

Curating Metadata FAQs

Contents of this Page

- [What is metadata?](#)
- [What is the difference between curation and harmonization?](#)
- [Help! I can't find the Data Elements that I'm looking for.](#)
- [What are my options when I find a Data Element that is nearly perfect for my use, but the definition is not quite what I need?](#)
- [How do I check a Data Element Concept \(DEC\) for completeness?](#)
- [What do I choose for the Conceptual Domain for my Data Element Concept or Value Domain?](#)
- [Can a non-enumerated Value Domain have a specific template, or format associated with it?](#)
- [I am a curator, and I was versioning a CDE that I own within my context. I created a new version in the Curation Tool, and then finished up by retiring the old version. I then noticed that other contexts have designated this DE for re-use. What do I do now?](#)
- [Which workflow statuses can be used in designation?](#)
- [What is a Value Domain? \(in the caDSR\)](#)
- [What exactly is a Permissible Value?](#)
- [Can I have more than one Value Domain per Data Element?](#)
- [Do DEC, VDs and their associated CDEs have to have the same name?](#)
- [How can we ensure that the applications using CDEs that are versioned don't end up with clinical data that is impossible to aggregate?](#)
- [What is the definition of Value Meaning?](#)
- [It is okay to have more than one Qualifier for an Object Class or Property?](#)
- [Can you clarify the relationship between a Conceptual Domain and a Classification Scheme?](#)
- [If I want to reference an EVS concept when I'm curating a Classification Scheme or Conceptual Domain, what do I do?](#)
- [Can Enumerated Value Domains Permissible Value lists contain values both from EVS and locally defined terms?](#)
- [What is the preferred Vocabulary for caDSR Concepts?](#)

What is metadata?

Metadata is the description of how information (or data) should be collected and stored, including a definition that describes the semantics (the meaning) of that data. The caDSR is an Oracle database that stores metadata.

Example: There is a question on a form that reads "What is the patient's age?", and there is a blank to fill in the answer, data is the number you put in the blank. If Jane is 35 years old, '35' is the piece of data that describes Jane's age. Metadata is information about how that piece of data should be collected and stored for any patient. In this case, the metadata defines the patient's age as an integer that gives the patient's lifespan in years.

What is the difference between curation and harmonization?

Curation is the process of creating/editing content in the caDSR. It involves characterizing the data that one intends to collect or has collected as an ISO /IEC 11179 data element.

Harmonization is a consensus process by which there is agreement by interested parties on how common data will be represented. After there is agreement, common data elements (or parts) may be curated in the registry and made available for reuse.

Help! I can't find the Data Elements that I'm looking for.

It is possible that some data elements have not yet been identified for use and so have not been created as metadata in the caDSR. However, here are some example searches that returned results:

Using an Advanced Search in the CDE Browser, I made sure everything was "All" defaulted in the search criteria, using the term *transcript* in the search field). I got two results.

In the Curation Tool, I did a search for Data Element using *primer* as the search term. I found one result.

What are my options when I find a Data Element that is nearly perfect for my use, but the definition is not quite what I need?

You can contact that Context Administrator and ask if the definition could be modified. If that is not possible, and the definition is the semantic equivalent of the existing definition, you can ask the Context Curator to add an alternate definition.

How do I check a Data Element Concept (DEC) for completeness?

A complete Data Element Concept contains at least one Object Class and one Property, a system-generated or semantically complete definition, an appropriately descriptive name, a Conceptual Domain, Workflow Status, Version and Begin Date.

What do I choose for the Conceptual Domain for my Data Element Concept or Value Domain?

Inspect the existing RELEASED Conceptual Domains, evaluating the name and definition. If none exist that seem appropriate, contact the CBIIT help desk to request a new one.

A context does not have to create a set of Conceptual Domains for their own use - all Conceptual Domains can be reused, regardless of the owning organization.

Can a non-enumerated Value Domain have a specific template, or format associated with it?

Yes, there are a number of Value Domain attributes, including Display Format, Minimum and Maximum Value, Minimum and Maximum length that can also be used to refine a non-enumerated Value Domain.

Value Domain Reference Documents should also be considered (reference document type = "Value Domain Reference") where a document can be attached and a specific URL referenced, as well as Value Domain examples.

"Examples" provide a way to help make any Value Domain less ambiguous, and in particular a non-enumerated one. Examples can be added by creating a Value Domain reference document with type, Example, and providing a short list in the Reference Document Text field.

I am a curator, and I was versioning a CDE that I own within my context. I created a new version in the Curation Tool, and then finished up by retiring the old version. I then noticed that other contexts have designated this DE for re-use. What do I do now?

Please contact the Context Administrators for the Using Contexts and let them know about the change.

Which workflow statuses can be used in designation?

Data Elements with any of the following workflow statuses may be designated as Used By a non-owning Context:

- Approved for Trial Use
- Committee Submitted Used
- Draft Mod
- Committee Approved
- Released-non-compliant
- Released

Refer also to the [caDSR Business Rules](#).

What is a Value Domain? (in the caDSR)

A Value Domain, in ISO 11179, is like a 'picklist.' It also has other common metadata attributes like the datatype, format, etc.

A Value Domain can be either enumerated or non-enumerated. If enumerated, it contains a list of permitted responses to populate a specific representation for a Data Element Concept. This combination of a particular list of values, for a particular datatype, to represent an instance of a particular Data Element Concept, is known as the Data Element.

Normally one or more concepts from a vocabulary might be in a 'picklist' for a Data element, and those concepts go into an enumerated Value Domain.

What exactly is a Permissible Value?

It is the value stored in an information system as one possible value of an enumerated Value Domain. The permissible value object in the caDSR records the value and its semantic definition using the following attributes:

- Value - a short label or name which is the value stored in an information system
- Value Meaning - a unique identifier, usually the EVS Preferred Name, for the concept
- Value Meaning Description - Value Meaning's concept definition, usually the EVS concept definition

Can I have more than one Value Domain per Data Element?

No. But a single Value Domain may have one or more permissible values.

Do DEC's, VD's and their associated CDE's have to have the same name?

No. Though the VD and CDE will frequently have the same names, they are not required to be the same or different. However, the DEC does not include a representation, and the CDE does, the CDE and DEC should NOT have the same preferred and long names.

How can we ensure that the applications using CDEs that are versioned don't end up with clinical data that is impossible to aggregate?

There are business processes and practices in place for the proposal, review, and acceptance of new CDE versions to minimize unnecessary changes. However, there will be situations where a version change is required for a CDE. The release of a new version of a CDE does not mean that an application or project must immediately update to using the new version.

For example, if a project is using version 1.0 of a CDE on a form, and that version will meet their needs for the duration of the trial, there is not necessarily a need for them to modify the form to use the new version - they can continue using version 1.0 for the duration. However, users should be aware of the new version, and make use of it as new projects are started and implemented. There might also be times when the new version of the CDE will have information that a user wants to make immediately accessible (i.e. a new value domain name with an additional set of values). In this case, the user would consider implementing the new version of the CDE as soon as possible.

At a minimum, use the caDSR Tools and APIs that are designed to help end users keep track of and understand such changes. The Sentinel Tool is used to create an Alert report so that Version changes to CDEs deployed in applications automatically cause a report to be sent to the Application Administrator. This report contained accurate historical metadata about the CDE that was versioned. Both the old and the new version remain permanently in the caDSR.

Also, the CDE Compare Tool, found on the CDE Browser home page, can be used to inspect and download both versions side by side. The application owner will need to have policies to make decision about when to deploy new versions of CDEs and also how to maintain accurate tracking of which CDE was used to collect a specific instance of the data when and if a different versions are supported over the course of one study. The ability to detect and inspect changes to different versions of CDEs is an advantage over a non-registered CDE environment.

What is the definition of Value Meaning?

The standard has two views of data, logical and physical. The logical view describes sets of concepts, each in a Conceptual Domain. Each member of a **Conceptual Domain** has a **Value Meaning** that distinguishes that member from other members. Value Meanings give semantic meaning to physical values. The physical view describes sets of values, each in a **Value Domain** where each member has a **Permissible Value**.

In order to move concepts and semantic meanings into physical implementation in software, a Value Domain is associated with a Conceptual Domain. The Permissible Values in the Value Domain are each associated with a Value Meaning in the Conceptual Domain.

A particular Value Meaning may have more than one means of representation by Permissible Values, each from a distinct Value Domain. Thus the Permissible Values 1, 2, and 3 in one Value Domain, and the Permissible Values Y, N, and U in another Value Domain may be mapped to the same Value Meanings, Yes, No, and Unknown. This makes it possible to compare 1 and Y, and find that they are the same.

It is okay to have more than one Qualifier for an Object Class or Property?

Yes. The use of compound concept qualifiers with Object Class, Property and Rep Terms is supported.

Can you clarify the relationship between a Conceptual Domain and a Classification Scheme?

Refer to this MS Word document: [Conceptual Domains vs. Classification Schemes](#)

If I want to reference an EVS concept when I'm curating a Classification Scheme or Conceptual Domain, what do I do?

Use the following conventions.

- Classification Schemes and Conceptual Domains: Use "Alternate Name". Use the caDSR Admin Tool to add the alternate name, using the "type" drop down list; select the appropriate type name depending on the source terminology. If you do not find an appropriate "type" code, send an email to [NCI CBIIT Application Support](#) asking for the type to be added. The request will be sent to the caDSR development team who can update the "type" name list almost immediately during regular business hours.
- Classification Scheme Items: "Comments" Field. Put the concept identifier from the source terminology into the Comments Field. Use a semi-colon as a separator character, along with the concept's source type name, and any other information that will help us resolve the concept id when the data is converted. The easiest references to resolve will be concept codes from terminologies already served by EVS. For those, we'd appreciate if you use the alternate name types already in use in caDSR; this will minimize research for data conversion. If the source is not an EVS source, include enough information so that a human can interpret your intent.

Can Enumerated Value Domains Permissible Value lists contain values both from EVS and locally defined terms?

Yes, this option is provided in the Curation Tool, however, the use of EVS concepts is important to assign the semantic meaning to the Permissible Value. If you aren't able to find an appropriate EVS concept, the Curation Tool provides a link to request new concepts.

What is the preferred Vocabulary for caDSR Concepts?

NCI Thesaurus, see [EVS and caDSR Content FAQs](#) for more information.