An Overview of the NCI EVS Services and Its Collaboration with CDISC

Presented by Dr. Jordan Li;
Clinical/Biomedical Information Specialist,
National Cancer Institute's Enterprise Vocabulary Services

Meet the Speaker

Jordan Li. Ph.D.

Title: Clinical/Biomedical Information Specialist

Organization: Guidehouse; NCI-EVS

- Subject matter and terminology expert from the National Cancer Institute's Enterprise Vocabulary Services (NCI EVS).
- Over 10 years of experience creating and publishing CDISC terminology.
- Directs delivery of multiple CDISC controlled terminology teams .
- Leads the CDISC Microbiology and Immunogenicity Submission Data Standards Development (SDS) subteam.
- CDaSH and SDTM standards experts and developed standards for several CDISC disease therapeutic area projects, including Cardiovascular Imaging, Kidney Transplant, Type 1 Diabetes and multiple Traditional Chinese Medicine projects.

Jordan holds a PhD in Pharmacology and Physiology from Georgetown University.

Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.*

National Cancer Institute - Enterprise Vocabulary Services (NCI - EVS) - What We Do

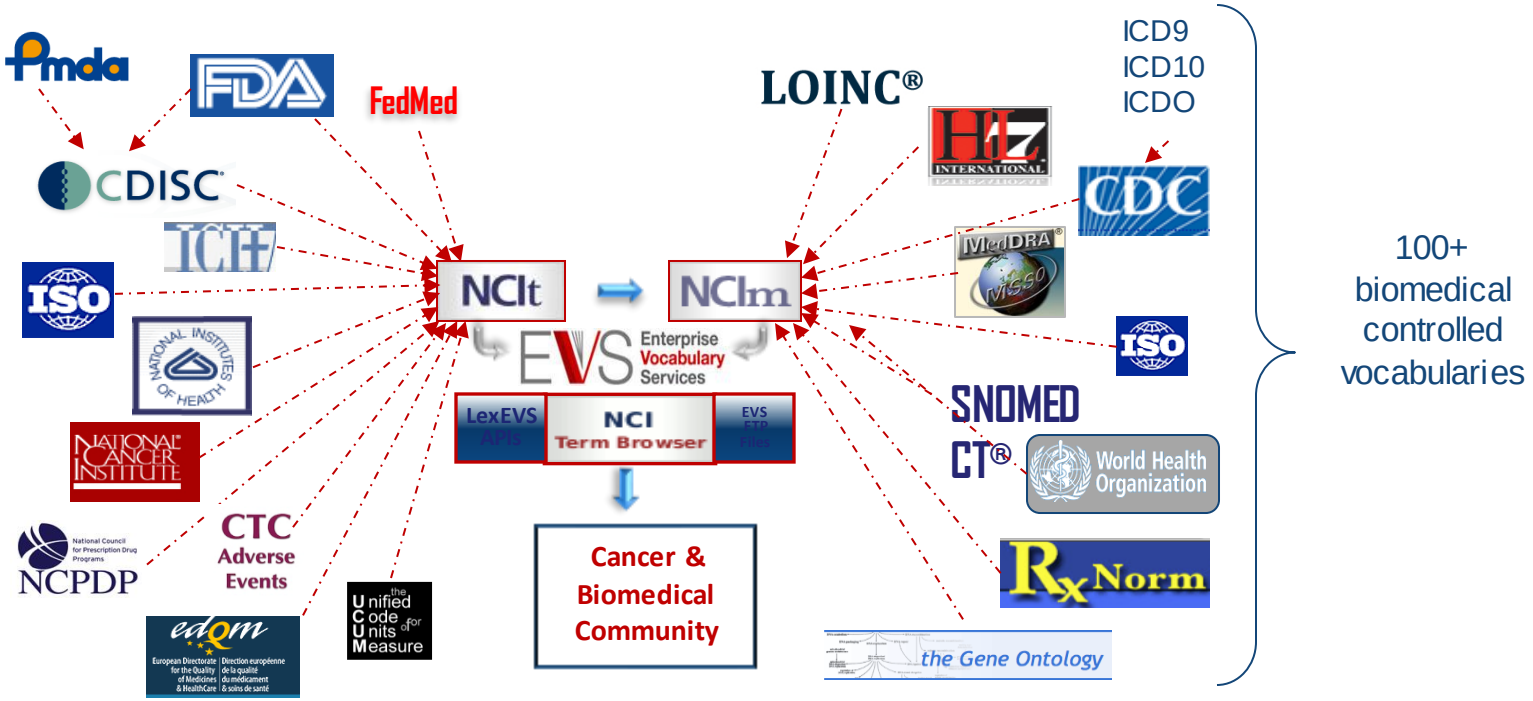
Mission: Development of services and resources that address the needs of NCI and many national and international partners for controlled vocabularies.

EVS provides an array of technology products and services to support the semantic infrastructure of NCI and our collaborating organizations.

- Subject Matter Expertise
- Definition Writing and Analysis
- Terminology Coding, Tagging and Subset bundling
- Publish, Maintain and Version Terminology
- Links to Other Controlled Terminologies
- Terminology Content Expansion & Facilitation of Terminology Requests
- Produce and Maintain Two Public Browsers - NCIt & NCIm

Source: NCI-EVS

NCI EVS:
Unified Terminology Services and Standards



NCI-EVS works with many national and international partners to develop, license and publish terminology, and also jointly develop software tools to support harmonization and shared standards.

Source: NCI-EVS

NCI EVS Production Engine: NCI Thesaurus (NCIt) Term Browser

- Free, Open Access 24/7

NCI Thesaurus (NCIt) Overview

- NCI's core reference terminology. It is a free, non-license restricted biomedical coding ontology system.
  - Main function of NCIt is to ***establish data semantics***.
- Adopted by FDA, CDISC and other partners as a shared standards development and coding environment.
- Allows participants to compare and harmonize with each other's content while taking advantage of full-text definitions, codes, and other features.
- User-driven concept expansion from NCI, NIH and other term source partners. It is also open to public request via the EVS term request system.
  - Anyone can suggest an addition or a change to existing concepts within the NCIt.
- Currently there are 50 source contributors tagged in NCIt concepts.
- As of April 2023, over 178,000 coded concepts stored in NCIt, including basic and translational research, clinical care, public information, healthcare and administrative activities.
- 400,000 relationships modeled between concepts.

Source: NCI-EVS

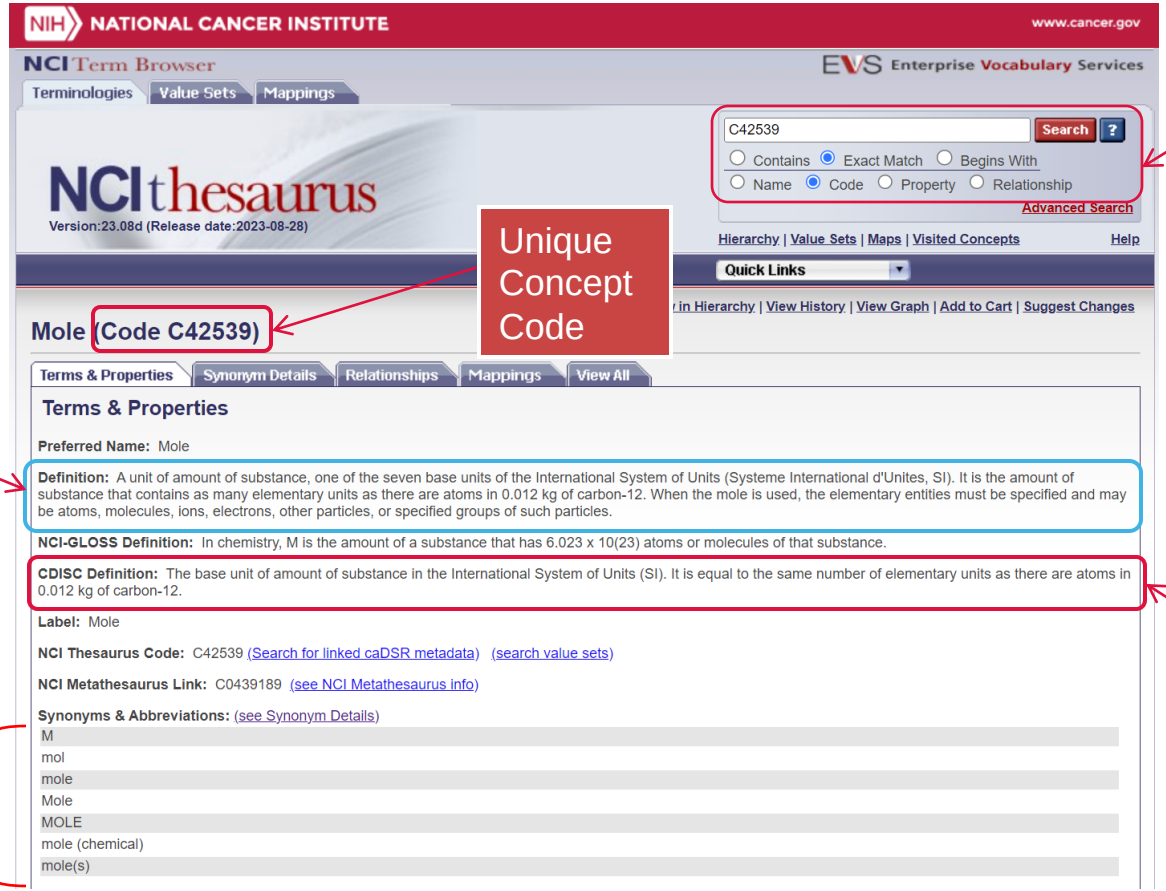
NCIt is used to build and enrich semantics for a concept

All concepts in the NCIt are given:

- **NCIt concept code (C-code):** a unique identifier assigned to each concept by EVS to *permanently* track a specific meaning.
- **Preferred Name**: the term chosen by EVS as most *unambiguous* and widely used in the biomedical community.
- **Synonyms**: additional term(s) chosen by NCI with meaning equivalent to the Preferred Term, this helps with *mapping to other* terminologies.
- **Definition**: a text statement of the *what the concept is.* This ensures that humans and systems have a common understanding of the concept and can therefore use it accurately, precisely and consistently across users and systems.
- **Hierarchical arrangement**: concepts are modeled using a *parent-child "is-a" relationships* in the NCIt ontology environment.

Source: NCI-EVS

Searching for a concept in the NCIt Term Browser



Search Box

Unique Concept Code

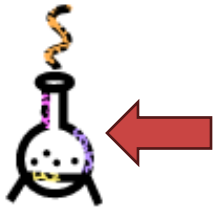
NCI Definition

Source-Specific Definition from CDISC

Common Synonyms

Source: NCI-EVS

The Importance of a Concept Code – Mole (the Unit)



NIH NATIONAL CANCER INSTITUTE www.cancer.gov

NCI Term Browser — EVS Enterprise Vocabulary Services

Terminologies | Value Sets | Mappings

C42539 Search ?

Contains Exact Match Begins With

Name Code Property Relationship

Advanced Search

NCI thesaurus

Version:23.08d (Release date:2023-08-28)

Hierarchy | Value Sets | Maps | Visited Concepts Help

Quick Links

View in Hierarchy | View History | View Graph | Add to Cart | Suggest Changes

Unique Concept Code

Mole (Code C42539)

Terms & Properties | Synonym Details | Relationships | Mappings | View All

Synonym Details

Term	Source	Type	Code	Subsource Name
mole (chemical)	NCI-GLOSS	PT	CDR0000659802	
mole (chemical)	NCI-GLOSS	SY	CDR0000659802	
mole(s)	ICH	PT	0196	
MOLE	FDA	PT	MOLE	Potency (C-DRG-00501)
mole	HL7	PT	12856	
Mole	NCI	PT		
mole	UCUM	PT		
Mole	CDISC	SY		
mole	FDA	SY		PQCMC
MOLE	FDA	SY		SPL
Mole	NCI	SY		
mol	HL7	AB	12856	
mol	ICH	AB	0196	
mol	NCI	AB		
mol	UCUM	AB		
mol	CDISC	PT		
mol	FDA	PT		PQCMC
mol	FDA	PT		SPL
M	NCI-GLOSS	PT	CDR0000659892	
M	NCI-GLOSS	SY	CDR0000659892	

Terms from Other Sources

Term Source

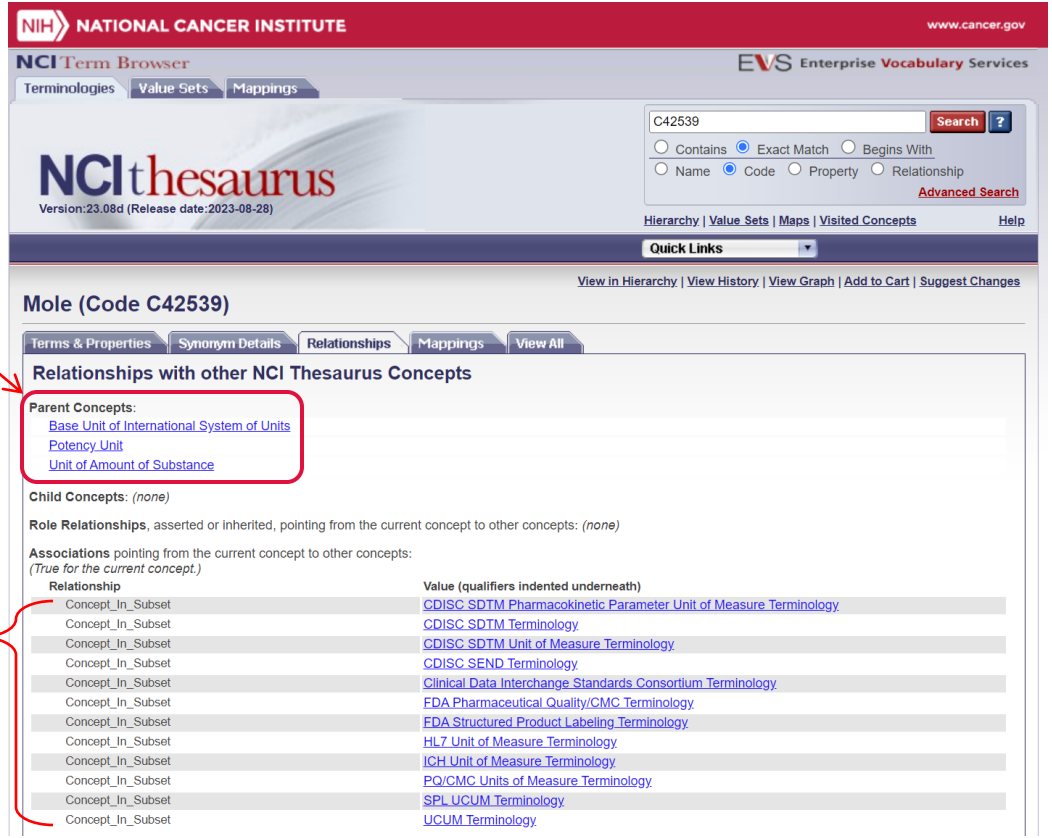
Additional Source Data

Source: NCI-EVS

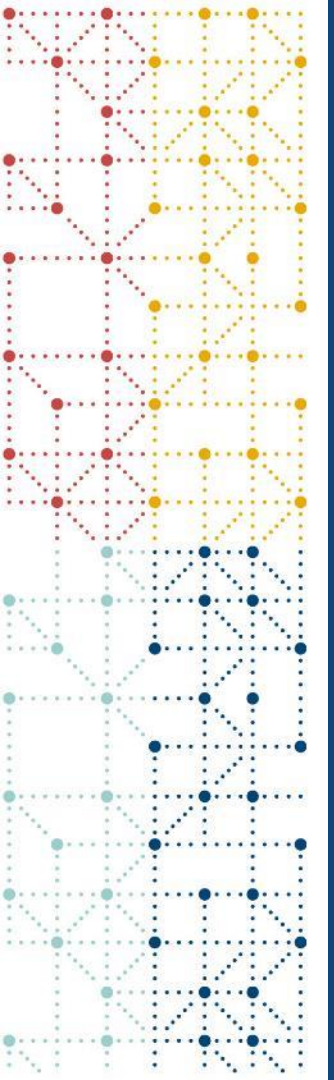
Relationships – Where else is "Mole" used? Association to Other Vocabularies and Codelist Bundling

NCIt Hierarchy

The "Concept_In_Subset" property indicates what other terminology source this concept is associated with. It also allows the bundling of concepts into supersets and value sets (codelists). Currently EVS maintains over 1500 value sets for 26 organizational entities.



Source: NCI-EVS



NCI-EVS and CDISC Terminology Program

NCI EVS Partnership





CDISC and EVS have been working closely together for more than 20 years.

- EVS provides dedicated terminology experts, subject matter experts.
- Provides established terminology infrastructure and standard operating procedures.
- Responsible for CDISC controlled terminology development, harmonization, publication and maintenance.

Required and preferred by global regulatory agencies.

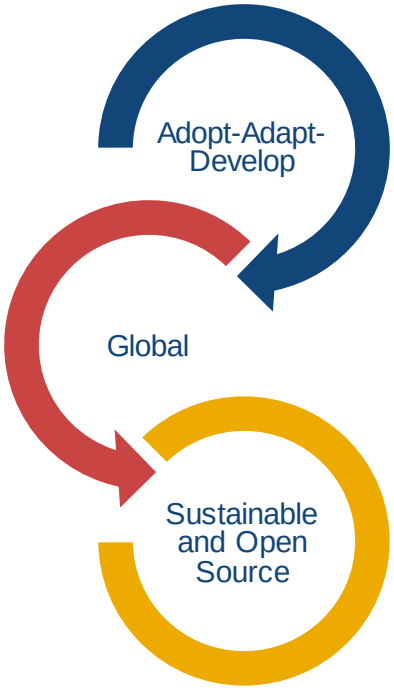
- Required by FDA and PMDA.
- Preferred NMPA and Recommended by EMA.
- Support both general clinical and non-clinical data submission.

>54,000 CDISC terms are coded and tagged in NCI Thesaurus (NCIt)

- All CDISC-tagged concepts contain a CDISC definition to ensure that all users understand and use the concept in the same way.
- 8 subsets of CDISC terminology, each with one-to-many codelists.
- Over 1300 published codelists associated with CDISC terminology in NCIt.

CDISC terminology standards average 35,000 downloads/month in more than 60 countries.

CDISC Terminology Objectives and Development Principles



- Support CDISC foundational and therapeutic area standards and CDISC teams. Fulfill CDISC user requests and support the user community.
- Develop CDISC submission values with clear definitions; Create codelists and concept coding; define relationships between and among CT and CDISC model elements.
- **Two-Way process:**
  - Use CDISC models to guide and drive terminology decisions.
  - Terminology analysis influences domain modeling and structuring.
- Consider the impact of changes.
- Communicate early and often with stakeholders.
- Focus on re-use of existing terminology.
- Commit to consistent decision-making; follow best terminology practice rules.

CDISC CT Are Made by Their Users

- Currently 16 active CDISC controlled terminology teams that develop CT to support various CDISC standards and to fulfill CDISC user requests.
  - New volunteers welcome! (Join us)
- Led and overseen by NCI-EVS, industry subject matter- and data standards experts with rich experience and extensive knowledge in cardiovascular medicine, pharmacokinetics/pharmacogenomics, oncology, laboratory sciences, human physiology, and immunology studies, etc.
  - CDISC terminology covers a wide variety of scientific and medical disciplines.
  - Used by and support both clinical and non-clinical studies.
- Made and consumed by their users: clear, consistent, practical, user-friendly.

Controlled Terminology to Support TMF Standards

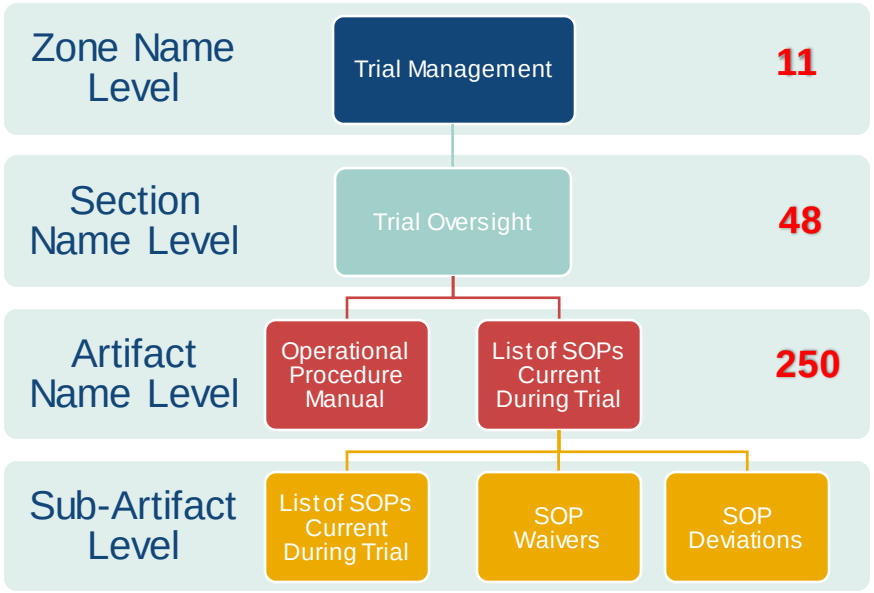
- Artifacts for TMF-RM
- Metadata Standards for eTMF-EM

Terminology Organization and Bundling Basics

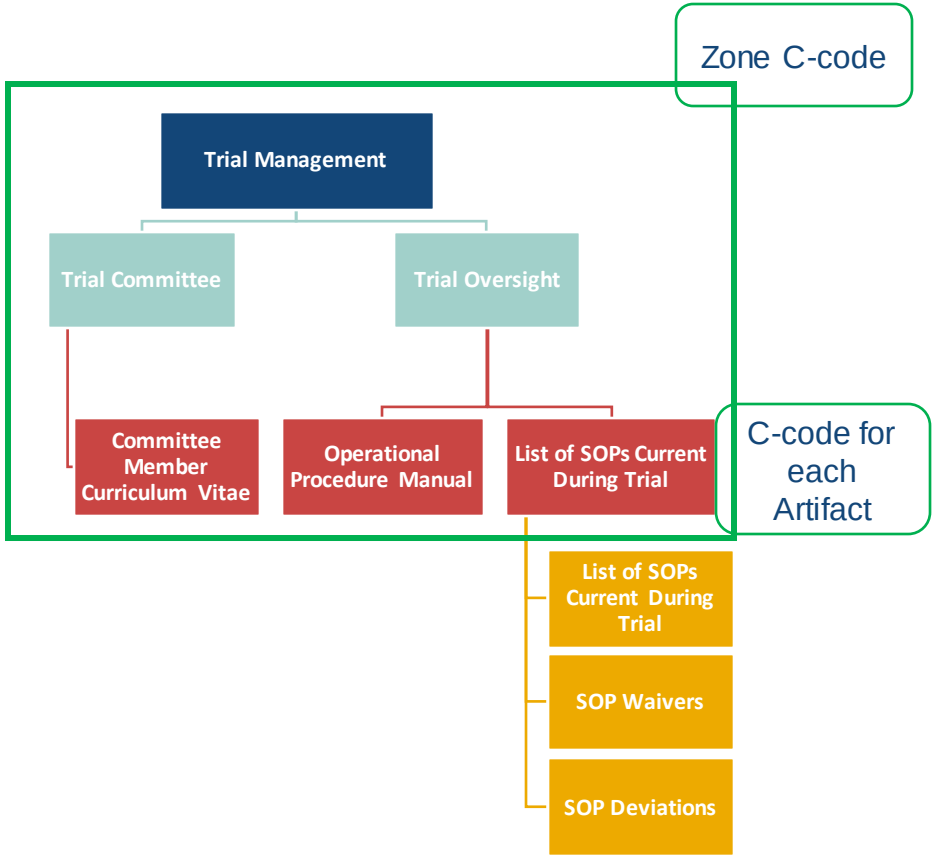
- The development and organization of controlled terminology are driven by the model structure it serves.
  - how do we organize the controlled terminology so it can best support the "overall structure" of the model and be the most user friendly?

- For example, some of the CDISC Terminology is grouped by CDISC Domains.
  - CDISC has what they call physiology/body system domains such as: Cardiovascular Domain/CV (for collection of heart assessment data), GI domain (for collection of GI assessment data), Respiratory/RE Domain (for collection of lung function test data).
  - Assessment Codelists are created for EACH domain.
    - ✓ Cardiac Output, belong in the CV codelist.
    - ✓ Number of Polyps is in the GI codelist.
  - If you are running a study on treating a heart disease, you are more likely to collect data related to heart findings:
    - Superset: weed through the GI and RE tests just to find the heart assessment you need?
    - Domain-specific Valuesets: Much easier to find what you need in the CV codelist for heart related tests. (CDISC CT approach)

Organization of TMF CT – Initial Plan for Artifacts in RM



- **Three-level Value Control:** Provide coding and semantics to the values from Zone, Section, and Artifacts. Each concept in these 3 levels will be given unique C-codes for permanent tracking as well as clear and consistent definitions.
  - Plans to also control sub-artifacts in the future.
- **Two-level Grouping**:
  - All Artifacts related to a **Zone.**
  - Artifacts related to a **Section**.



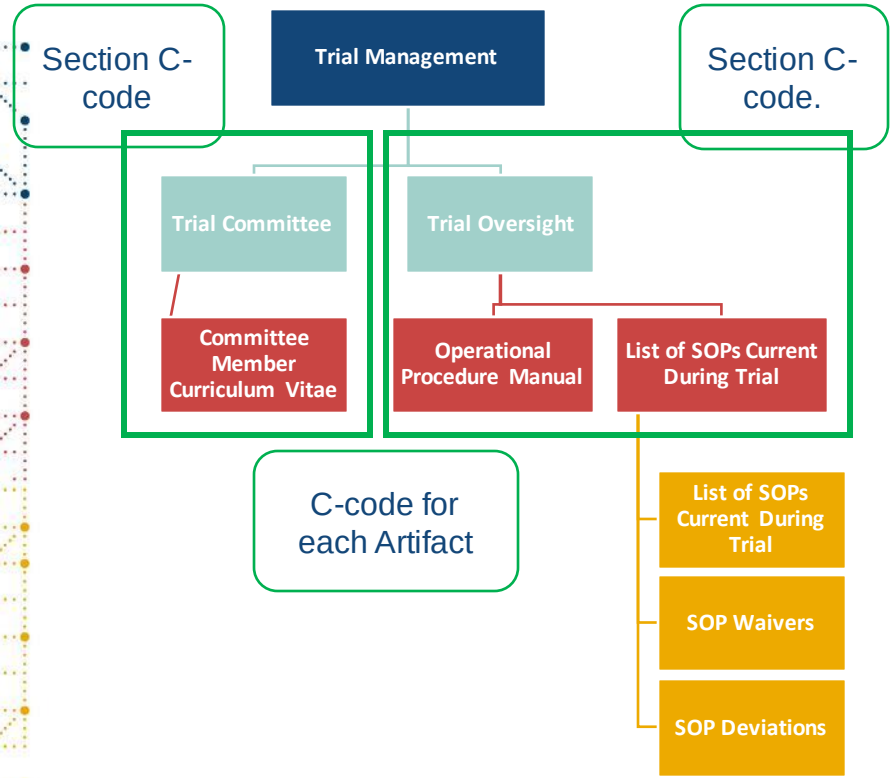
Group artifacts into their related Zone

Codelist Name: Trail Management Zone Related Artifacts

- Unique C-code and definition for this codelist.

All 37 related artifacts across 5 sections from this Zone will be grouped into this codelist.

- Unique C-code and definition for each artifact.



Group artifacts into their related Sections

Codelists Names:

- Trial management Zone - Trail Oversight Section Related Artifacts
  - Unique C-code and definition for this Section as a codelist.
  - 20 related artifacts mapped to this codelist with unique C-code and definitions.

- Trial management Zone - Trail Committee Section Related Artifacts
  - Unique C-code and definition for this Section as a codelist.
  - 7 related artifacts mapped to this codelist with unique C-codes and definitions.

Future Plans – Additional Valuesets to Support the TMF RM

- Grouping by artifacts needed and included at the Trail, Country, Site Levels.
- Grouping by artifacts needed and included at the Trail Level Milestone, Country Level Milestone and Site Level milestone.
- Value-set organization of the artifacts are driven by the model and how TMF users consume and use this data.
- Ability to retrieve artifacts and standards metadata only relevant to, and needed by a specific purpose(s), e.g. TMF documents included only at the trial and site levels.
- Customized (artifact) terminology to support the RM in the way you need.

TMF Level					
Trial Level Document	Trial Level MILESTONE/EVENT	Country/ Region Level Docume	Country Level MILESTONE/EVENT	Site Level Documen	Site Level MILESTONE/EVENT
X	01 First Country RA Approval	X	02 Clinical Infrastructure Ready		
X	01 First Country RA Approval	X	02 Clinical Infrastructure Ready		
X	01 First Country RA Approval				
X	01 First Country RA Approval	X	01 First Country RA Approval		
X	01 First Country RA Approval	X	01 First Country RA Approval	X	01 First Country RA Approval

Future Plans – Codify the Metadata Standards in “Exchange.XML”

Provide unique identifier codes and semantics for the metadata standards used by the eTMF-Exchange Mechanism.

Specifically, we will control the attributes under the 6 tags in the Exchange.xml:

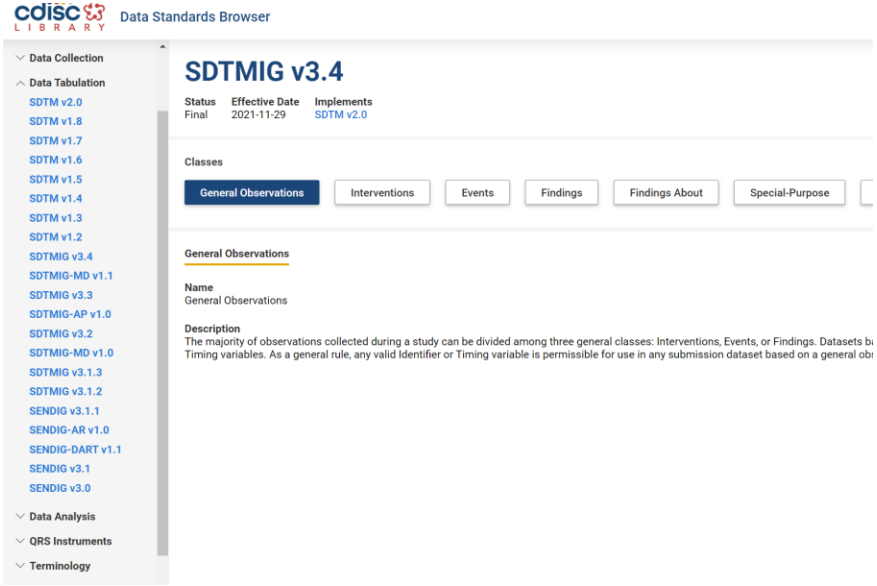
- <BATCH>Tag
- <OBJECT>Tag
- <FILE>Tag
- <SIGNATURE>Tag
- <AUDITRECORD>Tag
- <METADATA>Tag

5.3.2. <OBJECT> Tag

The <OBJECT> tag is used to encapsulate individual objects that are being exchanged. For each artifact included in the exchange there will be an <OBJECT> tag. An artifact is defined as a single record within the TMF. Each artifact must have its own <OBJECT> tag.

Attribute ID	Attribute Name	Mandatory	Description	Purpose of attribute	Attribute format
OBJECTID	Object ID	Yes	A system-readable identifier for an electronic record that represents a real-world TMF object. Generated by a source system and doesn't change over time.	To be able to identify the artifact and provide artifact traceability between the source and receiving system.	Text
OBJECTLEVEL	TMF RM Study Level	Yes	Identify the Artifact TMF study level as outlined in the current TMF reference model	Used to ensure that the artifact is filed at the correct level within the receiving system.	Text (Trial, Country, Site)
COUNTRYID	Study Country ID	Yes – For country or site level documents	Unique identifier for a study country	Used to file the artifact within the correct country within the receiving system.	ISO-3166 Alpha-3 i.e. CAN, FRA etc.
SITESYSTEMID	Study Site ID	Yes - For site level documents	Unique system identifier for a study site	Unique system identifier for a study site that is generated by the source system or receiving system. Used to file the artifact within the correct site within the receiving system.	Text
SITEID	Study Site Number	Yes – For site level documents	Unique human identifier for a study site	Unique human readable identifier for a study site. Could also be used to file the artifact within the correct site and could be identical to the SITESYSTEMID	Text
UNIQUEID	TMF RM Unique ID Number	Yes	Unique identifier for a specific TMF RM Artifact	Used to index objects against the correct artifact type within the receiving system.	Number on 3 digits i.e. 001, 010

Both the TMF Reference Model and Terminology will be stored in the CDISC Library



CDISC models and controlled terminology are stored in the CDISC library. You can:

- Access and easily browse multiple versions of the CDISC Foundational Standards' metadata and Controlled Terminology.
- Retrieve machine-readable standards metadata using multiple media-types: JSON, XML, ODM, CSV, Excel (moving away from spreadsheet for metadata consumption and simplify automation).
- Create Diff Reports, which deliver machine-readable (JSON, Excel) metadata comparisons that compute and display differences between the content in two versions of a standard.
- Model relationships between and within variable- and value- level metadata.
- Check conformance against standards metadata.

To learn more - Past Webinars:

- **CDISC Library: Ideas for using the CDISC Library**: https://www.cdisc.org/events/webinar/cdisc-library-ideas-using-cdisc-library-and-look-whats-coming-next
- **CDISC Library Virtual Workshop:** https://www.cdisc.org/events/webinar/cdisc-library-virtual-workshop

Source: NCI-EVS

Controlled Terminology Better Support the RM and the Exchange of TMF Content

Create Clear Standards and Formalize the model with regulatory authorities

- Concept coding at Zone, Section and Artifact levels – unique and permanent tracking for each TMF concept at each level.
- Definition proofing: clearly, consistently and meaningfully defined individual concepts (instead of plain language description now used by the reference model).
- Standardized TMF concepts will help to better support the Refence Model itself, which will further help to formalize the TMF-RM with regulatory agencies.
- Terminology analysis and placement influences RM modeling and structuring – users will have a better understanding of where to add and place a new or existing artifact.

Enhance the Trial Master File Exchange Mechanism Standard

- Codified artifacts and metadata standards will be used in the “exchange.xml” to support the automatic transfer of TMF content between sponsors, CROs and other stakeholders in the eTMF-EMS.
- Enhance communication among sponsors, partners, and CROs.

Maintain community engagement

- From the start: controlled terminology is made by their users: need to be clear, consistent, practical and useful.
- To the end: user community governance and quality assurance through public outreach, review and commenting.

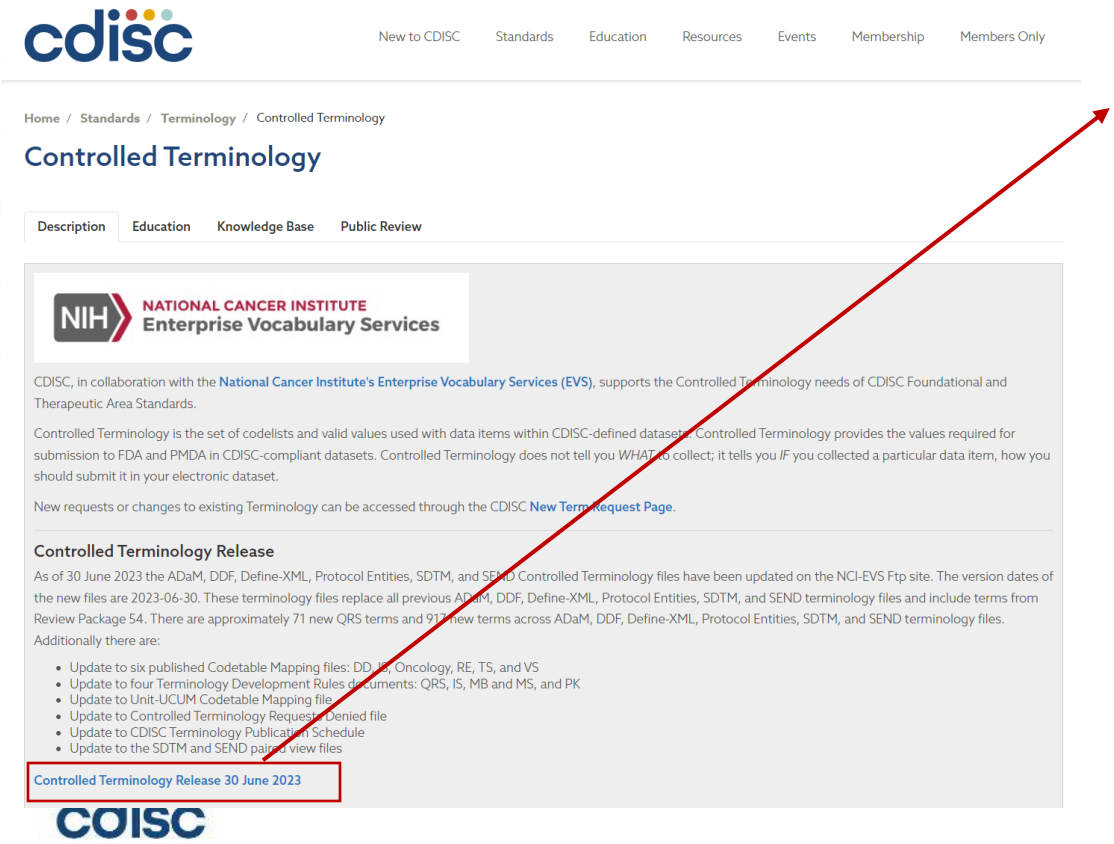
Imported into the CDISC Library

- Searching for specific artifacts, mapping relationships between artifacts.
- Control changes and version history.
- Retrieve versioned reference model and TMF terminology from the library.
- Create views of content and metadata for various types of trials.

cdisc

Additional Slides – More Info

CDISC Controlled Terminology Page



CDISC CT Landing Page - https://www.cdisc.org/standards/terminology/controlled-terminology

- Public announcement will be made about the new CT public review and publication (email blast from CDISC)
- High-level summary will be provided on updates made to specific CDISC terminology subsets.
- Link to the NIH CDISC Terminology Homepage for download.
- Other supporting documents (CT rules, etc.) will be published and hosted on this page.

Terminology Public Review and Comment Resolution – CDISC WIKI

Terminology Call for Public Review Package P55 - Comments Due by 21 July 2023

Created by Alana St. Clair, last modified on Jun 23, 2023

CDISC invites you to submit comments during the Public Review for Controlled Terminology Package 55, which consists of the following 19 documents:

- Controlled_Terminology_Requests_Denied_P55
- IS Terminology Mapping Codetable_P55
- QRS Naming and Business Rules
- Rules for IS
- Rules for MB and MS
- Rules for Lab, Unit, and MI
- Biospecimens
- Cell Phenotyping
- Devices
- ECG
- General
- Genomics
- Lab
- Microbiology-Immunology
- MRCT
- Oncology
- PK
- Protocol Entities
- UNIT

An * indicates that **changes or retirements of existing CDISC Submission Values** are included on the "Changes to Existing" tab in the document. Please review th

Comments due: 21 Jul 2023

Instructions for Reviewers

Register and/or Log in WIKI and in JIRA to provide comments. If you already have an account on Wiki or JIRA, our issue-tracking system, simply log in to your acco

Download the files below for review.

If you have no comments after review:

1. Click 'Like' at the bottom of the page. This will help us determine who has reviewed the documents.

To add comments from within JIRA:

1. Go to the **Controlled Terminology** project in JIRA at: http://jira.cdisc.org/projects/CT

- Instructions on how to enter comments through **CDISC JIRA – CT Project**.

cdisc

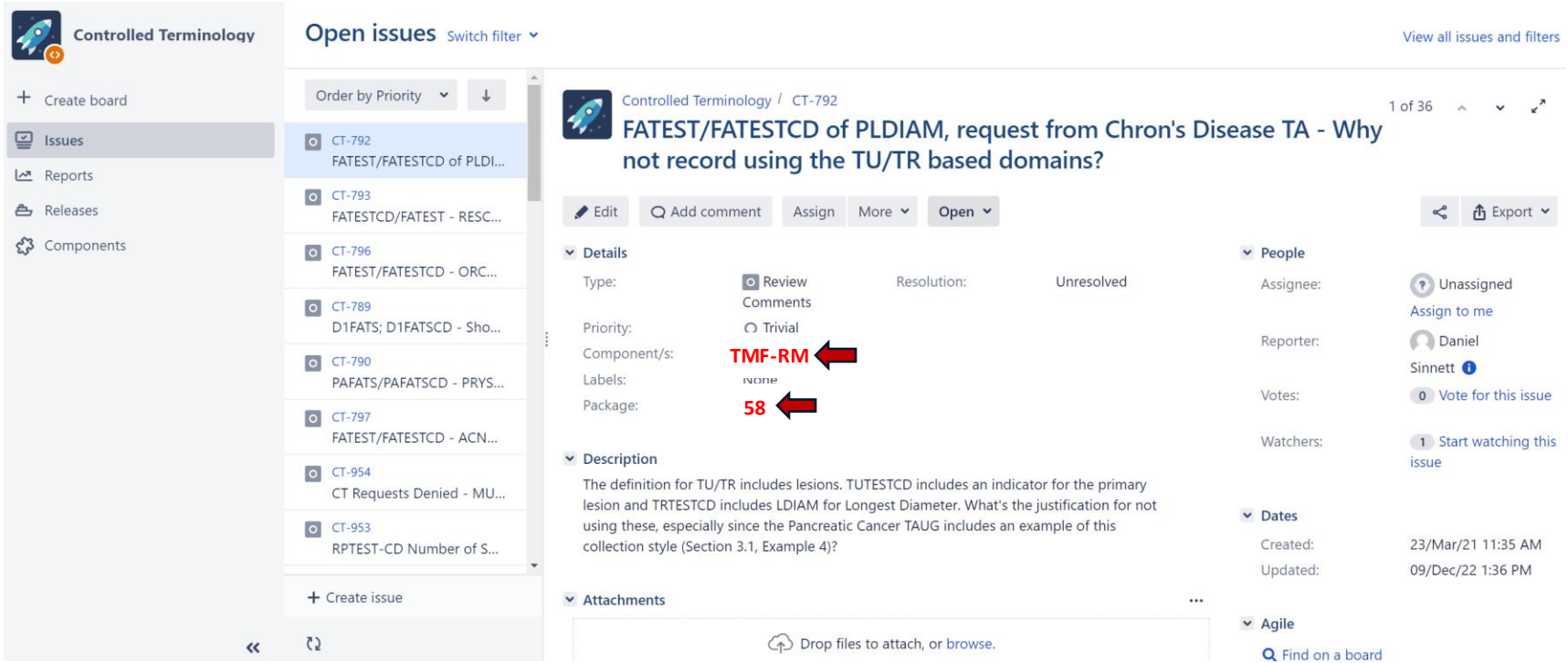
Public Review Documents

File

- Terminology_P55_Microbiology-Immunogenicity.xlsx
- Terminology_P55_MRCT.xlsx
- Terminology_P55_Oncology.xlsx
- Terminology_P55_PK.xlsx
- Terminology_P55_Protocol_Entities.xlsx
- Terminology_P55_UNIT.xlsx
- Controlled_Terminology_Requests_Denied_P55.xlsx
- IS Terminology Mapping Codetable_2023-09-29_P55 PR Version.xls
- QRS_Naming_and_Business_Rules_2023-09-29_P55 PR Version.xlsx
- Rules for IS_2023-09-29_P55 PR Version.docx
- Rules for MB and MS_2023-09-29_P55 PR Version.docx
- Rules_for_Lab_Unit_and_MI_2022-12-16_P55 PR Version.docx
- Terminology_P55_Biospecimens.xlsx
- Terminology_P55_Cell Phenotyping.xlsx
- Terminology_P55_Device.xlsx
- Terminology_P55_ECG.xlsx
- Terminology_P55_General.xlsx
- Terminology_P55_Genomics.xlsx
- Terminology_P55_Lab.xlsx
- image2023-6-23_10-21-36.png

> Terminology_P58_TMF-RM.xlsx

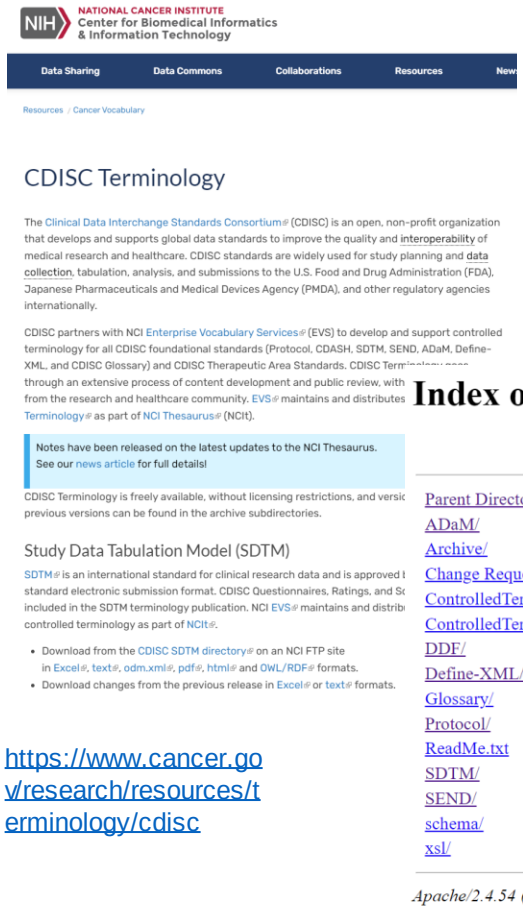
Terminology Public Review and Comment Resolution – CDISC JIRA CT Project



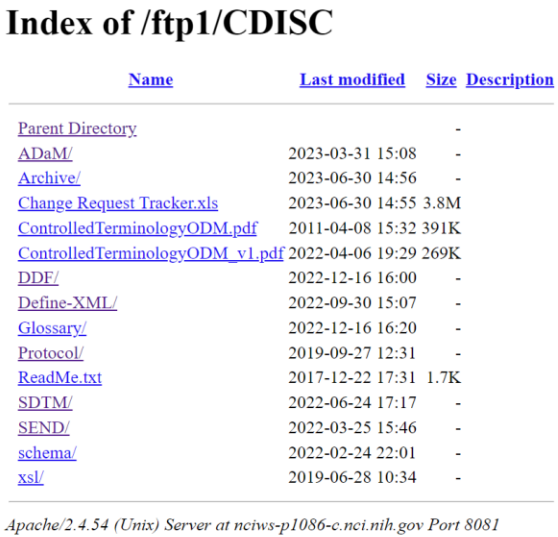
http://jira.cdisc.org/projects/CT

CDISC and TMF Terminologies are published on the NCI-FTP Site

- Plan to publish twice a year for TMF terminology; aiming for 2024 publication.
- **6 formats** can be directly downloaded from this page: .xls, .txt, .html, .odm-xml, .OWL/RDF, .pdf
- **CT Changes files** (from previous release) + Changes Program
- All files are stored on the **NCI Ftp**: https://evs.nci.nih.gov/ftp1/CDISC/
  - Including older published files that are stored in subset terminology Archives.
- Imported into the CDISC Library.



https://www.cancer.gov/research/resources/terminology/cdisc



Source: NCI-EVS

Join and Volunteer for a CDISC CT Team

- https://www.cdisc.org/volunteer/form
- Receive trainings.
- Fill out a form and indicate which standards and CT teams you'd like to join.
- Register and create accounts in CDISC WIKI and JIRA.
- Participate in public reviews/commenting for CDISC standards products.
- Join a team and talk to us directly.