Vocabulary Task Force Public Hearing September 1- 2, 2010

Testimony submitted by
Suresh Srinivasan, MS, Chief, Medical Language Branch
Office of Computer and Communications Systems
National Library of Medicine, National Institutes of Health, HHS

NLM's perspective is that of: (1) a producer and distributor of entire vocabularies (MeSH, NCBI Taxonomy, RxNorm) and value-added convenience subsets (e.g., CORE SNOMED CT Problem List, RxTerms); (2) a producer and distributor of concept-based terminology resources that link synonymous names and identifiers from - and provide a common distribution format for – many disparate terminologies (UMLS Metathesaurus, RxNorm); (3) a developer of vocabulary-related software, including sophisticated tools for creating and maintaining subsets from large vocabulary sources, browsing, natural language processing, etc. (4) a major funder of ongoing development and free dissemination of LOINC by the Regenstrief Institute; and (4) the US Member of the International Health Terminology Standards Development Organisation (IHTSDO). In its IHTSDO role, NLM supports US-wide access to SNOMED CT (in English and Spanish), serves as its US distributor (within the UMLS Metathesaurus and in its native formats), and is working on processes for streamlining high priority US additions and extensions to SNOMED CT.

NLM has facilitated the development and dissemination of some LOINC and SNOMED CT vocabulary value sets, e.g., for newborn screening, other patient assessment instruments, routes of administration of drugs. NLM has also supported the development and dissemination of draft use-specific mappings from clinical terminologies to HIPAA code sets (e.g. SNOMED CT to ICD9-CM, LOINC to CPT).

NLM currently provides direct access to all major vocabularies and code sets required for meaningful use (LOINC, RxNorm, SNOMED CT, CPT, ICD-9-CM, HCPCS) in value-added UMLS format and to some in their native format, either directly or via links to other robust distribution sites. NLM has current mechanisms for distributed license checking (currently necessary for access to some vocabulary resources required for "meaningful use") and for making new resources (e.g., new and updated versions of terminologies, convenience subsets, and value sets) available for basic downloading as they are created.

THEMES/QUESTIONS

We have responded to those questions for which NLM has the most directly relevant experience.

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?
   In NLM's view, the first target population for such a centralized access infrastructure should be developers (of terminology services and of EHR products). Viewed from the perspective of a developer and distributor (rather than a user) of terminologies, some minimum requirements for this user group are:
   - **Sustainability – a scope, architecture, pre-existing infrastructure, and funding approach that ensures continuing availability of whatever services are provided.**
   - **Reliable (essentially 24/7), secure and authenticated access to vocabulary resources.**
   - **At least one gateway site that allows users to easily identify, find documentation for, and navigate to places that (in combination) allow efficient downloading of the latest versions of all vocabulary resources required for "meaningful use", as soon as they are released.**
   - **Efficient electronic license checking for resources that require it. We may wish this was not necessary, but it is, at least for the foreseeable future**
   - **Validation that each new version of a resource adheres to its own technical specifications and satisfies well-defined quality assurance criteria (e.g., so that a value set of SNOMED CT identifiers does not include invalid SNOMED CT identifiers). NOTE: This service does not necessarily depend on all the vocabulary resources being served up from the same location.**
   - **Messaging service (e.g., RSS feed) that provides key information about "meaningful use"vocabulary resources, e.g., availability of updates, errors discovered, changes in vocabulary-related "meaningful use" requirements or product certification criteria, etc.**

- **Basic customer support – single point of entry with ability to answer basic questions quickly, route more advanced inquiries to experts, and direct those who need basic education about machine-readable terminologies and their use in EHRs to sources of such education. (NOTE: Our assumption is that other HITECH programs – e.g., new educational programs, Regional Extension Centers – will become the sources of such education over time.)**

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?
    - **Requirements identified under question 1 would be immediately useful and could be provided relatively rapidly by funding the development and maintenance of extensions and enhancements to existing infrastructure.**
    - **Additional functionality that would be highly desirable relatively soon includes:**
        - **Robust browser(s) for all resources**
        - **Rich set of interfaces that support both customized and standard searching, retrieval and manipulation of vocabulary resources**
        - **Efficient access to archives of previous versions of all resources**
        - **Easily customized mechanisms for automatically obtaining updates to terminologies, convenience subsets and value sets (e.g., the weekly and monthly releases of RxNorm).**

Detailed Questions

3. Where are you using value sets and subsets? For what domains? How many value sets and subsets?
4. In your experience with creating, disseminating, updating ~~and/or using value sets,~~ subsets, and entire vocabularies, what works and what does not work?
    - **What works – general dissemination:**
        - **Starting with a minimum set of services that can be sustained and delivered reliably - and then adding new ones gradually**

- **Developing new services based on input from *existing and new* users (*non-users* are not good at predicting what new services they will actually use.)**

- **What works – two specifics:**
  - **Remote license checking – NLM offers this electronic service now, for developers and others who wish to distribute SNOMED CT or other vocabularies currently distributed by NLM within the UMLS Metathesaurus, but who cannot do so on the open Internet due to license restrictions, and is developing a more robust version.**
  - **Basing convenience subsets on frequency of use/occurrence data, e.g.,**
    - **a.** CORE Problem List Subset of SNOMED CT: **http://www.nlm.nih.gov/research/umls/Snomed/core_subset.html**
    - **b.** Common Lab Orders LOINC Value Set: http://loinc.org/usage
- **What doesn't work:**
  - **Expecting (or waiting for) perfection in the vocabulary products/versions as released by producers, even along basic dimensions like agreement between the content of files and their documentation. We all make mistakes.**

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?

   NLM utilizes all of the above, and more (librarians, health policy analysts, lawyers, etc.) organized in cross-divisional teams. Some of these teams are more tightly structured than others.

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?

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

Our maintenance processes vary by product. Information about updates and changes to NLM vocabulary resources is available at:

- **MeSH -** **http://www.nlm.nih.gov/mesh/introduction.html#changes**

- **RxNorm -** **http://www.nlm.nih.gov/research/umls/rxnorm/index.html**
- **UMLS Metathesaurus –**
- **http://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/release/index.html**
- **http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=nlmumls&part=ch02#ch02.I26_Data_About_the_Metathesaurus**

8. What metadata do you maintain and how do you maintain versioning?
   See references under 7.
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?
10. How do you manage versioning in clinical decision support vs. changes in value sets?
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?
12. What are lessons learned about web links vs. storage of the vocabulary or other artifact in a physical repository?
    Access to vocabulary resources needs to be fast and reliable. Provided that the site at the end of a web link can support the volume of use essentially 24/7 (including during natural surge periods – as when new versions are released), a web link to the producer's official distribution site for an entire vocabulary has the advantage of identifying the most authoritative source; providing the most rapid access to new releases, corrections, updates, etc.; and avoiding problems that arise when it is difficult to determine which of slightly varying versions is in fact the "official release". (NOTE: This is *not* an argument against additional distribution sites which provide value-added features, e.g., additional distribution formats, integration with other resources and tools, to meet the needs of particular audiences.)

    There are relatively few entire vocabularies required under "meaningful use", however. A Web link approach to a large number of different sources for value sets needed for specific purposes may be problematic for users. One or a small number of

sites offering uniform robust access to value sets will probably be preferable.

13. How do you manage distribution of updates to multiple sites?
    Announce via Listservs and Web page posting the availability of updates and new resources to users who then come to download site.
14. Where is local customization appropriate and how much customization is acceptable?
15. How do you manage distribution of updates with local variations and optionality? Unique subsets? Local mappings?
    NLM provides UMLS licensees with the ability, on a limited basis, to create customized subsets of terminologies obtained from the UMLS Metathesaurus, using the MetamorphoSys tool.
16. What has to be local in an EHR implementation vs. what can be external in a vocabulary repository?
17. What functions are ~~required~~ **desirable** that users have not yet appreciated?
    - **At least some users will create customized views of standard vocabularies for their purposes (e.g., exclude sections not needed in their environment, suppress specific names of concepts not helpful to clinicians, create local extensions) and then update this customized view *efficiently* as new versions of the standard vocabularies are released (possibly with new content that duplicates their local extensions, involves major reorganization of content such that their exclusion strategies added features are not lost, etc.)**
    - **Tools that help users to produce the initial customizations, keep a computable record of what was done, and edit and reapply the formula when new versions appear would be highly desirable. (Some functionality of this kind is available in MetamorphoSys, a tool for customizing the Metathesaurus).**
    - **It would also be highly desirable if the "one stop shop" could provide early warning, guidance, tools, tips, etc. that would alert users to the kinds of changes that were highly likely to have a major impact on local customization**