

Health Information Technology Standards Committee

Clinical Operations Workgroup  
Vocabulary Task Force

**Science Applications International Corporation**

Testimony

Presented By Dixie Baker, Ph.D.  
Chief Technology Officer  
Health Solutions

February 23, 2010



Dr. Blumenthal, Mr. Ferguson, Ms. Humphreys, and members of the HIT Standards Committee and Vocabulary Task Force, I am Dixie Baker, and I serve as the chief technology officer 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 how vocabulary subsets and value sets should be created, managed, distributed, and supported to enable meaningful use of electronic health record (EHR) technology and electronic health information.

SAIC has been engaged in informatics and information technology development in the healthcare, public health, and life sciences domains for many years. In the early 1990s, we developed a clinical information system for the Department of Defense and deployed it throughout the world. We have continued to support that system and its successors ever since. 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. Our work with the Centers for Disease Control and Prevention (CDC) on the Public Health Information Network or PHIN, and on programs such as the National Electronic Disease Surveillance System (NEDSS), BioSense, and the National Healthcare Safety Network have significantly advanced the cause of interoperability and standards-based information exchange in the public health domain since 2001. Likewise, our work with the National Cancer Institute and the cancer bioinformatics grid (caBIG<sup>®</sup> [U.S. Department of Health and Human Services]) in developing domain models and value sets has improved translational research, collaboration and information exchange capabilities among cancer researchers in domains ranging from genomics to nanotechnology.

We have first-hand experience in the difficulties associated with adopting and using controlled medical terminologies. The terminology space is a broad and dispersed environment and getting one's arms around the numerous coding systems and the even more numerous value sets defined for specific use cases and messaging standards is a daunting challenge. To address this reality, for the past several years SAIC has sponsored an internal research and development (IR&D) project to develop software tools to facilitate better enterprise vocabulary management and distribution. We now have a working software solution that could serve as a model for the creation of a national capability for creating, managing and distributing coding systems, vocabulary subsets, and value sets that can serve the healthcare industry as we move aggressively toward ubiquitous adoption of EHR technology and electronic information exchange at a national level.

I would like to preface our answers to the specific questions the Vocabulary Task Force has posed with some general observations and comments based on our experience. In part, I suppose I am answering Question 9 first to establish some context from which to answer the other questions.

Looking back at the healthcare industry of the late 1980s and early 1990s, when healthcare delivery organizations were first acquiring clinical information systems technology and community health information networks (CHINs) were the wave of the future, it is astonishing to realize that we have yet to achieve broad adoption and use of semantically interoperable electronic health records systems – though we have gone through several conceptual iterations and renamings of the “CHIN.” We certainly underestimated both the complexity and cost of this undertaking. At the same time, we see that significant progress has been made. In particular, the tremendous body of work produced by our industry and its standards development organizations (SDOs) like HL7, the American Medical Association (AMA), Regenstrief Institute, and the International Health Terminology Standards Development Organization (IHTSDO), and profile development organizations, like Integrating the Healthcare Enterprise (IHE) and HITSP. These organizations and others have provided a solid foundation on which to build and extend EHR technology in the years to come. We owe a great debt of gratitude to these organizations and the hardworking volunteers and subject matter experts who have contributed so much to interoperability in our industry.

The development of healthcare information standards, both controlled medical terminologies and information exchange standards, is difficult and exacting technical work. Achieving consensus while developing standards can be a time-consuming process, often involving heated debates, as well as formal balloting and ballot comment reconciliation. Standards developers can have strong ideological biases that hamper consensus building, such as whether Logical Observation Identifiers Names and Codes (LOINC<sup>®</sup> [Regenstrief Foundation, Inc.]) or the Systematic Nomenclature of Medicine – Clinical Terms (SNOMED<sup>®</sup> CT<sup>®</sup> [International Health Terminology Standards Development Organisation]) is better suited for coding a particular concept. Users of standards can differ in their needs relative to granularity and precise purpose for even seemingly simple concepts like ethnicity, race and gender. As a result, during implementation, “standard” value sets may need to be customized to accommodate the needs and interests of the users.

In our view, the definition of value sets and subsets is generally use-case-specific and may be outside the responsibility of the SDOs themselves. Subsets and value sets often fall into the realm of implementation and application of a standard or an industry best practice. As an example, the lists of LOINC and SNOMED CT codes that define the laboratory results for infectious diseases that are reportable to public health, like tuberculosis, are defined by the Council of State and Territorial Epidemiologists (CSTE) working with the CDC and may be revised by each state before final adoption. State laws differ in terms of what diseases and conditions are reportable. When a reportable finding is observed in a clinical laboratory, the standard for electronically reporting that result to public health is an HL7 v2.5.1 ORU message. The message defines the structure and encoding requirements generally applicable across all reportable infectious disease conditions but is not intended to provide guidance on what is reportable. So here we have an example of a situation in which a standard, the set of reportable measures, and the value sets necessary to implement the standard and to report the measures, are developed by three different organizations!

The Vocabulary Task Force has recognized that the definition of quality measures frequently incorporates value sets in the form of lists of standard codes for inclusion and exclusion in the numerator and denominator of the measure. Each of these value sets is an important component to proper implementation of the measure and will be modified and changed from time to time as medical practice and science evolve. In this case, the quality measure developer is defining both the quality measure and the value sets that comprise that measure, and all must be easily accessible to EHR technology developers who must implement capabilities to support the measures and to providers who must report the measures to the Centers for Medicare and Medicaid Services (CMS).

I offer the following points in considering how the industry should move forward to more effectively facilitate the adoption and meaningful use of EHR technology through value sets and subsets of controlled medical terminology.

1. When value sets are developed, the needs of clinical users and patient care should be given priority over the needs of secondary users and reporting requirements – including meaningful use reporting.
2. Value sets should be developed and maintained by those most invested in their use and most qualified to define and maintain their composition.
3. Adoption of value sets and subsets can be accelerated by providing services to facilitate their publication, updating, and access.
4. Where consistency in use of a value set or subset is essential to interoperability or meaningful use, it should be specified as a standard in an EHR standard, implementation specification, or certification criteria.

I'll now address five (5) of the questions provided by the Vocabulary Task Force. I'd like to begin by addressing Questions 1 and 2 together.

1. ***Who should determine subsets and/or value sets that are needed?***
2. ***Who should produce subsets and/or value sets?***

We all understand that healthcare information must be portable. Healthcare information collected at one point of care must be easily understood and accurately interpreted at another point of care. The development of value sets is not a one-time task that is undertaken and completed. It is a continuous effort by healthcare delivery organizations and healthcare SDOs nationwide and worldwide. We believe this state of flux is unlikely to change. There is no single, central location where need is established and no central location where needs are satisfied. Our industry is supported by a large number of organizations equipped to apply open, consensus-driven processes to the establishment of new healthcare information standards. Needs will be identified from throughout our healthcare system. The terminology and vocabulary needs in an endocrinology practice will differ significantly from those of a cardiology practice. The needs of a clinical laboratory differ greatly from those of an imaging center. These domains are unique, and the practitioners in these domains are best qualified to understand how controlled medical terminologies and value sets

drawn from them can improve the quality of care and the interoperability of healthcare systems.

We believe that the best results in developing vocabulary sets will always be derived when clinical domain experts are brought together with informaticists with strong competencies in controlled vocabularies, data exchange standards, and software development.

### ***3. Who should review and approve subsets and/or value sets?***

Where subsets and value sets are developed as elements of an industry standard, regulation, or certification criterion, the authority to review and approve them is already established. The Vocabulary Task Force however suggests that EHR adoption rates can be improved, that EHR technology can be more meaningfully used, and the degree of encoded data within electronic health records can be increased by making more value sets available to healthcare providers. To do so, a framework and clearly defined processes and services must be established for submitting proposed value sets for review, adopting them for broad use, publishing them for easy access, distributing them to developers, and maintaining them over time. We believe this concept merits careful consideration and may fill a long-standing gap in our health IT standards space.

The Health Information Technology for Economic and Clinical Health (HITECH) Act amended the Public Health Service Act (PHSA) to improve healthcare quality, safety, and efficiency through the promotion of health information technology and electronic exchange of health information. Section 3004 of the PHSA redefines how the secretary of Health and Human Services (HHS) adopts standards, implementation specifications, and certification criteria. Under this authority, the secretary should adopt subsets and value sets as “standards,” with recommendations provided by the Office of the National Coordinator (ONC) and the National Library of Medicine (NLM). In addition, the HHS should provide facilities and services for enabling value-set developers to register value sets for consideration as standards and for making adopted value sets available to vocabulary consumers. We believe the NLM, with its legacy of providing the industry convenient access to medical vocabularies through its Unified Medical Language System (UMLS), is the most logical and qualified candidate to provide this function.

### ***4. How should subsets and/or value sets be described, i.e., what is the minimum set of metadata needed?***

SAIC recommends that metadata elements for a value-set repository conform to the ISO/IEC 11179 standard for metadata data registries (MDR). We offer the following list of items as necessary metadata for a value set.

- Unique Identifier
- Value Set or Subset Name
- Owner/Author

- Vocabulary source(s)
- Version
- Effective Date
- Created Date
- Updated Date
- Description or Purpose
- Group(s) (used to package value sets all called by the same standard)

5. *In what format and via what mechanisms should subsets and/or value sets be distributed?*
8. *What best practices/lessons learned have you learned, or what problems have you learned to avoid, regarding vocabulary subset and value set creation, maintenance, dissemination and support services?*

We have combined our responses to Questions 5 and 8. Regarding format, a wide variety of formats could be used to provide value sets. Today, simple spreadsheet files are the most accepted. We see three important considerations for the format of value-set distributions. First, the format must be able to accurately and completely convey all the metadata associated with the value set, including relationships, hierarchies and extended metadata properties that may be unique to a particular value set. Second, the format should be as simple and straightforward as possible so as not to require the value-set consumer to understand abstract vocabulary or ontological frameworks in order to load the value set into their healthcare system. Third, the format should be machine-readable.

SAIC has considerable experience in developing and distributing terminology subsets and value sets, having served as the prime contractor for a number of years on the CDC's Public Health Information Network Vocabulary Access and Distribution Service (PHIN VADS). Through PHIN VADS, CDC made value sets of importance to public health available via a Web site. The PHIN VADS Web site allowed the value sets to be browsed and manually downloaded. While PHIN VADS was a big step forward in making value sets accessible, it was still not an ideal solution for distributing value sets when considering the fact that a value-set consumer may have a need for a hundred or more value sets to accurately encode and share semantically interoperable health data. Even a messaging standard for a single specific use like an immunization record may require implementation of dozens of value sets. Requiring a value-set consumer to regularly visit a Web site to check the status of their more than 100 value sets to determine whether new versions have been published is unrealistic.

Through our experience with PHIN VADS, we discerned that as EHR technology employing controlled medical vocabulary was adopted more broadly, the need for processes and services to support vocabulary management and provisioning would become ubiquitous. This observation led us to propose to SAIC an internal research and development (IR&D) project to create a solution for providing terminology management and distribution services to the healthcare industry. Using the UMLS as

our authoritative source, the software suite that we built provides the capability for vocabulary consumers to subscribe to services that provide and manage the value sets that they require. The idea is that after subscribing to one or more value sets, the consumer automatically receives new versions of those value sets either periodically or as they become available, depending upon the consumer's preference. We have refined the service concept further by allowing consumers to subscribe to groupings of value sets – such as all demographic value sets or all value sets associated with a particular quality measure, for example. The subscriber can indicate the frequency at which to receive updates from the centralized service. In this way, a vocabulary consumer could subscribe to all value sets associated with the “Stroke 3” quality measure and indicate that they want updates to the Stroke 3 value sets sent to them every quarter. The subscription service would check quarterly for any updates to the Stroke 3 value sets and send those to the subscriber. The subscriber would have full control over the frequency of receiving value-set updates so that they coincide with their internal processes, such as those for software release cycles. Subscribers are given the choice of having value-set updates electronically “pushed” to them or receiving a notice that an update is available for download from the central site.

SAIC will be participating in the IHE Interoperability Showcase at the Healthcare Information and Management Systems Society (HIMSS) Conference in March as part of the Quality Reporting scenario sponsored by the CMS. In the demonstration, our terminology management and provisioning service will be integrated with a Healthcare Quality Measures Format (HQMF) processor and an EHR repository to produce the “Stroke 3” quality measure report. We have now completed three years of IR&D on this service, and we are exploring ways in which we can openly and productively share our concept, data model, software, and lessons learned with the informatics community.

**10. *What must the federal government do or not do with regard to the above, and/or what role should the federal government play?***

The federal government can exert a positive impact on the pace of health IT adoption and the realization of the associated benefits. The HITECH Act is directed largely at that goal. Specifically relating to terminology standards and value-set development for EHR users, we suggest the following roles and focus areas for the federal government:

- Federal policy is needed to shape and guide the adoption and use of vocabulary standards by the industry. The Consolidated Healthcare Informatics (CHI) initiative started this work, the HITSP used the CHI products, and this HIT Policy and Standards Committee should continue to build on these efforts.
- Federal funding will be necessary to eliminate barriers to adoption of standards. This includes providing national licensing for necessary standards as was done for SNOMED. The same needs to be done for HL7 and other standards involving restrictions on free access. Many SDOs derive a significant portion of their income from licensing or charging for access to

their standards. New avenues should be explored for compensating these SDOs so that the standards they develop are freely available in the U.S.

- The federal government should either provide an open, collaborative infrastructure for contributing, developing, maintaining, and distributing terminologies, value sets and subsets, or provide financial incentives to the healthcare market to develop and support that infrastructure.
- Finally, a comprehensive review should be undertaken of federal rules and regulations in the healthcare industry to identify existing regulations that may be barriers to the adoption of healthcare standards. As an example, particular aspects of enforcement of regulations in the Clinical Laboratory Improvement Amendments (CLIA) may be impeding adoption of vocabulary standards by laboratories, thus having an unintended impact on downstream systems such as EHRs. Today, a large percentage (approximately 90 percent) of laboratory microbiology results transmitted electronically are still formatted as text reports instead of being encoded using a standard such as SNOMED CT. One reason for this is that when the laboratory is evaluated for CLIA compliance, paper reports are compared to the electronic reports. The clinical content of the two are expected to be identical. The simplest approach for the laboratory to make certain this is the case is to use the identical text for both. Since coded information is not normally included in the paper reports, there is an obvious disincentive to include them in the electronic report.

Again, I thank you for giving us this opportunity to share our experiences, lessons learned, and opinions about how to facilitate the use of controlled medical terminology to create semantically interoperable EHR systems and information exchanges. You've heard me say that privacy and security can either make or break EHR adoption and healthcare information exchange – I would say the same about semantic interoperability, which is enabled by the effective use of vocabulary value sets.