MEDINFO 2007
HISA Ltd Australia
© 2007 The authors. All rights reserved.

Semantic Interoperability and SNOMED CT: A Case Study in Clinical Problem Lists

James R. Campbell[a], Alejandro Lopez Osornio[b]
Fernan de Quiros[b], Daniel Luna[b], Guillermo Reynoso[c]

[a]*University of Nebraska Medical Center, Omaha, NE*
[b]*Departamento de Información Hospitalaria, Hospital Italiano de Buenos Aires, Argentina*
[c]*Conceptum, Buenos Aires, Argentina*

Abstract

Sharing clinical data between institutions requires a shared semantic model and reproducible metrics for comparing post-coordinated clinical record data. We evaluated the interoperable features of SNOMED CT by merging and comparing SNOMED problem list dictionaries between primary care sites in Nebraska and Argentina. The two problem subsets had modest differences in semantics, but differed substantially in frequency of post-coordinated content. 64.3% of the final merged problem vocabulary was post-coordinated. 0.9% of the merged concept set was found to have semantic duplication. Classification employing SNOMED Normal Forms was found to effectively identify semantic equivalency in 65.2% of cases studied. The most frequent reasons for failure of the classification procedures included: post-coordination not compliant with SNOMED guidelines (28.8% of cases), conflicting guidance presented in SNOMED pre-coordinated content (3.6%), unclear definition of source clinical concepts (1.2%), and limited expressiveness of the SNOMED definitional model (1.2%). We conclude that interoperable post-coordination employing SNOMED CT is feasible, but that better dissemination of SNOMED editorial guidelines, refinement of the concept model and training of the user community will be required for reproducible and reliable implementation.

Keywords:

controlled terminology, Systematized Nomenclature of Medicine, semantic interoperability

Introduction

Semantic interoperability denotes the ability of two different clinical information systems to seamlessly share clinic record content and recognize meaning within and between those records. Semantic interoperability requires standardized coding of conceptual content as well as understanding of the information model employed by the clinical systems. Interoperability requires that the system be able to recognize equivalence between post-coordinated concepts developed at different clinical sites by different authors.

SNOMED CT is an advanced terminology supporting post-coordination. It is demonstrably the most comprehensive and useful clinical reference terminology available to vendors today. Under the governance of a new international standards organization, it is poised to support interoperability of clinical systems within and between those nations where it is deployed.

While organizing and characterizing the semantic features of a shared SNOMED CT problem list for primary care, we analyzed for equivalence in clinical concepts authored by separate clinicians working on two continents. From systematic study of classifier data, we identified similarities and differences in our practices, and discovered issues for the interoperable deployment and use of SNOMED CT.

Methods

Study sites

Clinics of the University of Nebraska Medical Center employ the GE Centricity Enterprise record which includes problem list and assessment functions with SNOMED CT as coding reference. Employing "just-in-time" post-coordination, UNMC has developed an extension to SNOMED CT which includes 6% post-coordinated concepts within a problem data dictionary of 9800 concepts and 14500 interface terms.

The Hospital Italiano of Buenos Aires is a non-profit university hospital, associated with a large network of ambulatory care services, located in Buenos Aires, Argentina. Hospital Italiano employs a health information system developed in-house including an institution wide electronic medical record supporting all levels of care. Clinical data entry uses a local interface terminology for diagnosis and procedures. Every concept from the terminology has been coded with SNOMED CT. Currently Hospital Italiano terminology has 24,800 concepts and more than 100,000 interface terms, with approximately 70% post-coordinated content.

In order to collaboratively develop a SNOMED CT subset to support problem lists for primary care, UNMC and Hospital Italiano merged and compared their SNOMED data dictionaries, evaluating for similarities of content. Frequency of use statistics for concept codes were further compared in order to identify useful SNOMED concepts for inclusion in a standardized problem subset. This merged problem subset comprised the source content used for this study.

SNOMED post-coordination and classifier software

Post-coordination refers to the definition (or modeling) of a concept that is required for clinical recording but which is not found in the standard SNOMED CT release. Post-coordination requires the concept to be compared to the SNOMED CT definitional model. Relationships and attribute values are chosen to encode the required meaning, and the new post-coordinated concept is employed in the vendor SNOMED extension.

Due to the richness of the SNOMED model, two or more alternative post-coordinated expressions are often possible. In order to support equivalence testing between these alternatives,

SNOMED CT provides technical guidance for the development of classifier software which will convert SNOMED expressions to Normal (Canonic) forms for comparison. Comparing the Normal Forms of two user concepts allows for a computable determination whether the two are synonymous or unique. Classification software systematically computes normal forms for all concepts within a set and determines the relationship of concepts to each other within the taxonomy while identifying synonymous content.

Problem list classification

We compared the separate SNOMED CT subsets representing problem list coding at the two institutions, first characterizing the shared pre-coordinated content and then analyzing post-coordination. Classifier software developed at UNMC was employed to compute Normal Forms for all concepts in the merged set. We identified synonymous concepts by comparing classifier results and assessed the validity of post-coordination of concept samples from both sites. We searched for masked synonymy employing exhaustive editorial search within the merged problem set.

Results

Problem list subsets at the two institutions were notably different in concept inventory and frequency of post-coordination. Only minor differences in local problem list recording habits were reflected in the semantics of the SNOMED CT use at the two institutions, reflecting more frequent recording of procedure history at the Nebraska site. Clinical disorders, findings and events were the majority of concepts in use, while "past medical history" and "family history" concepts, occurred regularly at both institutions.

The Normal Form was computed for 31,382 concepts in the merged subset. 293 concepts were identified as duplicating concept meaning with other subset members including 9/578 (1.6%) of Nebraska post-coordination, 281/16807 (1.7%) of Hospital Italiano post-coordination, and 3 cases in which a Nebraska concept was identical in meaning with a concept from Hospital Italiano.

Table 1 summarizes the issues we identified during detailed study of the classifier analysis of post-coordinated concepts.

Table 1 – Analysis of post-coordination / classification errors

Failure of post-coordination	
Vague source concept utterance	3(1.2%)
Limited expressiveness SNOMED model	3(1.2%)
Conflict of SNOMED guidance	9(3.6%)
Non-compliance with editorial guidelines	72(28.8%)
Total classification errors	87(34.8%)
Masked synonymy	3(1.2%)

Discussion

Clinical reference terminologies employing features of ontologies create new opportunities for interoperability of clinical systems, but they also lead to a new set of challenges. Experience at the two sites presented in this paper confirms that the organizational and information science approach of the enterprise can lead to rather different needs for support and maintenance of post-coordination.

Assuring that post-coordinated data sets interoperate requires attention to many details. It is especially challenging at this time when the technology of clinical systems such as SNOMED CT are evolving. While our research supports the utility of SNOMED Normal Forms for comparing disparate data sets, reliable and consistent results require that information system managers have well trained and consistent data modeling staff. This was the greatest single deficiency identified in our analysis and accounted for the largest variation from expected results. The SNOMED standards board must extend support to the enterprise employing post-coordination by addressing consistency within the SNOMED CT data model. But even more important for post-coordinated data management is dissemination and documentation of the SNOMED data model standard, promotion of open discussion and training in SNOMED use, and provision of clear guidance supporting reproducible and understandable post-coordination algorithms.