# Modeling and Integrating Terminologies into a French Multi-terminology Server

**Michel Joubert[a], Tayeb Merabti[b], Pierre-Yves Vandenbussche[c], Hocine Abdoune[a], Badisse Dahamna[b], Marius Fieschi[a] , Stefan Darmoni[b]**

*[a] LERTIM, EA 3283, Faculty of Medicine, University of Aix-Marseille II, Marseille. France*
*[b] CISMeF, Rouen University Hospital & LITIS, EA 4108, Biomedical Research Institute, Rouen. France*
*[c] INSERM, UMRS 872, eq. 20 & MONDECA, Paris. France*

**Abstract**

*The aim of the present work was 1) to design a terminology meta-model into which all terminology models can be integrated and 2) to design and to implement a process capable of integrating terminologies into a French-language medical multi-terminology server. This approach has the advantage of combining respect for each terminologies structure with a regrouping of the meta-data inherent in each terminology. Each specific class defined for a given terminology corresponds to a class specialization in the meta-model. This ensures the coherence of the overall system and provides the essential platform rendering the terminologies interoperable. Mappings between terminologies are then built on the basis of the UMLS Metathesaurus and are entered into the server.*

***Keywords:*** *Terminology, modeling, Unified Medical Language System, terminology mapping, semantic interoperability.*

## Introduction

Previous to the construction of a health multi-terminology server we achieved [1] no such server existed for French. Following this programme, the aim of the InterSTIS[1] project, supported by the French National Agency for Research, is to develop a French-language health multi-terminology server in which the different terminologies are interoperable. The ultimate purpose is the implementation of a shared semantic referential permitting effective and efficient interaction with a minimal loss of meaning. Semantic interoperability, in our view, requires a joint model for the representation of terms, whatever their terminology or referential of origin, as well as the resources to connect the terms in one terminology with the equivalent terms, whether direct or indirect, in other referentials.

Here, we first present a meta-model which acts as a reference enabling the terminologies to be integrated into the server. It is based on previous work in the same field [2, 3]. We then present a process developed in order to integrate the terminologies into the server. Finally, the mappings between terminologies are constructed on the basis of the UMLS and are entered into the server separately from the terminologies. The construction of these mappings draws upon previous work in this area [4-6].

## Materials and Methods

### Unified Medical Language System

One of the UMLS knowledge resources is its Metathesaurus [7]. More specifically, within the latter, we will be using:

- the MRCONSO table, which lists all the concepts incorporated in the UMLS with no duplication and in which every concept is attributed a unique identifier (CUI);
- the MRREL table which describes the relationships, if any, between concepts in the original terminologies;
- the MRMAP table which provides explicit mappings between integrated terminologies.

When an explicit mapping relationship exists (e.g. ICD-9-CM to SNOMED CT [8]) between two concepts $CUI_1$ and $CUI_2$, it is likely that all terms designating $CUI_2$ may be mapped to terms designating $CUI_1$, whatever the terminologies and whatever the language in which they are formulated. In other words, explicit mappings between two terminologies can be "reused" for other terminologies by means of the UMLS concept structure [6]. Our study was conducted using the 2008 AA version of UMLS.

### Modeling terminologies

The purpose of the present study was:

- to design a terminology meta-model into which every terminology model can be integrated [9] and
- to design a process capable of integrating terminologies into a health multi-terminology server.

This approach has the advantage of combining respect for the original structure of each of the terminologies with a re-

[1] Acronym for: Semantic Interoperability of Terminologies in French Health Information Systems.

grouping of the meta-data inherent in each terminology. The resulting terminology meta-model is shown in the UML diagram [10] in Figure 1.

As shown in figure 1, the meta-model is divided into four parts: *Terminological*, *Conceptual*, *Alignment* and *Group*. The most important class in the *Conceptual* part is *Concept* which is central to the entire model. The model allows extensions in order to express more specific entities for particular terminology structures. For instance, the *BroaderNarrower* relationship could be specialized in *Part_of* or *Is_a* relationship. The classes in the *Terminological* part are associated with *Concept* through its principal class *PreferredTerm*. The concepts in the different terminologies can be mapped out using different relationships (cf. below), as described in the *Alignment* part. A significant feature for terminology representation is its ability to describe all kinds of concept collections, sets, etc. Based on the SNOMED Clinical Terms model [11], we defined the notion of *Group* which can tackle most collection concerns.

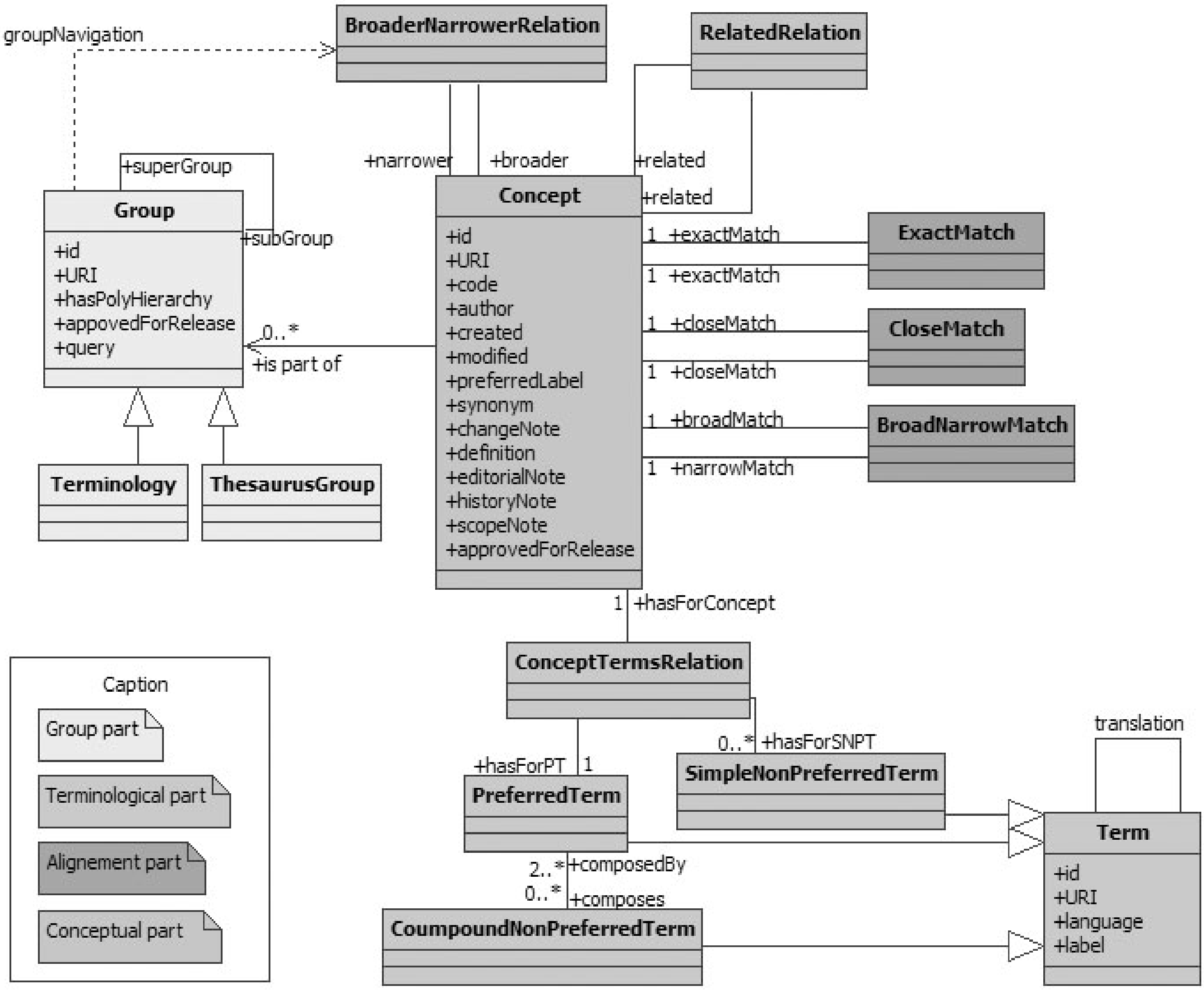


*Figure 1- UML class diagram of the terminology Meta-model.*

**Mapping terminologies**

The mapping method we used is inspired by SKOS definitions of mapping properties [12]. It is as follows: suppose two descriptors $t_1$ and $t_2$ of two terminologies $T_1$ and $T_2$, respectively; suppose $CUI_1$ and $CUI_2$, the respective projections of $t_1$ and $t_2$ in the Metathesaurus, then $t_1$ and $t_2$ are mapped if:

- $CUI_1$=$CUI_2$ (in MRCONSO), this corresponds to the SKOS transitive *ExactMatch* property; or
- there is a parent of $t_1$ or $t_2$ which maps $t_2$ or $t_1$ respectively (using MRREL), this corresponds to *BroadMatch* and/or *NarrowMatch* properties: these are used to state mapping links through hierarchies; or

- there is an explicit mapping between $CUI_1$ and $CUI_2$ (in MRMAP); this corresponds to the non-transitive *CloseMatch* property: two concepts are sufficiently similar that they can be used interchangeably.

The algorithm is carried out sequentially and stops if a candidate mapping is found.

As an application of this, even if an explicit mapping comes from other terminologies, e.g. ICD-9-CM and SNOMED CT not part of the terminologies under consideration, it still applies to $t_1$ and $t_2$ since it is established between $CUI_1$, to which $t_1$ is attached, and $CUI_2$, to which $t_2$ is attached.

We will discuss later the use of the SKOS *RelatedMatch* property which we plan to define as an associative mapping link between concepts.

## Results

The MONDECA ITM® platform [13] was chosen to integrate the terminologies. Each terminology model is in conformity with the meta-model. The terminologies are presented to the ITM® platform in the OWL format [14]. Mappings between terminologies are presented in the RDF format [15], which is both simpler and better adapted.

The method we adopted was as follows:

- creation of an OWL representation of each terminology;
- integration of each terminology into the ITM® platform database separately from the others;
- insertion of mappings between the integrated terminologies.

This approach offers a great deal of flexibility since a terminology can be integrated into the platform at any time. Its mappings with the other already integrated terminologies, if any, are then entered separately, once again at any time.

**Production of an OWL-formatted terminology**

Production of an OWL-formatted terminology requires two stages:

- a pre-processing phase in order to modelize the terminology and present it in a standard format;
- the writing of a parser capable of translating into OWL format the contents of the terminology described by its model.

The model of a terminology is operated in UML. Its content is modelized in SQL or XML. A manual operation is necessary to ensure the compliance of a model (e.g. the UML model) with the other one (e.g. the SQL/XML model). All that is illustrated schematically in Figure 2.

The second stage involves translating the UML model of the terminology into a set of Java classes that the parser can then exploit. For this, we use SWOOP software [16]. The parser will thus provide a representation of the terminology in OWL format.

A final process (serialization) involves instantiating the different classes of the OWL model with the data from the original terminology as permanent objects in the ITM® database.

**Production of inter-terminology mappings**

Inter-terminology mappings function two by two as previously described. They are performed in a relational database management system in order to produce three files in RDF format for each pair of terminologies: an *ExactMatch* file, a file describing the *BroadMatch / NarrowMatch,* and a final file describing the *CloseMatch.* These files are then submitted to ITM® for integration with the system.

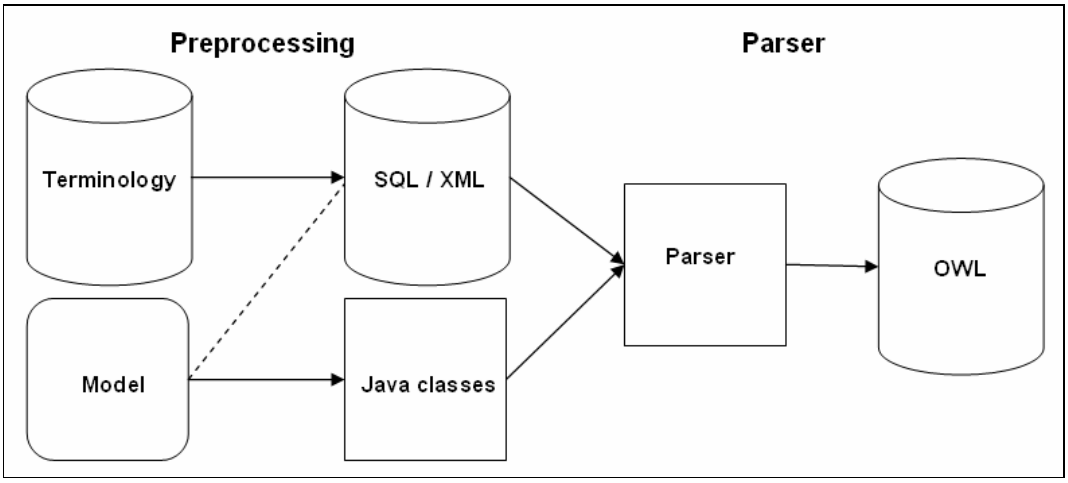


*Figure 2- Representation process of an OWL-formatted terminology.*

## Discussion

The model of each terminology integrated into the system is a specialization of the meta-model. This is true not only for the classes of objects but also for the classes of relationships. This ensures the coherence of the overall system and provides the necessary foundation to enable terminologies interoperability. The meta-model as well as the process integrating the terminologies into the server are sufficiently general and independent of the language to allow them to be deployed for languages other than French.

So far, not all the French-language terminologies, or their English-language equivalents, when they exist, have been integrated with UMLS. A case in point is CCAM [17] which is entirely French. This provides a classification of medical procedures which, following decomposition of a code, delivers its contents under three headings: action, morphology and technique. We decided to match each element resulting from the decomposition of a code in the Metathesaurus. To do that we used a translation table of French terms into English ones coming from a previous work [18] and Metamap [19] to match English terms to Metathesaurus concept names. In order to evaluate this approach, we began to decompose the SNOMED International [20] procedure codes and then to match them to Metathesaurus concept names by the means of Metamap too. In this way, we hope that the concepts recovered from the Metathesaurus after decomposition of the CCAM codes will match those obtained by decomposing the SNOMED procedure codes. This approach constitutes a particular mode of operating a *RelatedMatch* between two concepts from two terminologies. If the elements resulting from their decomposition are projected in a similar fashion in the UMLS Metathesaurus, one might well conclude that they are sufficiently close to one another to be mapped, in the same way that related articles are indexed using same key-words in a documentary system.

### Acknowledgments

This study forms part of the InterSTIS project which is partly funded by the French Agency for Research under grant: ANR-07-TECSAN-010. The authors wish to thank the U.S. National Library of Medicine for kindly providing versions of the UMLS.

## References

[1] Darmoni SJ, Joubert M, Dahamna B, Delahousse J, Fieschi M. SMTS®: a French Multi-terminology Server. *Proc AMIA Annu Symp* 2009. To appear.

[2] Masolo C, Borgo S, Gangemi A, Guarino N, Oltramari A. WonderWeb ontology library. *European IST Project: Ontology Infrastructure for the Semantic Web*, deliverable D18 (2003).

[3] Pathak J, Solbrig HR, Buntrock JD, Johnson TM, Chute CG. LexGrid: A Framework for Representing, Storing, and Querying Biomedical Terminologies from Simple to Sublime**.** *J Am Med Inform Assoc* 2009; 16: 305-315.

[4] Cimino J, Johnson S, Peng P, Aguirre A. From ICD9-CM to MeSH using the UMLS: a how-to guide. *Proc Annu Symp Comput Appl Med Care* 1993: 730-734.

[5] Bodenreider O, Nelson SJ. Beyond synonymy: Exploiting the UMLS Semantics in Mapping Vocabularies. *Proc AMIA Annu Symp* 1998: 815-819.

[6] Fung KW, Bodenreider O. Utilizing the UMLS for semantic mapping between terminologies. *Proc AMIA Annu Symp*. 2005: 266-270.

[7] National Library of Medicine, UMLS Metathesaurus. http://www.nlm.nih.gov/pubs/factsheets/umlsmeta.html

[8] Imel M. A Closer Look: the SNOMED Clinical Terms to ICD-9-CM Mapping. *Journal of AHIMA* 2002; 73: 66-69.

[9] Vandenbussche PY, Charlet J. Méta-modèle général de description de ressources terminologiques et ontologiques. *Proc Ingénierie de la Connaissance* 2009: 193-204.

[10] Object Management Group. Unified Modeling Language. http://www.uml.org/

[11] Stearns MQ, Price C, Spackman KA, Wang AY. SNOMED clinical terms: overview of the development process and project status. *Proc AMIA Annu Symp.* 2001: 662-6.

[12] World Wide Web Consortium. Simple Knowledge Organization System. http://www.w3.org/2004/02/skos/

[13] MONDECA. Intelligent Topic Manager. http://www.mondeca.com/index.php/en/applications/itm_t3

[14] World Wide Web Consortium. Web Ontology Language. http://www.w3.org/TR/owl-features/

[15] World Wide Web Consortium. Resource Description Framework. http://www.w3.org/RDF/

[16] Kalyanpur A, Parsia B, Hendler J. A Tool for Working with Web Ontologies. *International Journal on Semantic Web and Information Systems* 2005; 1: 36-49.

[17] Assurance Maladie. Classification Commune des Actes Médicaux, http://www.ameli.fr/accueil-de-la-ccam/index.php

[18] Rodrigues JM, Trombert-Paviot B, Baud R, Wagner J, Meusnier-Carriot F. Galen-In-Use: using artificial intelligence terminology tools to improve the linguistic coherence of a national coding system for surgical procedures. *Stud Health Technol Inform.* 1998: 623-7.

[19] National Library of Medicine. Metamap. http://skr.nlm.nih.gov/papers/index.shtml#MetaMap

[20] Lussier YA, Rothwell DJ, Côté RA. The SNOMED model: a knowledge source for the controlled terminology of the computerized patient record. *Methods Inf Med.* 1998; 37: 161-164.

## Address for correspondence

Michel JOUBERT
LERTIM, Faculte de Medecine, Universite d'Aix-Marseille II
27 bld Jean Moulin, 13005 Marseille. France

mjoubert@ap-hm.fr
http://www.lertim.org