	CODELIST_ CODE EXTENSIBLE	Codelist Code Extensible	CODELIST_CODE is in NCI CT and in Sponsor CT Library; however, EXTENSIBLE does not match.

Check #	Sponsor CT Library Variable(s) Used in Check	NCI Variable(s) Used in Check	English Version of Check
CT013	CODELIST_ CODE CODELIST_ DESCRIPTION	Codelist Code Codelist Name	CODELIST_CODE is in NCI CT and in Sponsor CT Library; however, CODELIST_DESCRIPTION does not match.
CT014	SUBMISSION_VALUE_CODE CODELIST_ CODE	Code Codelist Code	C_CODE is in NCI CT and in Sponsor CT Library; however, CODELIST_CODE does not match.
CT015	CODELIST_ DESCRIPTION CODELIST_ CODE	Codelist Name Codelist Code	CODELIST_DESCRIPTION is in NCI CT and in Sponsor CT Library; however, CODELIST_CODE does not match.
CT016	SUBMISSION_VALUE_CODE CODELIST_ DESCRIPTION	Code Codelist Name	C_CODE is in NCI CT and in Sponsor CT Library; however, CODELIST_DESCRIPTION does not match.
CT017	SUBMISSION_VALUE_CODE SUBMISSION_VALUE	Code CDISC Submission Value	C_CODE is in NCI CT and in Sponsor CT Library; however, SUBMISSION_VALUE does not match.
CT018	SUBMISSION_VALUE_CODE DECODE	Code NCI Preferred Term	C_CODE is in NCI CT and in Sponsor CT Library; however, DECODE does not match.
CT019	CODELIST_ DESCRIPTION SUBMISSION_VALUE_CODE	Codelist Name Code	CODELIST_DESCRIPTION is in NCI CT and in Sponsor CT Library; however, C_CODE does not match.
CT020	SUBMISSION_VALUE SUBMISSION_VALUE_CODE	CDISC Submission Value Code	SUBMISSION_VALUE is in NCI CT and in Sponsor CT Library; however, C_CODE does not match.
CT021	DECODE SUBMISSION_VALUE_CODE	NCI Preferred Term Code	DECODE is in NCI CT and in Sponsor CT Library; however, C_CODE does not match.
LISTING	SUBMISSION_VALUE_CODE CODELIST_ CODE EXTENSIBLE CODELIST_ DESCRIPTION SUBMISSION_VALUE DECODE	Code Codelist Code Extensible Codelist Name CDISC Submission Value NCI Preferred Term	Print out a list of all Sponsor CT Library CT that does not have a C_CODE. Use this list to cross reference NCI CT to determine if the CT is now accounted for in the NCI CT (MANUAL CHECK!)

Table 7. Example CT Compliance Checks for Sponsor CT Library

CT CHECKS

In preparation for running CT Checks, there are some tasks that need to be performed “every time”:

Every Time Tasks

1. Get files

   Retrieve any files needed, e.g., latest NCI CT version, in order to run the programs written to generate the files (per programming code)

2. Create (SAS) data sets to compare Sponsor CT Library to NCI CT

Run the code programmed above (see One Time Task #2 above) to generate two files: one for NCI CT and one for Sponsor CT Library (these may or may not be SAS data sets depending on how specifications are defined)

3. Run code to do comparison and output report(s)

   Code is based on specification documents. The general idea is to compare the NCI CT with the Sponsor CT Library and kick out any discrepancies. This can be run as new versions of the NCI CT are published.

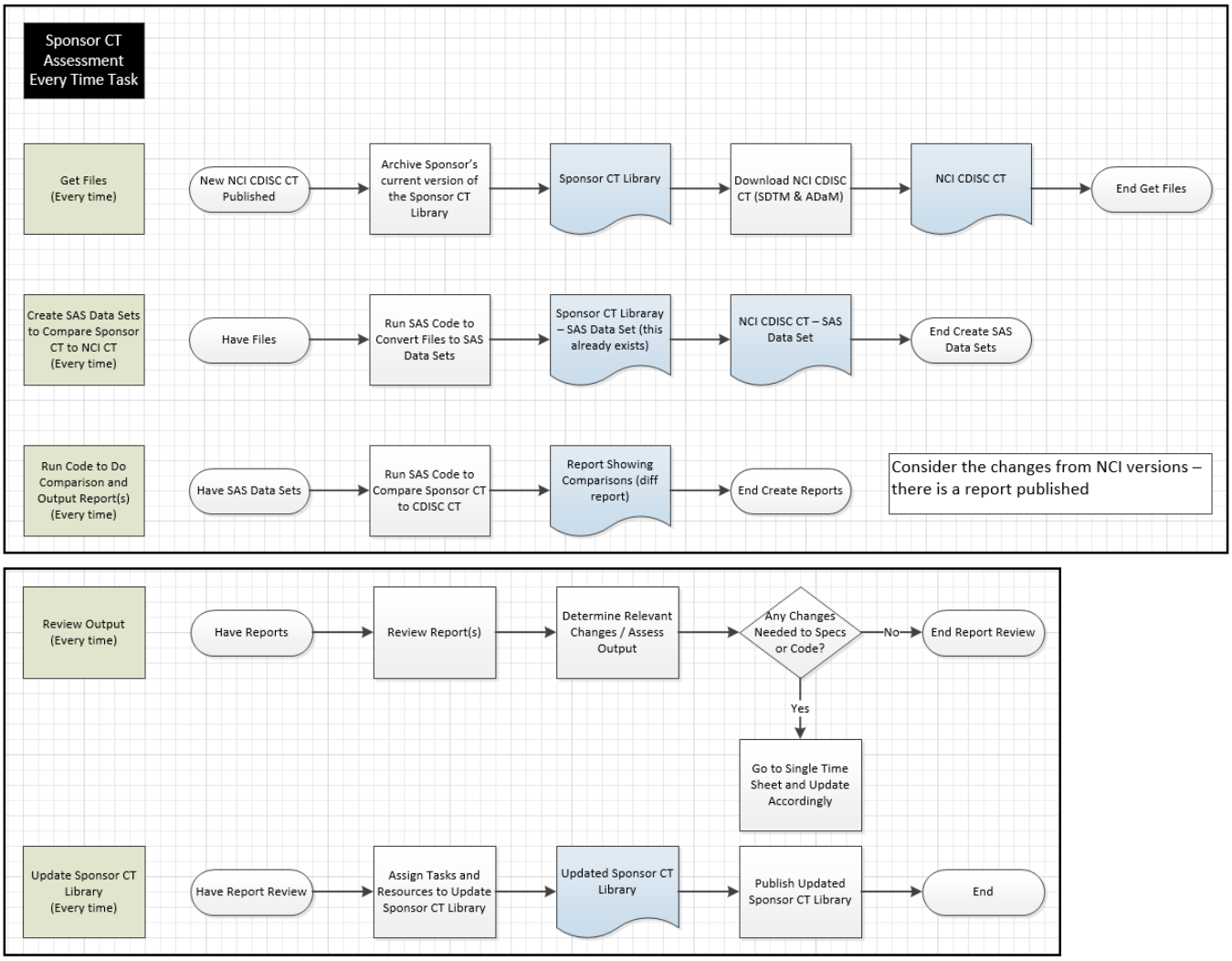
4. Review output

   This is a manual step requiring someone familiar with the specifications and how CT is maintained at the sponsor company (e.g., CT Steward). The CT Steward is tasked with determining which checks require updates and which checks do not require updates. An additional procedure has been developed to suppress checks from re-firing when a CT Steward has deemed a check acceptable. See “Process for Reviewing CT Output” below.

5. Update Sponsor CT Library

   As appropriate, update the Sponsor CT Library to ensure consistency with the latest version of the NCI CT. A sponsor may choose to continue this process and re-run the CT Checks until all discrepancies are accounted for.

Display 8



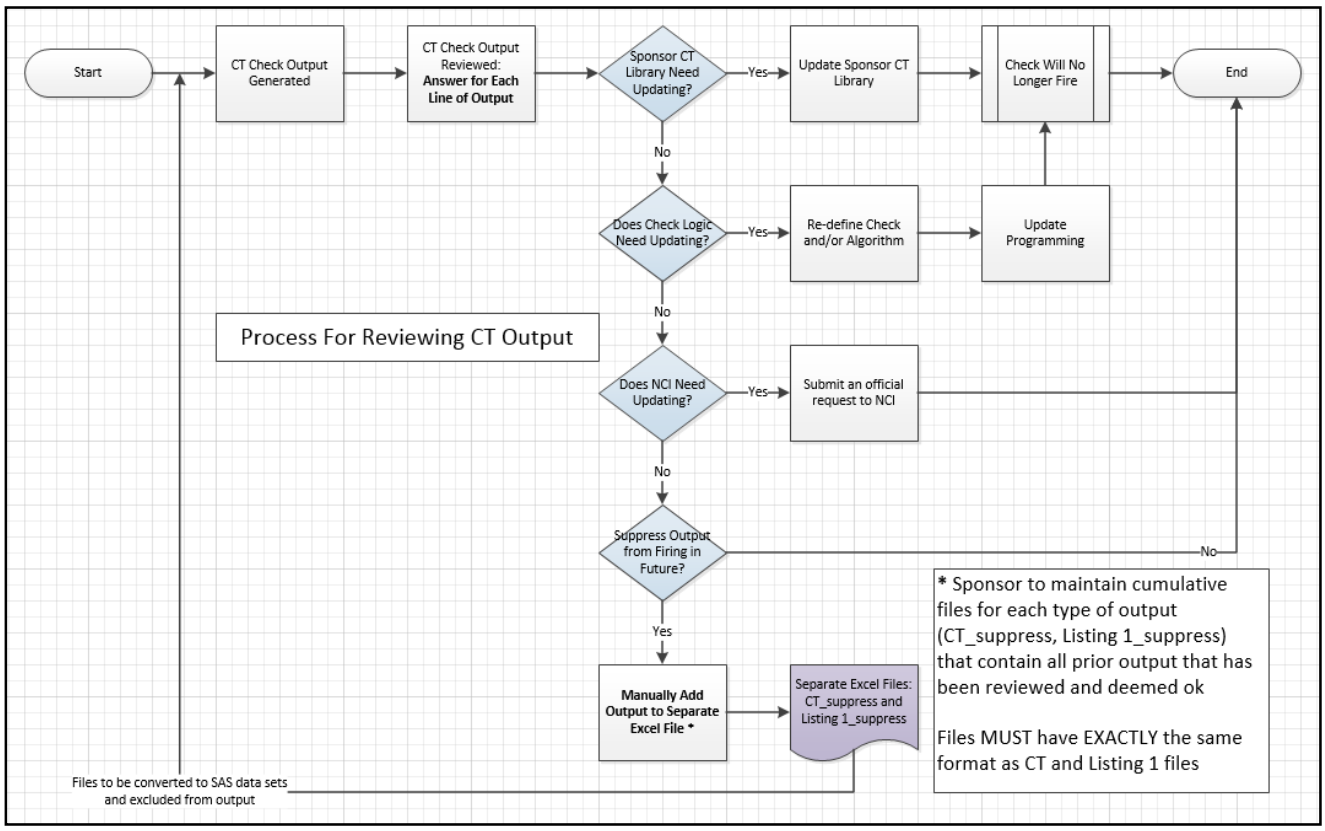
Display 8. Every Time Steps to Run CT Checks

UPDATING SPONSOR CT LIBRARY/REPOSITORY

Process for Reviewing CT Output

As mentioned above, in Every Time Tasks #4, there may be times when output is generated from running the CT Checks that may be acceptable to the sponsor. In these instances, a sponsor may choose to suppress these checks from kicking out each time the CT Checks are run. To do so, the following process may be implemented.

Display 9



Display 9. Process for Reviewing CT Output

ARCHIVING

Once the Sponsor CT Library has been updated to match a given NCI CT version, it is recommended that the sponsor archive each version. This process can be repeated quarterly as new NCI CT versions are published.

Display 10

Display 10. Dr. Seuss to End

CONCLUSION

To maintain the most updated CT and account for each new published version of the NCI CT and sponsor defined CT, it is imperative for sponsors to develop a process for maintenance, compliance, version control, and upkeep of a Sponsor CT Library(ies). While the process may vary from company to company, in general the process will include a Sponsor CT Library build upon NCI CT, a process for downloading new publications of NCI CT, compliance check specifications for CT checks, a program or programs to run compliance, a governance process to update the appropriate files based on the compliance check output, and a method for versioning Sponsor CT Libraries. This allows for proper identification of CT versions for integrating databases and/or submissions.

REFERENCES

CDISC web site. Study Data Tabulation Model (SDTM). V3.2. http://www.cdisc.org/

FDA Data Standards Catalog v4.4 (08-17-2015) – Supported and Required Standards. FDA. 17Aug2015. http://www.fda.gov/forindustry/datastandards/studydatastandards/default.htm

National Cancer Institute web site. CDISC Terminology. SDTM Terminology as Excel file. 18Dec2015. http://www.cancer.gov/research/resources/terminology/cdisc.

Providing Regulatory Submissions In Electronic Format – Standardized Study Data. Guidance for Industry. Department of Health and Human Services, FDA, CDER, CBER. December 2014. http://www.fda.gov/downloads/Drugs/.../Guidances/UCM292334.pdf

Study Data Technical Conformance Guide. U.S. Department of Health and Human Services, FDA, CDER, CBER. October 2015. Page 18-20. http://www.fda.gov/downloads/forindustry/datastandards/studydatastandards/ucm384744.pdf

CONTACT INFORMATION

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

Name: Shelley Dunn
Enterprise: Regulus Therapeutics
Address: 3545 John Hopkins Court
City, State ZIP: San Diego, CA 92121
Work Phone: 858-202-6300
Fax: 858-202-6363
E-mail: sdunn@regulusrx.com
Web: www.regulusRx.com