Unambiguous data modeling to ensure higher accuracy term binding to clinical terminologies

Rahil Qamar, (PhD), MSc, Jay (Subbarao) Kola, MBBS, (PhD), Alan L. Rector, MD, PhD
BHIG, University of Manchester, Manchester, U.K.

Abstract

Work in the field of recording standard, coded data is important to reduce medical errors caused by misinterpretation and misrepresentation of data. The paper discusses the need to ensure that the source of the data i.e. the clinical data model is unambiguous to increase the quality and accuracy of the data mapping to terminology codes. The study chooses one especially ambiguous data model and remodels it to make clearer both the structure of the data, as well as its intended use and semantics. By ensuring an unambiguous model, results of the data mapping increased in accuracy from 64.7% to 80.55%. The clinical experts evaluating the models found it easier working with the revised model and agreed on the mappings 93.1% times as against 48.57% times previously. The aim of the study is to encourage good modeling practice to enable clinicians to record and code patient data unambiguously and accurately.

Introduction

One of the key objectives of most national strategies for developing IT solutions for health care is to ensure that all Electronic Health Records (EHRs) in all hospitals share common standards for data, classifications, and coding systems [1]. For instance, the NHS in UK has agreed to adopt SNOMED CT as the standard terminology to code all health data [9]. The aim is to control the vocabulary used to record patient data to reduce the possibility of differing interpretation of information and the possibility of errors resulting from traditional paper records [9].

Efforts are also underway to promote the use of data entry forms for structured data capture in EHRs. Clinical data models help in defining the structure of the information to be stored in the EHRs [2] through the data entry forms. Therefore, work on coding data will need to extend to these models as well. Examples of clinical data models are HL7 V3 messages, and the European openEHR Archetypes.

A semi-automated mapping system called the Model Standardisation using Terminology (MoST) was developed to map clinical model terms to terminology codes. Discussion of the MoST system is beyond the scope of this paper but has been covered previously in [3].

Based on the MoST results and evaluation feedback, the paper discusses issues raised by the clinical experts, when performing the mapping exercise. One especially ambiguous model was then redesigned and the MoST process was repeated to determine whether there was any improvement in the mappings. The central idea is to highlight the need for authoring unambiguous and usable data models to have any chance of success at coding their data unambiguously. Dealing with terminology issues is a problem that requires to be addressed once issues with the source of the data i.e. the data models are resolved.

Methods

Data Source – openEHR Archetypes: openEHR Archetype Models were chosen as the source of the data. Archetypes are essentially nested hierarchies of data elements required to record a particular clinical event. These models specify pre-defined constraints on their data. The archetype approach is being standardized by the openEHR community for wider acceptance in the European healthcare domain. The openEHR Archetypes conform to the openEHR Reference Model [10]. The term ‘Archetypes’ is used in the paper to refer to openEHR Archetypes.

The archetype selected for the study is categorized as an observation by the author of the model. The other categories are evaluation, instruction, and action.

The terminology separation feature of Archetypes was the main reason for its selection. This feature separates the data from the terminology used to identify it. A local vocabulary along with a description of the data is provided separately in the ‘Terminology’ section of the model. In addition, external terminology code(s) can be mapped to these data elements at any stage of the archetype authoring process. This made it especially straight-forward to perform the term binding/mapping exercise.

Terminology Source - SNOMED CT: SNOMED CT or SNOMED aims to be a comprehensive clinical terminology that provides clinical content and expressivity for clinical documentation [4].
The reasons for selecting SNOMED, to code the archetype data were: (i) the size of the terminology. With over 300000 concepts there was greater scope of finding a semantically correct match, (ii) ease of using the terminology. The data files can be imported into database tables with minimum effort, and (iii) ease to query concepts. Information extracted consisted of concept definitions and their relationship to other concepts in the terminology.

Mapping Methodology: At first, the mapping methodology aims to auto generate a list of semantic and lexical SNOMED matches for the archetype data using the MoST system [3]. Next, the methodology proposes to select the most equivalent SNOMED codes for mapping using MoST and manual intervention by the clinical experts.

The term ‘modeler’ will be used for the original author of the archetype model, ‘clinical expert’/‘evaluator’ for the person performing the mapping, and ‘mapping’ for the term binding process.

Test Case Analysis: Archetype data from the ‘histology pap’ model were sent to MoST, which returned a number of SNOMED codes. The model was taken from the official openEHR website’s published list of archetypes [6] and is viewed in Figure 1 from the ‘Terminology’ section of the Archetype Editor developed at LiU, Sweden [7]. Although there are several other archetypes with ambiguous data, only one was chosen for this paper, as it had the lowest ratings amongst several others. The search results were evaluated by 10 clinical experts to determine the most equivalent codes for mapping. The experts were also asked to provide feedback on the quality of the archetype, SNOMED, as well as the MoST system results.

Evaluation: The ‘histology pap’ archetype had approximately 40 terms which were sent to the MoST system. The evaluators were required to score, on a scale of 0 to 10, the equivalence of all the SNOMED codes returned before and after the filtering process, as shown in Table 1. The filtering process included the elimination of codes that MoST considered to be nonequivalent to the archetype data terms. The matches found were of the order 1:n such that 1 term could have n codes returned as appropriate matches.

Table 1: Total number of archetype terms and the corresponding number of SNOMED codes returned

<table>
<thead>
<tr>
<th>(1:n matches)</th>
<th>Archetype data terms</th>
<th>SNOMED CT codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before filtering</td>
<td>40</td>
<td>397</td>
</tr>
<tr>
<td>After filtering</td>
<td>34</td>
<td>160</td>
</tr>
</tbody>
</table>

A code was deemed equivalent if the total score given by the 10 evaluators was above 50%. The final mapping accuracy was 64.7% (22/34*100). This means that of the 34 terms, 22 found at least one code with a total score above 50%. In addition to accuracy, the rating was also based on the ‘confidence’ quotient. This indicated the confidence the evaluators placed on a particular SNOMED code they thought was relevant for mapping to the archetype term. The higher the score (max. 10) given to a particular code, the higher the level of confidence and vice versa. The ambiguities present in the model resulted in an average confidence rating with a low of 6/10 (6 out of 10) given for each code deemed equivalent.

Although the accuracy rate of 64.7% might appear to be sufficiently high, it still highlights problems with adequacy of the archetype, SNOMED, and the MoST system. However, all the experts agreed that the primary reason for the lower accuracy and confidence ratings was the ambiguities in the archetype.

Results - Qualitative Review

Issues with the data source model: The ‘histology pap’ archetype was the source of the data chosen for the study. There were two main concerns that the experts based the issues on. First, what kind of
information is the archetype trying to capture? Second, can the archetype in its present state be safely used to map to terminologies without any conflict or error with its intended purpose? Based on these concerns the issues raised were:

1) Ambiguous categorisation of top level data terms: Although the ‘histology pap’ archetype belongs to the openEHR observation type, it is intended to record ‘cytological findings of a pap test’. In addition, the children of the root concept ‘cervical smear’ are used to represent different aspects of a test, which would categorise it as a procedure. E.g. the children include ‘specimen collection’, ‘quality (and all its children)’, ‘identification (and all its children)’, and ‘process date’. Therefore, a procedure type archetype would have resulted in lesser ambiguity. However, it is unclear whether the archetype would then belong to the openEHR action or instruction clinical entry type.

Most of the terms in this archetype were mapped to SNOMED findings and procedure, as shown in Table 2. However, some experts preferred to map to SNOMED observables although often no appropriate codes were available in this category for top level terms. Paper [8] discusses conflicting categorization of terms in other archetype models.

2) Ambiguously named terms: Often the labels of the archetype terms are inadequate to convey their intended meaning. E.g. the term ‘clinical’ is defined as ‘clinical notes sent with request’ in the model. Given that ‘clinical’ is nested under ‘specimen collection’, as shown in Figure 1, it is difficult to comprehend the intent of the modeler. In such a case, it is easy for a clinician to interpret ‘clinical’ as the clinical specimen collection process, especially since it has a sibling term ‘specimen’. The ambiguity could be avoided by using an explicit label such as ‘clinical notes’. Some other poorly labeled terms in the model are ‘macroscopic’ and its child ‘appearance’, ‘microscopic’, ‘process’, ‘identification’, ‘comment’, ‘laboratory’, and ‘preservative’, shown in Figure 1.

3) Similar or duplicate labeling of terms with inadequate definitions: At times, the labels assigned to the archetype terms were either duplicated or very similar to other labels in the model. Although the context of use differed, it was difficult to precisely ascertain their intended use as the term definitions lacked clarity. E.g. ‘specimen’ was used twice in the same model. In the first instance, it meant to record ‘details of specimen’. Later it was used to record the ‘problem with the specimen’. There were also other uses of the word specimen such as ‘specimen collection’, ‘specimen unlabeled’, ‘insufficient’ and ‘damaged’ specimen. Although the terms were placed in different sub-hierarchies, the context itself was not clear to justify its existence in the model. Another term was ‘image’ used for macroscopic and microscopic findings.

4) Insufficient separation of meta data information from core clinical recording information: The model does not sufficiently separate data that is meta (or general) to the recording of the ‘cervical smear’ from data that is more specific. Since 79% of the evaluators agreed that the archetype represents the recording of a procedure our comments will be based on procedure types. Terms such as ‘specimen collection’, ‘satisfactory assessment’, ‘quality’, ‘labeling’, and ‘preservative’ are common to any procedure and are therefore meta to the model. These terms need to be placed separately from data that is specific to the recording of cervical smear such as ‘low grade’ and ‘high grade’ epithelial abnormality, ‘negative for dysplasia or malignancy’, and ‘specific cytological finding’. In fact, since archetypes can reuse whole or parts of other archetypes, the meta data could be modeled in a separate archetype and included for use in this model. Even if the meta data is included with the more specific clinical data in the same archetype, it should be clearly separated to reduce ambiguity and enhance clarity.

5) Using post-coordination of terms adequately and in the right place: Some concepts in the model were needlessly split into fine-grained terms to the extent that their semantics was lost in the process. E.g. the term ‘specimen’ and its child ‘description’ both had the same definition - ‘details of the specimen’. Individually they did not make much semantic sense. However, when combined to form a composite term ‘specimen description’ the intent was clearer. Well-defined terms would help systems like MoST to lookup appropriate matches in SNOMED. Table 3 below shows the matches returned by MoST given

<table>
<thead>
<tr>
<th>Archetype term</th>
<th>Appropriate matches in SNOMED CT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Histology pap</td>
<td>Papanicolaou smear test (procedure) [119252009]</td>
</tr>
<tr>
<td></td>
<td>Cervical cytology test (procedure) [416107001]</td>
</tr>
<tr>
<td></td>
<td>Cervical smear result (finding) [269957009]</td>
</tr>
<tr>
<td></td>
<td>Cervical cytology finding (finding) [302796001]</td>
</tr>
<tr>
<td>Low grade epithelial abnormality</td>
<td>Low-grade squamous intraepithelial lesion (morphologic abnormality) [112662005]</td>
</tr>
<tr>
<td></td>
<td>Low grade histologic differentiation (finding) [399415002]</td>
</tr>
</tbody>
</table>

Table 2: Sample SNOMED for archetype terms. Matches include SNOMED category and code.

AMIA 2007 Symposium Proceedings Page - 610
the current model representation. Had the term been pre-coordinated i.e used one code to represent a single concept, the evaluator might have selected only the code 115267000 (code shown in Table 3).

<table>
<thead>
<tr>
<th>Archetype term</th>
<th>SNOMED CT Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specimen</td>
<td>Form and specimen details different (finding) [281322007]</td>
</tr>
<tr>
<td>Description</td>
<td>Description of specimen character (procedure) [115597007]</td>
</tr>
<tr>
<td></td>
<td>Specimen description (procedure) [115267000]</td>
</tr>
</tbody>
</table>

Table 3: SNOMED codes returned for the split terms.

On the contrary, at times post-coordination i.e composing concepts into two or more codes to convey the correct semantics was more suitable against pre-coordinating them. E.g ‘low grade epithelial abnormality’ might have been better represented in the form of post-coordinated terms, shown in Table 4.

Despite the ability to post-coordinate, as above, grading of the cytological abnormality on PAP smear can be represented better than the present low grade and high grade. This has been achieved in the revised model discussed later.

<table>
<thead>
<tr>
<th>Archetype term</th>
<th>Post-coordinated SNOMED match</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low grade epithelial abnormality</td>
<td>Epithelial cell abnormality (morphologic abnormality) [373886001] + Low grade (qualifier value) [349915008]</td>
</tr>
<tr>
<td></td>
<td>Histological grade finding (finding) [373372005] + Low grade (qualifier value) [349915008]</td>
</tr>
</tbody>
</table>

Table 4: Suggested post-coordinated SNOMED codes for archetype term.

Issues with the terminology source model: Despite the wide coverage of clinical areas by SNOMED several problems persist with its content. The problems relate to their definitions and subsumption hierarchies, as well as naming/labeling. In addition, SNOMED has problems classifying its concepts consistently despite being based on formal logic.

The focus of this paper is primarily to address issues in SNOMED w.r.t the chosen archetype. Issues with SNOMED, though discussed briefly, are part of a much larger problem to be presented in a separate paper. In this paper we present some of the concepts that were missing in SNOMED that were considered important for inclusion. Most of the missing concepts were common to clinical tests or procedures. Some of the codes found missing were:

- ‘Technically unsatisfactory’ to represent technical problems in any kind of test performed.
- ‘Appearance’ specific to histopathology examinations rather than general physical examinations. Also concepts for ‘macroscopic’ and ‘microscopic’ appearance of the tissue.
- ‘Adequate numbers of squamous cells present’ localizing it to PAP smear. The alternative is to post-coordinate Squamous epithelial cell (cell) [80554009] with a new general concept ‘adequate number of cells present’.
- ‘Slide’, ‘container’, and ‘specimen’ unlabelled instead of the generic ‘Sample unlabelled (finding) [125159002]’.
- Modifiers on ‘specimen’ such as ‘insufficient’, ‘damaged’, and ‘contaminated’ to qualify the quality of the specimen.
- Concept for ‘preservation of specimen’ with modifiers ‘insufficient’, and ‘incorrect’ etc. to code problems with the preservation of the specimen.

It is interesting to note that most of the missing concepts relate to general (meta) information on pathology procedures despite SNOMED starting out with an initial focus on pathology (SNOP)[5]. This list will possibly grow in size as more archetypes are evaluated for finding term mappings in SNOMED.

Revising the data source model

Based on the issues with the archetype model raised by the evaluators, the existing model was modified to address majority of the issues.

Revised model: Figure 2 presents the revised version of the original ‘histology pap’ archetype. The meta data common to any procedure was included in the same model to maintain some level of consistency with the original model.

Summary of changes:

(i) The data model was specified as describing elements of a procedure and its related findings.
(ii) Meta data related to any procedure/test was separated in the model from data specific to recordings of the PAP smear.
(iii) The model hierarchy was changed to reflect a general pattern to any clinical procedure or test with some data specific to the PAP smear procedure.
(iv) The labels and descriptions of the archetype terms were modified to reflect more clearly their intent. Suggestions from the evaluators were taken
into consideration when resolving ambiguous terms. E.g. the term ‘Clinical’ was modified to ‘Clinical Notes’ to describe the ‘Clinical history sent with the specimen’. A tabulated comparison of some of the changes made when revising the model is stated in Table 5 below.

<table>
<thead>
<tr>
<th>Original Model</th>
<th>Revised Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low grade epithelial abnormality</td>
<td>Mild epithelial abnormality</td>
</tr>
<tr>
<td>High grade epithelial abnormality</td>
<td>Moderate epithelial abn. Severe epithelial abn.</td>
</tr>
<tr>
<td>Macroscopic Appearance</td>
<td>Macroscopic findings</td>
</tr>
</tbody>
</table>

**Re-evaluation of revised model:** The revised model was sent to the MoST system and the results were evaluated by 5 clinical experts. Only one expert was used from the original study. The accuracy of the results increased from 64.7% to 80.5% (29/36*100). In addition, the scores given by the experts were higher, thereby increasing the confidence rating from an average of 6/10 to 9/10. There was also much better agreement on the mappings between the experts with all of them agreeing on the same term 93.1% of the times, as against 48.57% times previously. The improved ratings were the result of resolving ambiguous hierarchies and terms in the original model. In addition, clearer term definitions stating the intended semantics and use also helped improve the quality of the results. Some of the matches returned are shown in Table 6.

**Conclusion**

The issue of ensuring good modeling designs and strategies which are unambiguous in representation and intent was discussed in the paper. The main reason for the discussion was to encourage good modeling practice to (i) enable clinicians’ record patient data unambiguously, and (ii) establish reliable and accurate coding of data to standardise data in EHRs.

The experts found it simpler to re-evaluate the mapping results of the revised ‘histology pap’ archetype. Also the mapping accuracy increased from 64.7% to 80.5%. The higher accuracy score meant that not only did the term mapping tool, called MoST, return more appropriate SNOMED codes but also that more clinical experts agreed on the equivalency of the codes. Therefore, before being overly critical of terminology models it is imperative to check the quality and semantic appropriateness of data models that will be used increasingly to capture clinical data in health setups.

Therefore, it is critical for any work which brings together two or more structurally and compositionally different models to concentrate not only on the end result but also the source and the means to that end to ensure quality and precision.

**Acknowledgement**

This work is supported in part by the EU Funded Semantic Mining project and the UK MRC CLEF project.

**References**

8. Qamar R, Rector AL. Semantic Issues in Integrating Data from Different Models to Achieve Data Interoperability. Medinfo Conference 2007, Brisbane, Australia (Accepted).