Creating Interoperable Guidelines: Requirements of Vocabulary Standards in Immunization Decision Support

Karen M. Hrabak MSN RNC[1], James R. Campbell MD[2], Samson W. Tu MS[3], Robert McClure MD[4]

[1]The Nebraska Medical Center, Omaha, NE; [2] University of Nebraska Medical Center, Omaha, NE, [3]Stanford Medical Informatics, Stanford University, Stanford, CA; [4]Apelon, Inc., Ridgefield, CT

ABSTRACT

Interoperable support of electronic health records and clinical decision support technology are central to the vision of US information infrastructure. Efforts to implement interoperable clinical guidelines for immunization practice have been sparse.

We used the SAGE knowledge workbench to develop a knowledge base to provide immunization decision support in primary care. The semantic content required to completely capture CDC clinical decision logic required 196 separate concepts but was completely captured with SNOMED CT and LOINC.. Although 88% of concepts employed pre-coordinated codes, 6% of guideline concepts required expanded vocabulary services employing Boolean logical definition using two or more SNOMED concepts. Post-coordination requirements were minimal, representing just 6% of guideline semantic concepts. Heavy use of SNOMED CT hierarchical relationships at run-time improved model efficiency while supporting aggregation services.

We conclude that creation of interoperable knowledge bases employing federal clinical vocabulary standards is achievable and realistic. Employment of information model (HL7 RIM) and vocabulary (SNOMED CT, LOINC) standards is a necessary and feasible requirement to achieve interoperability in clinical decision support.

INTRODUCTION

An expert system employs an inference method to a domain ontology[1], ideally evaluates patient state data from the electronic health record, and issues recommendations for care. Advances in clinical vocabulary development[2] and agreement at a national level[3] have promulgated a core set of reference terminologies which offer to evolve into the comprehensive clinical ontologies needed for electronic health record (EHR) technology and decision support. SNOMED CT and NDF-RT are core reference terminologies within these recommendations. Laboratory LOINC is a controlled vocabulary with reference features also within the core. At this time however, implementation of these standards is infrequent despite elimination of barriers of acquisition cost, in part related to confusion regarding best deployment and difficulties with conversion of legacy data. While a single report[4] has argued that these terminologies may be insufficient to support guidelines, the functionality of these terminologies relative to representation of domain knowledge for a guideline expert engine poses complex issues not yet resolved.

Clinical practice guidelines seek to standardize care and facilitate the provision of evidence-based care. Historically published in free text formats, efforts to encode and implement guidelines within the EHR for clinical decision support face many challenges. These include ambiguity within the source publication and confusion regarding precise formulation of the guideline logic[5]. Concept modeling problems exposed by guideline encoding include differences in granularity and definition between the guideline and the domain ontology and interactions with the vendor information model.

The National Vaccine Advisory Committee guides best practices surrounding immunization administration.[6] Included in their recommendations is an emphasis on accurate vaccine recording practices and a method to send clinical reminders to patient and practitioners? Other investigators have recognized the importance of this domain and have worked to create decision support for immunization practice. These past efforts have successfully modeled immunization clinical decision support forecasting and reminder systems.[7,8] The IMM/SERV system supported childhood immunization forecasting and maintained a web based knowledge maintenance and testing environment[9]. The immunization reminder recall system[8] provided immunization decision support utilizing a modular architecture. Concepts in its knowledge base have

been mapped to medical entities dictionary (MED) employed at the New York Presbyterian hospital to provide integration with clinical records at that facility. Representation and maintenance of the knowledge domains in these systems employed tabular, rules based and procedural approach. Neither of these past implementations has employed clinical vocabulary standards. The decision models employed were unique to the environments in which they were developed.

The SAGE[10] consortium is a collaboration of academic and private sector interests with the shared goal of creating interoperable guideline decision support. In order to support interoperability, the knowledge bases created for SAGE employ only the NCVHS core vocabulary resources. Knowledge modeling occurs on an open source workbench created with the Protégé [11] knowledge tool.

The program interface between the SAGE inference engine and the vendor clinical information system communicates via an HL7 RIM compliant query engine termed the virtual medical record (vMR). For purposes of inference support the SAGE engine employs a suite of vocabulary services which bind the decision support software to the SNOMED CT structures. This binding provides ontologic features of subsumption and concept definition. Concept definition beyond the pre-coordinated scope of SNOMED CT is handled with vocabulary service extensions and post-coordination within the SAGE SNOMED CT extension. The end result is a knowledge construction and domain ontology (extension of SNOMED) which can freely interoperate with any other SAGE compliant system.

Given the clinical importance of immunization practices and the historical efforts of other decision support scientists, achieving interoperability through the use of standard terminologies is critical. We therefore organized, enumerated and characterized the vocabulary and knowledge services of the SAGE immunization guideline in order to inform the concerns of EHR vendors and emphasize the benefits of vocabulary standards compliance.

METHODS

Guideline clarification and logic modeling

The Center for Disease Control through the Advisory Committee on Immunization Practices issues guidelines with specific recommendations for vaccination of child and adult populations[12]. We obtained the immunization recommendations (Fall 2005) for children and adults for this analysis.

We compiled indications, contraindications and deferral criteria for each vaccine. Age specific criteria for eligibility, dosing intervals, and catch up rules for missed vaccine doses were identified. These criteria were used to formulate a knowledge base specification document containing logical IF-THEN statements which formalized the CDC logic while employing the source guideline concept statements..

From the logic base we compiled an inventory of concept references employing methods we have described[13] and worked with local clinical experts to disambiguate guideline statements which were unclear. For example, the guideline source concept ‘progressive neurologic disorder’ required clinical domain expert definition, resulting in a logical union of the concepts ‘Lennox Gastaut’, ‘Tuberous sclerosis’ and the nested union of concepts ‘Developmental delay’ and ‘Encephalopathy’.

We then compared our concept inventory against pre-coordinated concepts from SNOMED CT and LOINC. When source concepts were not pre-coordinated but could be formulated correctly with logical constructions of pre-coordinated concepts, we did so. Remaining concepts that were clearly outside of the scope of pre-coordinated SNOMED CT were modeled into an extension namespace for the guideline following editorial principles published by the College of American Pathologists[14].

All immunization logic rules and vocabulary concepts were then linked to a set of EHR queries employing the idealized record structure (vMR) which we have developed with the HL7 Clinical Decision Support Technical Committee. (Table 2 lists the primary objects of the vMR in the left hand column) This link involved an explicit assertion of the expected EHR records required and the attributes required to bind the decision logic to the clinical patient record.

Recognizing that guideline statements sometimes requested unique concepts for query from the record, while other statements implied retrieval from within a

set of concepts, we reviewed the guideline to clarify the vocabulary services required. We categorized all vocabulary service requirements on a scale reflecting the complexity of the concept relative to pre-coordinated NCVHS vocabularies, and the expected query function in the run-time environment.

Vocabulary service

Categorization of concept employment

Category 1: Concept instance only is directly referenced by the guideline logic. This category includes instances such as gender code, qualifier values (contraindicated, or true) and lab codes (e.g. Hepatitis B surface antigen, Measles virus IGG antibody).

Category 2: Concept with all specializations are implicitly referenced by the guideline. This category includes concepts that may have many conceptual variations within the EHR (such as "Diabetes mellitus" or "Hemoglobinopathy") and the guideline expects all more specialized children of the concept to be included at run-time query.

Category 3: Boolean constructions

3a: Guideline concept is represented by the logical Boolean 'OR' (Set union) of two or more category 2 references. For example "Functional or anatomic asplenia" is logically defined by the union of the concepts: "Splenectomy", "Functional splenectomy", "Congenital asplenia", "Sickle cell disease", "Asplenia syndrome" and "Hyposplenism" - including children - within the SNOMED CT ontology.

Category 3b: Concept requires the logical Boolean 'AND' (Set intersection). An example of a category 3b concept is 'bisexual male' which is the intersection of the concept sets 'Patient is male' and 'Bisexual'

Category 3c: Concept expression includes a Boolean 'NOT' (Set complement). Concepts in this category include concept expressions that exclude descendant concepts within a hierarchy. For example, the guideline concept of "Chronic respiratory disease" when clinically reviewed, was defined to exclude the SNOMED specialization concepts of "Chronic rhinitis" and "Chronic sinusitis" at run-time.

Category 4: Concept post-coordination required. This category includes concepts requiring extension development for SNOMED CT since they are not pre-coordinated and cannot be defined from logical statements employing pre-coordinated concepts (an example includes "day care worker" or "untreated active tuberculosis").

SAGE knowledge modeling

Employing the immunization rule logic and the concept inventory, we then proceeded to model the complete guideline using the SAGE guideline ontology[10]. Context of care, clinical workflow and organizational resources are elements of the SAGE ontology. All decision logic rules and vocabulary queries were bound and modeled employing SAGE formal criteria and SAGE actions linked to the vendor EHR. For comparison with previous immunization guideline work, we counted and summarized these execution elements.

The SAGE guideline workbench produces an XML knowledge base that can be shared between clinical systems. We validated the immunization knowledge base with a series of experiments including syntax checking of the XML and simulated run-time assessment employing test cases. We are now validating the knowledge base against actual clinical records in a test environment.

RESULTS

Vocabulary inventory

This "birth-to-death" immunization knowledge module was a complex construction. The 45 pages of clinical guideline publication were distilled into 75 separate "IF-THEN" statements in support of three clinical implementation scenarios proposed by the clinical team. The scenarios included vaccination advice at birth, the primary care office visit, and a survey scenario for population based reminders. An inventory of the source utterances from the guideline statements yielded 147 ACIP conceptual references. Disambiguation and expert clinical opinion was required with 7 concepts which were then defined within the SNOMED CT extension ontology. Table 1 provides a summarization of the pre-coordinated vocabulary concepts that were ultimately required to support guideline logic. These concepts were installed in vMR queries as data type restrictions which defined the value sets for retrieval of information from the EHR by the SAGE decision engine. Each query employed one or several coded

concepts from the distinct concepts tallied for the guideline.

Table 1
Pre-coordinated concepts by semantic type
(Category 1 and 2 concept complexity)

SNOMED domain	*n*
Context dependent category	2
Disorder	51
Finding	25
Observable entity	12
Occupation	7
Organism	1
Person	1
Procedure	18
Product (clinical drug)	33
Qualifier	16
Racial group	1
Substance	5
Total	172

Table 2
Concept inventory by vMR query class

vMR query	SNOMED	LOINC
Adverse reaction	41	
Agent	0	
Alert	0	
Appointment	0	
Encounter	1	
Goal	0	
Observation	47	5
Order	7	
Medication Order	21	
Problem	72	
Procedure	17	
Referral	0	
Substance administration	32	

Table 2 summarizes the final analysis of vMR queries to support rule logic with a total of the vocabulary concepts required.

Not all guideline concepts could be accurately modeled employing pre-coordinated SNOMED CT. Table 3 summarizes the complexity of the vocabulary model and services required to support immunization guidelines. This reflects run-time management requirements (only category 1 concept references do not require retrieval of data sets which include all children of the concept) as well as the requirements for post-coordinated vocabulary development (category 4 concepts).

Table 3
Concept inventory by complexity

Category	*n*
Category 1 (Concept entity)	35 (17.9%)
Category 2 (Subsumption)	137 (81.1%)
Category 3a&b (Booleans without NOT)	9 (4.6%)
Category 3c (Boolean with negation)	3 (1.5%)
Category 4 (Post coordination)	12 (6.1%)
Total	196

Characteristics of knowledge model

SAGE employs the frame-based knowledge modeling of the Protégé environment. Immunization logic criteria are formulated into frames which enforce a set of constraints on data query from the EHR. Criteria are employed within decision models which reproduce the source guideline logic and communicate with the vendor record via action specifications. The full immunization knowledge model required 236 Boolean criteria, 207 presence criteria and 161 comparison criteria. These were employed in 88 decision models which employed 82 action specifications.

DISCUSSION

The development of reference terminologies for clinical vocabulary standards has created utility but also poses new challenges for the knowledge information specialist. Previous studies[15] have documented the limitations of pre-coordinated terminologies, but a commitment to compositional forms means that procedures and methods for management of post-coordination must be developed.

In contrast to a previous report[4], we found that

comprehensive coding in support of our guideline was feasible, but that vocabulary services for the guideline engine had to be extended to include support for two services: 1) Boolean definitions of complex concepts and 2) integration of post-coordination within a SNOMED extension. Since management of large extension vocabulary sets requires new skills and software functionality such as description logic classifiers, this is a matter of developing understanding within the informatics community.

Run-time support provided by reference terminologies such as SNOMED CT are also important to decision support engines. Our experience clearly documents the significance of support for aggregation within record query activity. 80-90% of queries into the EHR were searching not just for a single concept, but for one within a related set. By providing for identification of all specializations of a concept with hierarchical relationships, SNOMED CT supplies knowledge structures which replace the need for exhaustive code list generation in knowledge bases. This defines a clear benefit resulting from standard reference terminology deployment, as well as an important use case for evaluating evolution of these vocabulary systems.

CONCLUSIONS

It is feasible to implement guideline decision support within a knowledge engine employing NCVHS content standard vocabularies. Effective use of these reference terminologies requires new procedures for vocabulary management and deployment. Benefits to the knowledge engineer include savings in domain knowledge development and true semantic interoperability.

Acknowledgement

This SAGE project has been supported by grant 70NANB1H3049 of the NIST Advanced Technology Program.

References

1. Musen MA. Scalable architectures for clinical decision support. Meth Inf Medicine. 1999; 38: 229-38.

2. Wang AY, Sable JH, Spackman K. The SNOMED clinical terms development process: refinement and analysis of content. Proc AMIA Symp. 2002:: 845-9.

3. http://www.hhs.gov/news/press/2004pres/20040506.html

4. Dykes PC, Currie LM, Cimino JJ. Adequacy of evolving national standardized terminologies for interdisciplinary coded concepts in an automated clinical pathway. J Biomed Inf. 2003. 36:313-325.

5. Tierney WM. Improving clinical decisions and outcomes with information: a review Int J Med Inform. 2001 Jun;62(1):1-9.

6. Poland G, Shefer A, McCauley M et al. Standards for adult immunization practices. Am J Prev Med 2003;25(2):144-150.

7. .Miller PL, Frawley SJ, Sayward, FG. Issues in computer-based decision support in public health illustrated using projects involving childhood immunization. J Public Health Management Practice, 2001, 7(6), 75-86.

8. Wang, D., Jenders, R., Dasgupta, B. Translating national childhood immunization guidelines to a computer-based reminder based reminder recall system within an immunization registry. Proc AMIA Symp. 1999:;994-8

9. .Miller, P.L., Frawley, S.J., Brandt, C. & Sayward, F. G. A prototype website for immunization maintenance website. Proc AMIA Symp. 1997;:293-7

10. Tu SW, Campbell J, Musen MA. The SAGE guideline modeling: motivation and methodology. Stud Health Technol Inform. 2004;101:167-71.

11. Gennari JH, Musen MA, Fergerson RW, et al. The evolution of protégé: An environment for knowledge-based systems development. Int J Hum Comput Stud 2003;58(1):89-123.

12. Center for Disease Control, Advisory Committee on Immunization Practice. http://www.cdc.gov/nip/ACIP/default.htm
Center for Disease Control, National Immunization Program. http://www.cdc.gov/nip/home-hcp.htm

13. Tu S, Musen MA, Shankar R, Campbell JR et all. Modeling guidelines for integration into clinical workflow. MEDINFO. 2004; 11(Part 1):174-178.

14. SNOMED Clinical terms Concept modeling/Style authoring guide. College of American Pathologists. October, 2004.

15. Campbell JR, Carpenter P, Sneiderman C, Cohn S, Chute CG, Warren J. Phase II evaluation of clinical coding schemes: completeness, taxonomy, mapping, definitions, and clarity.J Am Med Inform Assoc. 1997 May-Jun;4(3):238-51.