*Medical Informatics in a United and Healthy Europe*
*K.-P. Adlassnig et al. (Eds.)*
*IOS Press, 2009*

*© 2009 European Federation for Medical Informatics. All rights reserved.*
*doi:10.3233/978-1-60750-044-5-205*

# Functional Requirements of Terminology Services for Coupling Interface Terminologies to Reference Terminologies

Christel DANIEL [a,b,c,1], Antoine BUEMI [d], Laurent MAZUEL [a], David OUAGNE [a], Jean CHARLET [a]

[a] *UMR_S 872 team 20, INSERM, Paris, France*
[b] *UMR_S 872 team 20, Paris Descartes University, France*
[c] *Assistance Publique Hôpitaux de Paris, Georges Pompidou European Hospital, France*
[d] *Assistance Publique Hôpitaux de Paris, SiDoPa, France*

**Abstract.** Desiderata for interface terminologies (IT), designed to support interactions between humans and structured medical information, differ from desiderata for reference terminologies (RT). Terminology experts have recommended that IT be mapped to RT. The interface terminology of the Georges Pompidou European Hospital (GPEH-IT) contains more than 5,000 concepts, sometimes linked to ICD-10 but not yet to the SNOMED 3.5 VF, now available in France. Our objective was to use a formal characterization framework to compare GPEH-IT to SNOMED 3.5 VF and to define the functionalities of terminology services for managing both IT and RT and the mapping between them. We discuss the role of IT and RT in representing the meaning of clinical data.

**Keywords.** interface terminology, reference terminology, terminology server, terminology, controlled vocabulary, medical records systems

## 1. Introduction

The need for clinical terminologies to designate clinical statements in electronic health records is widely recognized. Clinical terminologies can range from simple code-name-hierarchies, e.g., ICD and MeSH, to formal description-logic-based ontologies such as FMA, SNOMED CT or Gene Ontology. Among clinical terminologies, interface terminologies (IT) are "systematic collections of clinically oriented phrases (i.e., terms) aggregated to support clinicians' entry of patient information in computer programs" [1]. Reference terminologies (RT) are designed to provide exact and complete representations of a given domain's knowledge, including entities and their relationships and typically optimized to support storage, retrieval and classification of clinical data. Evaluation methodologies are already available to describe and compare clinical terminologies in an objective and reproducible manner [2–6]. Desiderata for IT differ from desiderata for RT. Specific criteria of IT are balancing pre-coordination and

[1] Corresponding Author: Christel Daniel, INSERM UMRS 872 equipe 20, 15, rue de l'école de medicine, 75006 Paris, France; E-mail: christel.daniel@spim.jussieu.fr.

post-coordination; incorporating assertional knowledge i.e., information that provides nuance and context to a concept, but does not define it [1, 6] and mapping IT to RT. Terminology experts [1] as well as institutional policies [7] do not recommend to bind directly clinical statements to RT, such as SNOMED CT and support using standard RT concepts – and possibly semantic linkages – for representing commonly used IT. A consensus is emerging on requirements for terminology servers and the terminology services to manage RT [8]. Though Common Terminology Services version 2 specifies functions necessary to manage, search, and access terminology content applicable to both IT and RT [9]; the specific functions supporting developers and domain experts in authoring an IT and mapping it to an RT are not clearly defined in the literature.

At the Georges Pompidou European Hospital (GPEH), an interface terminology (GPEH-IT) was populated with "concepts" derived from data entry forms [10]. In addition we plan to also use RT, such as ICD-10 or SNOMED 3.5 VF (French version). Our objective was to use a formal characterization framework to compare GPEH-IT to SNOMED 3.5 VF and to define the functionalities of terminology services for managing both IT and RT and the mapping between them.

## 2. Methods

### *2.1. Comparative Analysis of GPEH-IT and SNOMED 3.5 VF*

Local terms of the GPEH-IT directly derive from the items of questionnaires. Each question is expressed in colloquial terms in the questionnaire and linked to one and only one term or expression of the GPEH-IT. For each coded entry, a value set of the predefined answers is created. Each term of the value set is linked to one and only one term/expression, preexisting in GPEH-IT or newly defined. It is possible to associate GPEH-IT terms or expressions to concepts selected in a reference terminology (ICD-10 and recently SNOMED 3.5 VF, distributed in France since 2008).

We considered different desiderata [4–6] and evaluation frameworks of clinical terminologies (CT) [1, 3]. For Cornet et al. [3], there are 6 types of CT: terminology (list of terms), thesaurus (alphabetically ordered terms with synonyms), classification (arrangement of concepts using is-a relationships), vocabulary (providing free-text or formal definitions of concepts), nomenclature (providing composition rules) and coding system. The authors defined an evaluation framework for CT that we extended with criteria from [1, 4–6] and used to compare GPEH-IT and SNOMED 3.5 VF.

### *2.2. Requirements for Terminology Servers*

Based on the results of the comparative analysis of GPEH-IT and SNOMED 3.5 VF, we investigated the terminology services specifications that could be useful to manage, search and access the terminology content of these two terminologies. We emphasized the specific requirements related to specific criteria of IT and especially mapping IT to RT.

## 3. Results

### *3.1. Evaluation Framework and GPEH-IT/SNOMED 3.5 VF Comparison*

The evaluation framework includes 39 criteria consisting in formalism-related criteria (n=27) and content-related criteria (in italics; n=12). Each criteria may be applicable to both interface and reference terminologies (IT, RT) or be specific to either interface (IT) or reference (RT) terminology.

**Table 1.** Comparison of GPEH-IT and SNOMED 3.5 VF. Criteria defined by: [a]Cimino [5], [b]Chute [6], [c]ISO [4], [d]Cornet [3]

| Criteria | GPEH-IT | SNOMED 3.5 VF |
|---|---|---|
| TERMINOLOGY | | |
| Distinction between concepts and terms[a,c,d] (RT) | Yes | Yes |
| *Total nb of concepts[d] /Total nb of terms[d]* | *5,426/7,693* | *100,731/145,532* |
| Term length restriction[d] (IT, RT) | Yes | None |
| *Covered areas[d]* | *See below* | *See below* |
| Concept obsoletion mechanism [a,b,c,d] (IT, RT): Concept status flag-Author tracked -Motivation | Yes-Yes-No | Yes-No-No |
| THESAURUS | | |
| Synonym representation[a,b,c,d] (IT, RT) | Yes | Yes |
| *Nb of synonyms by concepts (average)[d]* | *0.61* | *0.44* |
| Multilingual representation [b,c,d] (RT) | No | Yes |
| *Languages[b,d]* (RT) | *French* | *11 languages* |
| Synonyms for fragments[d] (IT, RT) | No | No |
| Synonym obsoletion mechanism[d] (IT, RT) | No | No |
| VOCABULARY | | |
| Free-text concept definition[d] (IT, RT) | Yes | No |
| *Nb of vague or ambiguous[d] / redundant concepts[d] (IT, RT)* | *ND/ND* | *ND/ND* |
| Formal concept definition[a,b,c,d] (RT) | No | Multiaxial definition |
| *% of concepts with free/formal definition[d] (IT, RT)* | *ND* | *16.64%** |
| Explicitly defined relationships[a,b,c,d] (RT) | No | No |
| Definition obsoletion mechanism (IT, RT) | No | No |
| Relationships used[d] (IT, RT) | No typed relat. | Axis type |
| CLASSIFICATION | | |
| Hierarchical relationships[d] / Polyhierarchies [a, b,c,d] (IT, RT) | Yes/No | Yes/No |
| Properties inheritance[d] (RT) | No | No |
| *Nb of parents per concepts (average)[d] (IT, RT)* | *1* | *1* |
| Hierarchical depth restriction[d] / breadth restriction[d] (IT, RT) | No/No | Yes/Yes |
| Classification inferred(/concept definitions)[d] (RT) | No | No |
| NOMENCLATURE: | | |
| Composition formalism[a,b,c,d] (IT, RT) | No | Yes |
| Detection of equivalent definitions[a,b,c,d] (RT) | No | No |
| Composition may change meaning[a,b,c,d] (IT, RT) | NA | Yes (absence of) |
| Assertional knowledge (value sets) (IT) | Yes | No |
| *Nb of refinable concepts, Degree Of Freedom[d] (IT, RT)* | *ND* | *ND* |
| CODE SYSTEM | | |
| Code generation mechanism[a,b,c,d] (IT, RT) | Manual | AxisID+4 to 6 digits code |
| Meaningful identifiers[a,b,c,d] (IT, RT) | No | Yes |
| Code length restriction [a,b,c,d] (IT, RT) | Yes | Yes |
| *% of concepts coded[d] (IT, RT)* | *<1%* | *100%* |
| Cross mappings with other code systems[d] (RT) | No | ICD9CM; ICD10 |
| *% of cross mappings (RT)* | *0%* | *0%* |

*Rate of concept with multiaxial definition according to SNOMED 3.5 VF axes: Chemistry/Biology: 5.17%; Living subjects: 0.80%; Functions: 4.52%; Topography: 10.59%; Morphology: 5.97%

GPEH-IT comprises 5,426 "concepts" among which 2,087 derive from the entries of the questionnaires and 3,339 from value sets. "Concepts" were organized according to the SNOMED 3.5 VF covered areas. Each "concept" of GPEH-IT is characterized by a preferred term or expression. A term can be preferred for no more than one concept, since the system recognizes redundancy. Each "concept" of GPEH-IT can be defined by a free text description but there is no formal representation of concepts and no explicit relationships between concepts. "Concepts" are organized into a mono-hierarchy and can be associated to meaningless local codes.

According to criteria specific to interface terminologies, GPEH-IT incorporates **assertional knowledge** since it is possible to attach a value set to each "concept" consisting of a list of allowable associated "concepts". For example, the assertional knowledge related to "hypertension" includes links to severity modifiers ("mild", "moderate", and "severe"), to contextual modifiers (e.g., "young patient hypertension"), to specific clinical form (e.g., "resistant hypertension" or "complicated hypertension"), to associated symptoms based on specific etiologies (e.g., "proteinury"). GPEH-IT does not provide any composition mechanism except **pre-coordination**. The **mapping between GPEH-IT and RT** is achievable in linking a "Concept" of GPEH-IT to a coded concept of either ICD-10 or SNOMED 3.5 VF but the cross mapping between GPEH-IT and these RT is far to be complete.

### *3.2. Functional Requirements of Terminology Services for IT Coupled with RT*

Common Terminology Services version 2 represents a set of functions necessary to manage, search and access the terminology content [9]. We extended the definitions of [9] with regards to terminology services required due to specific criteria of IT.

**Table 2.** Functional requirements of terminology services for managing IT coupled with RT

| |
|---|
| ***Mapping IT to RT***: "IT authors" may create IT "concepts" by identifying them or creating them from concepts contained in RT such as SNOMED 3.5 VF. The following functionalities may support "IT authors" in creating, querying and maintaining associations between IT and RT concepts.<br>• Searching candidate RT concept(s) to be associated to an IT term and supporting different types of associations: exact match, equivalent to, broader than, narrower than, etc.<br>• In cases of partial match, supporting IT term composition from RT concepts : selecting RT concepts and defining their relationships<br>• Specifying additional notes and information (such as versions of the IT and RT being used to create the association(s), additional comment explicating the association(s), if the association is one-to-one or one-to-many, etc)<br>• Retrieving definitions (free text definition or formal properties) of RT concept(s) associated to an IT term. Browsing descendants of RT concept(s) associated to an IT term. Translating a given IT term to other languages (available in RT).<br>• Mapping IT terms to concepts from more than one RT<br>• Identifying obsolete/deprecated RT term(s)/concept(s) associated to an IT term or obsolete/deprecated semantic relationships(s) in RT term(s)/concept(s) associated to an IT term. Updating IT terms when changes occur in the associated RT term(s)/ concept(s) |
| ***Assertional knowledge***: Managing assertional knowledge requires creating and maintaining value sets in the IT.<br>• Creating value sets by intension or by extension (with concepts that first exist in the IT and, as much as possible, are mapped to RT concept(s).<br>• Updating value sets: if an "IT user" identifies errors in a value set, "IT authors" may re-define the value set to be accurate to the understanding of the "IT user". |
| ***Balancing pre-coordination and post-coordination***:<br>• Creating meaningful pre-coordinated "concepts" in defining their relationships with atomic "concepts" of the IT. "IT authors" may create a new relationship type to be used to link two "concepts" if needed.<br>• Updating pre-coordinated "concepts": "IT users" may create post-coordinated "concepts" and these change requests may be reviewed by "IT authors" and when appropriate included in IT as pre-coordinated "concepts". |

## 4. Discussion

ITs provide healthcare professionals with a rich synonymy that allows them to represent clinical entities using the words or phrases that they prefer. In this study, we compared GPEH-IT and SNOMED 3.5 VF according to an evaluation framework based on [3] in order to identify their respective specific characteristics and to guide the specification of terminology services for managing GPEH-IT coupled with RT such as SNOMED 3.5 VF.

Rosenbloom investigated how to use SNOMED CT to represent two ITs (MEDCIN and CHISL) [11]. SNOMED CT provided concepts to represent 92.4% of MEDCIN and 95.9% of CHISL terms when no strict mapping for semantic linkages was required and 41% (MEDCIN) and 33% (CHISL) with such requirements. Rosenbloom et al. encourage terminology developers to create or enrich IT by using RT concepts as a starting point but stress out that assembling clinically meaningful compositions, appropriate synonyms and linkage between concepts (related concepts or modifiers) is a labor-intensive approach. With regards to MEDCIN, the authors have only investigated 500 terms among the 215,000 concepts of MEDCIN (0.2%) and do not clearly state if they intend to map the remaining MEDCIN concepts and do not describe the tooling and human resources that are required to perform this task [11]. Although GPEH-IT is much smaller than MEDCIN (5,000/215,000), its mapping to a RT such as SNOMED 3.5 VF is a challenging task and it is an important issue to provide terminology services that allow them to manage both interface and reference terminologies as well as the binding process between them. Searching functionalities for exact match but also for conceptual mapping (using synonyms) and total or partial post-coordinated mapping are particularly important in order to support IT terms composition from RT concepts.

## References

[1] Rosenbloom, S.T., Miller, R.A., Johnson, K.B. (2006) Interface terminologies: Facilitating direct entry of clinical data into electronic health record systems. *Journal of the American Medical Informatics Association* 13(3):277–288.

[2] Elkin, P.L., Brown, S.H. et al. (2002) Guideline and quality indicators for development, purchase and use of controlled health vocabularies. *International Journal of Medical Informatics* 68(1-3):175–186.

[3] Cornet, R., de Keizer, N.F., Abu-Hanna, A. (2006) A framework for characterizing terminological systems. *Methods of Information in Medicine* 45(3):253–266.

[4] ISO/TC251 WG3 (2000) Standard specification for quality indicators for controlled health vocabularies; 2000 July. *Report TS 17117*.

[5] Chute, C.G., Cohn, S.P., Campbell, J.R. (1998) A framework for comprehensive health terminology systems in the United States: Development guidelines, criteria for selection, and public policy implications. *Journal of the American Medical Informatics Association* 5(6):503–510.

[6] Cimino, J.J. (1998) Desiderata for controlled medical vocabularies in the twenty-first century. *Methods of Information in Medicine* 37:394–403.

[7] AHMIA, AMIA. *Healthcare Classification and Terminologies. An Action Agenda for the United States*. http://www.amia.org/files/terminologiesandclassifications.pdf.

[8] Cimino, J.J. (2001) Terminology tools: State of the art and practical lessons. *Methods of Information in Medicine* 40(4):298–307.

[9] HL7-OMG.CTS 2, http://hssp.wikispaces.com/cts2.

[10] Daniel-Le Bozec, C., Steichen, O., Dart, T., Jaulent, M.C. (2007) The role of local terminologies in electronic health records. The HEGP experience. In *Proceedings of Medinfo 2007*, *Studies in Health Technology and Informatics* 129(Pt 1):780–784.

[11] Rosenbloom, S.T., Brown, S.H., Froehling, D. et al. (2009) Using SNOMED CT to represent two interface terminologies. *Journal of the American Medical Informatics Association* 16(1):81–88.