Health Information Technology Standards Committee

Vocabulary Task Force

Science Applications International Corporation

Testimony

Presented By David O. Dobbs.
Director of Medical Informatics
Health Solutions

September 2, 2010

Ms. Humphreys, Mr. Ferguson and members of the HIT Standards Committee and Vocabulary Task Force, I am David Dobbs, and I serve as a Program Manager and Director of Medical Informatics for the health and life sciences businesses at Science Applications International Corporation (SAIC). Speaking on behalf of our medical informatics team, I want to thank you for giving me the opportunity to provide our thoughts on what constitutes the right set of requirements for a "one stop shop" vocabulary infrastructure, and which requirements should have the highest priority given the drive to achieve the goals of Meaningful Use of electronic health records (EHRs).

SAIC has been engaged in informatics and information technology development in the healthcare, public health, and life sciences domains for many years. We have been active participants in the development of standards within Health Level Seven (HL7) and the Health Information Technology Standards Panel (HITSP), and in the implementation of health information standards in healthcare, public health, and biomedical research organizations.

We have extensive experiences in developing and maintaining vocabulary subsets through our work for the National Cancer Institute and the cancer bioinformatics grid (caBIG® [U.S. Department of Health and Human Services]) where we are developing domain models and value sets that improve translational research, collaboration and information exchange capabilities among cancer researchers in domains ranging from genomics to nanotechnology.

Likewise, we have assisted the Centers for Disease Control and Prevention (CDC) on the Public Health Information Network, or PHIN, in developing and maintaining Implementation Specifications and their associated value sets. We were the original developers of the CDC's PHIN Vocabulary Access and Distribution System that makes values sets of interest to public health available through a web site.

SAIC has developed its own set of software tools to facilitate better enterprise vocabulary management and distribution. Our internally developed software, called the Terminology Management and Provisioning Service, or TMAPS, is an open source vocabulary management and distribution service that stores vocabularies, value sets and maps and enables vocabulary consumers to subscribe to, and automatically receive, vocabulary sets. These, as well as a myriad of other healthcare interoperability experiences, have enabled us to gain a keen insight into managing and distributing vocabularies and value sets.

Based on our experiences, the first step in developing a "one stop shop" for obtaining vocabularies and value sets is to identify the corpus of vocabularies and value sets that need to be distributed. A cursory glance at HHS ONC's EHR Meaningful Use Final Rule uncovers a relatively easy to manage set of vocabulary standards for problems, procedures, laboratory test results, medications, vaccines, race and ethnicity. A more thorough consideration uncovers the value sets associated with the 10 Content Exchange Standards and Implementation Specifications. These cover hundreds of value sets that have been developed and are being maintained by a multitude of separate organizations including HL7, the American Society of Testing and Materials (ASTM), the National Council for the Prescription Drug Programs (NCPDP) and the now defunct HITSP. Another key source of value sets are those associated with calculating the numerator, denominator and exclusion criteria for the 44 clinical quality measures identified by CMS' EHR Meaningful Use Final Rule.

In total, there are well over 500 value sets and scores of vocabularies that are required to meet all of the EHR Meaningful Use requirements. The desire to collect all these value sets and make them available from one place is certainly a worthy goal. Likewise, the complexity of collecting these value sets from a multitude of organizations and maintaining up-to-date versions in one place is daunting.

In addition to vocabularies and value sets, implementers will need concept mappings between standards-based concept codes in order to meet some EHR Meaningful Use criteria. An example that SAIC has recently encountered of the need for concept mappings is the generation of the HHS ONC selected HITSP C32 patient summary document. Generating HITSP C32 documents is one way to meet the Core Meaningful Use Objective for providing patient clinical summaries. The conformance rules that HITSP developed for the C32 document require that problems and procedure be encoded using <u>only</u> SNOMED CT® codes. In the real world, problems typically are encoded with ICD-9 codes and procedures with CPT codes. In order to conform to the HITSP C32 patient summary implementation guide, and to pass the National Institute of Standards' (NIST) C32 conformance testing requirements, SAIC has had to find mappings between ICD-9 and CPT codes and SNOMED CT® codes. These include literally thousands of individual concept mappings. These mappings, as well as other types of concept mappings such as those between HL7 Version 2 and HL7 Version 3 vocabularies, are needed to conform to some EHR Meaningful Use criteria.

Now I would like to turn my attention to the functionalities needed by a "one-stop-shop" for vocabularies, value sets and maps. In my ensuing remarks, I will refer to the "one-stop-shop" as a "vocabulary distribution source" and the collection of vocabularies, values sets and maps as "vocabulary objects".

Of great importance in developing a vocabulary distribution source is the need for a well crafted canonical data model for storing the requisite vocabulary objects. When SAIC developed its own vocabulary distribution software, TMAPS, we spent considerable time designing a canonical data model that took into account a number of very important considerations, including:

- Following HL7 vocabulary principles such as using Object Identifiers, or OIDs, for vocabulary objects
- Using the ISO/IEC 11179 standard for metadata data registries to identify appropriate metadata for vocabulary objects such as the source name, source release date, and source definition
- Clearly identifying the authority that is responsible for developing and maintaining the vocabulary object
- Versioning of all vocabulary objects, and
- Keeping the canonical data model simple, so as to be easily understood by those that needing to implement vocabulary within their EHR and healthcare integration applications, while still having enough information in the canonical data model to tie back to more powerful and complex tools such as the UMLS Metathesaurus, RxNorm and other ontologies.

Other important functionalities for a vocabulary distribution source include:

- Creating sets of vocabulary objects that support specific use scenarios, such as: those required to implement the 10 ONC identified Content Exchange Standards and Implementation Specifications; and those required for specific, or sets of, clinical quality measures.
- Allowing vocabulary users to subscribe to vocabulary sets and individual vocabulary objects. When changes occur to vocabulary objects, subscribers should automatically receive a notice of the change and, optionally, have those changes automatically sent to them.

- Providing a web based user interface for viewing, browsing, searching and downloading vocabulary objects.
- Having a flexible set of vocabulary download formats including XML, character delimited, OWL/RDF and Excel.

Finally, a vocabulary distribution source should include information on vocabulary principles and best practices to educate those who are trying to implement vocabulary into their EHR systems and healthcare integration applications. Lay vocabulary users need practical knowledge, such as: the different uses of object identifiers for coding systems and value sets; how to handle identifying coding systems in HL7 Version 2 and HL7 Version 3; and how to look up the meaning of a concept code from a medical document such as a HITSP C32 patient summary. Making vocabulary objects available through a vocabulary distribution source will do no good if those who have to implement vocabulary objects do not understand basic vocabulary principles.

Again, I thank you for giving us this opportunity to share our experiences, lessons learned, and opinions on what constitutes the right set of requirements for a vocabulary distribution source infrastructure. And, which requirements should have the highest priority given the drive to achieve the goals of EHR Meaningful Use. Having readily available and easy access to vocabulary that supports EHR Meaningful Use is vital to improving health care quality, preventing medical errors, increasing care coordination and reducing unnecessary healthcare costs.