

# caCORE Training Workbook

## Course 1040: Creating Well-formed Metadata and Metadata Business Rules



caCORE Training Website  
Help & Support

[http://ncicb.nci.nih.gov/NCICB/training/cadsr\\_training](http://ncicb.nci.nih.gov/NCICB/training/cadsr_training)  
[ncicb@pop.nci.nih.gov](mailto:ncicb@pop.nci.nih.gov) (please include "caCORE Training" in the subject)

# Table of Contents

INTRODUCTION .....	5
<b>1 COURSE DETAILS .....</b>	<b>5</b>
<b>2 LESSON 1: METADATA FUNDAMENTALS.....</b>	<b>5</b>
2.1 OBJECTIVES FOR THE LEARNER .....	5
2.2 REVIEW ISO/IEC 11179 ADMINISTERED ITEMS .....	6
2.3 LESSON 1 REVIEW.....	8
<b>3 LESSON 2: NAMING CONVENTIONS.....</b>	<b>8</b>
3.1 OBJECTIVES FOR THE LEARNER .....	8
3.2 DEVELOPMENT OF A NAMING CONVENTION .....	9
3.3 TYPES OF VOCABULARY LISTS .....	10
3.4 DIFFERENT TYPES OF DATA ELEMENT NAMES .....	11
3.4.1 Long Name Rules	11
3.4.2 Short Name Rules	12
3.5 LESSON 2 REVIEW.....	12
<b>4 LESSON 3: CREATING NAMES.....</b>	<b>13</b>
4.1 OBJECTIVES FOR THE LEARNER .....	13
4.2 STEP 1: CONSIDERING COMPONENTS OF THE DATA .....	13
4.3 STEP 2: DATA ELEMENT CONCEPT NAME .....	14
4.3.1 Finding the Object Class	14
4.3.2 Finding the Property	17
4.3.3 Created Data Element Concept Name	18
4.4 STEP 3: VALUE DOMAIN NAME.....	19
4.4.1 Selecting the Representation Term	20
4.4.2 Created Value Domain Name	22
4.5 STEP 4: DATA ELEMENT NAME.....	23
4.6 LESSON 3 REVIEW.....	25
<b>5 LESSON 4: COMPOSING DEFINITIONS .....</b>	<b>25</b>
5.1 OBJECTIVES FOR THE LEARNER .....	25
5.2 PURPOSE OF DEFINITIONS .....	25
5.3 DEFINITION GUIDELINES.....	26
5.3.1 Definitions Created by the Curation Tool	26

5.3.2	Explanatory Comments	29
5.4	LESSON 4 REVIEW.....	30
<b>6</b>	<b>LESSON 5: REUSING ADMINISTERED ITEMS.....</b>	<b>31</b>
6.1	OBJECTIVES FOR THE LEARNER .....	31
6.2	CONSIDER THE DATA BY ANALYZING THE QUESTION .....	31
6.3	SPECIFIC COMPONENTS.....	32
6.3.1	Data Element Concept	32
6.3.2	Value Domain	32
6.4	LESSON 5 REVIEW.....	36
<b>7</b>	<b>LESSON 6: DETERMINING WORKFLOW.....</b>	<b>36</b>
7.1	OBJECTIVE FOR THE LEARNER .....	36
7.2	WORKFLOW STATUS.....	36
<b>8</b>	<b>LESSON 7: REGISTRATION STATUS.....</b>	<b>38</b>
8.1	OBJECTIVE FOR THE LEARNER .....	38
8.2	REGISTRATION STATUS.....	38
<b>9</b>	<b>LESSON 8: VERSIONING ADMINISTERED ITEMS.....</b>	<b>40</b>
9.1	OBJECTIVES FOR THE LEARNER .....	41
9.2	TYPES OF VERSIONING CHANGES .....	41
9.3	VERSIONING A DATA ELEMENT CONCEPT .....	41
9.3.1	Major Change	42
9.3.2	Minor Change	42
9.3.3	No Version Change	42
9.4	VERSIONING A VALUE DOMAIN.....	42
9.4.1	Major Change	43
9.4.2	Minor Change	44
9.4.3	No Version Change	44
9.5	VERSIONING A DATA ELEMENT .....	45
9.5.1	Major Change	45
9.5.2	Minor Change	45
9.5.3	No Version Change	46
9.5.4	Versioning Common to All Administered Items	46
<b>10</b>	<b>LESSON 9: DERIVED DATA ELEMENTS.....</b>	<b>47</b>

10.1	OBJECTIVES FOR THE LEARNER .....	47
10.2	WHAT ARE DERIVED DATA ELEMENTS? .....	47
10.3	DERIVATION RULES .....	47
10.4	NON-DERIVED vs. DERIVED DATA ELEMENTS .....	47
10.5	EXAMPLE OF A DERIVED DATA ELEMENT .....	48
10.6	LESSON 9 REVIEW.....	49
<b>11</b>	<b>DOCUMENTATION / RECOMMENDED READING MATERIALS.....</b>	<b>50</b>
<b>12</b>	<b>CONTACT INFORMATION.....</b>	<b>50</b>
<b>13</b>	<b>COURSE REVIEW QUESTIONS.....</b>	<b>51</b>

# Introduction

---

Welcome to caCORE Training. The session is designed for caDSR Users and Metadata Consumers, including cancer Biomedical Informatics Grid (caBIG) Developers and other participants.

In order to receive credit for the mastery of this content, you will need to register for this course and complete the accompanying quiz in the caBIG Learning Management System.

To register, go here: <http://ncicbtraining.nci.nih.gov/TP2005/tp2000web.dll/NCICBTraining>

We want these sessions to be as effective as possible in meeting your needs so we ask that you complete the short training evaluation form (available when you register for a course) to share your feedback on the overall quality of the training process and materials.

## 1 Course Details

---

Course Category: caCORE  
Course Number: 1040  
Course Title: Creating Well-formed Metadata and Metadata Business Rules  
Course Level: Basic  
Audience: Metadata Curators and Users, Developers, and Context Administrators

This course is divided into 9 Lessons:

- Lesson 1: Metadata Fundamentals
- Lesson 2: Naming Conventions
- Lesson 3: Creating Names
- Lesson 4: Composing Definitions
- Lesson 5: Re-using Administered Items
- Lesson 6: Determining Workflow
- Lesson 7: Registration Status
- Lesson 8: Versioning Administered Items
- Lesson 9: Derived Data Elements

Creating metadata in the caDSR that is well-formed and strictly follows the published business rules is very important. Well-formed metadata is a means of sharing data across and within domains in respect to cancer clinical trials.

## 2 Lesson 1: Metadata Fundamentals

---

In this lesson, we will review the ISO/IEC 11179 administered items, cover the anatomy of a Data Element, and illustrate how metadata is a library of re-usable items.

### 2.1 Objectives for the Learner

---

By the completion of this lesson, the trainee will be able to:

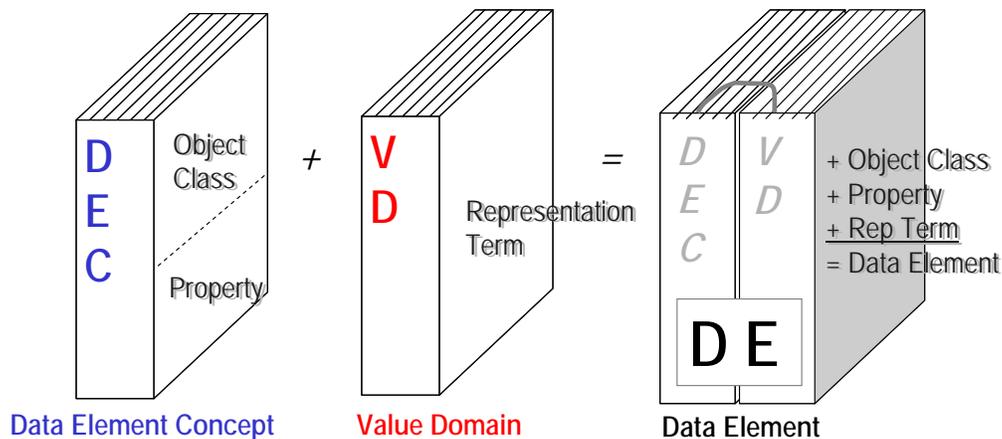
- Describe the purpose of a Data Element Concept (DEC)
- Describe the purpose of a Value Domain (VD)
- Describe the purpose of a Data Element (DE) or Common Data Element (CDE)
- Discuss how these Administered Items relate to one another

## 2.2 Review ISO/IEC 11179 Administered Items

ISO/IEC 11179 metadata standard defines an administered item as a registry item for which administrative information, such as start date, end date, etc., is recorded in an administration record. The administered items used to create Common Data Elements (CDEs) in the caDSR include:

- Data Element Concept (DEC) – A concept that can be represented in the form of a data element, described independently of any particular representation. A DEC is the “thing” in the real world you want to represent or the question asked on a case report form or other data collection instrument.
- Value Domain (VD) – A representation of the data collected, specified by a description or a list of valid values. A VD is how you want to represent or capture that “thing”.
- Data Element (DE or CDE) – A unit of data for which the definition, identification, representation and permissible values are specified by means of a set of attributes. A DE is the combination of (or relationship between) a DEC and a VD, or stated another way, the combination of a question asked and the legitimate responses.

Figure 1 is an illustration of the anatomy of a Data Element.



**Figure 1. Anatomy of a Data Element**

The required parts of a DEC are the Object Class and Property. The Object Class is a set of ideas, abstractions, or things in the real world that are identified with explicit boundaries and meaning and whose properties and behavior follow the same rules. The Property is a characteristic common to all members of the Object Class.

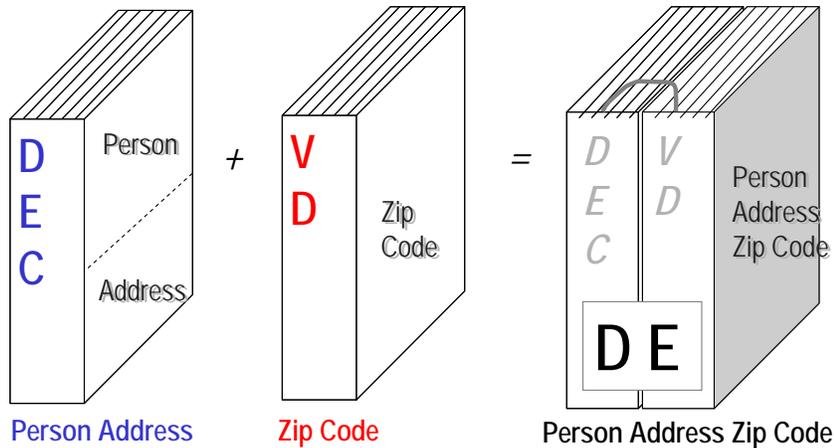
If there are specific descriptions of the metadata, then qualifiers are used to add specificity to the Object Class and/or Property. Qualifiers are not required items, but may be necessary to differentiate similar sub-items.

The required parts of the VD are the type, (enumerated or non-enumerated), Representation

Term and the Permissible Values (for an enumerated VD). The Representation Term is a term used to describe the form of the data. The caDSR has a list of acceptable Representation Terms and their definitions, to refer to when choosing how data is captured or represented on a form. Examples are Type, Name, Number, etc.

The DEC and VD (specifically the Object Class, Property, and Representation Term) together form a Data Element (DE or CDE).

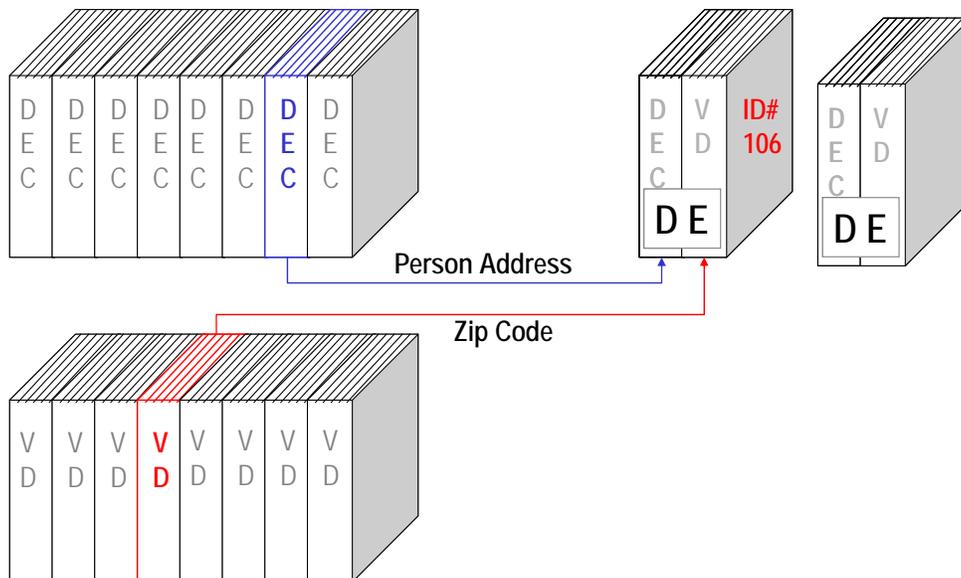
Figure 2 is an example of how a Data Element is formed.



**Figure 2. Data Element Example**

The Object Class (Person) and Property (Address) form the DEC (Person Address). The Representation Term (Zip Code) forms the Value Domain (Zip Code). Together the DEC and VD form the DE (Person Address Zip Code).

These metadata items form a library of re-usable items as shown in Figure 3 below.



**Figure 3. Metadata as Libraries of Re-Usable Items - Person Address**

The DEC of Person Address with the VD of Zip Code form the DE (ID# 106) called Person Address Zip Code. The same DEC (Person Address) can be used with a different VD to form a new DE as shown in Figure 4 below.

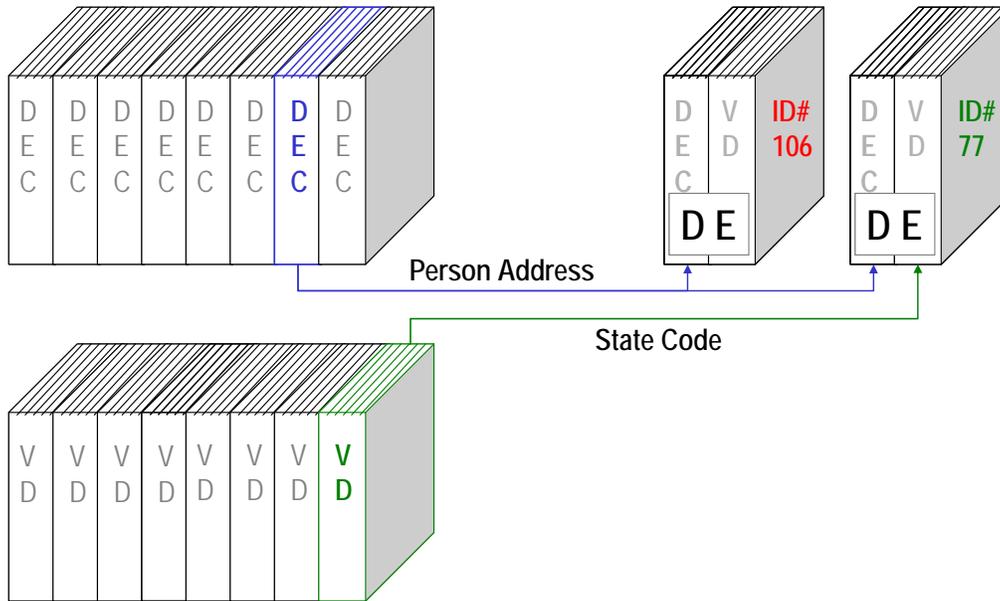


Figure 4. Metadata as Libraries of Re-Usable Items

By combining the same DEC (Person Address) with another VD (State Code) form a new DE (ID #77) of Person Address State Code.

## 2.3 Lesson 1 Review

In lesson 1, you learned that:

- A Data Element Concept (DEC) is idea concept that can be represented in the form of a data element, described independently of any particular representation. In short, a DEC is a thing in the real world you want to represent.
- A Value Domain (VD) is a set of attributes describing representation characteristics of data with or without permissible values. In short, a VD is how you want to describe or capture a thing in the real world.
- A Data Element (DE or CDE) is a unit of data for which the definition, representation and permissible values are specified by means of a set of attributes. In short, a DE/CDE is a combination of (or relationship between) a DEC and a VD.

# 3 Lesson 2: Naming Conventions

## 3.1 Objectives for the Learner

Upon completing this learning lesson, you will be able to:

- Identify six things to consider when developing a context naming convention
- Identify different types of vocabularies used for naming Data Elements
- Identify why a thesaurus is a good source for naming terms

- List the basics for a Data Element name
- Describe the rules for a Data Element long name
- Identify three types of Data Element short names and rules for their creation

## 3.2 Development of a Naming Convention

---

One of the basic tenets in the naming of Data Elements is to be consistent, within a context and across caDSR, in creating names that are understandable to users of the registry.

The ISO standard lists these six things to consider when creating a Data Element registry naming convention (ISO 11179-5:2005(E)):

- The scope of the naming convention, e.g. established industry name;
- The authority that establishes names;
- Semantic rules governing the source and content of the terms used in a name, e.g. terms derived from data models, terms commonly used in the discipline, etc.;
- Syntactic rules covering required term order;
- Lexical rules covering controlled term lists, name length, character set, language;
- A rule establishing whether or not names must be unique

In considering these guidelines, the following should be decided by the context considering registering metadata in the caDSR:

- Establish a scope for the convention. The scope of your context will determine the scope of your naming convention. Consider the terms you will use to name CDEs.
- Determine the authority that establishes the name. Identify who will have the authority to approve the names of the CDEs in your context. Will it be the Principal Investigators (PIs), curators, or the Clinical Coordinators? Teams should work together to determine where the terms will come from to create CDE names, whether word lists will be used, and for the final approval of names.
- Develop semantic rules for the source and content of words used in a name. What terms will be allowed as Object Classes, Properties and Representations?
- Formulate syntax rules for required word order. Use best practice rules to determine the structure of CDE names. If qualifiers are used, how many are allowed and in what order will they appear?
- Develop lexical rules covering controlled word lists, name length, character set, and language. Set additional rules to cover the vocabulary, words and definitions to be used by the context.
- Set guidelines on uniqueness of names in your context. Determine how names will be created to make them unique. Will specificity be added to DEC's, or VDs? How will the two be combined to create unique names?

Naming convention guidelines should be flexible enough to accommodate additions to the caDSR, but they also need give enough structure so that continual name changes are not necessary. This means that with thoughtful composition, you will be able to reuse your context's Data Elements, and will also offer Data Elements to be used by other contexts.

The framework for a naming convention is developed through discussion, consensus building, and the application of best practices/lessons learned. It will be difficult to develop a naming convention without experience with curation, training and most importantly, practice.

It may be helpful to look at various contexts' established conventions, business rules and word lists. Context Administrators can be contacted by referencing the caDSR Content Creation home page located at:

[http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore\\_overview/cadsr/curation](http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore_overview/cadsr/curation).

The caDSR implementation of the ISO/IEC 11179 model is the focus of the Content meetings of the bi-weekly NCI Context Administrators. Attendance at the meetings can also be an important source for information when developing naming conventions.

### 3.3 Types of Vocabulary Lists

---

The terms and concepts to use as naming components can come from a variety of sources including the following:

- **Vocabulary**: All the words of a language; the sum of words used by, understood by, or at the command of a particular person or group; a supply of terms and/or phrases.
- **Lexicon**: a stock of terms used in a particular profession, subject, or style; A word book describing language with definitions.
- **Ontology**: a set of representational terms. Definitions associate the names of entities in a logic grouping (e.g. classes, relations, functions or other objects) with human-readable text describing what the names mean and formal axioms that constrain the interpretation and well-formed use of these terms. An example is animal taxonomy.
- **Axiom**: an established rule, principle, or law.
- **Terminology**: the vocabulary of technical terms used in a particular field, subject, science, or art; nomenclature.
- **Code Sets**: a select list of terminology.

caDSR administered items are backed by the use of externally defined terminologies and controlled vocabularies. NCICB bases its data semantics on controlled terminology supplied by the NCI Enterprise Vocabulary Services (EVS) Project [<http://evs.nci.nih.gov/>]. EVS is a set of services and resources that address NCI's needs for controlled vocabulary. Within the NCICB, EVS provides the semantics used to create caDSR metadata.

The NCI EVS produces two terminology products: the NCI Thesaurus (NCIt), a cancer focused terminology, and the NCI Metathesaurus (Meta), a mapping between many terminologies. The NCI Thesaurus is the preferred terminology for creating caDSR metadata. caDSR tools allow users to create CDE names by selecting terms from NCI Thesaurus.

The benefits of using the NCI Thesaurus when naming CDE's include:

- Nonproprietary source of naming components
- Editable and maintained in real time
- Provides semantic linking of preferred terms
- Gives guidance in using homographs (one of two or more words that have the same spelling but differ in origin, meaning, and sometimes pronunciation)
- Shows equivalence, hierarchy, and association

Now that we have discussed the sources for words, terms and/or concepts used in naming CDEs, let's look at the ways a Data Element name can be represented in the caDSR.

## 3.4 Different Types of Data Element Names

---

An administered item's name can be defined as the word or combination of words by which a something is known. In the caDSR, there are three types of names: long name, short name and alternate name. We will focus on the long name and short name only in this session.

- The long name is a 255 character (maximum) fully annotated name describing an administered item.
- The short name is a 30 character (maximum) abbreviated name for an administered item. The short name may be generated by the database system, abbreviated by the system, or entered by a user.
  - NOTE: If using Oracle Clinical (OC) as the application to collect data, the short name will become the question name in the OC Table (global library). In order to allow OC users the flexibility for direct use of data elements, best practice is to enter a user entered short name with no more than 20 characters. This practice has been adopted across all contexts for caDSR data element short names.
- The alternate name is another meaningful name that identifies an administered item. Examples include: Dicom tags, SAS column names, and UML model name.
  - NOTE: If you load a UML Model into the caDSR, there will be two CDE alternate names. The first will be the name of the model package with the class and attribute (package.class.attribute). The second will be the class with each attribute (Class:Attribute). The long name of the Data Element will be the class (with any qualifiers preceding), the attribute (with qualifiers), and the name of the generic java data type (Class Property java.lang.xxxx), unless value domains are specified within the model. Questions regarding UML-based CDEs will be specifically addressed in the 1070 series of training classes, Creating Metadata from UML Models.

### 3.4.1 Long Name Rules

The long name is a readable and descriptive phrase describing an administered item. The long name is mixed case and the first letter of a major term is capitalized. There is a caDSR database maximum of 255 characters. Terms are separated with spaces. While you want the long name to be unique in your context, they need to be easily searchable as well.

The long name is composed of:

- Object Class
- Property
- Representation
- Qualifiers

In most cases the long name will be typed out in its entirety. If it is determined that an abbreviation is needed, the context should agree on the abbreviation while adopting existing abbreviations from other contexts whenever possible.

### 3.4.2 Short Name Rules

The short name is an abbreviated representation of the long name. There are three ways to create a short name: system generated by public IDs and versions, abbreviated system generated, and user entered.

If selecting a system generated short name, the caDSR Curation Tool will automatically concatenate the public ID and version of the DEC and VD separated by a colon, creating an alphanumeric short name, for example: 2145678v1.0:2356987v3.0. However, these names are not fully descriptive, and thus do not facilitate targeted searches for reuse of items.

If an alphanumeric short name is not desirable to your context, you may select the Abbreviated System Generated short name. The Curation Tool will take each term in the long name and truncate to 4 characters, mixed case, separated by underscores. An example of 'Clinical Stage Disease Text Name' would become 'Clin\_Stag\_Dise\_Text\_Name'.

If you are using an Oracle Clinical application, or if you would like to offer a descriptive short name for a Data Element, select 'User Entered' for the short name. Enter the standard abbreviation for each term from the long name in capital letters, separated by an underscore. The maximum character length for the short name is 20. If a standard abbreviation doesn't exist, create a new abbreviation by truncating the term to the first four letters of the long name term or an appropriate short term that has meaning (such as 'PT' for patient, or 'HX' for history). The caDSR Content Group maintains a list of agreed upon abbreviations for commonly used terms.

## 3.5 Lesson 2 Review

---

In Lesson 2, you have learned:

- Six things to consider when developing a context naming convention:
  - Establish a scope
  - Determine the authority that establishes a name
  - Develop semantic rules for the source and content
  - Formulate syntax rules
  - Develop lexical rules
  - Set guidelines on uniqueness
- Different types of vocabularies that can be used for naming Data Elements:
  - Vocabulary
  - Lexicon
  - Ontology of Program
  - Axiom
  - Terminology
  - Code Sets
- Why NCI Thesaurus is a good source for naming terms:
  - Source of name components
  - Provides semantic linking of preferred terms
  - Gives guidance in using homonyms and similar terms
  - Shows equivalence, hierarchy, and association
  - Allows use of a controlled vocabulary

- Three different Data Element names:
  - Long name
  - Short name
  - Alternate name
- The rules for creation of a Data Element Long Name:
  - Mixed Case - first letter caps
  - 255 characters maximum
  - Terms separated by a single space
  - Must have an Object Class, Property and Representation to form the DEC and VD
- Three types of Data Element short names and rules for their creation:
  - System Generated: alphanumeric DEC+VD Public IDs and versions separated by a colon.
  - Abbreviated System Generated: truncation (4 characters) of terms from the Long Name (30 character maximum, mixed case separated by \_).
  - User Entered: abbreviations from standard list for each term in the long name (all caps, separated by “\_”, 20 character maximum).

We’ve covered the Data Element components and discussed naming conventions and vocabularies; now let’s explore the process for creating names for DEC, VD, and CDEs.

## 4 Lesson 3: Creating Names

---

### 4.1 Objectives for the Learner

---

By the completion of this lesson, you will be able to:

- Summarize the process for creating Data Element Concept, Value Domain and Data Element names
- List the components needed for Data Element Concept, Value Domain and Data Element names
- Identify the caDSR tool used to create DEC, VD and CDE names
- Name a source of component terms
- Identify rules used in creating administered item short names

### 4.2 Step 1: Considering Components of the Data

---

Ask yourself: “What information do I want to capture to describe the data collected?” “If I wanted to look in the database for data, what words would I use for the most efficient search?”

For example, maybe we want to collect the race of participants in a protocol. We have a list of possible responses to the question, “What is your race?” Possible responses could include: White, Black, Asian, and Not Reported.

Consider the words you might need to find the metadata in the database, or in the caDSR.

These will be the keywords of the names created.

## 4.3 Step 2: Data Element Concept Name

---

To create a DEC name, you will need:

- Object Class – What is the concept or ‘thing’ described in the data being captured?
- Property – What is the characteristic of the Object Class that makes it identifiable?
- Qualifiers – Does the Object Class or Property need additional description?

Think back to grammar classes and remember the parts of a sentence, the object, verbs, and adjectives. Questions can often be analyzed in this way. The object of the question is often the Object Class of the Data Element Concept. The question verbs are similar to the Properties of the DEC and the adjectives can be thought of as DEC Qualifiers.

Consider the question, “What is your race?”

- What is the Object Class?
  - “Person” (the ‘thing’ or entity being described)
- What is the Property?
  - “Race” (the characteristic of the Object Class or ‘person’ being described)
- Do you need Qualifiers?
  - “No” (the Object Class and Property provide all the description needed)

The terms for the components are chosen from the NCI Thesaurus.

### 4.3.1 Finding the Object Class

Now that we know that we are looking for a term of “Person”, we will search the NCI Thesaurus for a concept for the Object Class. The CDE Curation Tool provides a link to search the NCI Thesaurus and will provide the mechanism to find an Object Class term.

The CDE Curation Tool is a web-based tool which requires a caDSR username and password. You can request a caDSR Training Account from NCICB Application support at [ncicb@pop.nci.nih.gov](mailto:ncicb@pop.nci.nih.gov). In the meantime, a set of training accounts have been set up for those enrolled in the caCORE training courses. These training accounts (username/password) are: Trainee01/Trainee01 through Trainee10/Trainee10. To access the Curation Tool, navigate to <http://cdecurate.nci.nih.gov>.

In this lesson, we will use the Curation Tool as a reference to cover selected concepts. A more in-depth demonstration is in the Course 1060 training sessions. For the purpose of this lesson, the CDE Curation Tool screen shots illustrate the links to the EVS vocabulary system via the search option.

The terms for the DEC name components are selected from the NCI Thesaurus. The search for terms to create a DE/CDE, DEC, and/or VD is a targeted one – the curator must decide on the best terminology to use in the creation of a new component before using any of the caDSR tools. The Curation Tool links to EVS for all naming components. Figure 5 below illustrates a search for an Object Class using the Curation Tool.

**Create New Data Element Concept**

\* Indicates Required Field

\* 1) Select Context  
caBIG

2) Select Data Element Concept Name Components

Object Class Long Name:

Property Long Name:

Qualifier Concepts:  [Search](#) [Remove](#)

Primary Concept:  [Search](#) [Remove](#)

Qualifier Concepts:  [Search](#) [Remove](#)

Primary Concept:  [Search](#) [Remove](#)

Figure 5. Using the Curation Tool to Find an Object Class

Click the search link for the Object Class Primary Concept. Figure 6 represents the Object Class search pane.

1) Search For: Object Class

2) Select EVS Vocabulary: NCI Thesaurus

3) Search In: caDSR

4) Enter Search Term: Person

5) Filter Search By: caDSR Owned By/Used By: All Contexts, Workflow Status: RELEASED

6) Display Attributes: Update

Search Results for Object Class - Person

Concept Name	Public ID	EVS Identifier	Definition	Definition Source	Workflow Status	Semantic Type	Context Vocabulary	caDSR Component Using	DEC's Using
<input checked="" type="checkbox"/> Person	2236731	C25190	A human being.	NCI	RELEASED		caBIG	caDSR	Object Class 163
<input type="checkbox"/> Person		C25190	A single human being.	NCI	Active	Human		NCI Thesaurus	
<input type="checkbox"/> Person Info		C41171	Person's first and last name and the connection to the party.	NCI	Active	Conceptual Entity		NCI Thesaurus	
<input type="checkbox"/> Person Observer		C69245	Human observer created the observations.	DICOM	Active	Intellectual Product		NCI Thesaurus	

How do I choose?

- Consider the Definition Source (NCI Thesaurus), the Workflow Status, and how others have used the term in the caDSR (DECs Using)

Figure 6. Object Class Search Pane

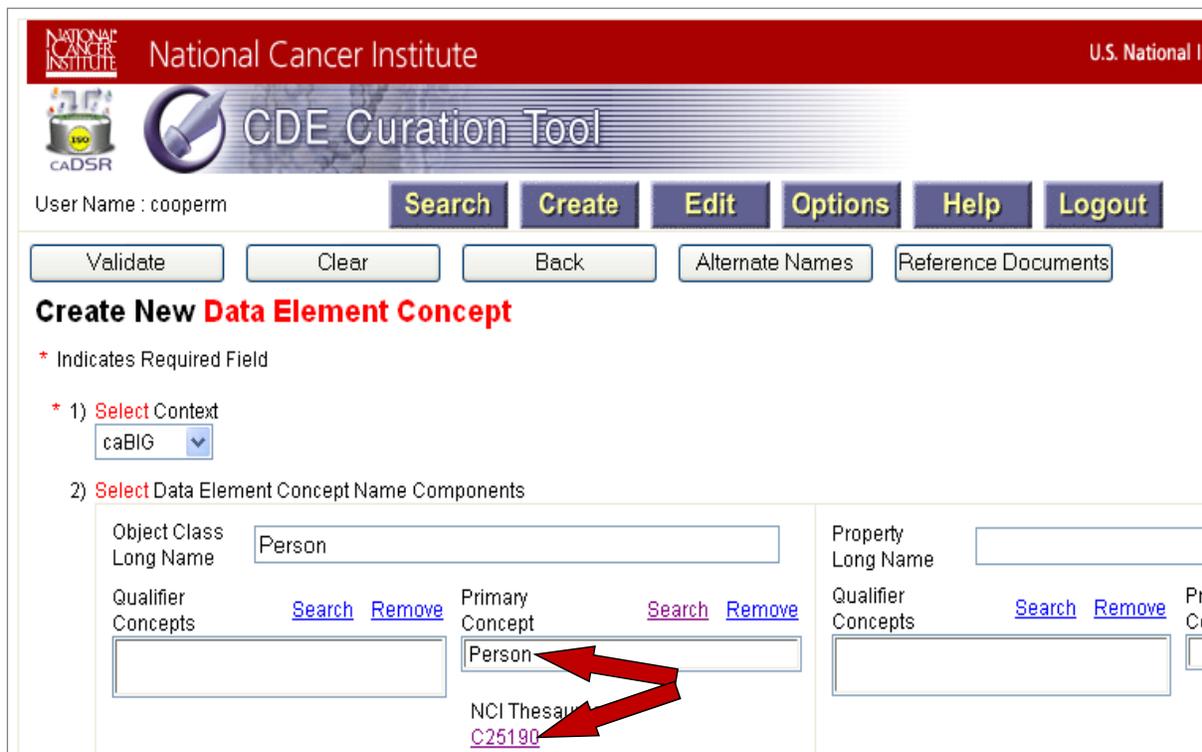
Figure 6 above represents a sample results set for a search for Object Class of “Person”. The Object Class term is entered in the search field. In our example, we are looking for “Person”. The Curation Tool will search EVS for terms to use as an Object Class. You need to select the correct term for your metadata and the context.

How do you choose? Consider the Definition Source (NCI Thesaurus), the workflow status (RELEASED), and how others have used the term in the caDSR (DEC’s Using). A best practice principle for caDSR curation is to choose terms from the NCI Thesaurus.

If a desired term is not in EVS, you can suggest a term and definition by using the ‘Suggest to EVS’ button at the top right of the screen as shown in Figure 6. When this happens, you will be notified when the term has been created in the NCI Thesaurus and uploaded into the caDSR. This takes about one business day.

According to the best practice rules, we have selected the first result for Person and clicked the Use Selection button.

Figure 7 illustrates the Create DEC screen with the Object Class field populated with the selected term.



The screenshot shows the 'Create New Data Element Concept' interface. At the top, it says 'National Cancer Institute' and 'U.S. National Institutes of Health'. Below that is the 'CDE Curation Tool' header. The user is logged in as 'cooperm'. There are navigation buttons: Search, Create, Edit, Options, Help, Logout. Below these are buttons for Validate, Clear, Back, Alternate Names, and Reference Documents. The main section is titled 'Create New Data Element Concept'. It has a legend: '\* Indicates Required Field'. Step 1) 'Select Context' shows a dropdown menu with 'caBIG' selected. Step 2) 'Select Data Element Concept Name Components' has two columns. The left column has 'Object Class Long Name' with the value 'Person'. Below it is 'Qualifier Concepts' with a search and remove button. The right column has 'Primary Concept' with the value 'Person' and a search and remove button. Below it is 'Qualifier Concepts' with a search and remove button. A red arrow points from the 'Person' in the Primary Concept field to the 'Person' in the Object Class field. Another red arrow points from the 'Person' in the Primary Concept field to the 'NCI Thesaurus C25190' link below it.

**Figure 7. Populated Object Class**

The Object Class is pasted in the Concept field along with the unique identifier (Concept Unique Identifier or ‘CUI’ pronounced ku-ee).

Question: What is the next thing we need to create a DEC name?  
 Answer: Property

### 4.3.2 Finding the Property

The same steps for finding an Object Class pertain to the Property. The search link is used to search for a concept to use as the Property.

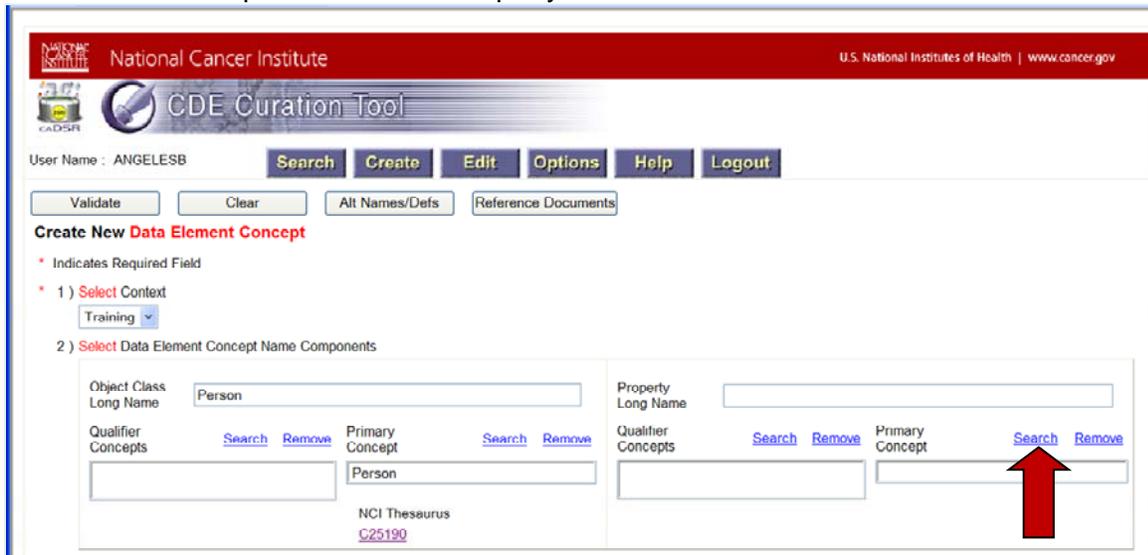


Figure 8. Using the Curation Tool to find a Property

The Property term needed is entered in the search field. In our example, we are looking for “Race”. The Curation Tool will search EVS for terms to use as a Property. You need to select the correct term for your metadata and the context. Figure 9 represents a sample search and results set for a Property of “Race”.

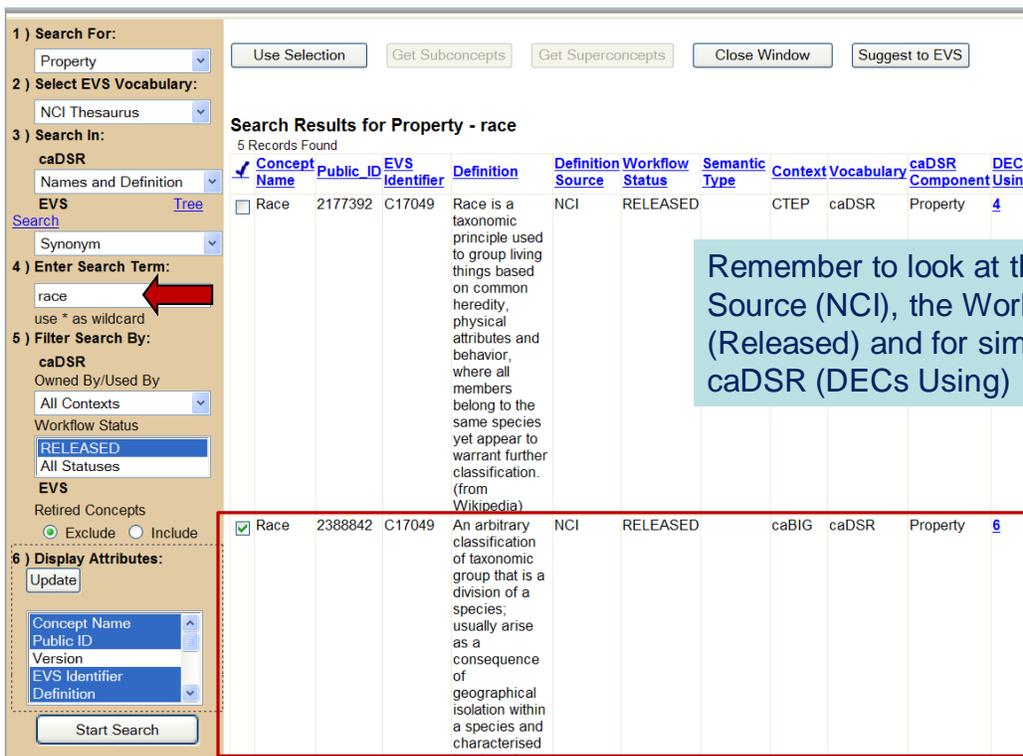
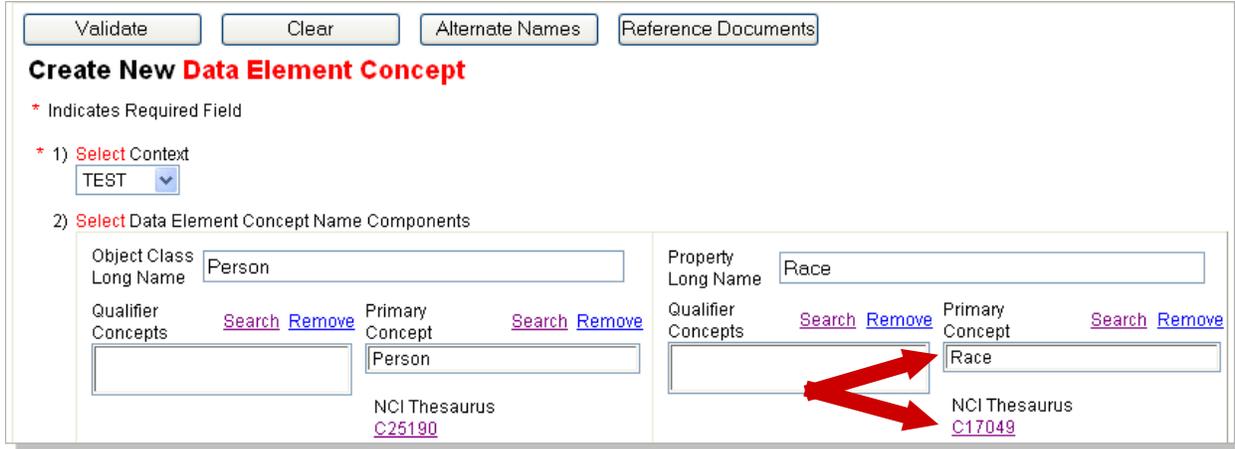


Figure 9. Property Search Pane

Select a desired term from the results set. Remember to look at the Definition Source (best practice is to use NCI), the Workflow Status (RELEASED) and for similar use in the caDSR (DECs Using). If a desired term is not found, click the ‘Suggest to EVS’ button. Since the 2<sup>nd</sup> result of “Race” is used in more DECs, we have selected that result and clicked the Use Selection button.

Figure 10 illustrates the Create DEC screen with the Property field populated with the selected term along with the unique identifier (CUI).



**Create New Data Element Concept**

\* Indicates Required Field

\* 1) Select Context  
TEST

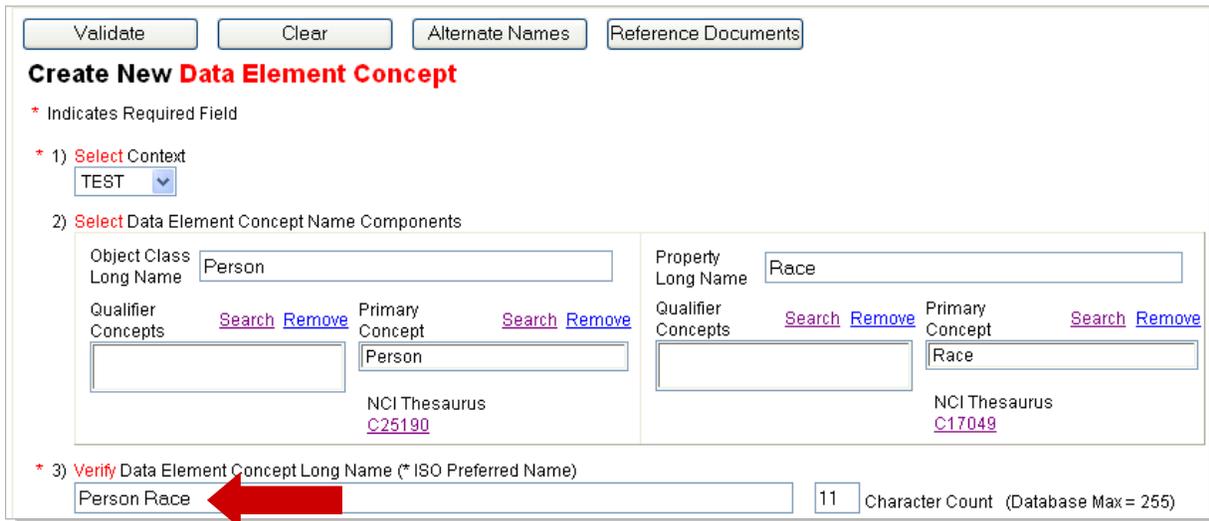
2) Select Data Element Concept Name Components

Object Class Long Name: Person	Property Long Name: Race
Qualifier Concepts: [Empty]	Qualifier Concepts: [Empty]
Primary Concept: Person NCI Thesaurus C25190	Primary Concept: Race NCI Thesaurus C17049

Figure 10. Populated Property

### 4.3.3 Created Data Element Concept Name

The Curation Tool concatenates the Object Class and Property terms together to form a DEC Long Name as shown in Figure 11.



**Create New Data Element Concept**

\* Indicates Required Field

\* 1) Select Context  
TEST

2) Select Data Element Concept Name Components

Object Class Long Name: Person	Property Long Name: Race
Qualifier Concepts: [Empty]	Qualifier Concepts: [Empty]
Primary Concept: Person NCI Thesaurus C25190	Primary Concept: Race NCI Thesaurus C17049

\* 3) Verify Data Element Concept Long Name (\* ISO Preferred Name)  
Person Race 11 Character Count (Database Max = 255)

Figure 11. DEC Long Name

NOTE: The Curation Tool confirms that both of these concepts, the object and property, are selected from the NCI Thesaurus, a basic best practice principle for metadata curation.

The DEC short name is created by selecting the ‘System Generated’, ‘Abbreviated’ or ‘User Entered’ radio buttons as shown in Figure 12 below.

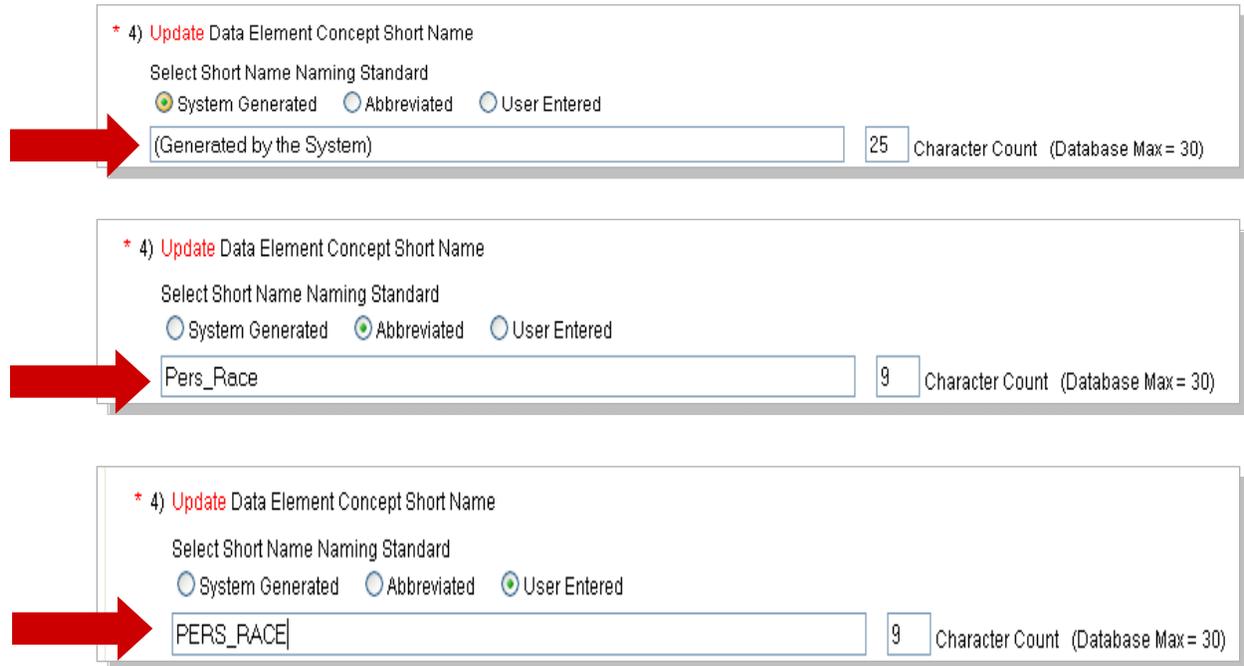


Figure 12. DEC Short Name

## 4.4 Step 3: Value Domain Name

Now that we have formed the DEC name, we need to repeat this process to name the Value Domain.

To create a VD name, you will need:

- Representation – What is the form of the data being collected? This attribute is different from its ‘data type’. A data type of ‘text’ or ‘string’ may include representations that are criteria, categories, or other characteristics that further describe the format of the data being collected.
- Qualifiers – Does the Representation term need additional description?

The semantics of the VD reside entirely within the representation term and qualifiers. Best practice is to build the name of the Value Domain using the Representation Term and Qualifiers exclusively.

Figure 13 below represents a sample of the caDSR Standard Representation Term list that can be found here: <https://wiki.nci.nih.gov/x/qgZy>

caDSR Public ID	Thesaurus Concept Code	Representation Long Name	Rep term definition
2519596	C13717	Anatomic Site	Anatomic site:Named locations of or within the body.
2433877	C25372	Category	This term is used informally to mean a class of things.
2261841	C25162	Code	A system, often of numbered categories, for representation of data.
2261845	C25463	Count	To determine the number or amount of something; the result of this activity; Count; the total number counted.
2233081	C25164	Date	The particular day, month and year an event has happened or will happen.
2423394	C37939	Date/Time	An expression of both date and time that an event has happened or will happened.
2432127	C25488	Dose	The total quantity or strength of a substance administered at one time; The amount of medicine taken, or radiation given, at one time.
2404653	C25330	Duration	The period of time during which something continues.
2433735	C48150	Float	A number that can have its decimal point in any position.
2443080	C25515	Frequency	The number of occurrences of something within a given time period.
2453188	C48309	Grade	A position on a scale of intensity or amount or quality.
2405190	C25364	Identifier	One or more characters used to identify, name, or characterize the nature, properties, or contents of a thing.
2421125	C38147	Ind-2	A response or indicator that can have a value of either yes or no.
2548260	C38148	Ind-3	A response or indicator that can have a value of yes, no, or unknown.
2694966	C49797	Ind-3b	A response or indicator that can have a value of either Yes, No, or Not Applicable.
2694968	C49486	Ind-4	A response or indicator that can have a value of Yes, No, Unknown, or Not Applicable.
2320047	C25180	Indicator	An event, entity or condition that typically characterizes a prescribed environment or situation and determines or aids in determining whether or not certain stated circumstances exist or criteria are satisfied.
2433737	C45255	Integer	A number with no fractional part.
2525196	C25543	Interval	The period of time or the distance separating two instances, events, or occurrences
2184312	C25209	Measurement	The determination of the size or magnitude of something. It is also used to refer to the result of this act.
2229718	C42614	Name	The words or language units by which a thing is known.
2233101	C25337	Number	Number; a concept of quantity derived from zero and units; a numeral or string of numerals that is used for identification; A numeral or string of numerals expressing value, quantity, or identification.
2433750	C38013	Range	The defined lowest and highest limits for a set of numeric data.
2405456	C25636	Rate	A measurement of degree, speed, or frequency relative to time.
2240632	C25638	Reason	An explanation of the cause of some phenomenon or action.
2476191	C20200	Result	A phenomenon that follows and is caused by some previous phenomenon.
2452834	C25664	Scale	An ordered reference standard used to measure incremental changes.
2193207	C25338	Score	A number or range of numeric values measuring performance, function, quality, or ability.
2529667	C25683	Source	Where something is available or from where it originates.
2390957	C25685	Specify	Be specific about something; define clearly.
2428501	C16899	Stage	The extent of a cancer within the body, especially whether the disease has spread from the original site to other parts of the body.
2404650	C25688	Status	A condition or state at a particular time.
2234203	C25704	Text	The words of something written.
2193249	C25207	Time	The continuum of experience in which events pass from the future through the present to the past; a specific point in this continuum or the interval separating two points.
2234813	C25284	Type	Something distinguishable as an identifiable class based on common qualities; Type; a subdivision of a particular kind of thing..
2233053	C25709	Unit of Measure	A named quantity in terms of which other quantities are measured or specified, used as a standard measurement of same kinds.
2414933	C25712	Value	A numerical quantity measured or assigned or computed.

**Figure 13. Standard Representation Term List**

Considering the question, “What is your race?” What would best represent the form of the values? Our sample question includes a list of possible responses which could include: White, Black, Asian, and Not Reported.

Of all the terms from Figure 13, Category seems to be the best fit for the Representation Term for our example.

Do we need Qualifiers for the Representation Term? - Yes, we will need qualifiers for Person and Race.

#### 4.4.1 Selecting the Representation Term

We use the Curation Tool to build the Value Domain Long Name by using the search to link to the EVS vocabulary. Figure 14 below shows the Create New Value Domain pane and where to search for a Representation Term.

Validate Clear Alt Names/Defs Reference Documents

**Create New Value Domain - Person Race [v1.0]**

\* Indicates Required Field

**Value Domain Details** | **Permissible Values**

1) Select Context  
Training

2) Select Value Domain Type  
Enumerated

3) Select Value Domain Representation

Value Domain Attributes

Rep Term Long Name  
Person Race

Qualifier Concepts [Search](#) [Remove](#) Primary Concept [Search](#) [Remove](#)

Person Race  
NCI Thesaurus  
C25190

4) Verify Value Domain Long Name (\* ISO Preferred Name)  
Person Race 11 Character Count (Database Max = 255)

Figure 14. Using the Curation Tool to Search EVS for a Representation Term

Figure 15 illustrates the Representation Term search pane.

1) Search For: Rep Term

2) Select EVS Vocabulary: NCI Thesaurus

3) Search In: caDSR Names and Definition EVS Tree Search Synonym

4) Enter Search Term:

5) Filter Search By: caDSR Owned By/Used By caBIG Workflow Status RELEASED All Statuses EVS Retired Concepts Exclude Include

6) Display Attributes: Update  
Concept Name Public ID Version EVS Identifier Definition

Start Search

Use Selection Get Subconcepts Get Superconcepts Close Window Suggest to EVS

**Preferred List of Rep Term Primary Concepts**  
37 Records Found

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Concept Name	Definition	Definition Source
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Anatomic Site	Named locations of, or within, the body.	NCI
<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Category	Category; used informally to mean a class of things.	NCI
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Code	A system of numbered categories for representation of data.	Source => Name: NCI, NCI
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Count	To determine the number or amount of something; the result of this activity.	NCI
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Date	a particular day specified as the time something has, or will, happen.	NCI
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Date and Time	An expression of both date and time that an event has happened or will happen. (NCI)	NCI
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Dose	The amount of medicine taken, or radiation given, at one time.	NCI-GLOSS
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Duration	The period of time during which something continues.	NCI
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Float	A number that can have its decimal point in any position.	NCI
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Frequency		NCI
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Grade		NCI
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Identifier		NCI
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Ind-3		NCI
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Ind-3b		NCI
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Ind-4		NCI
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Indicator	An event, entity or condition that typically characterizes a prescribed environment or situation and determines or aids in determining whether certain stated circumstances exist or criteria are satisfied.	Source => Name: NCI, NCI

The search field is automatically populated with the NCI-Approved EVS concepts for the Representation Term.

Select the desired result and click the Use Selection button.

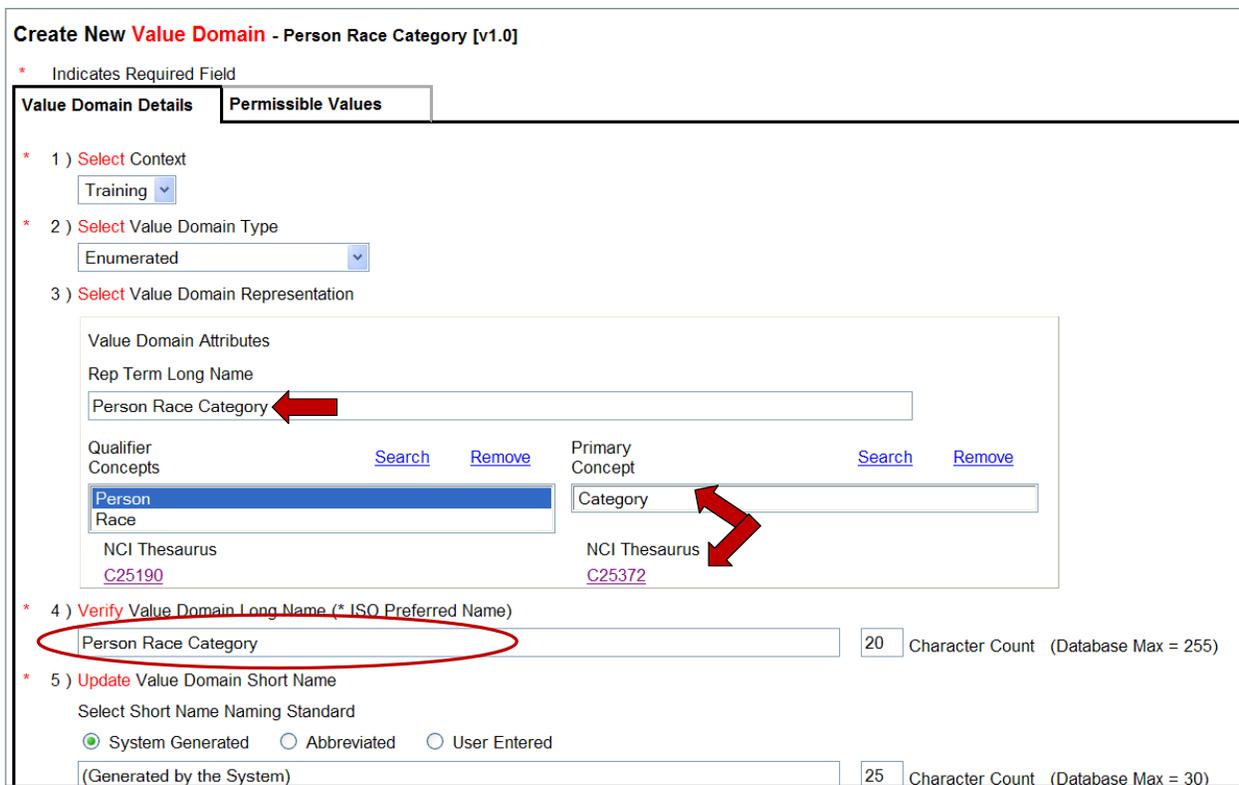
Figure 15. Representation Term Search Pane

Notice that that the Curation Tool automatically displays the set of 37 approved concepts to be use as Representation Term primary concepts. If the desired term is not displayed, submit the term to EVS using the 'Suggest to EVS' button.

Of the 37 listed, the term "Category" fits our needs for our Representation Term. Select the Category term and click the Use Selection button.

Figure 16 illustrates the Create VD screen with the Representation Term field populated with the selected term.

### 4.4.2 Created Value Domain Name



**Create New Value Domain - Person Race Category [v1.0]**

\* Indicates Required Field

**Value Domain Details** | **Permissible Values**

\* 1) Select Context  
 Training

\* 2) Select Value Domain Type  
 Enumerated

3) Select Value Domain Representation

Value Domain Attributes

Rep Term Long Name  
 Person Race Category

Qualifier Concepts	Search	Remove	Primary Concept	Search	Remove
Person Race NCI Thesaurus C25190			Category NCI Thesaurus C25372		

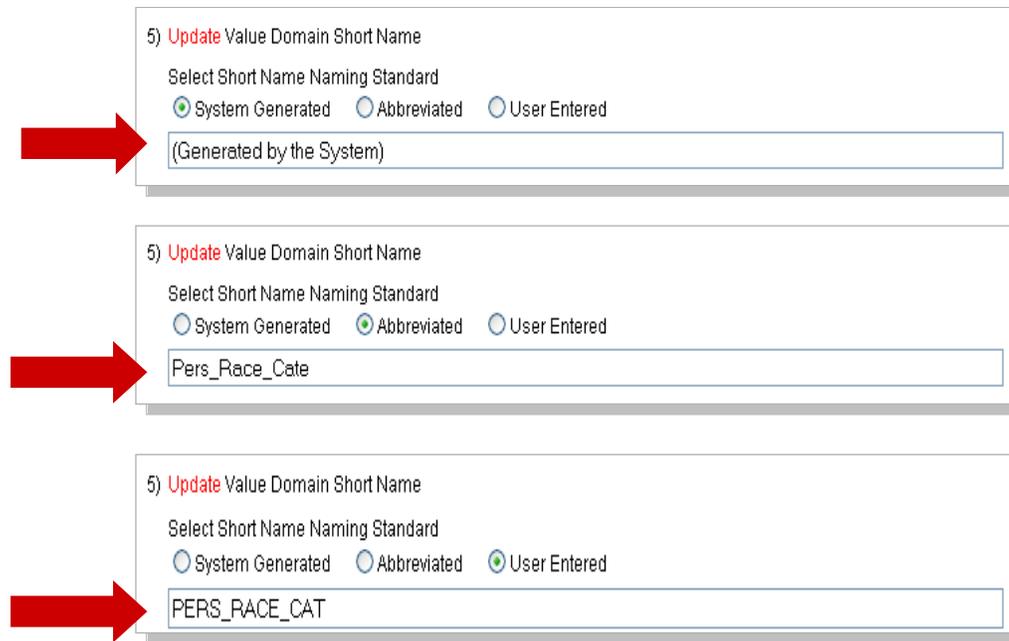
\* 4) Verify Value Domain Long Name (\* ISO Preferred Name)  
 Person Race Category 20 Character Count (Database Max = 255)

\* 5) Update Value Domain Short Name  
 Select Short Name Naming Standard  
 System Generated  Abbreviated  User Entered  
 (Generated by the System) 25 Character Count (Database Max = 30)

**Figure 16. Populated Representation Term and VD Long Name**

The term (Category) and its CUI (C25372) are populated in the Value Domain screen. Qualifier concepts are added for the creation of a well-formed Value Domain name. The Value Domain Long Name of Person Race Category is created.

The Short Name is created by selecting the 'System Generated', 'Abbreviated' or 'User Entered' radio buttons as show in Figure 17 below.



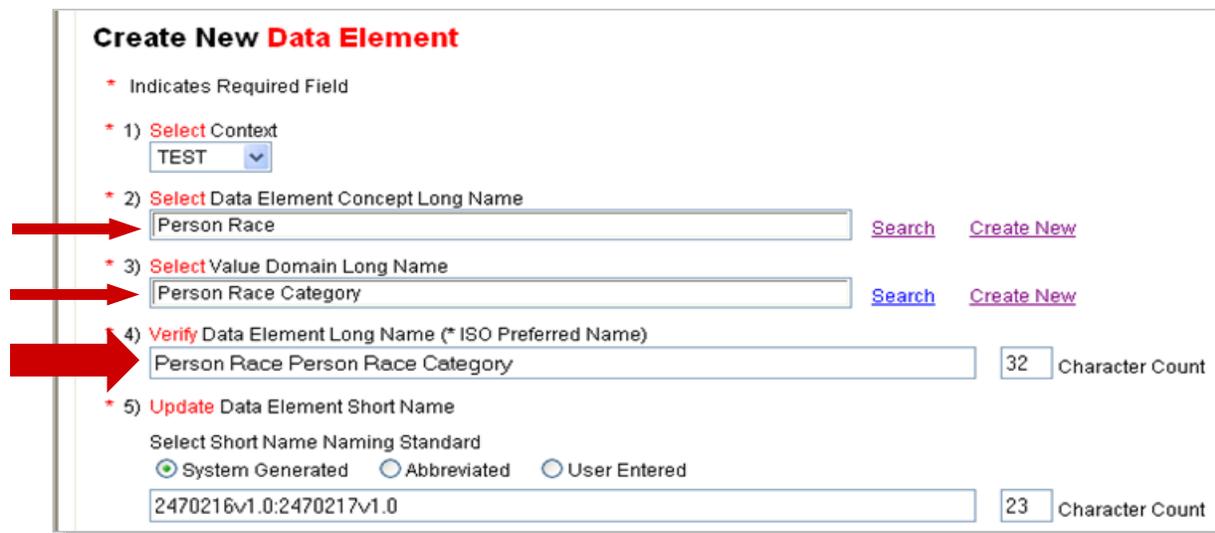
**Figure 17. Value Domain Short Name**

It is recommended to keep the default of System-generated, but not required. You may select any of the 3 options for the Short Name.

## 4.5 Step 4: Data Element Name

After creating the DEC and VD names, the next step is to compose the name of the Data Element. To do this, we need the names of the DEC and VD.

When the DEC and the VD have been created, the name of the CDE will be populated in the Long Name field as shown in Figure 18 below.



**Figure 18. Creating the Data Element Name**

The terms for the DEC and VD components from the NCI Thesaurus are used. Notice the short name for the CDE has the system generated public IDs and versions in the appropriate field.

Do you see anything wrong with the long name of the CDE in Figure 18?

There are redundant terms (Person and Race). The long name will need to be edited to remove the second set of Person Race terms. Figure 19 shows the edited long name.

\* 2) **Select** Data Element Concept Long Name  
 [Search](#) [Create New](#)

\* 3) **Select** Value Domain Long Name  
 [Search](#) [Create New](#)

\* 4) **Verify** Data Element Long Name (\* ISO Preferred Name)  
  Character Count (Database Max = 255)

**Figure 19. Editing the Data Element Long Name**

The long name has been edited to remove the second set of Person Race terms.

NOTE: If the DEC and VD short names are system generated, the abbreviated CDE short name will try to reflect the short names of the components. If you want to create an abbreviated short name, you must choose 'User Entered' and enter the abbreviations from the standard list. The edited user-entered short name is shown in Figure 20.

**Create New Data Element**

\* Indicates Required Field

\* 1) **Select** Context

\* 2) **Select** Data Element Concept Long Name  
 [Search](#) [Create New](#)

\* 3) **Select** Value Domain Long Name  
 [Search](#) [Create New](#)

\* 4) **Verify** Data Element Long Name (\* ISO Preferred Name)  
  Character Count

\* 5) **Update** Data Element Short Name  
 Select Short Name Naming Standard  
 System Generated  Abbreviated  User Entered

Character Count



**Figure 20. User Entered Short Name**

## 4.6 Lesson 3 Review

---

In lesson 3, you have learned:

- The required components needed for Data Element Concept and Value Domain names:
  - DEC: Object Class and Property
  - VD: Representation Term
- The caDSR tool used to create DEC and VD names:
  - Curation Tool
- The source of component terms in the caDSR:
  - EVS NCI Thesaurus
- The types of short names
  - System Generated, Abbreviated or User Entered
- The components needed to create a Data Element name:
  - DEC and VD
- The rules used in creating CDE long and short names:
  - CDE Long name can't have duplicate terms. Short names will follow the form of short name of the DEC and VD.

## 5 Lesson 4: Composing Definitions

---

Now that the components have been determined, we will use the underlying concepts to create definitions for the DEC, VD and CDEs.

### 5.1 Objectives for the Learner

---

By the completion of this lesson, the trainee will be able to:

- Describe the purpose of a metadata component definition
- List six ISO guidelines for an effective metadata component definition
- Name the tool used for creating definitions and the source for Administered Item definitions
- Compose meaningful definitions for Data Element Concepts, Value Domains and Data Elements

### 5.2 Purpose of Definitions

---

The purpose of a Data Element definition is to define a CDE with words or phrases that describe, explain, or make clear its meaning. Good definitions promote metadata standardization and reuse of Data Elements, which facilitates data sharing and interoperability of information systems. This data sharing is the purpose of a metadata registry. Without standardization and sharing, we would continue to build and maintain local databases, but there wouldn't be a frame of reference to exchange data between them. The **challenge** is to create a definition that is specific enough to meet a study or organization's needs; but, generic enough to be used across a community in order to promote harmonization.

## 5.3 Definition Guidelines

---

ISO provides guidelines for creating metadata definitions. A summary of the guidelines identifies the following as desirable characteristics for metadata definitions, along with explanations:

- Unique – Distinguishable from every other definition within the registry.
- Singular – Express the definition as a single instance.
- A statement of the concept, not the negative. Do not exclusively say what it is not.
- A descriptive phrase or sentence. Consider how specific you need to be when thinking of reuse in the caBIG community.
- Commonly understood abbreviations. Use only widely known abbreviations that could be interpreted across contexts. Explain all abbreviations and acronyms important to retain in the definition.
- Not use embedded definitions. The concept should not appear in the definition. Definitions shouldn't include examples. When there is a need to embed a definition for clarity, that piece of the definition can be used in a different form called an Explanatory Comment. This is covered in more detail in section 6.3.2 of this document.

### 5.3.1 Definitions Created by the Curation Tool

The CDE Curation Tool has functionality that builds a default definition based upon what is selected for the Object Class, Property, Qualifiers, and Representation Term.

As the Object Class, Property, and Representation are selected, a definition is built by concatenating the definitions of the Administered Items. The definitions are from the concepts chosen through the EVS/NCI Thesaurus links. In some cases the default definition is entirely appropriate, in other cases; modification is needed to make the definition human readable and understandable.

#### 5.3.1.1 Data Element Concept Definition

Figure 21 represents the default definition created for the Data Element Concept from our example question of "What is your Race?"

2) **Select** Data Element Concept Name Components

Object Class Long Name Person	Property Long Name Race
Qualifier Concepts [ ]	Qualifier Concepts [ ]
Primary Concept Person NCI Thesaurus C25190	Primary Concept Race NCI Thesaurus C17049

\* 3) **Verify** Data Element Concept Long Name (\* ISO Preferred Name)  
 Person Race [ 11 ] Character Count (Database Max = 255)

\* 4) **Update** Data Element Concept Short Name  
 Select Short Name Naming Standard  
 System Generated  Abbreviated  User Entered  
 (Generated by the System) [ 25 ] Character Count (Database Max = 30)

\* 5) **Create/Edit** Definition  
 (Changes to naming components will replace existing definition text.)  
 Person; a human being. Major living subspecies of man differentiated by genetic and physical characteristics. There are four racial groups: Australoid, Caucasoid, Mongoloid, and Negroid.

**Figure 21. Building a Default DEC Definition**

The default definition is created from the definitions for the Object Class (Person) and Property (Race). The default definition created from these terms is: Person, a human being. Major living subspecies of man differentiated by genetic and physical characteristics. There are four racial groups: Australoid, Caucasoid, Mongoloid, and Negroid.

When the default definition doesn't meet the needs of a context, it could be restructured to provide a human readable clarification. This means a concise description of the Object Class along with a description of how the Property provides differentiation to the Object Class.

A modified definition might be: "A person's self-declared racial origination".

NOTE: When a manually-curated (more human readable) definition is created for a Data Element Concept or Value Domain, it should be stored as an Alternate Definition. This is to ensure that the system-generated definitions are readily available when using a DEC or VD to make a Data Element. Once the Data Element is created, a manually-curated (human readable) definition can then be stored as the Preferred Definition as long as the system-generated definition is stored as an Alternate Definition. In any case, be sure to append the more human readable definition with '[Manually-curated]'.

### 5.3.1.2 Value Domain Definition

For the Value Domain, the Representation Term Primary concept definition along with any qualifiers, are concatenated together to form the definition. Figure 22 shows the default definition for our example of "Person Race Category".

Default Definition: A single human being.\_An arbitrary classification of taxonomic group that is a division of a species. It usually arises as a consequence of geographical isolation within a species and is characterized by shared heredity, physical attributes and behavior, and in case of humans, by common history, nationality, or geographic distribution.\_Category; used informally to mean a class of things.

3 ) **Select** Value Domain Representation

Value Domain Attributes

Rep Term Long Name

Qualifier Concepts [Search](#) [Remove](#) Primary Concept [Search](#) [Remove](#)

NCI Thesaurus [C25190](#) NCI Thesaurus [C25372](#)

4 ) **Verify** Value Domain Long Name (\* ISO Preferred Name)

Character Count (Database Max = 255)

5 ) **Update** Value Domain Short Name

Select Short Name Naming Standard

System Generated  Abbreviated  User Entered

Character Count (Database Max = 30)

6 ) **Create/Edit** Definition  
 (Changes to naming components will replace existing definition text.)

**Figure 22. Default Value Domain Definition**

The default definition built from the components is: A single human being.\_An arbitrary classification of taxonomic group that is a division of a species. It usually arises as a consequence of geographical isolation within a species and is characterized by shared heredity, physical attributes and behavior, and in case of humans, by common history, nationality, or geographic distribution.\_Category; used informally to mean a class of things.

This definition provides the unambiguous semantic clarity of the concepts used in the Value Domain name and but doesn't include the concise description of the Representation Term and any qualifiers. Clarification may be provided in a human readable summary of the concatenated definition. An appropriate definition may be created by looking what data is being captured. The representation along with a reference to any qualifiers should be included in the definition.

The modified (manually-curated) definition may be: The classifications that describe a person's self-declared racial origination.

### 5.3.1.3 Data Element Definition

The Curation Tool concatenates the definitions of the DEC and VD to create a default definition for the DE. It is a best practice to retain the system-generated definitions for the DEC and VD until the DE is created. As in the Data Element Name, there will often be repetitive information that needs to be modified or deleted. Figure 23 illustrates the default definition created for our example of "Person Race Category".

\* 2) **Select** Data Element Concept Long Name  
 [Search](#) [Create New](#)

\* 3) **Select** Value Domain Long Name  
 [Search](#) [Create New](#)

\* 4) **Verify** Data Element Long Name (\* ISO Preferred Name)  
  Character Count (Database Max = 255)

\* 5) **Update** Data Element Short Name  
 Select Short Name Naming Standard  
 System Generated  Abbreviated  User Entered  
  Character Count (Database Max = 30)

\* 6) **Create/Edit** Definition  
 (Changes to naming components will replace existing definition text.)

 A single human being.\_Major living subspecies of man differentiated by genetic and physical characteristics. There are four racial groups: Australoid, Caucasoid, Mongoloid, and Negroid.\_A human being.\_An arbitrary classification of taxonomic group that is a division of a species; usually arise as a consequence of geographical isolation within a species and characterised by shared heredity, physical attributes and behavior, and in case of humans, by common history, nationality, or geographic distribution.\_Category; used informally to mean a class of things.

**Figure 23. Default Data Element Definition**

The default definition formed from the selected terms is: A single human being.\_Major living subspecies of man differentiated by genetic and physical characteristics. There are four racial groups: Australoid, Caucasoid, Mongoloid, and Negroid.\_A human being.\_An arbitrary classification of taxonomic group that is a division of a species; usually arise as a consequence of geographical isolation within a species and characterised by shared heredity, physical attributes and behavior, and in case of humans, by common history, nationality, or geographic distribution.\_Category; used informally to mean a class of things.

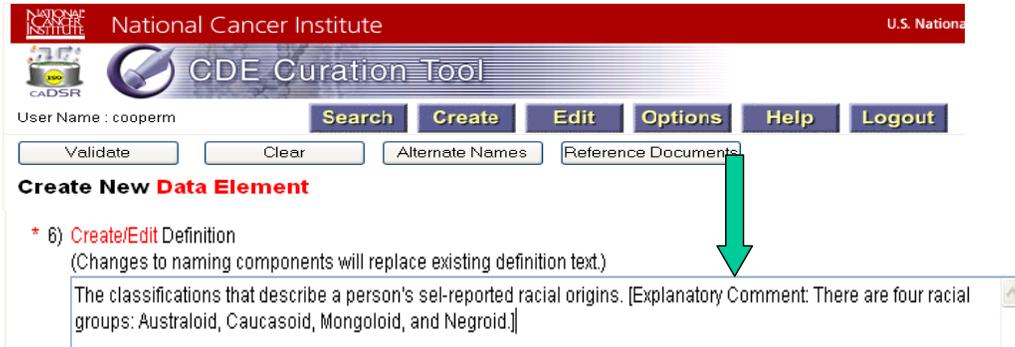
If a human readable (manually-curated) definition is preferred, the definition may be modified by summarizing the concatenated definition. The concatenated (system-generated) definition can be stored as an ‘alternate definition’ with the notation ‘[System-generated] immediately following the definition, with the manually curated (preferred) definition being marked as ‘[Manually-curated]’ or ‘[Standard-based]’, depending on the case.

A modified (manually-curated) definition for the example above might be: The classifications that describe a person’s self-declared racial origination.

### 5.3.2 Explanatory Comments

When the definition requires additional information to provide uniqueness or clarity, that information may be added in an Explanatory Comment. Explanatory Comments can provide examples of any broad concepts in the DEC or VD. Explanatory Comments should be included in the Comment field of the CDE.

Figure 24 below illustrates a sample Explanatory Comment for a Data Element in the Curation Tool.



**Figure 24. Sample Explanatory Comment in Curation Tool**

Along with being placed in the definition, the explanatory comment should also be built under comments for the Data Element in the Admin Tool. Figure 25 below illustrates the explanatory comment in the Admin Tool.



**Figure 25. Explanatory Comment in Admin Tool**

## 5.4 Lesson 4 Review

In Lesson 4, you have learned:

- The purpose of a metadata component definition:
  - To define a CDE, DEC or VD with unambiguous phrases
  - To promote standardization and interoperability across local databases.
  
- Six ISO guidelines for an effective definition:
  - Unique
  - Singular
  - Not negative
  - Descriptive
  - Common Abbreviations
  - No Embedded Definitions.
  
- The tool used for creating and the source for Administered Items definitions:

- Curation Tool links to EVS
- How to compose meaningful definitions for Data Element Concepts, Value Domains and Data Elements
  - Summarize and consolidate default definitions
  - Curate and register alternate definitions when appropriate.

## 6 Lesson 5: Reusing Administered Items

---

Metadata can be envisioned as a library of re-usable items. To reuse items you must first analyze the metadata to identify areas where reuse is possible and appropriate, and the effects of reuse on the resultant data.

### 6.1 Objectives for the Learner

---

By the end of this lesson, the attendee will be able to:

- Identify areas where reuse of administered items is possible and appropriate
- Discuss the effects of reuse on the resultant metadata

### 6.2 Consider the Data by Analyzing the Question

---

Using these two questions below, we'll review how DEC, VD and CDE names and definitions are created. We'll also identify areas where reuse of items are possible and the effects of reuse on the resultant metadata.

**Question 1:** How many times have you mixed pesticides?

Responses: Never  
 < 50  
 > 50

**Question 2:** How many times have you mixed household cleaners?

Responses: Never  
 < 50  
 > 50

Although these two questions are similar, they have unique differences.

**For Question 1:**

What would be the Object Class?

Pesticide

What would be the Property?

Mixing

Is there a Qualifier?

No

**For Question 2:**

What would be the Object Class?

Cleaner  
 What would be the Property?  
 Mixing

Is there a Qualifier?  
 Yes. Household is an Object Class Qualifier

## 6.3 Specific Components

### 6.3.1 Data Element Concept

Figure 26 illustrates the specific DEC metadata components for the two questions above.

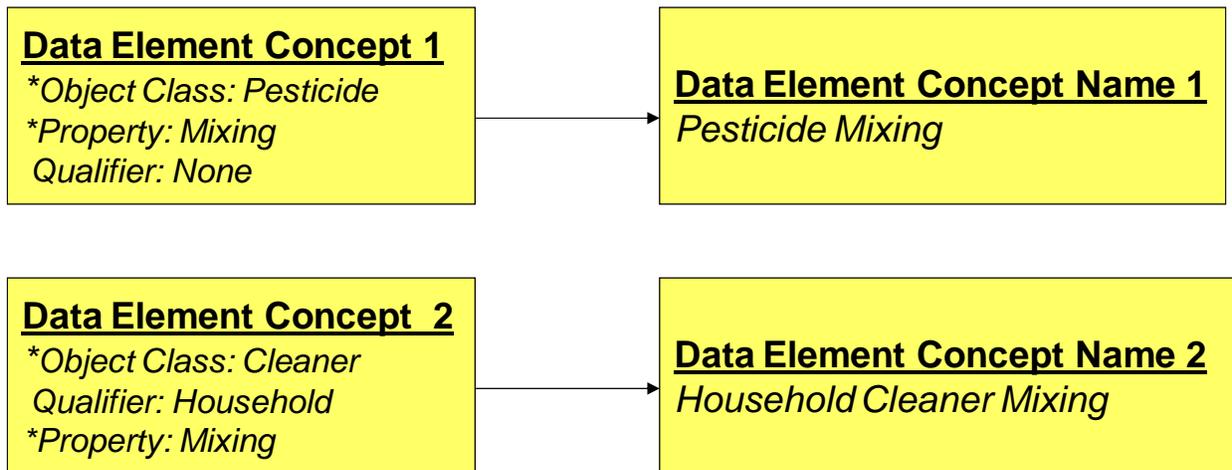


Figure 26. DEC Components

These DECs are different.

### 6.3.2 Value Domain

The responses for the 2 questions above have the same values and value meanings:

Responses: Never  
 < 50  
 > 50

Do we need to create 2 different Value Domains? NO

Figure 27 illustrates the specific components of the Value Domain to use as metadata for our responses for the 2 questions above.

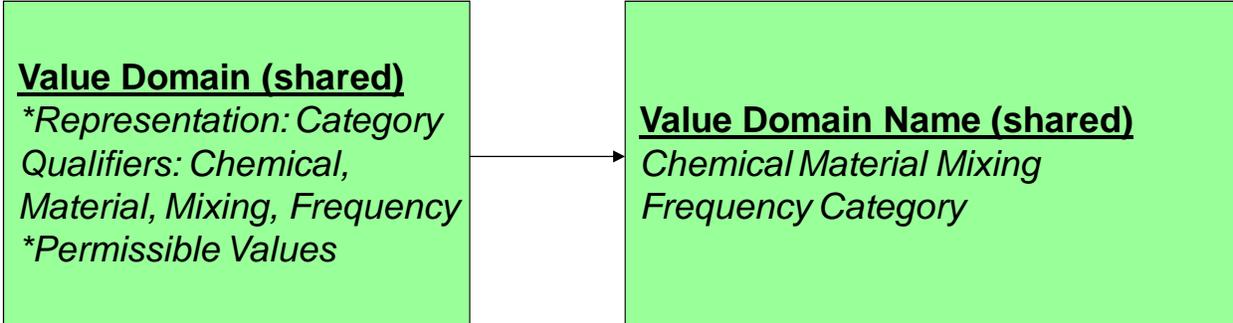


Figure 27. VD Components

Let's look at a choice for Representation Term. The list of Permissible Values appear to be in categories. We can use the Standard Representation Term of "Category" (This term is used to informally to mean a class of things). The qualifier of "Frequency" is used because the questions ask, "How many times?"

We can identify a generic term that describes both Object Classes in the two DEC: Chemical Material. We would create one Value Domain with these terms as qualifiers to the representation term.

What's the next step in creating the CDE Name?

- Before creating a new DEC or VD, first search for caDSR components that can be reused. We have already used the Curation Tool to search for existing components.

What other caDSR tool can be used to search?

- The CDE Browser

Figure 28 illustrates the search fields in the Curation Tool.

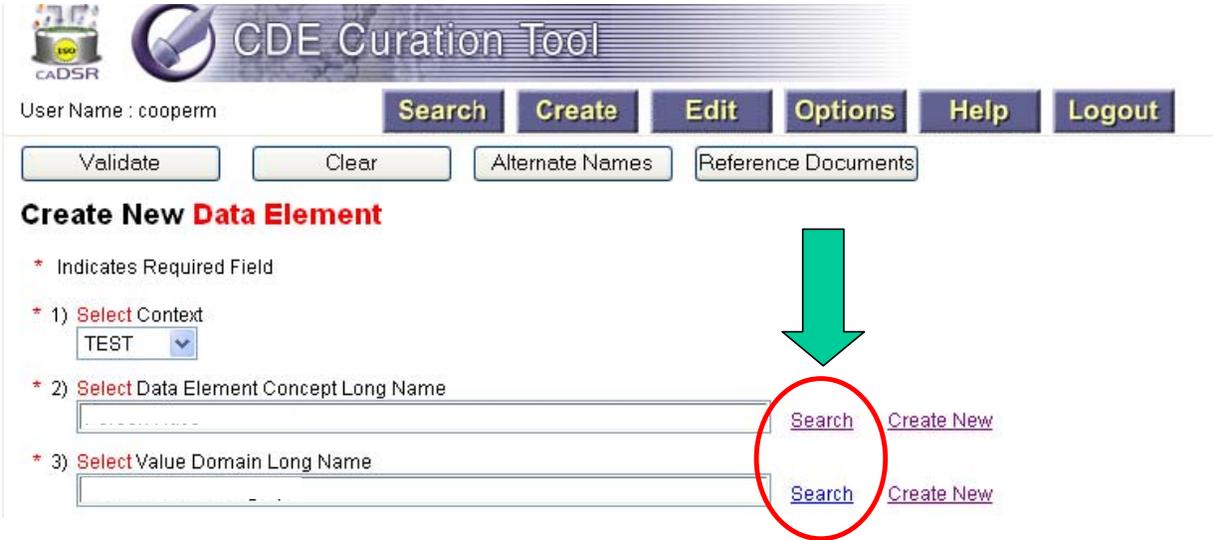


Figure 28. Reuse in the caDSR

Within a context, a DEC and VD can be paired only once to make a unique CDE. If a VD is shared, you have to pair it with two different DEC. The result is two different CDEs as

shown in Figure 29 below.

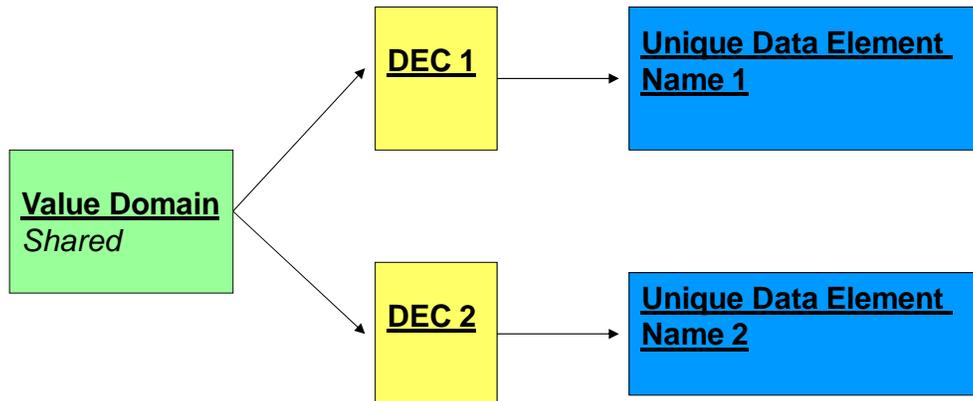


Figure 29. Combine Components to Create Unique Pairings

The converse is true: If a DEC is shared within a context, you must pair it with two different VDs. The result is two different CDEs.

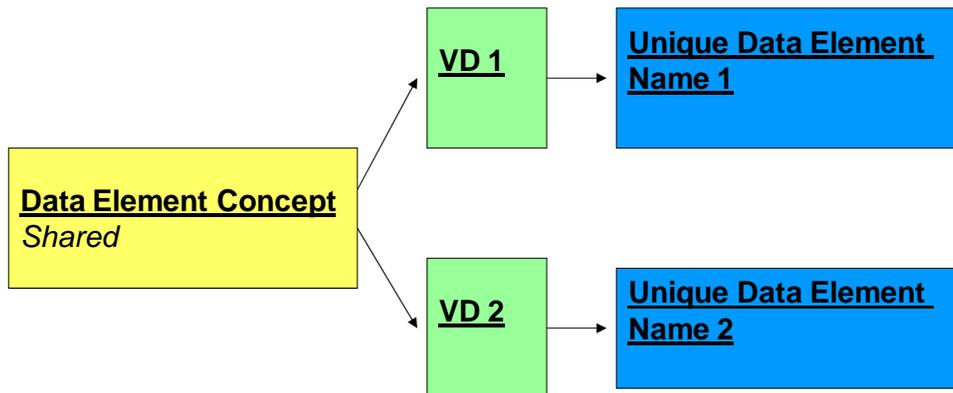
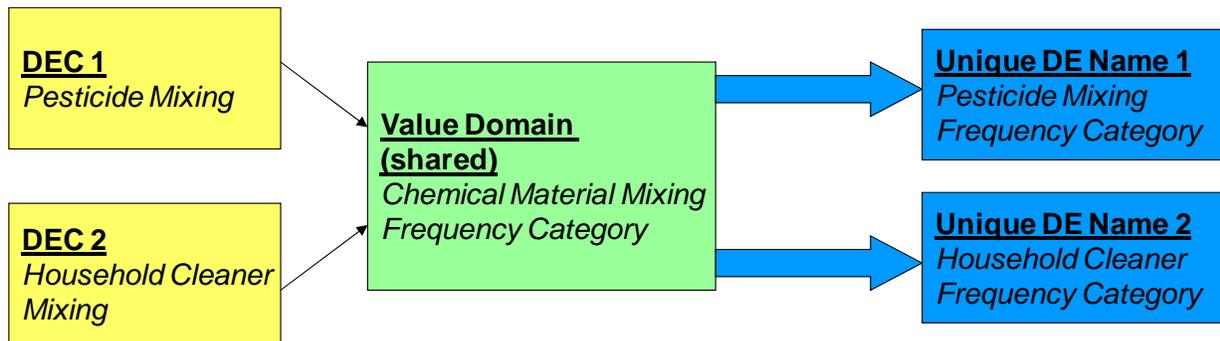


Figure 30. Combine Components to Create Unique Pairings

CDE 1 = Pesticide Mixing **Chemical Material Mixing** Frequency Category



CDE 2 = Household Cleaner Mixing **Chemical Material Mixing** Frequency Category

Figure 31. Creating the Data Element Name

Two DEC's combined with a shared VD will create how many CDEs? How will the CDEs be named? The resultant two unique CDEs retain their DEC Object Class and Property, but share

the Representation and Qualifier from the VD. The redundant terms from the shared Value Domain, with the Qualifier, is deleted from the names of the CDEs.

Let's look at another example for re-using components. Figure 32 below represents two questions regarding performance status. We will identify what components can be shared, re-used or created.

Karnofsky Performance Status Score ?	Lansky Performance Status Score?
100 Fully active, normal	100 Normal, no complaints, no evidence of disease
90 Minor restrictions in physically strenuous activity	90 Able to carry on normal activity; minor signs or symptoms of disease
80 Active, but tires more quickly	80 Normal activity with effort; some signs or symptoms of disease
70 Both greater restriction of and less time spent in play activity	70 Cares for self, unable to carry on normal activity or to do active work
60 Up and around, but minimal active play; keeps busy with quieter activities	60 Requires occasional assistance, but is able to care for most of his/her needs
50 Gets dressed, but lies around much of the day; no active play, able to participate in all quiet play and activities	50 Requires considerable assistance and frequent medical care
40 Mostly in bed; participates in quiet activities	40 Disabled, requires special care and assistance
30 In bed; needs assistance even for quiet play	30 Severely disabled, hospitalization indicated. Death not imminent
20 Often sleeping; play entirely limited to very passive activities	20 Very sick, hospitalization indicated. Death not imminent
10 No play; does not get out of bed	10 Moribund, fatal processes progressing rapidly
0 Unresponsive	0 Dead

**Figure 32. Performance Status Use Case**

In this particular case, there are two separate questions and if you look closely, two distinct set of answers. The Values are the same, but the Value Meanings are different. Since there are two sets of Permissible Value/Value Meanings, two separate Value Domains need to be created.

Looking at Figure 33 below, notice the VD contains the specificity and the DEC is generic and reused. What is the result of the VD/DEC Pairings? Two different CDEs are created.

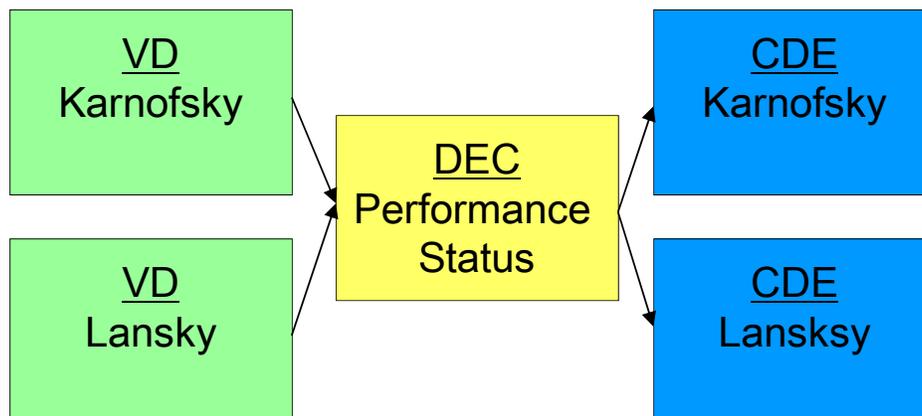


Figure 33. Performance Status Shared Data Element Concept

## 6.4 Lesson 5 Review

---

In Lesson 5, you have learned that:

- Whenever possible and appropriate reuse the DEC, VD, and/or DE.
- When creating DEC's and VD's consider generic terms to promote reuse.
- Within a context, a Data Element is created by a unique pairing of a VD and a DEC.
- The DEC, VD, and DE must have all required components. The selection of component terms will create well-formed metadata.

# 7 Lesson 6: Determining Workflow

---

## 7.1 Objective for the Learner

---

By the completion of this lesson, the attendee will be able to:

- Identify and define metadata component workflow statuses

## 7.2 Workflow Status

---

Workflow Status relates to the development process a CDE. The most common workflow statuses that are used within the caDSR are listed here:

- **Draft New** - Status assigned to new Administered Items created during protocol review or in response to another request. Modifications can be made without versioning. This workflow status is used when a data element has not been reviewed and confirmed, and is not a candidate for designation or reuse.
- **Released** - Status assigned to Administered Items that have been reviewed and approved by the owning context's review process (by committee or other decision-maker). Released Administered Items must be fully specified and compliant with the caDSR naming guidelines and conventions as well as the ISO 11179 model. Versioning rules apply. These CDEs may be designated and reused by non-owning contexts.
- **Approved for Trial Use** - Status assigned to Draft New Administered Items included on CDE-compliant forms. Promotion to this status occurs when the Administered Item has been fully specified and CRFs with which the Administered Item is associated are deeming CDE-compliant.
- **Retired Archived** - Status assigned to old versions of Administered Items that have been replaced with new versions. Retired Archived Administered Items should not be used on future trials and should not be sent through the owning context's review process again. This status is assigned to superseded versions of DEs.
- **Retired Deleted** - Status assigned to Administered Items that were created in error.

Retired Deleted Administered Items are not be viewable in the CDE Browser and are not candidates for reuse.

- **Retired Phased Out** - Status assigned to new Data Elements or modified versions (not old versions) of Data Elements (and associated Administered Items) that have been rejected for general release during the owning Context's review process (by committee or other decision maker). Retired Phased Out Data Elements should not be sent through another round of the owning context's review process, but may be used only as a last-resort on future protocols, if the need is determined by the CRF reviewer.
- **Retired Withdrawn** – Not reviewed by owner; removed from CRFs. These can be reused after approval and change in workflow status. Status assigned to Data Elements, and their associated Administered Items, that have not been subjected to the owning Context's review process (by committee or other decision-maker), including those created for grandfathered protocols (those approved prior to the creation of Released Data Elements for the disease under study). This also includes Draft New or Draft Mod Data Elements, and their associated Administered Items that are removed from submitted CRFs prior to the forms being deemed CDE-compliant and Data Elements, and their associated Administered Items that are removed from consideration prior to the owning Context's review process. Retirement may only occur if the Data Element is not being used on other CRFs. Retired Withdrawn Data Elements may be used on future CRFs as Draft New.

The additional workflow statuses below are included solely for informational purposes:

- **Draft Mod (Modified)** - Status assigned to new versions of existing Administered Items created to accommodate requested changes. The new version will have the same public Id as its predecessor.
- **Committee Submitted** - Status assigned to new or modified Data Elements (and associated Administered Items) created and submitted for the owning context's review.
- **Committee Submitted Used** - Status assigned to new or modified Data Elements (and associated Administered Items) that have approved for use on a trial and submitted for owning context's review. Committee Submitted Used Administered Items must be fully specified.
- **Committee Approved** - Status assigned to 'Committee Submitted' or 'Committee Submitted Used' Data Elements (and associated Administered Items), that have been approved for general use by a context. Promotion to this status occurs upon consensus of the validity of the Administered Item by the owning context.
- **Released Non-Compliant** - Temporary status assigned to Released Administered Items that are not fully specified or compliant with the ISO 11179 standards and guidelines; these Administered Items will be specified and assigned the 'Released' workflow status.

## 8 Lesson 7: Registration Status

---

### 8.1 Objective for the Learner

---

By the completion of this lesson, the attendee will be able to:

- Identify and define metadata component Registration Statuses

### 8.2 Registration Status

---

Workflow status and registration status are related attributes of metadata. Workflow status relates to the development process a CDE. Registration status relates to usage and acceptance of CDEs within the registry, and is a measure of the reuse potential of a single element of metadata.

Example caDSR registration statuses include:

- **Qualified** - has all the necessary metadata attributes and ready to be reviewed for potential usage more widely than in a single context or group.
- **Proposed** – met all quality requirements, ready to be proposed for usage in all contexts.
- **Candidate** - has all the necessary attributes, met all quality requirements, been proposed for usage across all contexts and by repository users.
- **Standard** - has all necessary metadata, met all quality requirements, been approved for usage across all contexts and by repository users.
- **Standardized Elsewhere** - Adopted for standard usage outside of NCI caBIG.
- **Suspended** - Considered for standard status, not selected for promotion.
- **Superseded** - Replaced by another preferred Administered Item and no longer recommended for NCI use.
- **Retired** - No longer recommended for use as a NCI standard.
- **Application** - Part of a computer program or application and may not have all the metadata generally required by the registry. Components with this registration status may have been registered into the repository as part of an activity to register the metadata in a legacy application.

The registration statuses of 'Proposed', 'Candidate', and 'Standard' apply to DEs that have gone through the caBIG Data Standards Governance Process at:

[http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore\\_overview/cadsr/curation/data\\_standards](http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore_overview/cadsr/curation/data_standards)

The Registration Statuses of Proposed and Candidate apply to CDEs that are being considered for advancement to Standard Status within the caBIG community. These CDEs are subject to review by the NCI Context Administrators, Vocabularies and Common Data Elements

Workspace and caBIG User communities. They should be used / reused with the understanding that revisions agreed upon by these groups may result in versioning of the CDEs. CDEs with Registration Status of Standard have been reviewed and adopted as the definitive metadata to be used / reused in applications. Standard CDEs should be given preference when considering use / reuse.

Figure 34 below illustrates the pairing of dynamic Registration (Standard, Candidate and Qualified) and Workflow Statuses.

Registration Status	Workflow Status
Standard	Released
Candidate	Released
Qualified	Released Committee Approved Committee Submitted Approved for Trial Use

**Figure 34. Pairing of Dynamic Registration and Workflow Statuses**

The registration status of ‘Qualified’, ‘Candidate’, and ‘Standard’ will be assigned to Administered Items that are considered for usage across all NCI contexts. These three dynamic registration statuses address progression toward acceptance for general usage.

A ‘Qualified’ component may have the following workflow statuses: ‘Released’, ‘Committee Approved’, ‘Committee Submitted’, or ‘Approved for Trial Use’.

In order for a component to have a registration status of Proposed, Candidate or ‘Standard’, it must have a workflow status of ‘Released’.

A Data Element in the ‘Proposed’ status indicates that it has been proposed for standardization review by the NCI Context Administrators.

A Data Element in “Candidate” status has been accepted for promotion by the NCI Context Administrators and is being reviewed by the Vocabularies and Common Data Elements Workspace and caBIG user community for promotion to Standard Status.

A Data Element in the ‘Standard’ status indicates that the contexts confirm that the Administered Item is of sufficient quality and of broad interest for use in all groups.

Figure 35 below illustrates the pairing of Registration and Workflow Statuses.

Registration Status	Workflow Status
Retired	Released Retired Withdrawn Retired Phased Out Retired Archived
Superseded	Released Retired Archived
Standardized Elsewhere	Released
Application	Released-non-compliant
Suspended	Released Committee Approved Committee Submitted Approved for Trial Use

**Figure 35. Pairing of Static Registration and Workflow Statuses**

The following registration statuses will be assigned to Administered Items that have reached a stage where they are no longer being recommended for usage across contexts (Retired, Superseded, and Suspended) or have been recorded in the registry from another source (Standardized Elsewhere, Application). These registration statuses are considered to be static with additional changes not anticipated.

For any of the ‘Retired’ or ‘Suspended’ type registration statuses to be valid, the workflow status must have been in a ‘Released’ status.

A ‘Retired’ Administered Item should not be used in new applications or studies but may continue to be used in applications or studies that were using the item before it was Retired. Retired Administered Items are retained in the caDSR for historic reference purposes.

An Administered Item in the ‘Superseded’ status indicates that the Context Administrators have decided that the Administered Item is no longer recommended for use in the registry community. A ‘Superseded’ Administered Item may be used but the successor-Administered Item is the preference for use. Such Administered Items are retained in the caDSR Registry for historic reference purposes.

A ‘Standardized Elsewhere’ Administered Item carries its identification and meta-attributes from its originating registry, if available.

‘Application’ Administered Items may have originated from application systems and could be missing some registry metadata attributes

The ‘Suspended’ Administered Item status has been considered for standard status in the past and was not selected for promotion. This status will record historical decisions on a candidate standard.

## 9 Lesson 8: Versioning

# Administered Items

---

## 9.1 Objectives for the Learner

---

By completion of this lesson, the attendee will be able to:

- List and describe three types of versioning
- Apply versioning rules to use cases
- Discuss how versioning a single item can affect associated items

## 9.2 Types of Versioning Changes

---

There are three types of versioning:

- Major – Administered Item needs to be versioned to its next whole number (e.g., 2.3 -> 3.0).
- Minor – Administered Item needs to be versioned by a tenth of a number (e.g., 2.8 -> 2.9 -> 3.0).
- No Change – Administered Item does not need to be versioned.

Versions are created to represent modifications to an existing Data Element that is in use or has been released. When an Administered Item is versioned, the in use/'Released' version remains, while a new version is created and assigned the 'Draft Mod' status. The new and old versions will be considered side-by-side and result in one version being rejected and retired and the other version being approved and released. Should it be determined that both Data Elements are useful and distinct, the old version should be assigned a status of 'Released'; the new version should be retired and then copied and released to create a distinct Data Element.

## 9.3 Versioning a Data Element Concept

---

Figure 36 is an in-depth look at a Data Element Concept, its items, and associated items that affect versioning.

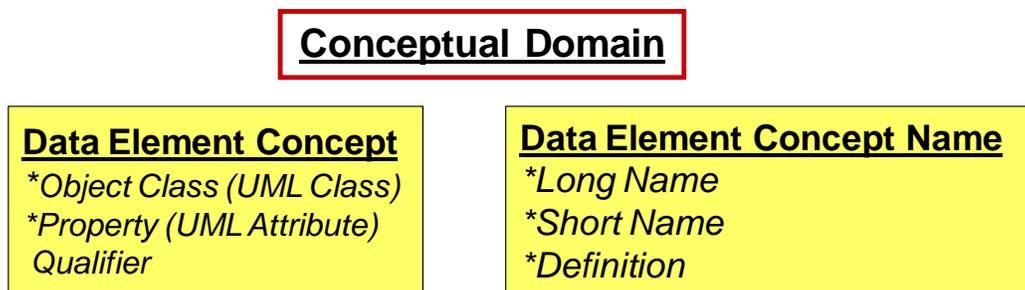


Figure 36. Data Element Concept Overview

A DEC is a member of a Conceptual Domain (CD). The DEC is composed of an Object Class, Property and any Qualifiers. The DEC has a long name, short name and definition.

### 9.3.1 Major Change

A Data Element Concept must be versioned, with its Version number incremented to the next whole number (e.g., 2.3 to 3.0), when:

- Its Preferred Name or Long Name is to be changed.
- It is to be associated with a different Conceptual Domain or a different version of the previously associated Conceptual Domain; an exception is made if the previously associated Conceptual Domain was “<Context name> (e.g., “CTEP”, “CCR”, “DCP”).

### 9.3.2 Minor Change

A Data Element Concept must be versioned, with its Version number incremented by one-tenth (e.g., 2.3 to 2.4), when:

- A substantive or semantic change is to be made to its Definition.
- It is to be changed to accommodate requests made through the CRF CDE compliance review process; in this situation, the new version reflective of the requested changes would be assigned a Workflow Status of “Draft Mod”.

### 9.3.3 No Version Change

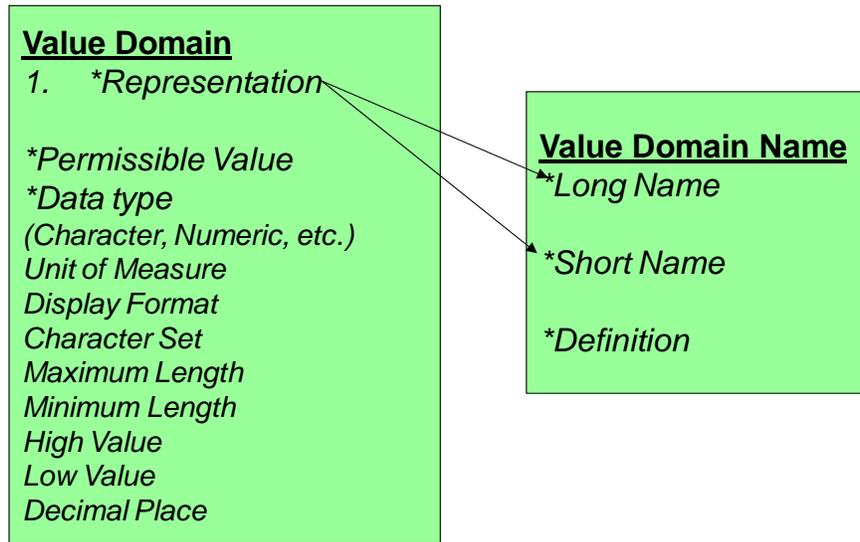
A Data Element Concept should not be versioned when:

- Minor changes, such that a new version of the Conceptual Domain is not required, are to be made to its associated Conceptual Domain.
- Typographical or grammatical errors are to be corrected in its Definition.
- If previously undefined (e.g., Definition reads “Please provide definition” or the Data Element Concept’s name), the Data Element Concept is to be defined.
- One or more of the following of its previously unspecified attributes is to be defined: Object Class, Property.
- Its Workflow Status is to be changed.
- A Designation is to be added or deleted for the Data Element Concept or any of its associated Administered Items.
- A Change Note is to be added or deleted for the Data Element Concept.

## 9.4 Versioning a Value Domain

---

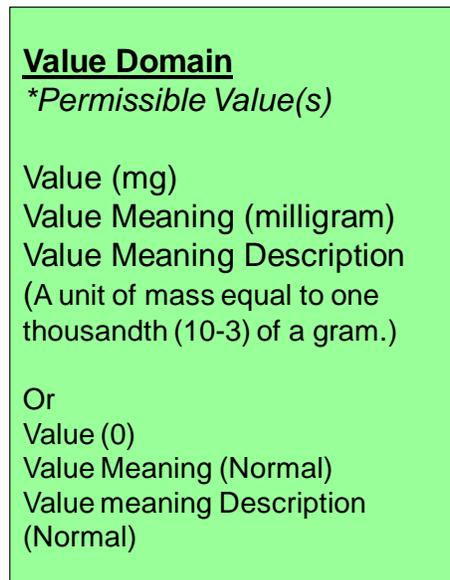
Figure 37 below illustrates an in-depth look of a Value Domain.



**Figure 37. Value Domain - An In-Depth Look**

The top two items make up the Name of the Value Domain, but there other items associated with a Value Domain, including Permissible Values, Data type and their associated Value meanings.

Figure 38 is an in-depth look of Permissible Values.



**Figure 38. Permissible Values - An In-Depth Look**

### 9.4.1 Major Change

A Value Domain must be versioned, with its Version number incremented to the next whole

number (e.g., 2.3 to 3.0), when:

- Its Preferred Name or Long Name is to be changed.
- It is to be associated with a different Conceptual Domain or a different version of the previously associated Conceptual Domain; an exception is made if the previously associated Conceptual Domain was "<Context name>" (e.g., "CTEP", "CCR", "DCP").
- Its Domain Type is to be changed.

### 9.4.2 Minor Change

A Value Domain must be versioned, with its Version number incremented by one-tenth (e.g., 2.3 to 2.4), when:

- A substantive or semantic change is to be made to its Definition.
- It is to be changed to accommodate requests made through the CRF CDE compliance review process; in this situation, the new version reflective of the requested changes would be assigned a Workflow Status of "Draft Mod".
- One or more of the following of its attributes, unless previously unspecified, is to be changed: Data Type, Unit of Measure, Display Format, Character Set, Maximum Length, Minimum Length, High Value, Low Value, Decimal Place, and Representation.
- One or more of its Valid Values are to be deleted.
- One or more Valid Values, representing a substantial change in the quantity, source or standard, or specificity of the Valid Values, is to be added.
- One or more Valid Values is to be associated with a different Value Meaning; an exception is made if the previously associated Value Meaning was "Undefined in conversion" or "<Valid Value>".
- The Value Meaning Description, unless previously undefined, of an associated Value Meaning is to be changed.

### 9.4.3 No Version Change

A Value Domain should not be versioned when:

- Minor changes, such that a new version of the Conceptual Domain is not required, are to be made to its associated Conceptual Domain.
- Typographical or grammatical errors are to be corrected in its Definition.
- If previously undefined (e.g., Definition reads "Please provide definition" or the Value Domain's name), the Value Domain is to be defined.

- One or more of the following of its previously unspecified attributes is to be defined: Data Type, Unit of Measure, Display Format, Character Set, Maximum Length, Minimum Length, High Value, Low Value, Decimal Place, and Representation.
- One or more Valid Values, not representing a substantial change in the quantity, source or standard, or specificity of the Valid Values, is to be added.
- One or more Value Values is to be associated with a different Value Meaning, if the previously associated Value Meaning was undefined.
- A Value Meaning Description, where previously undefined, is to be added to an associated Value Meaning.
- Its Workflow Status is to be changed.
- A Designation is to be added or deleted for the Value Domain or any of its associated Administered Items.
- A Change Note is to be added or deleted for the Value Domain.

## 9.5 Versioning a Data Element

---

A DE/CDE is composed of a DEC and VD. Therefore, whenever a VD or DEC requires a version change, the Data Element will have resultant versioning requirements. A Data Element will require a version change based on:

- Rules that apply solely to the DE.
- Rules that are dependent upon VD or DEC changes.

NOTE: Versioning requirements go into effect only AFTER an item (VD, DEC, and/or DE/CDE has reached 'Released' workflow status or its equivalent.

### 9.5.1 Major Change

A Data Element must be versioned, with its Version number incremented to the next whole number (e.g., 2.3 to 3.0), when:

- Its Preferred Name or Long Name is to be changed.
- It is to be associated with a different Value Domain or a different version of the previously associated Value Domain.
- It is to be associated with a different Data Element Concept or a different version of the previously associated Data Element Concept.

### 9.5.2 Minor Change

A Data Element must be versioned, with its Version number incremented by one-tenth (e.g., 2.3 to 2.4), when:

- One or more of its Document Text entries (of type “Long Name” or “Historic Short CDE Name”) is to be changed; exceptions are made for the correction of a typographical error or a change to an unused element (i.e., not yet used on any CRFs) of “Draft New” status.
- A substantive or semantic change is to be made to its Definition.
- It is to be changed to accommodate requests made through the CRF CDE compliance review process; in this situation, the new version reflective of the requested changes would be assigned a Workflow Status of “Draft Mod”.

### 9.5.3 No Version Change

A Data Element should not be versioned when:

- Minor changes, such that a new version of the Value Domain is not required, are to be made to its associated Value Domain.
- Minor changes, such that a new version of the Data Element Concept is not required, are to be made to its associated Data Element Concept.
- Typographical errors are to be corrected in one or more of its Document Text entries.
- One or more Document Text entries are to be added.
- Typographical or grammatical errors are to be corrected in its Definition.
- If previously undefined (e.g., Definition reads “Please provide definition” or the Data Element’s name), the Data Element is to be defined.
- Classification scheme items assignments are to be added or deleted.
- Its Workflow Status is to be changed.
- A Designation is to be added or deleted for the Data Element or any of its Administered Items.
- A Change Note is to be added or deleted for the Data Element.
- Alternate Definitions or Document Text are added to the Date Element.

### 9.5.4 Versioning Common to All Administered Items

Common changes to all Administered Items result in the same versioning as described below.

Major Version change:

- Long or short name changes
- Associated with new version of CD, DEC or VD

- VD – changes domain type

Minor Version change:

- Change definition
- Associated with new, or new version of CD, DEC or VD
- CRF CDE Compliance Review – Draft Mod
- DEC – change Object Class or Property concepts
- VD – change attributes or PV
- CDE – change document text

No Version change:

- Typographical or grammar corrections
- Addition of definitions when not previously defined
- Change workflow status or designation
- Addition of a change note

## 10 Lesson 9: Derived Data Elements

---

### 10.1 Objectives for the Learner

---

By the completion of this lesson, the attendee will be able to:

- Define Derived Data Elements
- Discuss Derivation Rules for metadata
- Distinguish Between a Non-Derived and Derived Data Element

### 10.2 What are Derived Data Elements?

---

Derived Data Elements are two or more Data Elements with a relationship based on a rule.

### 10.3 Derivation Rules

---

ISO/IEC 11179 defines a derivation rule as the specification of actions for creation of metadata. A rule may range from a simple operation such as subtraction to a very complex set of actions. Derivation rules are not limited to arithmetic and logical operations. A derivation is defined as a relationship between a derivation rule and an input set upon which it acts.

### 10.4 Non-Derived vs. Derived Data Elements

---

In a non-derived Data Element, a question creates a single DE.

In a derived Data Element the question has initial input Data Element(s) where a rule is applied to produce an output Data Element(s).

Figure 39 below is a visual representation of a non-derived vs. derived Data Element.

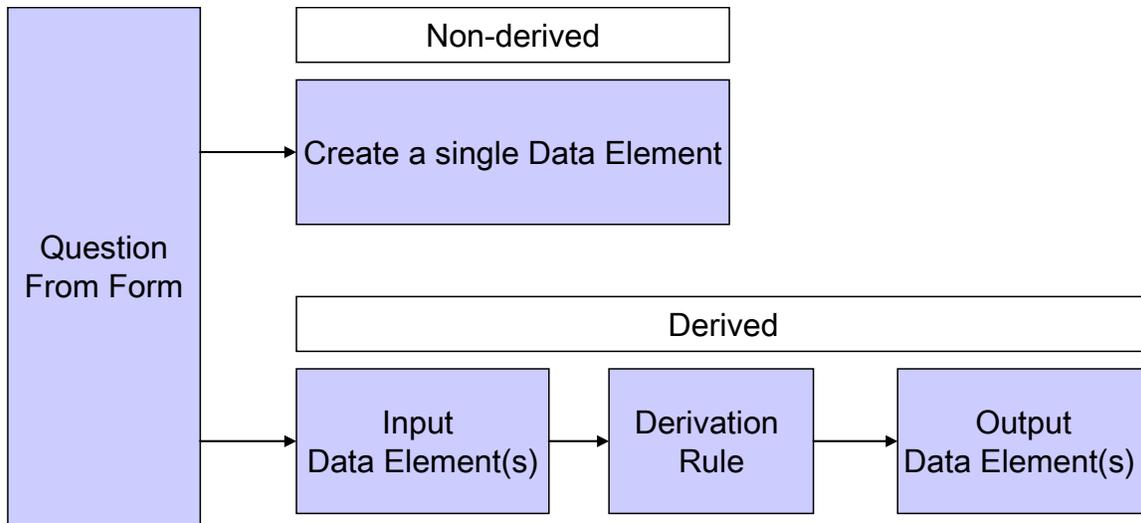


Figure 39. Non-Derived vs. Derived Data Element

## 10.5 Example of a Derived Data Element

Figure 40 below illustrates how a Derived Data Element is formed.

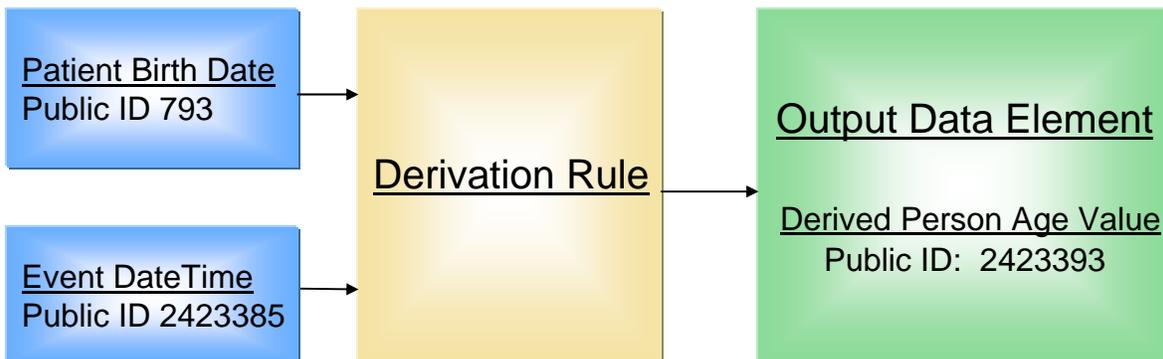


Figure 40. Derived Data Element

Derived Data Elements start with an input of existing two or more Data Elements and are combined with a Derivation Rule. The output is a new Derived Data Element.

As Figure 40 illustrates, two CDEs, Patient Birth Date and Event Date Time are combined with a derivation rule to form the Derived Data Element of Derived Person Age Value.

Figure 41 is an example of a Derived Data Element in the caDSR viewed by the CDE Browser.

Selected Data Element

<b>Public ID:</b>	2423393
<b>Version:</b>	1.0
<b>Long Name:</b>	Derived Person Age Value
<b>Short Name:</b>	DERIV_PRSN_AGE_VAL
<b>Document Text:</b>	
<b>Definition:</b>	The length of time that one has existed in completed years, months and days at the time of the event. [Explanatory Comment 1: Age is calculated as the Event Date - Person Birth Date. The transfer format of Year, Month and Day (Y.M.D) includes values for year between 0 and 999, for month between 0 and 11, and for day between 0 and 3. Explanatory Comment 2: The majority of external standards do not represent, transmit, or store the concept of Age. Microsoft provides instructions on the most common calculations and an algorithm is included to provide guidance for the computation of age value.]
<b>Workflow Status:</b>	RELEASED

Figure 41. Derived Data Element Example

Figure 42 illustrates the details of a Derived Data Element located within the caDSR. Please note the derivation type, rule, and method.

Data Element Derivation Details

<b>Derivation Type:</b>	CALCULATED
<b>Rule:</b>	Age = Event Date - Date of Person's Birth
<b>Method:</b>	-Line-listed data with uncategorized age is preferred. The reporting of age categorizations is not recommended to facilitate integrated analysis of data. Summary data should follow Census age categorizations. -Fractional ages, e.g. 1.5 years, are not recommended to be consistent with self-reported age. All age units should be incremented after reaching the next full unit, e.g. .5 minutes or 30 seconds = 0 minutes, 1.5 years = 1 year. -Mixed unit ages, e.g. 1 year and 3 months, are not recommended. The next smaller age unit should be used if necessary, e.g. 1.5 years would become 18 months. -If rounding is done when calculating age, e.g. years from age in months, it should be documented. Note: Collection of date of birth is recommended with age as suggested due to the selected data element components.
<b>Concatenation Character:</b>	

Component Data Elements

Display Order	Long Name	Context	Workflow Status	Public ID	Version
1	Patient Birth Date	CTEP	RELEASED	793	4.0
2	Event DateTime	caBIG	RELEASED	2423385	1.0

Figure 42. Details of a Derived Data Element

## 10.6 Lesson 9 Review

In Lesson 9, you learned:

- Derived Data Elements are complex Data Elements based on a rule.
  - Derivation rules are the specification of actions for creation of data. A rule may range from a simple operation such as subtraction to a very complex set of actions. Derivation rules are not limited to arithmetic and logical operations.

- The difference between a non-derived and derived Data Element:
  - In a non-derived Data Element, a question creates a single DE.
  - In a derived Data Element the question has initial input Data Element(s) with a rule applied to produce an output Data Element(s).

## 11 Documentation / Recommended Reading Materials

---

Below is a list of links to documentation used to create this session and of recommended reading materials.

caDSR Homepage:

- [http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore\\_overview/cadsr](http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore_overview/cadsr)

caDSR Training Home Page:

- [http://ncicb.nci.nih.gov/NCICB/training/cadsr\\_training](http://ncicb.nci.nih.gov/NCICB/training/cadsr_training)

caCORE Developers Guide:

- [https://gforge.nci.nih.gov/docman/view.php/148/8650/caCORE%20SDK%204.0%20Developer's%20Guide\\_101007.pdf](https://gforge.nci.nih.gov/docman/view.php/148/8650/caCORE%20SDK%204.0%20Developer's%20Guide_101007.pdf)

caCORE SIW and UML Loader Technical Guide:

- [http://gforge.nci.nih.gov/docman/view.php/16/12647/caCORE\\_SIW\\_UML\\_Loader\\_3.2\\_Technical\\_Guide.pdf](http://gforge.nci.nih.gov/docman/view.php/16/12647/caCORE_SIW_UML_Loader_3.2_Technical_Guide.pdf)

caBIG<sup>®</sup> Learning Management System:

- <http://ncicbtraining.nci.nih.gov>

caDSR Business Rules:

- [http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore\\_overview/cadsr/business\\_rules/](http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore_overview/cadsr/business_rules/)

caDSR\_Users Listserv Subscription:

- [https://list.nih.gov/archives/cadsr\\_training-l.html](https://list.nih.gov/archives/cadsr_training-l.html)
- <http://list.nih.gov>

Send Request for caDSR Account to:

- [ncicb@pop.nci.nih.gov](mailto:ncicb@pop.nci.nih.gov)

## 12 Contact Information

---

NCICB Application Support:

[ncicb@pop.nci.nih.gov](mailto:ncicb@pop.nci.nih.gov)

Instructor:

Jennifer Brush

[brushj@mail.nih.gov](mailto:brushj@mail.nih.gov)

NCICB Liaison:

Dianne Reeves

[reevesd@mail.nih.gov](mailto:reevesd@mail.nih.gov)

## 13 Course Review Questions

---

Listed below are a set of questions that you should be able to answer after reviewing the material of this workbook. To get credit for this course, you must register and pass the accompanying quiz hosted in the caBIG<sup>®</sup> Learning Management System at: <http://ncicbtraining.nci.nih.gov>.

1. The preferred terminology for creating Common Data Elements in the caDSR is:
  - a. NCI Thesaurus
  - b. NCI Metathesaurus
  - c. SNOMED
  - d. LOINC
2. One of the benefits of using the NCI Thesaurus as a source for CDE names is that it is editable and maintainable in real time.
  - a. True
  - b. False
3. The “short name” is an abbreviated name for an Administered Item (e.g. Data Element).
  - a. True
  - b. False
4. The required parts of a Data Element Concept include the Object Class and Property.
  - a. True
  - b. False
5. To create a Data Element Concept name, you will need the following:
  - a. Object Class and Property
  - b. Data Element
  - c. Value Domain
  - d. Representation Term
6. The CDE Curation Tool provides links to search the NCI Thesaurus for Object Class, Property and their Qualifiers.
  - a. True
  - b. False
7. What do you do if you can't find the term you need in EVS?
  - a. Make up your own term
  - b. Use the Admin Tool instead
  - c. Suggest a new term to EVS
  - d. Give up
8. To create a Value Domain name, you will need the following:
  - a. Data Element Concept
  - b. Data Element
  - c. Conceptual Domain

- d. Representation Term
9. There are three options for selecting a short name for an Administered Item. Which option do you choose if you want to enter a customized abbreviation of the name?
- System Generated
  - Abbreviated
  - User Entered
10. CDE Long names can have duplicate terms in the name.
- True
  - False
11. Good definitions promote metadata standardization and reuse of Data Elements, which facilitates data sharing and interoperability of information systems.
- True
  - False
12. A metadata definition should be unique, singular and not contain embedded definitions.
- True
  - False
13. What caDSR Tool can be used to search for existing Data Element Concepts and Value Domains?
- UML Loader
  - caDSR Sentinel Tool
  - Semantic Integration Workbench
  - CDE Curation Tool
14. Within a context, a DEC and VD can be paired only once to make a unique CDE.
- True
  - False
15. An example of a major version change for a Data Element is:
- v2.0 → v3.0
  - v2.0 → v2.1
  - v2.0 → v2.0
16. An example of a minor version change for a Data Element is:
- v2.0 → v3.0
  - v2.0 → v2.1
  - v2.0 → v2.0
17. If you change the Value Domain Type from Non-enumerated to Enumerated, you should:
- Version the Value Domain to the next whole number
  - Version the Value Domain by one tenth (e.g. 2.1 to 2.2)
  - No versioning is necessary
18. Correcting typographical errors in a Data Element definition requires a new version.
- True
  - False

19. The workflow status assigned to Data Elements (and associated Administered Items) that have been reviewed and approved by the owning context's review process is:
  - a. DRAFT MOD
  - b. RELEASED
  - c. APPROVED FOR TRIAL USE
  - d. RETIRED PHASED OUT
  
20. A minor version change is required when a semantic change is made to the definition of a Data Element Concept.
  - a. True
  - b. False
  
21. A major version change is required when a Value Domain is associated to a different Conceptual Domain.
  - a. True
  - b. False
  
22. Changing the max length of a Permissible Value constitutes a minor version change in the Value Domain.
  - a. True
  - b. False
  
23. Changing the Value Domain in a Data Element requires a minor version change for the Data Element.
  - a. True
  - b. False
  
24. Adding a Permissible Value from the same source as the other values to a Value Domain constitutes a major version change.
  - a. True
  - b. False
  
25. Changing the wording of the document text constitutes a minor version change to the Data Element.
  - a. True
  - b. False