

LexEVS 6.5.1 Value Set Architecture

Contents of this Page

- [Observations](#)
- [Meeting Minutes](#)
- [Requirements](#)

This page is for capturing the discussion around updating the workflow for value set definition creation.

Observations

- LexEVS model and implementation are more complex than requirements of most NCI value sets
- NCI Thesaurus may have enough assertions to adequately describe value sets without external modeling

Meeting Minutes

Value set requirements gathered through meetings with the NCI stakeholders. Minutes from the meetings are here:

- [LexEVS Meeting Minutes - Value Set Architecture Planning Session - 2017.01.06](#)
- [LexEVS Meeting Minutes - Value Set Architecture Planning Session - 2017.01.09](#)
- [LexEVS Meeting Minutes - Value Set Architecture Planning Session - 2017.01.19](#)
- [LexEVS Meeting Minutes - Value Set Architecture Planning Session - 2017.02.21](#)
- [LexEVS Meeting Minutes - Value Set Architecture Planning Session - 2017.03.02](#)
- [LexEVS Meeting Minutes - Value Set Architecture Planning Session - 2017.03.27](#)
- [LexEVS Meeting Minutes - Value Set Architecture Planning Session - 2017.04.24](#)

Requirements

Requirements for the new value set work flow architecture

Requirement	Priority	Notes
Better logging to help determine if there were any failures for resolving the 760+ value sets.		
Resolving all 760+ value sets should be able to finish over night (Rob).		
leafOnly=false • We can write a value set loader that ignores these. • We can ignore targetToSource=false		2017.04.24 VS Arch. Meeting

Proposed or Possible Requirement	Priority	Notes	Is Requirement? (Yes/No)
Remove Dependency on value set definition files for NCIt defined Value Sets		There are still some value sets that need to be approached with the old value set method.	yes
Generate All NCIt sourced value set coding schemes from NCIt source in LexEVS (DB)		There are still some value sets that need to be approached with the old value set method.	yes
Generate value set URI from NCIt source based on source and hard coded structures.		- There is information of the agency is in the annotation on that concept. This information can be used to create the unique URI that represents the agency.	yes
(Browser) Auto generate value set definitions from the NCIT source		• In browser, value set path, group by standards authority/ sourced terminology, click on collapse. • In order to create the top nodes, need to find that annotation. • Kim queries against these for Value Set metadata	yes
Resolve discrepancies between number of value set definition files and value sets defined in NCIt		This is not necessary	no
Maintain value set functionality for those few value set definitions which define leafOnly as false			yes

Provide acceptable substitutions for value set URI's and other metadata that is defined in the source (List in other rows as necessary)			yes
Maintain Resolved Value Set Coding Scheme API as interface		• There are several users. We should keep this for 6.x • Re evaluate if needed after 6.x	yes
Provide concurrent value set loading capability		This was originally a suggestion on how to speed the load up. This could still be a possibility that we should look at going forward,	no
Provide programmatic access to value set definition XML files		We need efficient way to retrieve this day (Kim)	no
Do we need to define A8 as an association each time			no
(Browser) Provide efficient way to retrieve label (URI) and version of all resolved value set coding schemes		• Return time should be in seconds	yes
(Browser) Efficient resolution of VS graphs as it pertains to hierarchies		Seconds	yes
(Browser) Efficient results to VS resolutions		sub second for most	yes
(Browser) Efficient retrieval VS definition metadata resolution calls. Faster queries against the definitions themselves.		1 second or less	yes
(Browser) Efficient search by code or name - query for value set that this concept matches		This needs further definition/discussion with Kim. Kim's notes: Search value sets by code or name that matches with any member (concept) of a value set using a user-specified (exactMatch, startsWith, or contains) algorithm.	yes

(Browser) Source specific information - need to call LexEVS API for each vs Iterator without having to further query	Kim mentioned this when you click the value set button. Needs further definition and investigation. Kim's notes: Provide source specific value set resolution data through an iterator. Formats: (Case 1) Value sets with a non-NCIt default coding scheme (e.g., NDF-RT). Code Preferred Name Coding Scheme Name, Namespace (Case 2) Value sets with a NCIt default coding scheme and has no non-NCI supportedSource, NCIt Concept Code NCIt Preferred Term NCIt Synonyms NCIt Definition (Case 3) Value sets with a NCIt default coding scheme and has at least one non-NCI supportedSource (e.g., FDA), NCIt Concept Code Source Name (e.g., FDA Name) NCIt Preferred Term NCIt Synonyms Source Definition (e.g., FDA Definition) NCIt Definition (Note: If there are multiple supportedSource, then use the first supportedSource.)	yes
Build source (i.e., standards authority) view and terminology view of value set hierarchy efficiently from the NCIt Source, the NDFRT source and value set definitions as they exist the database.	We need to understand the values and procedures needed to support this better First, a terminology Termonology_Value_set.owl is constructed for supporting the creation of value set hierarchies. Each concept in the Termonology_Value_set terminology is assigned a unique TVS code, for example, TVS_FDA (see below). The hierarchical relationship among concepts in the Termonology_Value_set is uniquely defined by the subClassOf relationship. One can view this hierarchical structure as a graph with each node representing a bin which holds value sets. Value sets are placed into these bins in accordance with the value of the source tag as appearing in its corresponding value set definition xml file. For example, the value set, FDA CDRH GUID Terminology, will be placed into the bin with an id or tag called TVS_CDRH_GUID_Component because in the XML file, there is a source tag with value TVS_CDRH_GUID_Component: <source>TVS_CDRH_GUID_Component</source>	This informs our value set meta data resolution performance requirement, but is not currently a requirement on it's own.

Note that a value set can be placed into multiple bins.

(A) A sample class in the terminology Termonology_Value_set.
owl:

```
<owl:Class rdf:ID="TVS_CDRH_GUDID">

  <rdfs:subClassOf>

    <owl:Class rdf:ID="TVS_FDA"/>

  </rdfs:subClassOf>

  <Source rdf:datatype="http://www.w3.org/2001/XMLSchema#string"

    >FDA_CDRH_GUDID</Source>

  <Display rdf:datatype="http://www.w3.org/2001/XMLSchema#boolean"

    >false</Display>

  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string"

    >FDA CDRH GUDID Terminology</rdfs:label>

  <Description rdf:datatype="http://www.w3.org/2001/XMLSchema#string"

    >A set of terminology created to support the efforts of the FDA CDRH Global Unique Device Identification Database project.</Description>

  <Preferred_Name rdf:datatype="http://www.w3.org/2001/XMLSchema#string"

    >FDA CDRH GUDID Terminology</Preferred_Name>

</owl:Class>
```

(B) A sample Value set definition XML file, FDA CDRH GUDID Terminology:

```
<?xml version="1.0" encoding="UTF-8"?>

<valueSetDefinition xmlns="http://LexGrid.org/schema/2010/01/LexGrid/valueSets" isActive="true"
```

```

status="1" valueSetDefinitionURI="http://evs.nci.nih.gov
/valueset/C106039"

valueSetDefinitionName="FDA CDRH GUID Terminology"
defaultCodingScheme="NCI_Thesaurus"
conceptDomain="Intellectual Product">

<ns1:owner xmlns:ns1="http://LexGrid.org/schema/2010/01
/LexGrid/commonTypes">NCI</ns1:owner>

<ns2:entityDescription xmlns:ns2="http://LexGrid.org/schema
/2010/01/LexGrid/commonTypes">

A set of terminology created to support the efforts of the
FDA CDRH Global Unique Device Identification

Database project.</ns2:entityDescription>

<mappings>

<ns3:supportedCodingScheme xmlns:ns3="http://LexGrid.
org/schema/2010/01/LexGrid/naming"

localId="NCI_Thesaurus"

uri="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.
owl#"

isImported="true">NCI_Thesaurus

</ns3:supportedCodingScheme>

<ns4:supportedConceptDomain xmlns:ns4="http://LexGrid.
org/schema/2010/01/LexGrid/naming" localId="Intellectual
Product">Intellectual Product

</ns4:supportedConceptDomain>

<ns5:supportedNamespace xmlns:ns5="http://LexGrid.org
/schema/2010/01/LexGrid/naming" localId="NCI_Thesaurus"
uri="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#"
equivalentCodingScheme="NCI_Thesaurus">NCI_Thesaurus

</ns5:supportedNamespace>

<ns6:supportedSource xmlns:ns6="http://LexGrid.org
/schema/2010/01/LexGrid/naming" localId="CDRH">

CDRH

</ns6:supportedSource>

<ns7:supportedSource xmlns:ns7="http://LexGrid.org
/schema/2010/01/LexGrid/naming" localId="FDA">

FDA

</ns7:supportedSource>

</mappings>

<source>TVS_CDRH_GUID_Component</source>

<properties/>

<definitionEntry ruleOrder="0" operator="OR">

<entityReference entityCode="C106039"
entityCodeNamespace="NCI_Thesaurus" referenceAssociation="
Concept_In_Subset" transitiveClosure="true" leafOnly="true"
targetToSource="true"/>

</definitionEntry>

</valueSetDefinition>

```

There are two views of value set hierarchies in the NCI term browser.

	(See https://ncitterms.nci.nih.gov/ncitbrowser/ajax?action=create_src_vs_tree&nav_type=valuesets&mode=0) (1) Value sets grouped by Standards Authority. (2) Value sets by Source Terminology. At term browser initialization, value set metadata containing the data in all value set definition XML files are retrieved from the database through the LexEVSAPI and used for constructing the above two value set hierarchies. The first hierarchy (Value sets grouped by Standards Authority) is constructed by the value of the source, such as TVS_CDRH_GUID_Component above. The second hierarchy is constructed using the same method with an additional step that bins are partitioned in according to the value of the value of defaultCodingScheme as shown in the value set definition XML files.	
--	--	--

Discussion Points	Notes
Who are the stakeholders and end users of value sets	FDA, CDISC, and others is only through files. Otherwise this is through the editors who write value sets into the NCIt. They are browser users, and users of the ReportWriter. (LexEVS supports these) caDSR, CTRP, Cancer.gov site developers.
Define what work flow end user interface is (Shell script, Rest Service, Browser based GUI)	Shell script is acceptable
Define performance or other considerations require a move to triple store or OWL API (For Example: Do value sets need full OWL expressivity)	2017.04.24 VS Arch. Meeting - Gilberto and Larry would have to answer this question. This remains an open question following last Wednesday's meeting.
In particular do we need reification for end users so that we can understand whether queries or API's need triple store or OWL API support.	
Will non NCIt sourced value sets continue to use legacy value set definitions? (more a scope statement question) Yes.	Resolved and see notes above. At this point yes.
What considerations/requirements drive the development of an architecture that encompasses hierarchical value sets and new resolution mechanisms?	Should be able to see these as hierarchies in the browser. We need an extension for hierarchical value sets. Loading and expressing through an API. Some value sets will be in the hundreds of concepts. At least one is 8000 members large. Flavors of Hierarchical Value Sets: Source to Target Association Value Sets NICHD parent values sets. Association is read out and generates an external file. CDRH parent value sets. Association is read out to an external file. SubClassOf Based Value Set Neoplasm. Also subClassOf, is now a flat list. Should be hierarchical. Alternative Value Sets: Source to Target that are transitive restrictions found inside the thesaurus. Anatomical_Structure_is_Physical_part of is an example. LexEVS will provide an extension that loads these value set use cases as value set coding schemes with their own relational assertions. Whether this is a coding scheme or coding schemes is an open question. (Requirement)
Create OWL source for some/all values sets from LexEVS api or other source? (OWL export of value sets)	Performance based consideration, but not a requirement.
What user needs around the report writer generate requirements for LexEVS or the LexEVS team	No requirements for this – moving to sparql

Does Excel spread sheet generation fall into the scope of LexEVS value set resolution or otherwise generate requirements for the LexEVS team	Not our concern
Do the users/stakeholders in the value set API have any new requirements beyond those already stated	Better, more tailored result sets from CTS2 Rest. Not a value set requirement.
What does it mean, in terms of requirements, to provide support for Neoplasm like value sets (Hierarchical)	See above