

Creation of a Local Interface Terminology to SNOMED CT

Alejandro Lopez Osornio ^a, Daniel Luna ^a, Maria Laura Gambarte ^a, Adrian Gomez ^a,
Guillermo Reynoso ^b, Fernán González Bernaldo de Quirós ^a

^aMedical Informatics Department, Hospital Italiano de Buenos Aires, Argentina

^bConceptum, Buenos Aires, Argentina

Abstract

This paper describes the steps followed in the creation of a local Interface Terminology to SNOMED CT (as reference terminology) with a strong focus on user acceptability. The resulting list of terms is used for clinical data input by physicians and nurses at the Hospital Italiano in Buenos Aires, Argentina. Description includes data model, mappings to SNOMED CT and classifications, subsets definitions and extensibility mechanisms. The Interface Terminology is currently used in the recording of diagnosis and procedures in inpatient discharge summaries and its coverage is improving from user feedback. Its current size is 24,800 concepts, 67% of them needed post-coordination for appropriate semantic representation, due to a very flexible policy that allows the use of any number of modifiers on concepts.

Keywords:

interface terminology, terminology services

Introduction

Interface terminologies (IfT) had been used for a long time in electronic medical records, and consisting in a list of local terms, familiar to the user, that are used for clinical data entry[1]. The name “interface” comes from its role as a link between user’s vocabulary and a stricter, standardized, list of terms used to describe clinical data. Each term in the IfT is mapped to a term in any standard terminology used as a “reference”. This strategy isolates the user from the complexity of standard classifications or nomenclatures used as reference terminologies, whose terms may not be appropriate for the setting, may have an arbitrary level of detail or may have too many rules for code selection.

SNOMED CT®[2] is currently regarded as the most advanced terminology system for storing clinical information, and is a government mandated standard in many countries[3-8]. SNOMED CT includes mappings to several standard classifications as ICD-9CM and ICD-10. In this paper we will review the approach of the Hospital Italiano of Buenos Aires for implementing an Interface Terminology mapped to SNOMED CT in a hospital wide Health Information System.

Background

The need of a local interface terminology for SNOMED CT

SNOMED CT includes IfT capabilities, with a list of different terms applied to each concept. The advantage of using a local IfT is to provide easier adaptation to local content and to create a permanent, list of terms, isolated from periodic SNOMED CT changes.

Localization involves two basic functions: hiding not desired content of SNOMED CT and including new terms and concepts needed in the local setting but not present in the standard terminology. These objectives can be achieved using two SNOMED CT functions, the “subsets” for restricting content and the “extensions” for adding new content.

Defining subsets for content restriction can define different scopes, using large subsets for extensive topics like “Problems List” and smaller ones for specific purposes like data input from templates, etc. The local IfT consists in all the subsets created for local use, including a mix of standard terms and local terms, created in an extension of SNOMED CT.

The main drawback of creating a local IfT is the additional work required for its development and maintenance. In our case, conforming to usability demands from our user base was a key point in our implementation, so the extra effort in creating and maintaining an IfT was deemed acceptable.

Setting

The Hospital Italiano of Buenos Aires is a 650-bed non-profit university hospital located in Buenos Aires, Argentina. More than 150,000 outpatient visits and 3,000 hospitalizations are registered every month. It is affiliated with a Health Maintenance Organization (Plan de Salud) that takes care of a population of 140,000 patients.

Since 1998, a full scale HIS has been gradually implemented, including ambulatory Electronic Medical Record (EMR), inpatient discharge summary, administrative systems, scheduling systems, inpatient tracking systems, pharmacy systems and complementary studies report and visualization. Several health informatics standards had been implemented, including HL7, CDA Version 2, ICD-9CM, DRG, ICD10, and ICPC.

Existing experiences

Several health systems created local IFTs for their electronic medical records, mapped to necessary billing classifications like ICD-9CM and other standard terminologies, some examples are shown in Table 1. Sometimes, IFTs were created from scratch and others were based on previously existing standards.

Table 1 - Examples of interface terminologies

Interface Terminology	Site
MED (Medical Entities Dictionary)[9]	Columbia University and the New York Presbyterian Hospital
VA terminology Lexicon[10]	U.S. Department of Veterans Affairs
Convergent Medical Terminology[11]	Kaiser Permanente
UNMC Lexicon[12]	University of Nebraska Medical Center
ICPC Plus[13]	Australia, General Practice
Mayo Master Sheet Index[3]	Mayo Clinic in Rochester, Minn.

Terminology Services had been described as a way to implement centralized terminology access in enterprise wide electronic health information systems[14, 15]. HL7 Common Terminology Services (CTS) specification proposes a full set of features that a terminology server should provide.

Recently, Rosenbloom et al. published a review about IFT features and challenges[1], which we used to adjust and re-think some parts of our project, and we are now applying some of these concepts to structure this paper.

Objectives

The key objective of our project was to build a local Interface Terminology that allows users to record clinical data choosing options from a list of familiar terms but storing information SNOMED CT -compatible.

The IFT should provide adequate coverage for our reality. Users should have the ability to refine terms, choosing a more specific option of a given term, and propose new terms to improve coverage. The system should provide the equivalence of a given local term in standard classifications.

A “Terminology Team” would be in charge of the maintenance of the IFT, the system should also provide the tools required by this team.

System description

Concepts are unique cognitive representations of objects or situations in the reality in people’s minds. Concepts are represented

in textual form by terms, also called descriptions in this setting. Terms included in an IFT are arranged in concepts, expressing synonymy, and concepts are linked to other semantically related concepts in hierarchies.

Scope of the local IFT covers most aspects of medicine in a setting like ours. For the purposes of this paper, we will describe the development of the “Diagnosis Subset”, ‘Procedures Subset’, “Drugs Subset” and a very specific “Causes of liver failure Subset” as an example.

The creation of the local IFT included the steps described in the following sections using the “Diagnosis Subset” as the main example.

Definition of IFT data structure

Our IFT was designed to take full advantage of SNOMED CT semantic structure and description logics, so our data structure is also tied to the standard SNOMED CT structure. SNOMED CT provides a data structure[16] (Figure 1) that allows storing local data following simple guidelines; local data can be differentiated from original SNOMED CT standard data by a “namespace” identifier included as part of each concept, term or description identifier.

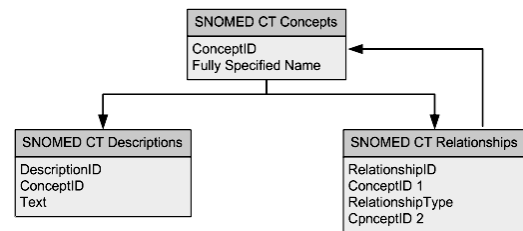


Figure 1 - Overview of SNOMED CT core data structure

This strategy is appropriate for small additions to SNOMED CT. In our case, we aimed to create an extensive local IFT, as independent as possible from the reference terminology. We decided to duplicate SNOMED CT data structure, and fill the empty model with our local data. The concept and description tables will contain our local concepts and descriptions, regardless if they are also present in standard SNOMED CT data or not. The relationship table will work as the link between the two models, linking from each local concept to standard SNOMED CT concepts (Figure 2).

Concepts that exist in both terminologies are linked with a “same as” relationship. New concepts in the local IFT are linked with a “is-a” relationship with the appropriate super-types in SNOMED CT structure. The defining characteristics of the new concept, those that make it different from its super-type or father concept, are represented using other “attribute” relationships to standard SNOMED CT concepts, following official SNOMED CT modeling guidelines. The definition process using additional attributes is called post-coordination, which is done by experts from the Terminology Team based on user request of new terms. In this way, our local relationship table always links from local concepts to standard SNOMED CT concepts, providing semantic information to the local model.

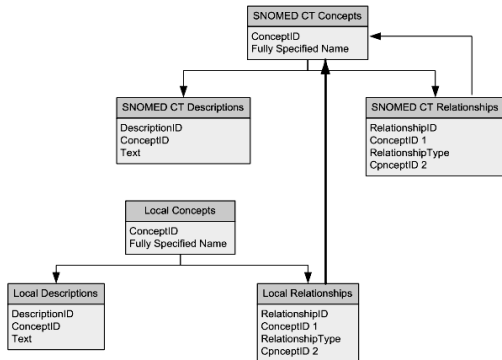


Figure 2 - Relationship between local IFT data structure and SNOMED CT core data structure

All concepts will be unique and each will have a unique “Fully Specified” description that fully describes the concept meaning, a preferred term for displaying to the user and a list of synonyms (descriptions).

Unique identifiers for concepts, descriptions and relationships are independent numbers, conforming to the SNOMED CT concept ID specification, using a local namespace. Applications use these identifiers to store data in the clinical repository and to retrieve subsets for structured entry in user interfaces.

Applications will code clinical data using the identifier for the description selected by the user, not the concept identifier. In this way, whenever the terminology team changes the relationship between a concept and some of its terms, pointing to another concept, there is no need to recode previously stored data in the clinical data repository.

We also decided that terms would be unique in our IFT, so each term is a synonym of only one concept. Ambiguous terms, defined as terms that possibly would relate to more than one concept were treated separately, as will be described later.

Subsets are implemented using SNOMED CT subset mechanism for enumeration of members (Figure 3).

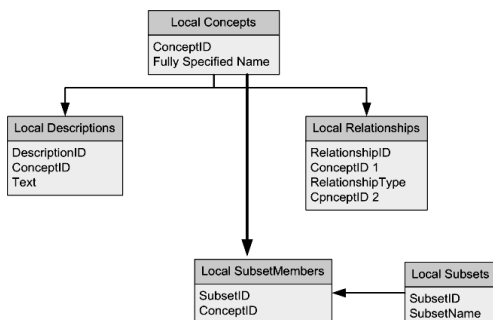


Figure 3 - Local Subsets data structure

Local terms collection and conceptual aggregation

Some IFT’s domains are expected to have more local variation and need to be more adjusted to our setting, such as

terms for the diagnosis subset. On the other hand, some smaller issues, like drugs, have a much more limited number of textual variations and any standard list can be presented to the user as the IFT.

The Hospital Italiano previous vocabulary control system for the problems list consisted in physicians entering free text descriptions in the electronic medical record, which were later assigned classification codes by a group of coders. This process generated a backlog of more than 2 million free text phrases describing medical problems, entered by the professionals of the institution. We used this data for creating the terms included in the “diagnosis” subset of our IFT.

After applying a normalization process to these data, a list of more than 250,000 different short text phrases was created; about 110,000 descriptions appeared with 10 repetitions or more, accounting for more than 90% of the free text expressiveness variation of our setting. These 110,000 terms were manually revised, and synonyms were grouped together. We adopted the definition that synonyms are a group of terms that describe the same unique concept. After this process, the 110,000 terms were arranged in 24,800 concepts. We included these concept and synonym data into the “diagnosis” subset of our Interface Terminology. Only 33% of these concepts are direct links to SNOMED CT, a number of additional attributes where needed to represent the remaining 67%.

Terms for “Drugs” subset were imported from our Computer Physician Order Entry (CPOE) software databases.

For specific data-entry subsets like the “Cause of liver failure Subset” content was provided by specialists, describing the possible valid inputs.

Defining invalid terms, refinement and disambiguation rules

During manual revision of the free text entries, several terms were flagged as “Invalid” when their use was not appropriate in each sub-domain. Examples of invalid terms are uncompressible strings of text or terms of concepts not considered valid entries in the sub-domain, like entering procedures in a diagnosis field. In the implementation of the IFT, user interfaces should not allow terms included in this list to be added to the clinical record.

Some concepts were flagged as “Mandatory Refinement”, when represented too general concepts and its clinical use would be improved with more detail. An example of that is the term “Diabetes Mellitus”, which would be much more useful if refined to “Diabetes Type I”, “Diabetes Type II”, etc. In the implementation of the IFT, if the user selects one of these concepts he/she should be enforced to select a more specific concept, and a set of appropriate options should be presented, according to the semantic relationships in the local IFT.

Physicians were found to record data using less detail than they actually have obtained during a consultation or examination. Therefore, it must be possible to refine every concept to its more detailed sub-types, and these options should be presented to the user as non-mandatory alternatives.

Ambiguous terms cannot be assigned as valid descriptions of any concept and should be manually disambiguated by the users. These terms were stored separately and were related to the possible concepts in the Interface Terminology, so user interfaces should present the options to the users and force them to specify the concept they wish to use.

Adding semantic definitions through SNOMED CT mapping

Each local concept was modeled either as a direct equivalence to SNOMED CT or its meaning was described with a set of attribute relationships following SNOMED CT guidelines. The resulting relationships were also used for detecting synonymy, as there is a good chance that two terms that share the same definition refer to the same concept.

At this point, our Interface Terminology consisted only in a list of unrelated concepts in each subset, each of them with one or more textual descriptions. In our model all semantic information is derived from SNOMED CT. That means that it would be possible to understand the relationship between two terms in the Interface Terminology by following SNOMED CT semantic structure as can be interpreted from our data structure (Figure 2). This is equally possible for our post-coordinated terms, as they share the same semantic structure with original SNOMED CT concepts.

Definition of cross-maps to standard classifications

Mapping from our local IfT to standard classifications like ICD-9CM, ICD-10 or ICPC2 is provided through the SNOMED CT standard cross-maps mechanism.

Concepts included in the local IfT that are not present in the original SNOMED CT distribution, are mapped through its super-types, the more general, standard concepts used for represent their meaning.

This strategy was preferred instead of a direct linking in order to provide a full access to available standard cross-maps sets from SNOMED CT to classifications.

Defining strategies to maintain domain coverage

The coverage of the Interface Terminology can be described at two levels, conceptual level, and term level. The conceptual level refers to the proportion of the locally used medical concepts that are included in the terminology. Term level coverage refers to the proportion of the different text descriptions or terms used to describe those concepts which are included in the Interface Terminology.

The coverage is originally determined by the accuracy of the initial process of data recollection for inclusion in the Interface Terminology. Later, this coverage is maintained and adjusted by the expansion of the terminology, based on user experience.

According to its definition, the expansion process had to be based on direct user suggestions that would be manually evaluated by the Terminology Teams prior to its formal incorporation in the interface terminology. Software user interfaces would enforce user participation with suggestions and commentaries.

Definition of strategies for updating to new versions of SNOMED CT

The IfT should provide an adequate solution to the management of new SNOMED CT versions. New versions may carry ambiguity, including new standard concepts that were previously included in the local extension. In this case, the local IfT concept should map directly to the new SNOMED CT concept instead of the super type and the attributes used to model its meaning. Classification algorithms can automatically detect these cases and remap the concepts maintaining consistency.

Another problem is SNOMED retiring concepts or terms in new versions, in order to correct problems in its representation. When those concepts were used in the representation of the local IfT, concepts would be pointing to inactive codes in the standard. Current SNOMED CT mechanisms do not support automatic selection of new, valid, codes to replace retired ones. The IfT should provide a local valid list of concepts and terms whose representation is affected by changes in the new version of SNOMED CT to manually correct its representation.

Current status

Implementation of the local IfT

The local IfT was implemented in our health information system on June 2006 using a set of Terminology Services. The first area was the inpatient structured discharge summary input. Subsets for Diagnosis and Procedures are used in a user interface that allows text input and search of related terms. The user interface uses the rules for dealing with invalid, ambiguous and refinement rules.

During the initial 6-month deployment, the user proposal of terms is enforced and free-text input is an alternative available to the users. These inputs are redirected as new terms proposals.

The IfT administration software provides tools to support all the design objectives described before.

Coverage

Coverage is estimated by subset, the incidence of "new terms" proposal issued by users. We measured this during a week after 3 month of implementation; using the Diagnosis input interface new terms were proposed in 12% of the interactions by users, meaning that in 86% of the cases a good option was selected from the IfT. Using the Procedures input interface, proposal rate was 30% of the cases.

New terms proposals were evaluated by the terminology team, and accepted as new terms in the IfT in less than 40% of the cases for Diagnosis and 80% for Procedures. Reasons for rejections were the proposal of invalid terms or already existing terms.

Discussion

Our model performance was very satisfactory for our extensibility needs and the creation of local subsets. The duplication of data models with SNOMED CT adds complexity to data access, but once the basic mechanisms were created and tuned they did not impact on system performance or software development speed.

Our aim is to gradually enforce the user to only use terms included in the IfT, allowing suggestions but in a different process outside the data entry workflow. From the results is clear that the Diagnosis domain has achieved a much more mature status than Procedures. New improvements in search algorithms and contents are still needed until final implementation of a more restrictive interface.

The desired level of user expression freedom we wanted to support included describing laterality in findings and procedures. This led to a combinatorial “explosion” of thousands of new concepts to model outside SNOMED CT and an important increase in manual work by the terminology team. This is the cause of the 67% of post-coordination, which puts us on the need of a very strict quality assurance process over those more than 16,000 post-coordinated concepts. This will be the focus of a future paper.

Additionally, body parts are not always represented as lateralized concepts in SNOMED CT, for example different codes exist in SNOMED CT (1-2006) for left and right hand or breast, but not for great toe. We used a non-official “Laterality” attribute applied to the finding or procedure instead of modeling new concepts for body parts in SNOMED CT. In order to provide interoperability this can be easily solved at the moment of sharing data.

The procedures subset proved to be a much more complex domain than findings (diagnosis). Surgeons in our setting used a complex language, often combining different procedures in the same sentence, or adding details on the intention or result of the procedure. The rate of proposals of new terms by the user remains high in this domain; perhaps a structured, compositional approach to the user interface would be useful.

Analysis of user proposals was very important for introducing changes in the search algorithms, as in many cases the term already existed and was not found by the user.

In this experience we implemented post-coordination to SNOMED CT by expert authors of the Terminology Team. We are testing end-user interfaces to post-coordinate new concepts in specific settings, like the “Family History” subset. In this way, any user can create a new term in the IfT expressing that is a “Family History” of a given existing local concept, and selecting in which family member occurs. The proposed model is compared to existing “Family History” concepts. In case it has not previously appeared, a new concept is created, as well as the automatic creation of relationships to the corresponding standard SNOMED CT concepts.

A test of interoperability sharing post-coordinated data with other institutions with a similar SNOMED CT-based IfT is being tested and will be the subject of a future paper.

We are also testing a model for dynamic subset definitions, a set of rules that allows the definition of a subset in terms of their relationships with SNOMED CT concepts. In this way we can define in the local IfT a “Diabetes Subset” with all the terms related to the concept “Diabetes Mellitus” in SNOMED CT or any of its subtypes. Dynamic subsets are

updated nightly in order to include recently added local concepts, and are the way to select data from the data repository with a clinical focus, including a wide range of terms, from the standard concept “Diabetes Mellitus Type II” to the very specific, post coordinated, local concept “Diabetes Mellitus Type II with mild diabetic neuropathy”.

Future implementation of the system in the outpatient ambulatory setting and general restriction of free text input will prove how much work is required to maintain good domain coverage of the local IfT.

References

- [1] Rosenbloom, S.T., et al., Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. *J Am Med Inform Assoc*, 2006. 13(3): 277-88.
- [2] Snomed, SNOMED Clinical Terms User Guide. 2005: College of American Pathologists.
- [3] Elkin, P.L., et al., Evaluation of the content coverage of SNOMED CT: ability of SNOMED clinical terms to represent clinical problem lists. *Mayo Clin Proc* 2006; 81(6): 741-8.
- [4] Elevitch, F.R., SNOMED CT: electronic health record enhances anesthesia patient safety. *Aana J*, 2005. 73(5): p. 361-6.
- [5] Penz, J.F., et al., Evaluation of SNOMED coverage of Veterans Health Administration terms. *Medinfo*, 2004. 11(Pt 1): 540-4.
- [6] Wasserman, H. and J. Wang, An applied evaluation of SNOMED CT as a clinical vocabulary for the computerized diagnosis and problem list. *AMIA Annu Symp Proc*, 2003: p. 699-703.
- [7] Dykes, P.C., L.M. Currie, and J.J. Cimino, Adequacy of evolving national standardized terminologies for interdisciplinary coded concepts in an automated clinical pathway. *J Biomed Inform*, 2003. 36(4-5): p. 313-25.
- [8] McClay, J.C. and J. Campbell, Improved coding of the primary reason for visit to the emergency department using SNOMED. *Proc AMIA Symp*, 2002: p. 499-503.
- [9] Cimino, J.J., From data to knowledge through concept-oriented terminologies: experience with the Medical Entities Dictionary. *J Am Med Inform Assoc*, 2000. 7(3): p. 288-97.
- [10] Lincoln, M.J., et al., U.S. Department of Veterans Affairs Enterprise Reference Terminology strategic overview. *Medinfo*, 2004. 11(Pt 1): p. 391-5.
- [11] Dolin, R.H., et al., Kaiser Permanente's Convergent Medical Terminology. *Medinfo*, 2004. 11(Pt 1): p. 346-50.
- [12] Campbell, J.R., Semantic features of an enterprise interface terminology for SNOMED RT. *Medinfo*, 2001. 10(Pt 1): p. 82-5.
- [13] Britt, H., A new coding tool for computerised clinical systems in primary care--ICPC plus. *Aust Fam Physician*, 1997. 26 Suppl 2: p. S79-82.
- [14] Chute, C.G., et al., Desiderata for a clinical terminology server. *Proc AMIA Symp*, 1999: p. 42-6.
- [15] Rector, A.L., et al., A Terminology Server for medical language and medical information systems. *Methods Inf Med*, 1995. 34(1-2): p. 147-57.
- [16] Snomed, SNOMED Clinical Terms Technical Reference Guide. 2005: College of American Pathologists.