

PINNACLE²¹

CONTROLLED TERMINOLOGY BEST PRACTICES

Harmonizing data to promote consistencies across
clinical trials

Sarah Angelo, Principal Consultant
sangelo@pinnacle21.com

PHUSE US Connect 2021



P²¹

AGENDA

- Definition
- Importance of Controlled Terminology
- Common incorrect extensions
- Pinnacle 21 Validation & Fix Tips
- New Validation Checks
- Best Practices



CONTROLLED TERMINOLOGY DEFINITION

WHAT IS CONTROLLED TERMINOLOGY?

- ▶ Data standards
 - ▶ Provide consistent meaning to data
 - ▶ Across information systems, programs, and agencies
 - ▶ Uniform structuring, format, definition
- ▶ Terminology Standards
 - ▶ Controlled terms for submissions
 - ▶ FDA MDR Adverse Event codes
- ▶ CDISC CT Terminology standards
 - ▶ Grouped into subsets: SDTM, CDASH, ADaM, SEND, Define.xml, Protocol
 - ▶ Updated quarterly
- ▶ External Dictionaries
 - ▶ MedDRA, WHODrug, SNOMED, MED-RT, UNII, LOINC

CDISC CONTROLLED TERMINOLOGY FILES

▸ NCI Enterprise Vocabulary Services

Study Data Tabulation Model (SDTM)

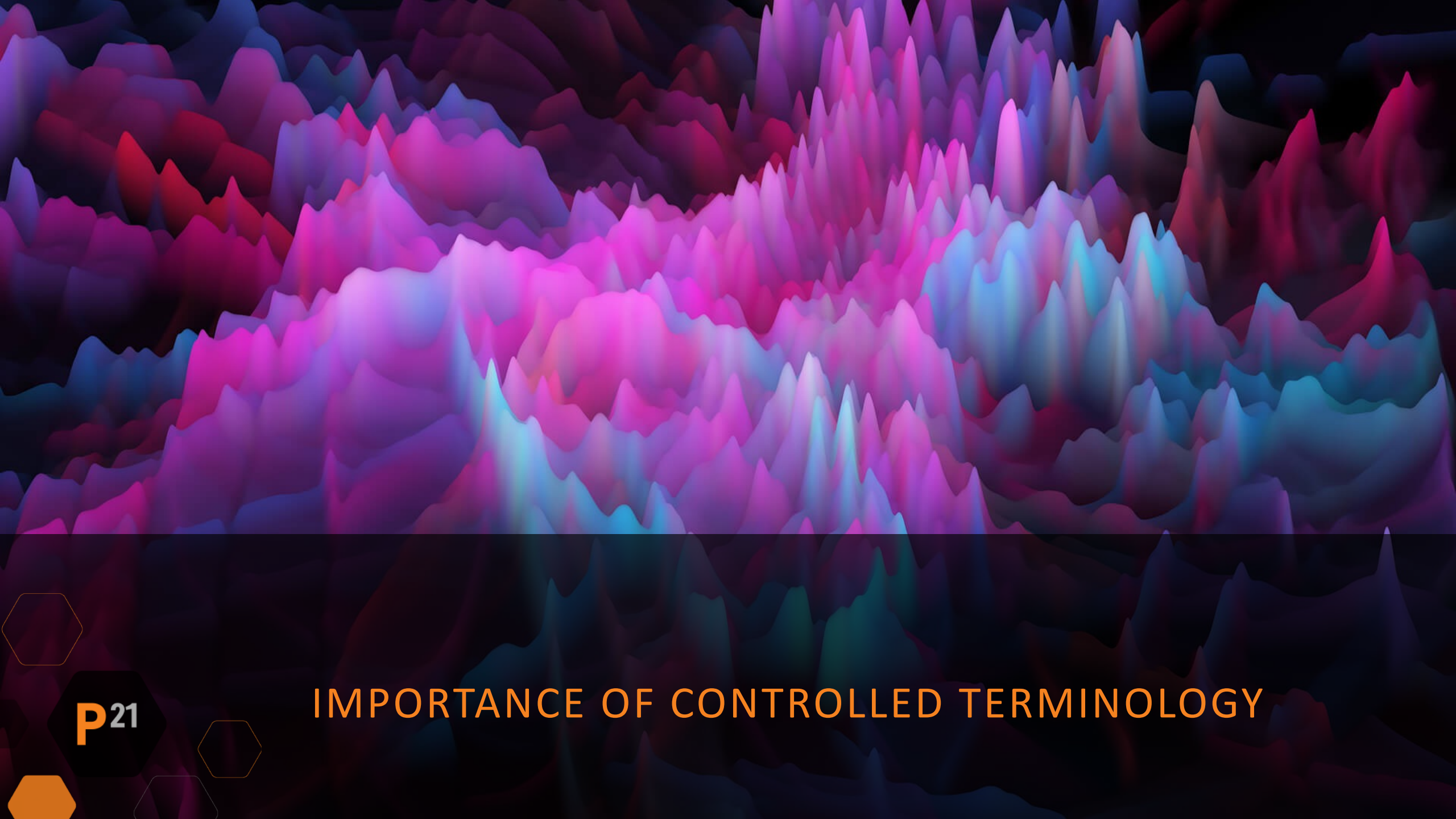
[SDTM](#) is an international standard for clinical research data and is approved by the FDA as a standard electronic submission format. CDISC Questionnaires, Ratings, and Scales terminology is included in the SDTM terminology publication. NCI EVS maintains and distributes SDTM controlled terminology as part of NCIt.

- Download from the [CDISC SDTM directory](#) on an NCI FTP site in [Excel](#), [text](#), [odm.xml](#), [pdf](#), [html](#) and [OWL/RDF](#) formats.
- Download changes from the previous release in [Excel](#) or [text](#) formats.

Clinical Data Acquisition Standards Harmonization (CDASH)

CDISC leads the [CDASH](#) project, which develops clinical research study collection standards in

- <https://datascience.cancer.gov/resources/cancer-vocabulary/cdisc-terminology>



IMPORTANCE OF CONTROLLED TERMINOLOGY

P²¹

IMPORTANCE OF CT

- ▶ FDA and PMDA
 - ▶ Required for submission
 - ▶ Part of CDISC-compliance
 - ▶ Assist reviewers with predictability, consistency, transparency, efficiency
- ▶ CDISC
 - ▶ Provide submission-controlled terms and definitions
 - ▶ Harmonize data
 - ▶ Aid with data exchange and interpretation
 - ▶ Consistencies across systems and studies
- ▶ Execute analytic tools
 - ▶ Automate validation, review, and analysis
- ▶ Cross-study comparative analyses
 - ▶ Improve patient safety and outcomes

REQUIREMENTS

▸ AGENCY REQUIREMENTS

▸ FDA Technical Conformance Guide

- Latest version
- Same MedDRA version for ISS
- Exact case and spelling of terms

▸ PMDA Technical Conformance Guide

- Same version used within Clinical Study
- Exact case and spelling of terms

▸ CDISC

▸ Non-extensible vs. Extensible codelists

- Action Taken with Study Treatment - Non-extensible
- Route of Administration Response - Extensible

▸ Extending codelist

- CDISC submission value rules



COMMON INCORRECT EXTENSIONS

SOURCES OF INCORRECT EXTENSIONS

- Free-text data entry
- Data collection design
- SDTM Mapping
- Extension of CDISC codelists
- Use of synonyms
- Misspelling and casing
- External dictionaries unmatched coding
 - E.g. MedDRA, WHODRUG

SYNONYMS

- Should not be used
- Provided to help map to the correct submission value
 - Example:
 - *mg/mL* - incorrect
 - *g/L* - correct

LBORRES	Result or Finding in Original Units	Char		Result Qualifier	Result of the measurement or finding as originally received or collected.	Exp
LBORRESU	Original Units	Char	(UNIT)	Variable Qualifier	Original units in which the data were collected. The unit for LBORRES. Example: "g/L".	Exp
LBORNRL0	Reference Range Lower Limit in Orig Unit	Char		Variable Qualifier	Lower end of reference range for continuous measurement in original units. Should be populated only for continuous results.	Exp

Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Synonym(s)
C71620		Yes	Unit	UNIT	Unit
C42576	C71620		Unit	g/L	Gram per Liter; Kilogram per Cubic Meter; Microgram per Microliter; Milligram per Milliliter; g/L; kg/m3; mg/mL ; ug/uL

“OTHER”

- ▶ Mapping to “OTHER” not recommended
 - ▶ FDA TCG “6.1.2.1 Use of the Specific Controlled Term ‘OTHER’”
 - ▶ Search for other possible terms
- ▶ --TERM fields mapped to a --DECODE of “OTHER”
 - ▶ Explain in the RGs

VALIDATION

Overview

Issues

Datasets

Data Reports

SDTM

Home > Xanomeline > CDISCPLOT01 - End User Training > SDTM

Export Report

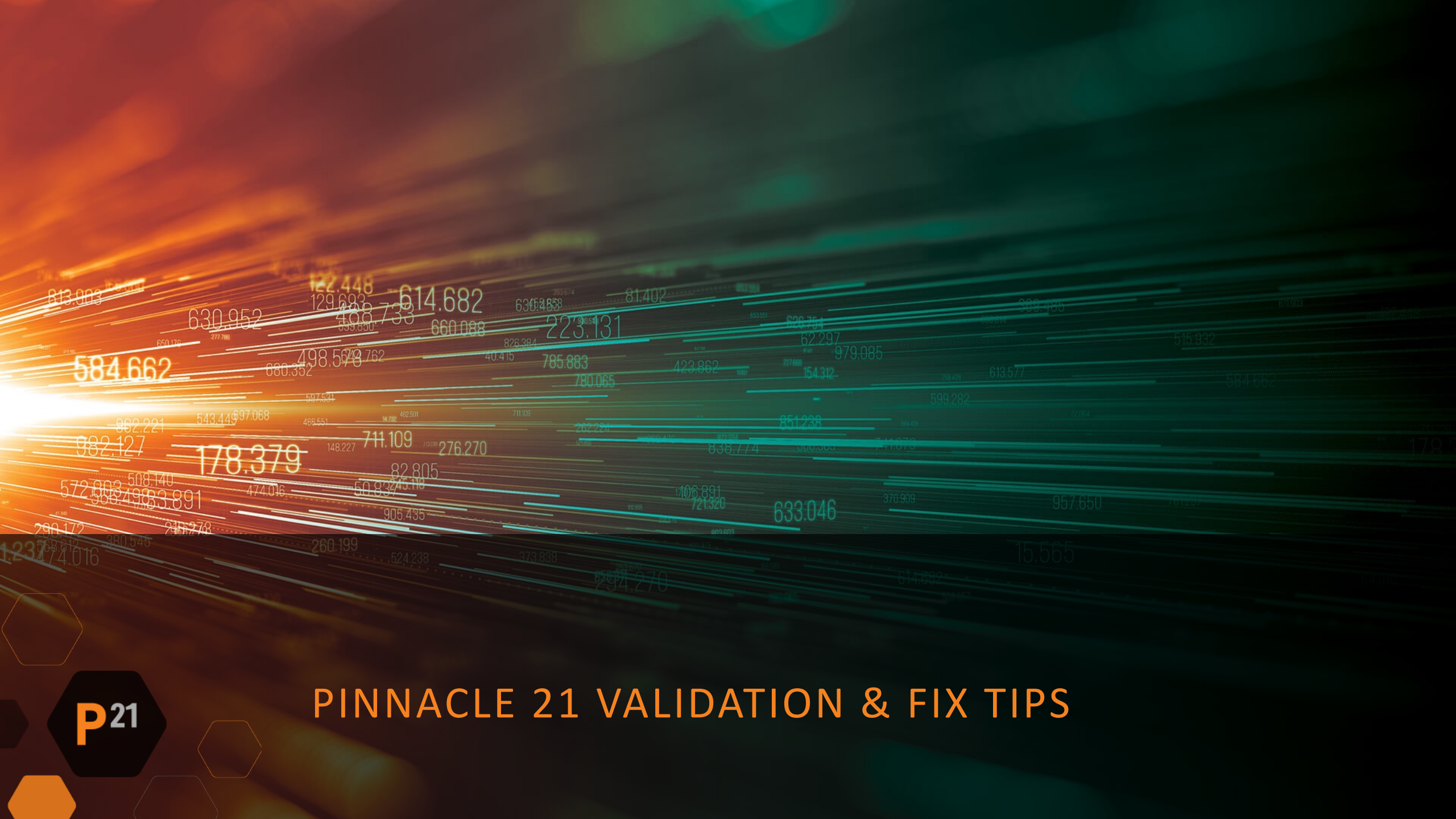
Data Composition	Data Quality	Coding Quality
Character Values Top 30 'text' values per variable 1,958 records	Inconsistent Lab Units Inconsistent standard units for tests 0 records	Adverse Events Similarity score/details of AETERM to AEDECOD/AELLT 249 records
Suppqual Character Values Top 30 'text' values per QNAM/QLABEL pair 12 records	Liver Labs with Ranges over ULN Summary of LBSTNRHI over Upper Limit of Normal 4 records	Disposition Similarity score/details of DSTERM to DSDECOD 40 records

Disposition

Search

Reported Term (DSTERM)	Standardized Term (DSDECOD)
Withdrew due to AE	OTHER

CDISC Submission Value	CDISC Synonym(s)	CDISC Definition	NCI Preferred Term
ADVERSE EVENT		Any untoward medical occurrence in a patient or clinical investigation subject administered a pharmaceutical product and which does not necessarily have	Adverse Event



PINNACLE 21 VALIDATION & FIX TIPS

CDISC VALIDATION RULES

► CDISC Terminology rules

ID ▲	Message ◆	Description ◆
CT2001	Variable value not found in non-extensible codelist	Variable must be populated with terms from its CDISC controlled terminology codelist. New terms cannot be added into non-extensible codelists.
CT2002	Variable value not found in extensible codelist	Variable should be populated with terms from its CDISC controlled terminology codelist. New terms can be added as long as they are not duplicates, synonyms or subsets of existing standard terms.
CT2003	Coded and Decoded values do not have the same Code in CDISC CT	Paired variables such as TEST/TESTCD must be populated using terms with the same Codelist Code value in CDISC control terminology. There is one-to-one relationship between paired variable values defined in CDISC control terminology by Codelist Code value.
CT2004	Variable value not found in non-extensible codelist when value-level condition occurs	Variable must be populated with terms from its CDISC controlled terminology codelist, when its value level condition is met. New terms cannot be added into non-extensible codelists.
CT2005	Variable value not found in extensible codelist when value-level condition occurs	Variable should be populated with terms from its CDISC controlled terminology codelist, when its value level condition is met. New terms can be added as long as they are not duplicates, synonyms or subsets of existing standard terms.
CT2006	Coded and Decoded values do not have the same Code in CDISC CT when value-level condition occurs	Paired variables such as TEST/TESTCD must be populated using terms with the same Codelist Code value in CDISC control terminology. There is one-to-one relationship between paired variable values defined in CDISC control terminology by Codelist Code value within the same value level condition.

CT2001

- Variable value not found in non-extensible codelist
 - Verify value in your raw data
 - Check CDISC CT terms
 - Update SDTM data value
 - Match the CDISC Submission Value (including case and formatting, no typos, etc.).
 - Example:
 - *Not Done* - incorrect
 - *NOT DONE* - correct

CT2001 PRSTAT value not found in **Not Done** non-extensible codelist

Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Synonym(s)
C66789		No	Not Done	ND	Not Done
C49484	C66789		Not Done	NOT DONE	

CT2001 (CONT.)

- Verify value in your raw data
- Check CDISC CT Submission values
- Review Requests Denied Document
- Update CRF design for future studies
- Update SDTM mapping
 - (including case and formatting, no typos, etc.)

▸ Examples of CRFs

▸ Action Taken with Study Drug

- None
- Drug discontinued
- Dose Reduced
- Dose Postponed -?

▸ Legacy: Action Taken with AE

- None
- Drug discontinued
- Hospitalization -?
- CM given -?

Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Synonym(s)
C66767		No	Action Taken with Study Treatment	ACN	Action Taken with Study Treatment
C49503	C66767		Action Taken with Study Treatment	DOSE INCREASED	
C49504	C66767		Action Taken with Study Treatment	DOSE NOT CHANGED	
C150826	C66767		Action Taken with Study Treatment	DOSE RATE REDUCED	
C49505	C66767		Action Taken with Study Treatment	DOSE REDUCED	
C49501	C66767		Action Taken with Study Treatment	DRUG INTERRUPTED	
C49502	C66767		Action Taken with Study Treatment	DRUG WITHDRAWN	
C48660	C66767		Action Taken with Study Treatment	NOT APPLICABLE	NA; Not Applicable
C17998	C66767		Action Taken with Study Treatment	UNKNOWN	U; UNK; Unknown

CT2002

- ▶ Variable value not found in extensible codelist
 - ▶ Verify value in your data
 - ▶ Check CDISC CT Submission for possible values
 - ▶ Ensure synonyms are not used
 - ▶ Ensure consistency in your study and submission
 - ▶ Avoid free text in data collection
 - ▶ Submit your extended terms to CDISC (as early as possible)
 - ▶ Document in Reviewer's Guide

CT2002 EPOCH value not found in Epoch extensible codelist

Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value
C99079		Yes	Epoch	EPOCH
C125938	C99079		Epoch	BASELINE
C102255	C99079		Epoch	BLINDED TREATMENT

Check ID	Diagnostic Message	Dataset	Count (Issue Rate)	Explanation
CT2002	EPOCH value not found in 'Epoch' extensible codelist	LB	1 (7.14%)	New terms were added to extensible codelist Epoch (C99079) for the study protocol needs. • BLINDED TREATMENT 1 • BLINDED TREATMENT 2

CT2003

- ▶ Coded and Decoded values do not have the same Code in CDISC CT
 - ▶ Review the CDISC CT
 - ▶ Ensure not using synonym
 - ▶ Avoid misspelling
 - ▶ Check SDTM mapping
 - ▶ Update so the NCI codes match for the paired values

Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Synonym(s)
C74683	C65047		Laboratory Test Code	CYTYRO	Tyrosine Crystals
C74683	C67154		Laboratory Test Name	Tyrosine Crystals	Tyrosine Crystals
C74756	C65047		Laboratory Test Code	CYTRPHOS	Triple Phosphate Crystals
C74756	C67154		Laboratory Test Name	Triple Phosphate Crystals	Triple Phosphate Crystals

INCORRECT

STUDYID	USUBJID	LBTESTCD	LBTEST
CDISCPILOT01	01-701-1015	CYTYRO	Triple Phosphate Crystals

Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Synonym(s)
C74683	C65047		Laboratory Test Code	CYTYRO	Tyrosine Crystals
C74683	C67154		Laboratory Test Name	Tyrosine Crystals	Tyrosine Crystals

CORRECT

STUDYID	USUBJID	LBTESTCD	LBTEST
CDISCPILOT01	01-701-1015	CYTYRO	Tyrosine Crystals

CT2004, CT2005, CT2006

- Similar to CT2001, CT2002, CT2003
 - Except for Value-level data

OTHER DICTIONARIES AND FORMATS

- ▶ MedDRA
 - ▶ SD0008, SD00082C, SD1114, SD1114C, ... (12 rules)
- ▶ SNOMED
 - ▶ SD2240, SD2257, ... (8 rules)
- ▶ MED-RT / NDF-RT
 - ▶ SD2243, SD2263, SD2264, SD2265
- ▶ WHODRUG
 - ▶ SD1344, SD1345, SD1346
- ▶ LOINC
 - ▶ SD2285
- ▶ ISO 3166 (Country)
 - ▶ SD1322, SE1323, SE2339
- ▶ ISO 2190 (Null Flavor)
 - ▶ SD2283



NEW VALIDATION RULES

LOINC

- ▶ SD2285: --LOINC value is not found in LOINC system
 - ▶ FDA requires LOINC **if it was provided by Labs**
 - ▶ Sponsor should not populate LOINC by themselves

WHODRUG VALIDATION

- Validation against the WHODrug dictionary available
- Supports **B3** format of WHODrug
 - Version listed in FDA's Data Standards Catalog
 - --DECOD **Drug Name**
 - --CLAS **ATC Text**
 - --CLASCD **ATC Code**

SD1344	Value for the Standardized Medication Name (--DECOD) variable must be populated using a Drug Name from the WHO Drug dictionary version specified in the define.xml.
SD1345	Value for the Medication Class (--CLAS) variable must be populated using ATC Text from the WHO Drug dictionary version specified in the define.xml.
SD1346	Value for the Medication Class Code (--CLASCD) variable must be populated using ATC Code from the WHO Drug dictionary version specified in the define.xml.

200+ CHARACTERS



CMDECOD in B3- and C3-format

The B3- and C3-formats were introduced in March 2017. These formats are designed to remove the workaround for retrieving generic names in B2- and C-formats. To obtain the generic name in the B3- and C3-formats, simply use the Preferred drug name from drug name field in MP.txt or DD.txt, depending on format used. Based on company conventions the Preferred Name is the Drug Code ending with 01 001 (Preferred Base Name) or ending with 001 (Preferred Salt Name).

CMDECOD is longer than 200 characters

For drugs with many ingredients, the generic name is longer than 200 characters. The SAS export format has a limitation to 200 characters per field, if this format is used for submission, the supplemental dataset needs to be utilised. Note that the guidelines state that the text should be truncated between words, in the case for long generic names the text should be truncated after the semicolon closest to 200 characters. Illustrations of the ordinary and supplemental datasets are shown in table 1 and 2.

Table 1. Illustration of SDTM dataset where CMDECOD is longer than 200 characters.

USUBJID	CMSEQ	CMTRT	CMMODIFY	CMDECOD	CMCLAS	CMCLASCD
AB-21-01	1		...	Ascorbic acid;Biotin;Calcium;Carbohydrates nos; Chloride;Choline;Chromium;Colecalciferol; Copper;Cyanocobalamin;Docosahexaenoic acid; Fats nos;Folic acid;Fructooligosaccharides; Iodine;Iron;Magnesium;		

Table 2. Illustration of supplemental dataset for CM domain where CMDECOD is longer than 200 characters.

USUBJID	RDOMAIN	IDVAR	IDVARVAL	QNAM	QLABEL	QVAL
AB-21-01	CM	CMSEQ	1	CMDECOD1	Standardized Medication Name 1	Manganese;Nicotinic acid;Pantothenic acid;Phosphorus;Phytomenadione; Potassium;Proteins nos;Pyridoxine; Retinol;Riboflavin;Selenium;Sodium; Thiamine;Vitamin e nos;Zinc



BEST PRACTICES

UNITS- SPLIT OUT CODELISTS

- ▶ UNIT Codelists in Define.xml
 - ▶ --ORRESU, --STRESU, and --DOSU
 - ▶ Split out and create separate codelist

NOT RECOMMENDED - COLLAPSED

ID	Name	NCI Codelist Code	Data Type	Order	Term	NCI Term Code
UNIT	Unit	C71620	text	1	%	C25613
UNIT	Unit	C71620	text	2	10^12/L	C67308
UNIT	Unit	C71620	text	3	g/dL	C64783
UNIT	Unit	C71620	text	4	mg/dL	C67015
UNIT	Unit	C71620	text	5	U/L	C67456
UNIT	Unit	C71620	text	6	g/dL	C64783
UNIT	Unit	C71620	text	7	mEq/L	C67474
UNIT	Unit	C71620	text	8	CAPSULE	C48480
UNIT	Unit	C71620	text	9	mg	C28253
UNIT	Unit	C71620	text	10	TABLET	C48542

RECOMMENDED- SPLIT OUT

ID	Name	NCI Codelist Code	Data Type	Order	Term	NCI Term Code
LBORRESU	Original Units for LB	C71620	text	1	%	C25613
LBORRESU	Original Units for LB	C71620	text	2	10^12/L	C67308
LBORRESU	Original Units for LB	C71620	text	3	g/dL	C64783
LBORRESU	Original Units for LB	C71620	text	4	mg/dL	C67015
LBORRESU	Original Units for LB	C71620	text	5	U/L	C67456
LBORRESU	Original Units for LB	C71620	text	6	g/dL	C64783
LBORRESU	Original Units for LB	C71620	text	7	mEq/L	C67474
CMDOSU	Dose Units for CM	C71620	text	1	CAPSULE	C48480
CMDOSU	Dose Units for CM	C71620	text	2	mg	C28253
CMDOSU	Dose Units for CM	C71620	text	3	TABLET	C48542
EXDOSU	Dose Units for EX	C71620	text	1	mg	C28253

BEST PRACTICE FOR EXTENSIBLE CT

- Control data collection process
- Maintain consistency
 - within study, submission, and organization
- Check denied terms document from CDISC CT Team
- Submit request to CDISC CT team

<https://www.cdisc.org/standards/terminology/controlled-terminology>

Controlled Terminology Release

As of 18 Dec 2020 the Protocol Entities, SDTM, SEND, CDASH, Glossary, and Define-XML Controlled Terminology files have been updated on the NCI-EVS Ftp site. The version dates of the new files are 2020-12-18. These terminology files replace all older Protocol Entities, SDTM, SEND, CDASH, Glossary, and Define-XML and include terms from Review Package 44. There are approximately 307 new QRS terms and 235 new terms across Protocol Entities, SDTM, SEND, CDASH, Glossary, and Define-XML. Additionally there are updates to 6 published Codetable Mapping files: DD, DS, RP, SC, TS, and VS.

[Controlled Terminology Release 18 Dec 2020](#)

Supplemental Files

[NCI FTP Links](#)

[Resources](#)

[Rules](#)

[Codetable Mapping Files](#)

[Unit-UCUM Mapping File](#)

[LOINC to LB Mapping Files](#)

[Paired Codelists](#)

[New Term Request Page](#)

[Multiple Term Request](#)

[CT Requests Denied](#)

[Publication Schedule](#)

REVIEWER'S GUIDE EXPLANATIONS

- Providing detailed explanation in Issue Summary

Example of insufficient explanation for CT2002

Check ID	Diagnostic Message	Dataset	Count (Issue Rate)	Explanation
CT2002	CMDOSU value not found in UNIT codelist	CM	3 (<0.1%)	Codelist is extensible.

Example of appropriate explanation for CT2002

Check ID	Diagnostic Message	Dataset	Count (Issue Rate)	Explanation
CT2002	CMDOSU value not found in UNIT codelist	CM	3 (<0.1%)	This rule fired for 'OTHER', 'DROPS', and 'UNK' values in CMDOSU. These values could not be mapped to CDISC CT but the UNIT codelist is extensible.



THANK YOU ;)