

# Value Sets, an HL7 Perspective

W. Ted Klein, MS

Co-chair, Vocabulary Workgroup, Health Level Seven International

Member, Clinical LOINC Committee

President/CEO, Klein Consulting, Inc.

Testimony to The HIT Standards Committee, a Federal Advisory Committee

Vocabulary Task Force

September 1 and 2, 2010

## Background

Health Level Seven International (HL7) is a not-for-profit, ANSI-accredited standards developing organization dedicated to providing a comprehensive framework and related standards for the exchange, integration, sharing, and retrieval of electronic health information that supports clinical practice and the management, delivery and evaluation of health services. It was founded in 1987, and in 23 years has grown to have over 2300 members including approximately 500 corporate members who represent more than 90% of the information systems vendors serving healthcare. It is international in scope, with members and affiliates in 29 nations, and growing. HL7 has published standards for passing messages between healthcare systems, integrating applications, specifying decision support rules, structuring documents, defining clinical content of claims attachments, providing infrastructure services for vocabulary management, and many other areas. Some of the original founders of the organization are still actively involved in the activities of HL7, and available for continuity of knowledge and expertise to the thousands of members involved worldwide.

The HL7 Vocabulary Workgroup has developed a powerful and robust meta-model for Value Sets, and how they should be used in the models underlying messages, documents, and EHRs in HIT today. The model has been through many rounds of balloting, and is in wide use in Version 3 and CDA constructs throughout the world.

Klein Consulting, Inc. is a for-profit small business providing expert assistance and training to organizations, both public and private, in the Healthcare Information Technology industry, primarily in North America. It was founded in 2002, and has been involved primarily in projects in Public Health since that time. Expertise is provided to the HIT community in clinical vocabularies (primarily LOINC and SNOMED) and the analysis, design, and implementation of HL7 interfaces and their use of vocabularies.

The perspectives from both a Standards Development Organization, and an industry developer using those standards provide a set of key insights into the balances and requirements for widespread standards, and the advantages and challenges to the adoption of these standards by industry and government.

Following are observations and points of view as both a developer of standards and a user of those standards, on the questions posed for this testimony.

## Introduction to Answers

The HL7 Vocabulary Workgroup has been developing effective and deployable solutions for Value Sets for over 9 years. We have had considerable input from international groups that have been implementing robust modern HIT infrastructures for several years, as well as domestic organizations with in-depth experience in the solutions to the challenges faced in broadly managing terminology. These include Intermountain Healthcare, Kaiser Permanente, Mayo Clinic, New York Presbyterian, Partners Healthcare, and many other well known institutions. Many of the answers to the questions below use nomenclature from, or reference to, the Vocabulary Model developed over this time by HL7, and standardized in the Version 3 and the Common Terminology Services specifications. The most critical and different element of this is the concept of the difference between a “Definition” of a Value Set, and an “Expansion” of a Value Set.

A “Value Set Definition” is a machine-processable set of 1 or more formalisms that permit a specific collection of coded concepts at a given point in time to be reliably reproduced. This is absolutely required because a Value Set might be unbounded, or underlying code systems upon which it draws may change over time in ways that the cost associated with the administrative maintenance of the Value Set as a list of codes becomes a needless waste of valuable resources. Once this definition is machine-processed, it yields a “Value Set Expansion” which is a set of valid concept identifiers. The purpose of such a set in HL7 is to explicitly enable the capability where any concept identifier in a database, message, or document can be tested to determine whether it is a member of the Value Set at a specific point in time. A concept identifier in such a set may be a single concept code or a post-coordinated expression of a combination of codes.

The reason for this very precise definition and separation is because the specification of the semantics of a particular collection of vocabulary for a specific use should be kept separate from the rendering and usage aspects of the collection.

## Overall Questions

1. What are the requirements for a centralized infrastructure to implement “one-stop shopping” for obtaining value sets, subsets, and vocabularies for meaningful use?

HL7 has developed and standardized both the first generation (CTS) and second generation (CTS 2) of required functions for managing vocabulary, both of which are standards for Common Terminology Services. The first generation of the CTS infrastructure services are ANSI and ISO standards, and OMG is now working with CTS 2 developed from lessons learned from CTS. Many important requirements are defined in this standard. A summary of some key requirements that the infrastructure must be capable of for wide use are:

- Codes must have their validity in vocabularies and value sets checked;
- There must be no need for special software for access – should use common browsers and technologies;
- Access to Value Sets should be via the same syntax used to publish the Value Set identifiers in Implementation Guides;
- Subsetting of vocabularies or Value Sets should not require proprietary tools, or tools requiring specialized skill sets;
- Intellectual Property (IP) limitations on certain vocabularies should be mitigated or eliminated to ease access;

- Value Set Expansions (the lists of actual codes) should be distributed in a standard form, not any special or proprietary form;
  - Value Set Definitions should be expressed in a standard form, not any special or proprietary form;
  - Version tracking, resolution, and control of all vocabulary items, library and stored;
  - Correct and unambiguous handling of concepts expressed using post-coordination;
  - Extensions of Value Sets with local or custom content, or content that is not expressible using standard vocabularies;
  - Extensions of Value Sets with information to support user interface needs;
  - Mapping of vocabulary;
  - Clear differentiation between persisted and versioned Value Set Definitions, and generated Value Set Expansions.
2. Which requirements or functionalities are urgent, i.e., absolutely required to support “meaningful use”? Which would be most useful immediately? What would be a staged approach over time to get to the desired end state?

The following requirements are immediately urgent and absolutely required:

- All Value Set identifiers and versions (where applicable) must be clearly and straightforwardly expressed and indexed in the Meaningful Use guidelines, with easy single-click or copy-paste operations (nothing more complex) to access the content of the referenced Value Sets;
- Clear and direct access published to services that can supply the Value Sets so referenced;
- Clear and direct access to the definition of the form in which the content of the Value Sets are expressed (since the standards are only just emerging);
- Mapping of vocabulary.

Items that may be staged later include:

- Mitigation of IP restriction issues;
- Point-in-time generation of Value Set Expansions;
- Handling of post-coordination;
- Access and update of Value Set Definitions;
- Version identification, control, and tracking of Value Set Definitions as well as versions of codes in the Expansions.

### **Detailed Questions**

3. Where are you using value sets and subsets? For what domains? How many value sets and subsets?

The HL7 Version 3 ballot has created and persisted Value Set Definitions for over 10 years, and currently is publishing 1753 of them in the September 2010 ballot round. This library is maintained, with updates and modifications, 3 times every year, with 3 full releases. The following domains have standard models developed and published by HL7, and have defined value sets to support them:

accounting & billing, claims & reimbursement  
blood, tissue, and organ  
care provision  
clinical document architecture (CDA)  
clinical genomics  
clinical statement  
common product model  
decision support  
diagnostic imaging  
immunization  
informative public health  
laboratory operations

materials management  
medical records  
pharmacy and medication  
orders & observations  
patient administration  
personnel management  
public health and public health reporting  
quality measures  
regulated products & studies  
scheduling  
specimen handling and tracking  
therapeutic devices

4. In your experience with creating, disseminating, updating and/or using value sets, subsets, and entire vocabularies, what works and what does not work?

In many years of these operations with vocabulary in HL7 ballots, a number of key design decisions in the HL7 Vocabulary model have exhibited their strengths:

- Permitting sophisticated formalisms to include collections of codes in Value Sets rather than limiting them to flat enumerated lists;
- Distributing the schema (an XSD file in HL7's case) of the vocabulary model with the content;
- Generating human-readable content and documentation from the underlying XML store programmatically rather than attempting to maintain separate prose documentation;
- Making sure that Informaticists, Clinicians, Modelers, and IT staff work together to maintain the vocabulary, and facilitating good trade-off discussions during the process;
- Separating the association (binding) of a Value Set and a Data Element from both the Value Set and the Data Element, and documenting the association separately;
- Working very hard to have rapid frequent turnaround for vocabulary maintenance and publishing (3 times per year);
- Refusing to create HL7-specific vocabulary for concepts where standard and well-maintained vocabularies already exist (e.g. SNOMED, LOINC, ICD, CPT, ICF, etc.) or to copy subsets of the vocabulary into Value Sets, instead referring to them with the Value Set Definition formalisms.

Some things that we have found work poorly or not at all:

- Not having Model experts, Informaticists, and Domain Experts working together leads to failure;
- Rolling out clever but hard to learn and use non-standard tooling is an obstacle;
- Requiring special or specific architectures inhibits effective use.

5. What human resources does it take to implement and manage value sets, subsets, and entire vocabularies? Informaticists? Clinicians? IT people? How are you organized?

Experience has shown that a multidisciplinary team is required for the implementation and management of vocabulary, whether it be Value Sets, subsets, entire vocabularies, or mappings. HL7 has both administrative management and domain/semantics management sets of processes. We have 3 individuals with IT background, two of whom are Informaticists, who are the primary administrative managers of the vocabulary in HL7.

Each Domain has a mixture of Clinicians and Informaticists who develop the domain models, and submit requests for addition or modification to the HL7 vocabulary and Value Sets. It should be noted that excellent results require these participants to have some level of experience with vocabulary use.

6. What national resources and services could be leveraged to reduce the level of effort required for local implementations ? What is the irreducible minimum of local work at an implementation site, or within an organization or system?

There are a number of needs that could benefit from the application of national resources and services:

- Mapping from local to standard vocabularies;
- Education and training in best practices;
- Central FAQ resource for questions from usage to IP issues to Best Practices;
- Assistance with the huge cost of modifying existing systems.

All local sites have existing systems, the vast majority of which handle vocabulary poorly and in non-standard ways. Many vendors do not have an ROI to update the systems, and the sites do not have the resources to change the systems themselves. Some incentive-based system is needed to address this, as modifications to systems (even just changing the content of the vocabulary libraries) is an irreducible minimum task for everyone. Another is the requirement for mapping, if to do no more than be able to access historical data.

7. What is your maintenance process? How do you manage updates?

Maintenance of HL7 vocabulary (Value Sets and HL7 Code Systems) is done in 3 cycles annually. Each workgroup in HL7 constructs models which make use of, and often require modifications to, the HL7 vocabulary and Value Sets. These workgroups submit maintenance requests for the vocabulary, documented on checklist forms with content in XML or Excel. The 3 primary maintainers (the Vocabulary administrative managers) and one expert from each Domain submitting vocabulary change requests meet together in a weeklong workshop to review and either accept, modify, or reject each request, which is reviewed in detail. Then the approved requests are applied to produce a new vocabulary release using a variety of technical tools and significant manual labor. The tools and techniques employed are all HL7 'home-grown', as they began to be developed over 9 years ago, when serious vocabulary maintenance activities in HL7 began. There were few, if any, available tools to satisfy HL7 vocabulary requirements at that time. This process is repeated 3 times per year, and the new vocabulary is made available for distribution and viewing in an XML form and a rendered HTML set of documents. At this time, the entire vocabulary set is released as a new version, with a change log also released that identifies all those items that were updated from the previous release.

8. What metadata do you maintain and how do you maintain versioning?

The HL7 vocabulary model currently has rich sets of metadata for all vocabulary objects. There is a set of metadata for a Value Set Definition, a set for a Code System, a set for the content of a Code System (the codes, identifiers, relationships, display strings, and other entities published with a Code System), a set of data for the primary identifier of a Code System or a Value Set, and a set of metadata for instances of concept codes that are used in

clinical records and transactions in systems. The metadata includes semantic entities (definitions, relationships, categories, etc.) as well as administrative entities (version identifiers, intellectual property information, publishing organizations, etc.).

Every Code System uses a different mechanism and syntax for reporting its published versions. The HL7 model recognizes that these differences must be resolved through a common-denominator mechanism so that EHRs and other HIT systems can handle versioning in a single way across many different Code Systems. The mechanism HL7 employs is through a timestamp; one of the critical infrastructure requirements is to provide the answer to the question (via an API of some kind) “What version of this vocabulary object is current here as of this timestamp?” and variants of this. HL7 maintains the versions of Code System content that it produces with an ‘effective date’ and ‘retire date’ pair of timestamps for every versioned vocabulary object so that versioning can be determined. This is critical, as Value Sets may contain codes from disparate code systems, and the Value Set Definition is permitted to avoid specifying a version (this is very important for many code systems which are in wide use, such as MIME types, Country Codes, Language Codes, and other vocabularies where an administrative and processing burden of explicitly handling versioning is neither desired nor justified), or particular versions may be defined in a binding. This use of timestamp permits a data element to be bound to “the version of the Code System content version in effect at this time” regardless of the number of disparate Code Systems referenced in the Value Set. This also permits predictable behavior when local code extensions are made to Value Sets.

9. Is there a difference between versioning for clinical documentation vs. versioning for reported measures, i.e., when do you go live with a change in the EHR vs. when do you use the new version for measures?

HL7 does not implement systems, as it is a standards developing organization. In the HL7 Vocabulary Model, there is no difference in versioning between documentation and measures information; all use the same versioning model for documents, measures, messages, and databases.

10. How do you manage versioning in clinical decision support vs. changes in value sets?

The HL7 Vocabulary Model permits implementers to ‘lock’ certain Value Sets to restrict versioning impact on Clinical Decision Support. HL7 has found that it is poor practice to have changes in Value Sets be uncontrolled in an environment. For critical applications, such as decision support, it is recommended that all vocabulary items be identified with a minimum of a triplet: (code, Code System, version). In this way policies can be developed appropriate to site-specific behaviors of updates to Value Sets, and synchronized with updates to Clinical Decision Support logic.

11. How does an application know which value set is for which purpose? How is the specific context for a value set maintained at the message data element level of specificity? How is the English language intent of the value set context documented and maintained?

HL7 models contain a number of coded data elements. Each of these is explicitly linked to a specific Value Set through the use of Terminology Binding, which associates each coded data element with a conformance statement that includes Value Sets, Versioning

information, and information on the local extensibility of the Value Set for this particular coded data element. Each of these bindings has annotation notes that document the purpose of the Value Set for this specific use for this specific coded model element. The separation of this information from both the data element and the Value Set permits the use of the data element in different contexts and the use of the Value Set in different contexts. The maintenance of the binding between a Data Element and the Value Set, along with its native language notes (HL7 International has annotation comments in a number of different human languages besides English) is a separate activity, and may be done by a different group than the group maintaining either the Value Set or the data elements.

12. What are lessons learned about web links vs. storage of the vocabulary or other artifact in a physical repository?

Central vocabulary resources on the web represent a single source of truth, single access point, eased maintenance, and guarantee that multiple facilities are accessing the same vocabulary. However they have a dark side: more complex network connectivity and topology often present challenges with firewalls and network access. They are also vulnerable to network outages, and DNS problems unrelated to the facilities trying to get access and other single-point failures. However, local storage of artifacts in local physical repositories usually present greater challenges for update, maintenance, extension, and funding, as well as requiring a level of IT expertise that is sometimes lacking in smaller facilities.

13. How do you manage distribution of updates to multiple sites?

HL7 maintains a single central web access point for Value Sets and Vocabulary, and employs a 'pull' architecture for multiple sites to access the resources. The resources are not designed to be used in a 'run-time' high performance way in live systems; the 'pull' is used to refresh local repositories at sites that use the vocabulary.

14. Where is local customization appropriate and how much customization is acceptable?

Local customization is needed in virtually every location. States and Localities often have requirements that are above and beyond Federal requirements for data capture and reporting, and thus extensions must be made to any Federal standards requirements. Many vocabularies, such as LOINC, have human-readable display names for concepts that differ from those in use at many facilities. It has been shown that changing the names of lab tests can lead to Patient Safety events and there are some JCAHO requirements relating to the display of clinical items that absolutely require some customization at local sites. That having been said, the only customization that should really be put in place is the minimum necessary to ensure usability, Patient Safety, and conformance to regulations. More than that tends to interfere with interoperability.

15. How do you manage distribution of updates with local variations and optionality?

Unique subsets? Local mappings?

HL7 has a mechanism for extension of published Value Sets, but does not archive or distribute such extensions; these are all done locally. The Value Set binding mechanisms permit local extensions to be segregated into separately maintained and versioned Value Sets, and both the standard Value Set and the local variant may be bound to a data element.

Optionality is addressed with the mandatory Coding Strength attribute of a binding, which specifies whether codes that are not included in a pre-defined Value Set may be populated in a data element or not. All subsets of either whole vocabularies, or existing value sets are themselves a Value Set; this recursion permits the Value Set machinery to handle Unique Subsets in a standard way, and does not require anything special for them.

16. What has to be local in an EHR implementation vs. what can be external in a vocabulary repository?

Local vernacular display terms (in the local human language) should be local to the EHR implementation, whereas the lists of the concepts should be external so that they can easily interoperate when the contents of the EHR need to be shared across institutions. Mappings to local terminologies for local business partners should also be kept local. It appears that new concepts whose need surfaces in one local facility should probably be included in an external repository, as we keep finding that as things are discovered locally, their broad applicability and wider use later becomes apparent.

17. What functions are required that users have not yet appreciated?

There are a number of functions that are beginning to show their strengths in locations where the use of the HL7 Vocabulary model is further along, such as in the UK in the NHS efforts, and in Canada at Canada Health Infoway. Users have yet to fully appreciate the need for:

- Storage of coded vocabulary items with their version identifiers;
- Separation of Value Sets, data elements, and the bindings that connect them;
- Value Set Definitions that are more robust than flat code lists;
- Administrative costs of updating mappings in the face of code system updates;
- Distribution of vocabulary in standard encodings, such as XML with a stable published XSD defining it;
- Informaticists must participate with Clinicians and IT staff in design and decision sessions involving vocabulary.

Thank you very much for the opportunity to present the viewpoint of Health Level Seven International in these hearings.