Serveur Terminologique

Consolidation d'un noyau de base et d'une méthodologie de développement d'un Vocabulaire Médical Contrôlé (CMV) dans le cadre de la mise en place future d'un serveur belge de terminologie dans le secteur de la santé.

Ref. Ctr-n10-05-serveur terminologique-MIM.2, financé par le SPF Santé Publique

Partie 3/4 : Terminological Markup Framework (TMF) norm

Section A : Application de la norme ISO 16 642 (TMF) au CMV belge.

Version : 1.0

31 octobre 2011

Auteurs : Joseph Roumier[1, 2], Robert Vander Stichele[1]

[1] *Heymans Institute of Pharmacology, Ghent University, Belgium*
[2] *Centre of Excellence in Technology and Communication (CETIC), Belgium*

Comité d'accompagnement : Etienne De Clercq (président de la MIM),
Luc Nicolas (représentant SPF Santé publique)

Note : cette partie du rapport contient deux sections

Section A : Application de la norme ISO 16 642 (TMF) au CMV belge

Section B (uniquement sous format excell) : European Multilingual Glossary for popular and technical medical terms linked with the proposed Terminological Markup Language

Editeur. : MIM

Editorial note : /

Analysis of the application of

ISO 16 642 norm

Terminological Markup Framework (TMF)

for the Belgian Reference Terminology

Authors:

Joseph Roumier[1,2]

Robert Vander Stichele[1]

[1] *Heymans Institute of Pharmacology, Ghent University, Belgium*
[2] *Centre of Excellence in Technology and Communication (CETIC), Belgium*

Date:

28th October, 2011

Contractor :

Belgian Medical Informatics Association,
on behalf of the Belgian Federal Department of Health

Steering committee:
E. De Clercq (MIM – president)
L. Nicolas (SPF, Federal department of Health)

Version
Final version 1.0
ref.contratMIM_Ghent_U.

Next page - picture: Dandelion 037 - Dandelion in the sun, cygnus921,Attribution 2.0 Generic (CC BY 2.0)

Executive Summary

Depending on the time, language, country, culture, school of thought, whether you are a patient or a physician, a specialist or a general practitioner, whether the goal is medical care or public health management, different terms are used to represent the same set of concepts that relate to health, body, care, etc.

These different terms have many reasons to be valid. To seek a unity of language and words would be detrimental to the health of patients. But this makes the exchange of information (semantic interoperability) difficult, for example the exchange of medical reports written in different times, places, and languages. Achieving correct semantic interoperability would be useful for traveling patients, but also for collecting data to inform health care policy decisions.

To meet this double requirement of taking into account the diversity of terms and to achieve semantic interoperability, a Belgian reference terminology for storage and retrieval of medical information is needed.
In this report we present a methodology to structure and implement a concept-oriented and multilingual terminology compatible with its publication on the Internet.

The proposed approach relies on two standard technologies :

- The ISO standard to develop models for multilingual terminologies that take into account the diversity of terminology sources while preserving interoperability and sustainability (Terminological Markup Framework ISO 16642).
- The W3C standards for the publication of semantic data on the Internet.

In this report we present the methodological choices for the structure of a national reference terminology, which is :

- multilingual
- multidisciplinary
- Connected to conceptual medical nomenclatures (SNOMED-CT and Unified Medical Language System)
- Linked to several international classifications (e.g. ICPC and ICD)
- Integrated with lexical resources, oriented to end-users (all medical disciplines, but especially general practitioners and patients).

Such an approach allows to control the size and thus the maintainability of a reference terminology while promoting the independent existence of nomenclatures, terminologies, glossaries, specialized glossaries whose evolution will not be hampered by a resource trying to include everything at the risk of becoming unmanageable.

The technical solution chosen to publish it as Linked Data provides the opportunity to link the reference terminology with other useful resources.

The results presented include :

- the rationale for the proposed approach

- the methodology of the approach
- an XML Terminological Modeling Language for a reference terminology terminology, based on the ISO standards of the domain
- a typology of end-users (or medical knowledge domains)
- a typology of health care concepts
- a conversion table for the exchange of terminological results between past and future projects
- a proposal for the publication of the reference terminology as Linked Data
- a report on the most salient points of discussion that emerged during the preparation of the report.

Résumé Exécutif

Selon l'époque, la langue, le pays, la culture, l'école de pensée, selon que l'on est patient ou médecin, spécialiste ou généraliste, que l'objectif est la gestion médicale ou économique de la santé, des termes différents sont employés pour représenter le même ensemble de concepts qui se rapportent à la santé, au corps, aux soins, etc.
Ces termes différents ont des raisons d'être souvent parfaitement valables. Chercher à obtenir une unicité de langage et de termes serait in fine dommageable à la santé des patients. Mais cela rend difficile l'échange d'informations (l'interopérabilité sémantique), par exemple la prise en compte de rapports de soins rédigés en des temps, des lieux et des langues différents. Arriver à une interoperabilité sémantique correcte serait pourtant utile pour les patients en déplacement, mais également pour décider de politiques de santé.
Pour répondre à cette double exigence de prise en compte de la diversité des termes et d'interopérabilité sémantique, une terminologie belge de référence doit être mise en place pour l'indexation et la recherche d'information.
Le présent document propose une méthodologie pour la structuration et la mise en place d'une telle terminologique conceptuelle et multilingue, compatible avec une publication sémantique sur le Web.

L’ approche proposée s'appuie sur deux axes technologiques standardisés :

- d'une part l'existence de standards ISO permettant de développer des modèles terminologiques multilingues respectant la diversité des sources terminologiques qui soient néanmoins interopérables et pérennes (Terminological Markup Framework ISO 16642);
- d 'autre part l'existence de standards W3C pour la publication de données sémantiques sur le Web (Linked Open Data).

Dans ce rapport, la méthodologie choisie est présentee et une structure est proposée pour une terminologie de référence nationale, qui est :

- multilingue
- pluridisciplinaire
- connectée à des nomenclatures médicales conceptuelles (SNOMED-CT et Unified Medical Language System)
- liée à plusieurs classifications internationales (par example ICPC et ICD)
- liée à des resources lexicales orientées vers les utilisateurs finaux (toutes spécialités comprises, en particulier les médecins généralistes et incluant les patients).

Une telle approche permet de maîtriser la taille et donc la maintenabilité d'une terminologie de référence en favorisant l'existence autonome de nomenclatures, terminologies, lexiques, glossaires spécialisés dont l'évolution ne sera pas freinée par une ressource englobante et rendue inerte par la lourdeur de sa maintenance.

La solution technique de publication sur le Web sémantique sous forme de Linked Open Data permet de lier la terminologie de référence avec les autres ressources utiles.

Les résultats présentés sont :

- l'argumentaire de l’ approche proposée
- la méthodologie de l’ approche proposée
- un langage XML pour un modèle terminologique de terminologie de référence, basé sur les standards ISO du domaine
- une typologie d’ utilisateurs (où de domaines de connaissances médicales)
- une typologie de concepts médicaux
- un fichier de correspondance pour échanger les résultats de travaux terminologiques antérieures et futurs.
- des propositions pour la publication de la terminologie de référence en Linked Open Data
- une discussion des alternatives propres à la gestion terminologique.

Korte Inhoud

Naargelang de tijd, de taal, het land, de cultuur, de denkrichting, naargelang het gaat om een patiënt of een arts, een huisarts of een specialist, de kwaliteit van de zorg of het management van de zorg is, altijd zullen er verschillende termen worden gebruikt om concepten te verwoorden die te maken hebben met gezondheid, licham en zorgverstrekking.

Die verschillen in het gebruik van termen hebben een bestaansreden. Een keurslijf van eenheid van taal en woorden zou afbreuk doen aan gezondheid van patiënten. Het bemoeilijkt wel de uitwisseling van informatie (semantische interoperabiliteit), bijvoorbeeld de uitwisseling van medische rapporten, geschreven op verschillende tijdstippen, verschillende plaatsen, landen of talen. Het realiseren van correcte semantische interoperabiliteit is nuttig voor patiënten die reizen in verschillende landen, maar ook voor de inzameling van gegevens voor het beleid in de gezondheidzorg.

Om de dubbele uitdaging van de diversiteit van termen en de noodzaak voor semantische interoperabiliteit te realiseren, is er een Belgische Referentie terminologie nodig, die zowel het opslaan als het opzoeken van gezondheidsinformatie ondersteunt.

In dit rapport stellen wij een methode voor om een concept-geörienteerde, multilinguale, terminologie te bouwen en te publiceren in het semantische web.

De voorgestelde aanpak steunt op twee technologische standaarden :

- The ISO standaard voor het datamodel van multilinguale terminologiën, waarin rekening wordt gehouden met de diversiteit van terminologische bronnen, zonder interoperabiliteit en ondeherhoudbaarheid te ondermijnen (Terminologocial Markup Framework ISO 16642).
- The W3C standaarden voor publicatie in het semantische web.

In dit rapport stellen wij de methodologische keuzes voor bij het structureren van een nationale referentie terminologie met de volgende kenmerken :

- multilinguaal
- multi-disciplinair
- verbonden met conceptuele medische nomenclaturen (SNOMED-CT en Unified Medical Language System)
- gelinkt aan verschillende internationale classificaties (bv. ICPC and ICD)
- geïntegreerd met lexicale bronnen, die op eind-gebruikers zijn gericht (all medische en paramedische disciplinses, maar vooral op de huisarts en ook op de patënt).

Een dergelijke aanpak laat toe de grootte en dus de ondehoudbaarheid van een referentie terminoligie in de hand te houden, naast de apart bestaande nomenclaturen, terminologiën, en gespecialiseerde glossaria, waarvan de evolutie niet zal gehinderd worden door een systeem dat probeert alles in te sluiten en daardoor onbeheersbaar wordt.

Er is gekozen voor de technische oplossing om de referentie terminologie te publiceren als Linked Open Data, wat de mogelijkheid biedt om te linken met anderen nuttige informatiebronnen in het semantische web.

De volgende resultaten worden hier voorgesteld :

- de rationale voor de aanpak
- de methodologie voor de aanpak
- een XML Terminological Modeling Language voor de referentie terminologie, gebaseerd op de heersende ISO standaarden
- een typologie van eind-gebruikers (of van medische kennisdomeinen)
- een typologie van concepten in de gezondheidszorg
- een conversietafel voor de uitwisseling van terminologische gegevens tussen lopende and komende projecten
- een voorstel om de referentie terminologie te publiceren als Linked Open Data
- een overzicht van de belangrijkste discussiepunten, gerezen tijdens de opstelling van dit rapport

Table of Contents

Illustration Index

1 Introduction/Motivation

1.1 Context

Clinical medicine has evolved in the 21th century to a stage where information technology is pervasively present in every day practice. Health care activities are continuously recorded and kept for later consultation and knowledge extraction, and guided by the provision of factual and conceptual evidence-based knowledge. Almost all workplaces of health care workers are equipped with powerful computers, capable of storing a high number of patient records in sophisticated electronic health care records. Ergonomic devices permit recording of clinical data at point of care but also the consultation of remote bibliographic information, or evidence-based information, specifically tailored for point-of-care consultation. The production, maintenance and dissemination of such scientific information repositories is now facilitated through the internet, and increasingly integrated in what is called “the Semantic Web” [Horrocks, 2007].

Health care workers communicate with each other increasingly in electronic formats, with standardized telematic messages, delivered trough secured, central eHealth platforms. The forum of this exchange has become global, multidisciplinary, and multilingual, inducing new demands for semantic interoperability [Anonymous, 2010].

All this calls for radical changes in the attitude, knowledge and capabilities of health care workers. Most of them are now convinced of the necessity to continuously record their activities, and to do so in a structured, standardized and semantic interoperable way. In their vocational training, more attention is been given to knowledge transfer of the appropriate classification systems and techniques to facilitate information storage and there is more training in appropriate record keeping techniques and privacy protection methods. Most health care workers now have capabilities to analyze their own clinical records and to extract new knowledge. Their search capabilities and performance in information retrieval from knowledge databases on the web have been strengthened.

Nevertheless, there is a huge need in the current generation of health care workers for extensive technological support in performing information storage and retrieval activities. This support is needed both in developing intuitive human-machine interfaces, as well as in developing terminological support to translate human thought into machine-readable concepts.

In the past century international classification systems, nomenclatures and thesauri have been developed for primary and specialized care. In addition, reference and end-user terminologies have been developed to facilitate the interaction of the user with these information systems. The latter resources however are still poorly developed, lack interoperability, and are mostly uni-lingual.

1.2 Description of the mission

In Belgium, a country with 3 official languages (French, Dutch, German) the need for a multilingual medical terminology has been recognized for years. Terminologies (3BT : Belgian Bi-coded, Bilingual Thesaurus [3BT, 2006], and LOCAS: Logiciel de Codage et d' Acquisition de Synonymes [Jamoulle and Roland, 1993]. [Roland et al., 2000]) have been developed by family physicians and are used in EHR systems and in referral letters from primary care emergency services. These were based on and mapped to ICPC (the International Classification of Primary Care [Okkes et al., 1999], [Okkes et al., 2002], [ICPC-2-R, 2005]) and ICD (International Classification of Diseases [ICD-10, 2010],

[Tuderache et al., 2010]). In the hospitals, reference terminologies have been developed for structured discharge letters and administrative billing purposes (e.g. VUB Terminology Server). Some of them were based on the SNOMED-CT[1] nomenclature or on ICD.

At an early stage of the development of the eHEALTH platform, the need was recognized for a state-of-the art, semantic interoperable, multilingual, multidisciplinary terminology system, both for information storage as well as for information retrieval purposes[2].

On behalf of the Belgian Department of Health, a contract was made between the Belgian Association for Informatics in Medicine and the Heymans Institute of Pharmacology, Ghent University, which had experience with the development of an European Multilingual Glossary for popular and technical medical terms, for patients and health care professionals in the realm of drug information[3].

1.3 Aim of the project

The request was to define the functionalities and structure of a reference terminology for health care professionals, as the key resource of a terminology server, maintained by a terminology center within the Belgian eHEALTH Platform.

1.4 Principles of the approach

1.4.1 Choice for a reference terminology for professionals

Health care professionals are routinely called to formalize thoughts into concepts and must interact with international classification systems in languages other than their native tongue. Their familiarity with technical terms is far greater than that of lay man. Therefore it is possible to build for professionals an onomasiological interface terminology [Rosenbloom et al., 2006], [Rosenbloom et al., 2008], based on concepts, with one reference term for each concept. This makes it easier to map to international classifications and to manage multilingualism. See discussion sections for more details.

For laymen (and to some extent also for professionals) the interface terminology should have a more lexically-oriented end-user terminology, of a semasiological nature, with possibly multiple senses of the same term, connected to much larger lexicons, and routed in natural language processing applications.

It is hardly feasible to combine all these functions in one resource.

So an option was made to focus on the structure of a reference terminology for professionals, as a mapping tool to international classifications.

In a later stage, a proposal for the structure of true end-user terminology for patients should be developed, but this is not the direct purpose of this project.

1.4.2 Choice for a standardized terminology

When building a national terminology system, it is wise to adhere to international ISO standards in terminology, especially when the ambition is to become the prototype for an international terminology

1 http://www.ihtsdo.org/snomed-ct/
2 http://www.health.belgium.be/eportal/Healthcare/Telematics/Studygroups/SEMINOP/index.htm?fodnlang=en
3 http://users.ugent.be/~rvdstich/eugloss/welcome.html

system. Moreover, to describe the information within the reference terminology, the use of internationally accepted Data Categories and conceptual domains (values accepted for a given Data Category) is to be preferred.

1.4.3 Choice for a multilingual terminology

For a Belgian initiative, the choice of a multilingual capable terminology is obvious, as the country has 3 official languages, and hence, a great sensitivity for the importance and the subtleties of language.

1.4.4 Choice for a double functionality : information storage and information retrieval

From the viewpoint of the technical user, it is not practical to conceive two different reference terminology systems, one for information storage (medical registration) and one for information retrieval (bibliographic searches, analysis of health care data). Therefore, the reference terminology needs to be functionally mapped to an array of international information resources: bibliographic thesauri such as MeSH (Medical Subject Headings) [Lipcomb, 2000], nomenclatures such as UMLS (Unified Medical Language System) [Lindberg et al., 1993], [Bodenreider, 2004] and SNOMED-CT (Systematized Nomenclature of Medicine – Clinical Terms), and to classification systems, such as ICPC and ICD.

1.4.5 Choice for a multidisciplinary terminology

The reference terminology must be useful for medical professionals and for allied health personnel (nurses, physiotherapists, etc.). It will for pragmatic purposes be focused on primary health care but must integrate services for the different medical specialized domains.

A multidisciplinary reference terminology must respect the use of different classification systems for different professions.

1.4.6 Choice for mapping to multiple international classifications

Some terms may relate to concepts in different classification systems (eg. obstipation can be a symptom, a diagnose, a side effect, a contra-indication).

No classification system will serve all purposes, and depending on a typology of concepts, mappings to one or several classification systems need to be made. For primary care a double connection to ICPC and ICD will be necessary (as indicated by precursor terminologies such as 3BT).

1.4.7 Choice for connecting to an ontology based nomenclature

As the reference terminology for professionals is built as an onomasiological, concept-driven system, mapping to a comprehensive, international nomenclature is necessary.

1.4.8 Choice for pragmatic restraint

Reference terminologies are at danger to extend indefinitely and to become impossible to manage.

Investments in concept selection, in multilingualism, and in quality of mapping demands that the size of the system remains at a reasonable level, as it is not the intention to replicate international attempts to comprehensiveness such as SNOMED.

Moreover, the human mind may not be able to handle a myriad of concepts in all domains. It will be more feasible to familiarize with a reference terminology, when its size is kept at a reasonable size.

A balance between general universal concepts and precise specialized concepts needs to be found.

The number of concepts should not exceed 20.000, and preferably be lower. In this sense, the reference terminology is a controlled vocabulary, meaning that the number of concepts within the system is limited and that one preferred term in each language is chosen for each concept.

1.4.9 Applying Semantic Web Technologies

Knowledge databases available on the web are quickly restructuring to new information technologies, which not only link universal locators to documents, but also to terms, concepts and data within documents. Semantic properties of and relations between data are captured in triplets (subject, predicate, object) and used to build innovative, associative applications.

In the field of health care, there has been an explosion of mapping between international classifications and nomenclatures, using the technologies of Web 2.0 and Web 3.0.

It is important for a reference terminology to be in the future integrated to the web of knowledge by publishing itself as a Linked Open Data[4] resource.

1.4.10 Bridging to lay language end-user terminologies

Semantic interoperability not only encompasses human-machine, multilingual, multidisciplinary interaction, but also interaction between the patient and the health care professionals.

Most patients will want access to their clinical information stored in records and understand the content, and search for scientific and evidence-based information, wherever it is available and trustworthy.

Hence, builders of a reference terminology for professionals must keep in mind that a bridge will have to be made between technical concepts and lay language. In the structure of the reference terminology this bridge to a lexical, semasiological end-user terminology for the patient must be anticipated.

4 http://linkeddata.org/

2 Methods

To build a standardized, multilingual, ontology-based resource, linkable to national and international classifications, a series of choices have been made, that are taken into account in the following methods.

2.1 Presentation of the TMF meta-model of the ISO norm 16642 ISO (Terminological Markup Framework)

The TMF Meta-model[ISO 16642, 2003], [Romary, 2006] , coupled with isoCat[ISO 12620, 1999] answer well many of the choices made for the belgian reference multilingual terminology server:

- Standardized: TMF and isoCat are both international standards.
- Multilingual: TMF has been conceived to structurally address multingualism.
- Mapped to International Classifications: the terminological entries are dealt with as concepts and the mapping to the International Classifications is done at this same level.
- Ontology-based: the entries of the reference terminology are concepts that can be unambiguously linked to an international nomenclature which is providing a ontological structure.

2.1.1 Starting with the meta-model

The reference standard for multilingual terminologies within the ISO-norm family is ISO norm 16642.

This standard basically provides a meta-model, which decomposes the organization of a terminological database into basic components as shown in figure 1. The model is in keeping with the traditional onomasiological, concept-oriented view of a terminological entry dating back to Wüster's early works [Picht & Schmitz, 2001]. It consists of 5 consecutive nodes :

- the terminological resource (Terminological Data Collection)
- the concept (Terminological Entry)
- the language chosen (Language Selection)
- the term(s) in the language chosen (Term Entry)
- components of the term (Term Component Section)

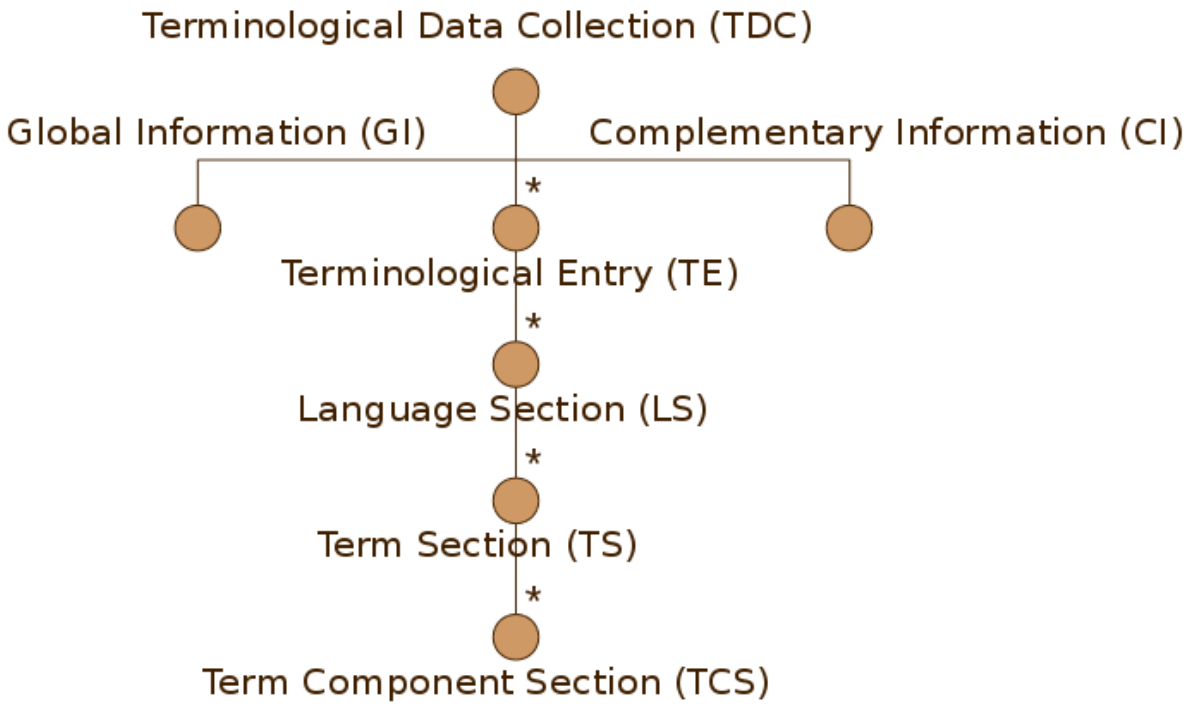

Illustration 1: TMF Christmas tree - Romary, L.

2.1.2 Choosing Data Categories to be hooked into the meta-model specification

This crude meta-model needs now to be further made more explicit, by defining a number of Data Categories (fields in a database or elements in an XML schema).

Within the ISOcat organization there is a rich collection of predefined Data Categories, pertaining to linguistic and terminological databases, and compliant with the ISO/IEC 11179 [ISO 11179-3] family of standards (http://isocat.org).

The aim of such standard collection of descriptors is to ”Provide uniform naming and semantic principles for such Data Categories so as to facilitate the interoperability or leveraging of language resources across applications and approaches”[Kemps-Snijders et al., 2009].

The challenge is to choose a limited set of Data Categories from this extensive collection, which are pertinent to the functionalities of the application to be constructed : i.e. a reference terminology for health care professions.

The approach in this project was to study the set of Data Categories chosen in two terminological applications, namely TBX[5] (a model for applications for translators) and TERMSciences[6] (an application of the French INIST)[Khayari et al., 2006], as first guidance for reduction of the set of Data Categories. In addition, we analyzed the methodology and database fields of a merger project in

5 http://www.lisa.org/standards/tbx (currently offline)
6 TermSciences - Portail Terminologique Multidisciplinaire : http://www.termsciences.fr/

Belgium, which had the aim to concatenate two existing terminologies (3BT and the UZ VUB terminology server) and link them to SNOMED-CT. The Data Categories potentially to be selected were discussed by a knowledge engineer (Joseph Roumier) and a physician/ information scientist (Robert Vander Stichele). Data Categories were judged upon their pertinence for the functionalities to the reference terminology under construction, and rejected or selected by consensus.

In a last step, the selected Data Categories were hooked to one or several pertinent nodes of the meta-model, and provided with standardized or ad hoc domain values. This step was also performed by the research team by consensus.

The result of this process is a Data Category Subset (DCS), specific to this application.

2.1.3 Building Content Mapping in a Terminological Markup Language and serialization into a TMF-conformant XML-representation

After choosing the TMF meta-model and building the Data Category Registry that ornates its nodes, one must create a language (for example in XML). This comprises the mappings between the Meta-model, the Data Categories and the vocabularies used to express them (e.g. as an XML element or a database field), and the domain values of the Data Categories. A generic mapping tool is necessary to assure interoperability with other instantiations of the TMF meta-model in terminological markup languages.

A search for existing software applications to act as Terminological Markup Language and Generic Mapping Tool was initiated, through the Internet and the French Institutions INIST[7] and Loria[8] with the editor of the ISO Norm 16642.

One example of a TML is GENETER[9], a markup language for terminology works, compliant with TMF, and publicly available as Geneter_V01.

We found within the toolkit of TermSciences a publicly available XML schema in RelaxNG[10] to express our own specific terminological Markup language, specifically geared to the project of the Belgian Reference Terminology. In Annex 6.1 - “RelaxNG Terminological Model”, an example is given of an adaptation of the RelaxNG schema, which could be used in this project.
The Generic Mapping Tool (GMT), is based on a set of XML elements and attributes, which serve as containers for:

- the nodes of the structural skeleton (identified by <struct> tags)
- the Data Categories (identified by <feat> tags).

This Generic Mapping Tool can also be used for exchange between the resource under development and other existing terminologies, such as Geneter (http://www.geneter.org/) and MSC Terminology[11] (see fig2).

7 http://www.inist.fr/
8 http://www.loria.fr/
9 http://www.geneter.org/
10 http://www.relaxng.org/
11 http://xml.coverpages.org/ISO16642-200207.pdf

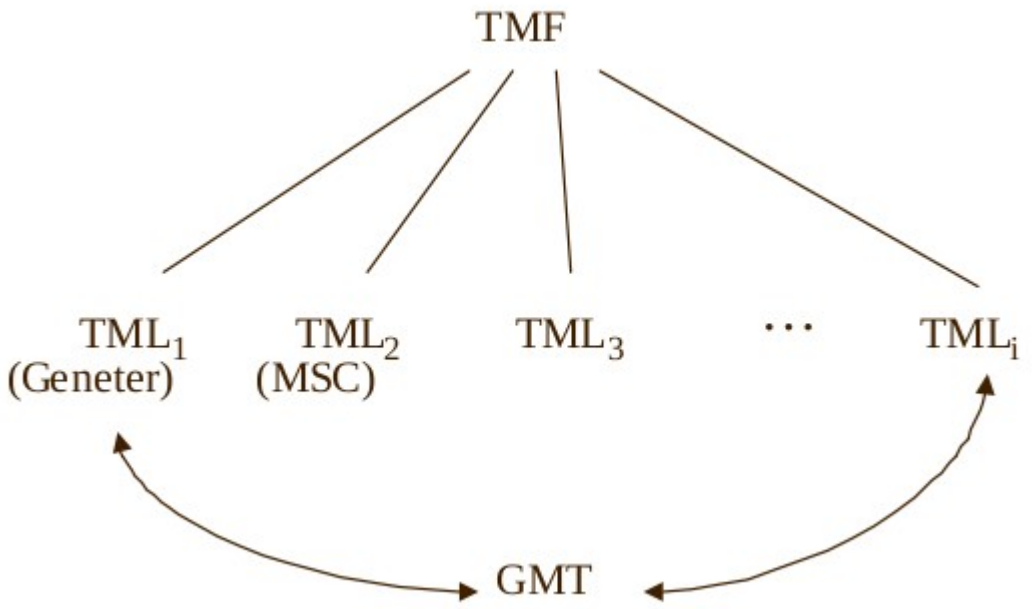

Illustration 2: GMT mechanism - Romary, L.

In addition of the study of the ISO norms and the available tools, an extensive literature search was performed, scouting for articles on existing mappings between terminologies and international classifications, and between international classifications [Bodenreider, 2008], [Cardillo et al., 2008], [Cardillo, 2011], [Zeng et al., 2006], [Zeng et al., 2007]. Also, the literature on end-user terminologies and reference terminologies was studied [Rosenbloom et al., 2006], [Cardillo et al., 2009a], [Cardillo et al., 2009b].

Based on these literature searches and on the experience of the Belgian Project for merging two existing terminologies (3BT and UZ VUB terminology Server), criteria were developed for the selection of concepts to be represented in the Belgian Reference Terminology.

In addition a preliminary procedure or work process was proposed for adding a new concept to the reference terminology (see annex 8.3 - Management of a Reference Terminology as a controlled vocabulary).

The two researchers in this project also set out to develop a typology of concepts for the reference terminology, by studying the types of concepts covered by existing classifications.

Finally, the research team has build a pragmatic list of potential users for the reference terminology, based on the different certified medical specialists in Belgium and on a list of certified allied medical personnel, as a substitute for defining sub domains for the reference terminology.

2.2 Methodology to publish the terminology as Linked Data

The Linked Data paradigm is able to cope with the following subset of choices taken in the approach section:

- Standardized: the Linked Data paradigm uses a technology stack that is standardized by the W3C.
- Information Storage and Retrieval: Linked Open Data allows the use of sophisticated queries over web-available resources for information retrieval.
- International Classifications: a growing number of these are available as Linked Open Data, the reference terminology would be a great addition to these.
- Ontology based: the Linked Open Data paradigm is an outcome of the Semantic Web initiative by the W3C[12] and uses the standards for the publishing of ontologies on the web.

The basic principle of Linked Data[13] [Bizer et al., 2010] is to link data from various sources through the semantic web technologies. It is possible to link sources simply based on hyperlinks. The Linked Data approach recommends that one at least provides the data as semantic web triplets, adding relations to the data.
A set of rules to publish data as linked data has been written by Tim Berners-Lee[14]:

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
4. Include links to other URIs, so that they can discover more things

For example, if entry 22298006 in SNOMED-CT deals with the same disease as the entry 10039912 in 3BT, a linked-data system can be set-up easily to exemplify the concept.

First, publish the data with URIs:

http://data.snomed-ct.org/identifier/2298006

and

http://data.3bt.org/id/10039912

Links to other URIs will be provided as useful information using semantic web triplets.

A triplet consists of three elements : a subject, a predicate and an object.

In this case, the subject would be a concept in one terminology, and the object a similar concept in another terminology, to which a map is to be created. The Predicate will then express the quality of that mapping, e.g. by expressing it as with the OWL (Web Ontology Language) [Dean et al., 2004]

12 http://www.w3.org/standards/semanticweb/
13 http://www.w3.org/standards/semanticweb/data
14 http://www.w3.org/People/Berners-Lee/

predicate #sameAS.

A link to show the sameness between the entries can be easily described using an triple:

Subject: http://data.snomed-ct.org/identifier/2298006
Predicate: http://www.w3.org/2002/07/owl#sameAs
Object: http://data.3bt.org/id/10039912

As a result, a software agent, interpreting the sameAs predicate as “there is a new source of information about the same thing” will be able to collect more information about Myocardial Infraction automatically. In this project, domain values will be created for this Data Category to express the quality of mapping between a given concept in the reference terminology with a similar concept in another international (or local) terminology system will be constructed, by questioning international experts in semantic web mapping between classifications.

2.3 Methodology for the Integration of a Reference Terminology in a wider terminology system

The integration of a Reference Terminology in a wider terminology system is based on the fulfillment of some of the choices made in principles of the approach section:

- Storage and Retrieval: The reference terminology makes it possible to have a really common terminology. For storage, the obvious link is SNOMED-CT for which it is possible to have a 1—1 relationship between the concepts in the reference terminology and the concepts of SNOMED-CT nomenclature at the desired level of abstraction. On the other hand, full link to bibliographic retrieval systems and well-known international classification systems becomes possible.
- Multidisciplinary: The reference terminology can serve as a bridge between terms and concepts of neighboring disciplines.
- Pragmatic Restraint: Being given the possibility to make use of external sources through unambiguous conceptual links to external international nomenclatures and classifications allows to choose to keep the size of the reference terminology between reasonable boundaries and avoid the need to create and maintain a new ontology.
- Lay Language: Linked Open Data offers mechanisms to link resources that have been developed independently, such as the end-user lexicons and terminologies that were created for specific purposes in all the health related institutions, while not impeding their autonomous development.

The research team discussed the possibility to link to nomenclatures, Classifications, and bibliographic thesauri.

The team explored solutions to link rather than integrate local nomenclatures to the reference terminologies, by associating data structures of current initiatives with the data models being developed for the terminology server.

Finally, a thorough analysis was made for the best way to bridge between the reference terminology and lay language terminological resources, such as the Multilingual Medical Glossary and the ICMV.

3 Results: Application to the Belgian Reference Terminology System

3.1 From TMF and the data-categories to a Terminological Markup Language

3.1.1 Deciding on whether and how to use the Terminological Model Framework (TMF) meta-model

Given the principle to work with international standards, the choice for an ISO – norm was obvious.

Because ISO-Norm 16642 ISO (Terminological Markup Format) specifically and formally deals with multilingualism in terminology, this specification was chosen. In addition, the application of this norm was checked for, by studying existing projects, such as Geneter, MSC terminology, and the French project TermSciences.

We considered but did not retain the use of the alternative SKOS (Simple Knowledge Ontology System) [Issac et al., 2009] (see discussion).

Within the TMF meta-model, the necessity to provide the 5th node in the model (TermCompomentSection) was discussed. This could be useful to split composite terms or "phrases" into its components. However, this might be less important when building a reference terminology, coupled to other classifications and nomenclatures such as SNOMED-CT (where similar functions are already elaborated). Hence, it was decided on the one hand to provide a minimal set of Data Categories for this node, but on the other hand, to postpone the filling of these Data Categories with data to a later moment.

3.1.2 Building the language

3.1.2.1 Building on the experience of ongoing terminological project in Belgium

Since the beginning of 2011, the Belgian authorities have started a project to merge existing medical terminological systems, such as 3BT and the UZ VUB Terminology Server. This was recommended by the expert panel on semantic interoperability (SEMINOP) (http://www.health.belgium.be/eportal/Healthcare/Telematics/Studygroups/SEMINOP/index.htm?fodnlang=en)

In this project, a selection of terms from both systems are matched to a SNOMED-CT concept. In the experience from this project, it was observed that when applying the rules for translation of US-English lexical representations of SNOMED-CT concepts, sometimes terms are obtained, which do not sound familiar to the health care professionals of the country. Another,more frequently used local term would be more suitable. Hence, a solution for this problem in the Data Categories is needed.

The structure of the Excel sheet used in this project for data entry by the validating terminologists was studied and to check whether all the fields and elements of this practical application were represented in the new set of Data Categories to be constructed. Fields specific to the task of validation were not included.

3.1.2.2 Choosing the right Data Categories and their association with the relevant TMF levels

First, we analyzed TBX (TermBase eXchange). This LISA (Library and Information Science Abstracts) standard, which was revised and republished as ISO 30042, allows for the interchange of terminology data including detailed lexical information. The framework for TBX is provided by three ISO standards: ISO 12620, ISO 12200 and ISO 16642. ISO 12620 provides an inventory of well-defined "Data Categories" with standardized names that function as data element types or as predefined values. ISO 12200 (also known as MARTIF) provides the basis for the core structure of TBX. ISO 16642 (also known as Terminological Markup Framework) includes a structural meta-model for Terminology Markup Languages in general.

In addition we analyzed the Data Categories selected for the French TermSciences project.

In these examples a variety of Data Categories exist to represent internal cross references between concepts within a given terminological system : broader/narrower/partitive. We decided not to use these elements, as the option was to deal with ontological relations within the system through the maps to external international terminological systems.

It was also decided to avoid elements dealing with the management of linguistic and semantic variants of lexical representations of concepts, such as homonyms, synonyms, abbreviated forms. These aspects will be dealt with through bridges with language-specific lexical systems See further justifications in the discussion.

3.1.2.2.1 Data Categories

The following figure summaries the main data-categories as well as the levels chosen.

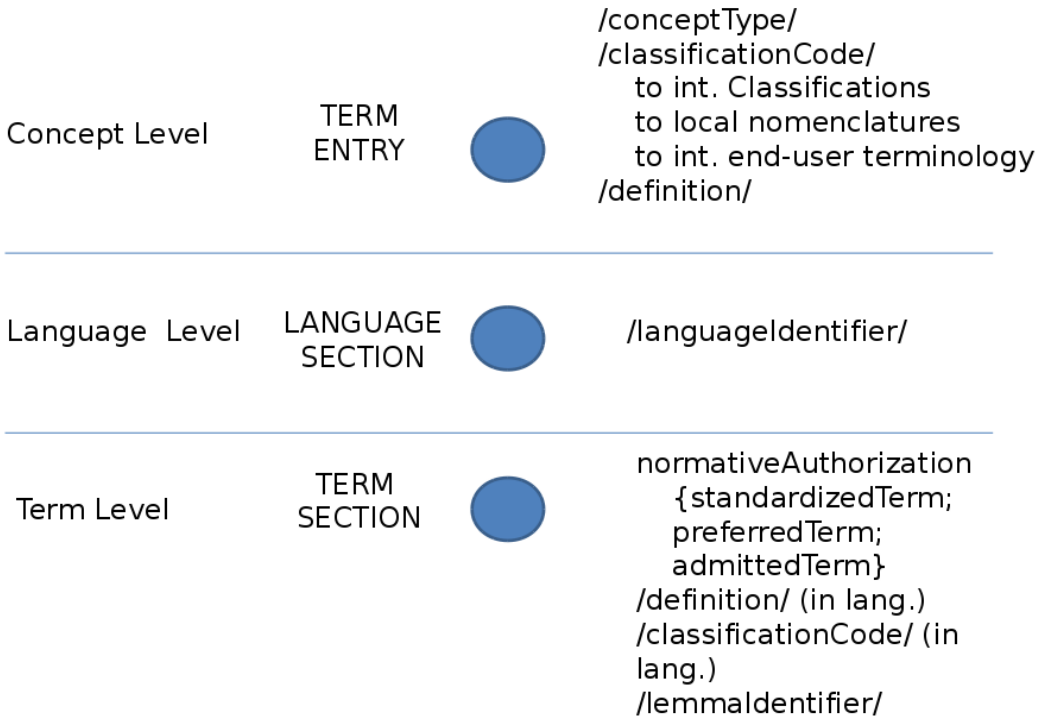

Illustration 3: TMF and Data Categories, essentials

In the following figure, all the Data Categories, sub-data-categories and levels are collected:

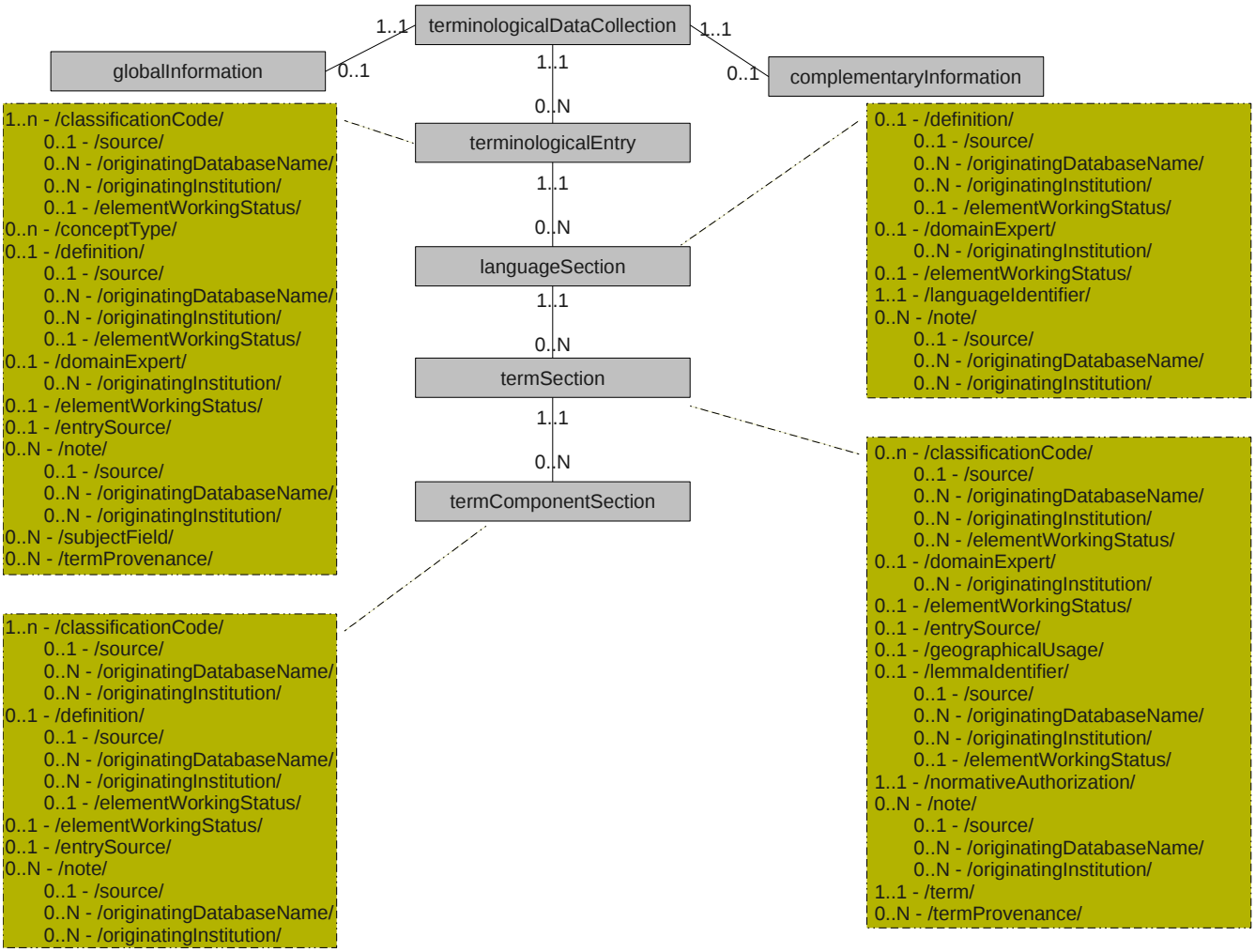

Illustration 4: TMF and Data Categories for TE, LS, TS, TCS

3.1.2.2.2 Description of the chosen Data Categories and the TMF levels

In this section, each chosen data-category is described, and the TMF levels chosen indicated.

Some of the definitions and explanations are taken from ISOcat.org.

/classificationCode/ TE|TS|TCS 1..n

Is for links to all external systems (ICPC, Snomed-ct, thesaurus, local nomenclatures), provided that they are specified.
They are specified with the (see later) /originatingInstitution/ and /originatingDatabaseName/ and grouped with a <brack>. At the level of terms section, this domain category is used for mappings to translations of international classifications..

/conceptType/ **TE** **0..n**

Each concept may pertain to one or more specific categories of concepts (see further for the categorization in annex 6.5 - Cross table : concept type / classification).

/definition/ **TE|LS|TCS** **0..1**

Definitions are relevant at the conceptual level, but also each language requires a translation of the definition.

/domain expert/ **TE|LS|TS** **0..1**

A domain expert is useful to write and control the definition of the concept at the TE level; at LS level to write and control a translation of it; and at the TS Level to control the target language equivalent.

/elementWorkingStatus/ **TE|LS|TS|TCS** **0..1**

Useful to indicate how far the element is in the work-flow of approval.

Conceptual domain: {pending, validated}

/entry source/ **TE|TS** **0..1**

A database or format from which data is imported.

/geographical usage/ **TS** **0..1**

Use to permit regional differences of the local preferred term usage. Example: differences in dutch between the Netherlands and Flanders or differences in French between Wallonia, Canada and France. It is only used in the Term Section for a /term/ having the /normativeAuthorisation/ of preferred term. The International ISO 3166-1 alpha-2 norm for countries is used.

/languageIdentifier/ **LS** **1..1**

The language of the Language Section, borrowed from the TermSciences.fr TML. The identifiers specified in ISO 639-1 should be used: en = English fr = French ru = Russian (russki) de = German (Deutsch) es = Spanish (Espanol) and so on [Gillam et al., 2007].

/lemmaIdentifier/ **TS** **0..1**

Use to indicate the link to a lexical entry in an external (for example popular) lexicon.

/normative authorization/ **TS** **1..1**

Indicating if the term is admitted (here for the term used in the terminology systems for lay language), deprecated, legal, preferred (here for locally preferred terms), standardized (here for the literal translation according to IHTSDO rules of lexical representation of SNOMED-CT concepts) or superseded.

Three of these domain values are used within this system with a specific purpose :

- standardizedTerm : for the literal translation according to IHTSDO rules of lexical representation of SNOMED-CT concepts
- preferredTerm : for the occasional national alternative terms. In addition, an extra rule is required to take into account cases in which two terms are identical. Then one of two has to be changed.
- admittedTerm : for the popular term used in the terminology systems for lay language.

/note/ **all** **0..n**

Some freedom for the terminologists.

/originatingDatabaseName **all** **0..n**

Used associated with other Data Categories, borrowed from TermSciences.fr TML.

/originating institution/ **all** **0..n**

Used associated with other Data Categories.

/source/ **all** **0..1**

Source is only used associated with originatingDatabase and originatingInstitution and provides a complete citation of the bibliographic information pertaining to a document or other resource..

/subject field/ **TDC|TE** **0..n**

A field of special knowledge. Here are indicated for which field the TDC and the TE are relevant.

/term/ **TS** **1..1**

A verbal designation of a general concept in a specific subject field.

/term provenance/ **TE|TS** **0..n**

Explanation on the methodology used to select the concept or the locally preferred term. Example for concept: “proposed from the Italian vocabulary, coming from the nurse study”. Example for preferred term: “a choice was made between two preferred terms, different from the standardized term based on frequency in a phrase enabled term extraction study”.

3.1.2.3 Serialization into a TMF-conformant XML-representation

3.1.2.3.1 Creating an XML schema for the terminological model language

RelaxNG is a schema language for XML, comparable with XML Schema and DTD.

The created RelaxNG schema for the TML contains the information about the Data Categories as well as the levels of TMF. As result, it can be used to check that a given Terminological Data Collection is compatible with the TML.

Excerpts from the RelaxNG schema for the reference terminology TML, created based on TermSciences schema and GMT (the full schema is in annex 6.1 - RelaxNG Terminological Model):

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- Reference Terminology RelaxNG v0.9 - 20110912 -->
<!-- adapted from RelaxNG niveau DataCollection TermSciences V2 - le 11/05/2007 -->
<grammar xmlns="http://relaxng.org/ns/structure/1.0">
    <start>
        <ref name="termDatCol_element"/>
    </start>
    <define name="attLang">
        <optional>
            <attribute name="xml:lang">
             <choice>
                 <value>nl</value>
                 <value>fr</value>
                 <value>en</value>
                 <value>de</value>
                 <value>it</value>
                    </choice>
            </attribute>
        </optional>
    </define>
[…]
       <define name="termSec_element_DCures">
    <element name="feat">
      <choice>
        <ref name="classificationCode_DC"/>
        <ref name="term_DC"/>
        <ref name="subjectField_DC"/>
```

3.1.2.3.2 XML example

Simple example based on the excel sheet. Full example in annex 6.1 – XML example.

```
<?xml version="1.0" encoding="UTF-8"?>
<struct type="terminologicalDataCollection">
[...]
 <struct type="terminologicalEntry" xml:id="TE.5465">
```

Taking into account several classifications:

```
  <brack>
    <feat type="classificationCode">22298006</feat>
    <feat type="originatingInstitution">IHTSDO</feat>
    <feat type="originatingDatabaseName">SNOMED-CT</feat>
  </brack>
  <brack>
    <feat type="classificationCode">I21.9</feat>
    <feat type="originatingInstitution">WHO</feat>
    <feat type="originatingDatabaseName">ICD10</feat>
  </brack>

[...]
```

Definition and source of the definition:

```
    <feat type="source">Dorland, 27th ed</feat>
    <feat type="definition" xml:id="TE.5465.DE.1" xml:lang="en">Gross necrosis of the myocardium, as a result of interruption of the blood supply to the area.</feat>
  </brack>
```

Indicating the expert responsible for the entry:

```
  <brack>
    <feat type="domainExpert">Michael Card</feat>
    <feat type="originatingInstitution">University Hospital</feat>
  </brack>
```

Indicating the concept Type of the entry:

```
<feat type=”conceptType”>diseases</feat>
```

Indicating the status of the entry:

```
<feat type="elementWorkingStatus">pending</feat>
```

Indicating the fields from which the entry is taken:

```
<feat type="subjectField">cardiology</feat>
<feat type=”subjectField”>general practitioner</feat>
```

And the way the entry was identified:

```
<feat type="provenance">existing classification</feat>
```

Then the different languages:

```
<struct type="languageSection">
  <feat type="languageIdentifier">en</feat>
```

Nested Term Sections→:

```
  <struct type="termSection" xml:id="TE.5465.TS.1">
    <feat type="term">myocardial infarction</feat>
    <feat type="normativeAuthorization">standardizedTerm</feat>
  </struct>
</struct>
```

The language Sections re-occur for each language of the terminology:

```
<struct type="languageSection">
  <feat type="languageIdentifier">nl</feat>
  <brack>
    <feat type="definition" xml:id="TE.5465.DE.2" xml:lang="nl">Een hartinfarct, is het afsterven van een deel van de hartspier door onderbreking van de bloedtoevoer ervan door de kransslagaderen.</feat>
    <feat type="source">http://nl.wikipedia.org/wiki/Hartinfarct</feat>
  </brack>
```

Nested Term Section(s):

```
    <struct type="termSection" xml:id="TE.5465.TS.3">
```

Direct translation of the SNOMED-CT term is mandatory but is not necessarily the preferred term. A local version of this term is preferable.

```
      <feat type="term">myocardinfarct</feat>
    </struct>
  </struct>
[…]
```

3.2 Publish the terminology as Linked Data and use it

The basic principle of Linked data is to link data from various sources through the semantic web technologies. It is possible to link sources based on hyperlinks. The Linked Data approach recommends that one at least provides the data as semantic web triplets, adding relations to the data.

The translation of the set of rules is for the first – and no complete - iteration of the model the following:

First, a URI for the Linked Data server: http://referenceterminology.be/

http://referenceterminology.be/TE.5465

On the side of the Data Categories the task of defining the URIs has already been done:

/subjectField/ → http://www.isocat.org/datcat/DC-489

Then the triplet consists of:

subject:	http://referenceterminology.be/TE.5465/
predicate:	http://www.isocat.org/datcat/DC-489
object:	cardiologist

In the case of linking to SNOMED-CT, it is possible to used the predefined OWL[15] predicate “owl:sameAs” as we know that en entry in the reference terminology is necessarily taken from SNOMED-CT.

As a result (using here a fictitious URI for snomed-ct):

subject:	http://referenceterminology.be/TE.5465/
predicate:	http://www.w3.org/2002/07/owl#sameAs

15 OWL: Web Ontology Language - http://www.w3.org/TR/2009/REC-owl2-primer-20091027/

object: http://fictiouslinkto-snomed-ct/22298006

In the case of linking to other external source like ICPC-2 or ICD-9, and even though it will require more work from the experts, we identified the necessity to qualify the kind of link to classifications by using two other qualifiers:

- http://referenceterminology.be/predicate/narrowerConcept → if the object from ICPC-2 is a narrower concept than the one from SNOMED-CT.
- http://referenceterminology.be/predicate/broaderConcept → if the object from ICPC-2 is a broader concept than the one from SNOME-CT.

In addition to that,it is worth mentioning that the existence of “bags” for unclassified concepts in ICPC-2 require extra care as by nature the classification of the entries they contain is not as clear as for the others.

3.3 Integration of a Reference Terminology into a wider terminology system

The research team discussed the possibility to link to nomenclatures (SNOMED-CT, UMLS), Classifications (ICD, ICPC, ATC, NANDA, LOINC, etc), and to bibliographic thesauri (MeSH).

For information storage purposes the obvious choice is SNOMET-CT. The consequence is that for each concept in the reference terminology an exact map to a concept in SNOMED-CT is to be found. It will therefore also not be necessary to develop an ontology within the reference terminology.

A cross table has been produced between the list of types of concepts (see annex 6.5 - cross table : concept type / classification) and the classifications to be linked to. In addition to these, many linked data sources involved in the eHealth domain are good candidates to link with and to be linked to.

For example, Diseasome[16] describes diseases and their relationships with genes. The Linked Open Data server publishing it[17] already links it with several other linked data sources dealing with drug description, side effects, genes.

The software offering to publish linked data on the web is now quite large, many as free and open-source software. Among these, Jena[18], Sesame[19], Virtuoso[20].

The result of the analysis is that it is possible to link efficiently the entries of the reference terminology to external and local nomenclatures, glossaries, lexicons. As a result the benefit/cost ratio of linking instead integrating data is interesting. This applies for example to the INAMI medical acts, the communes codes, and also the lay language terminological resources such as the Multilingual Medical glossary and ICMV.

For these two resources oriented towards patients, the general problem of linking onomasiological and semasiological resources was discussed, as well as the potential usefulness of SKOS and the ISO norm Lexical Markup Framework for lay language terminological resources. The result of the discussion is the possibility to use at the level of the Term Section the /lemmaIdentifier/ Data Category to link the reference terminology to the lexicon, without breaching the terminological principles.

16 Diseasome: http://www.nd.edu/~alb/Publication06/145-HumanDisease_PNAS-14My07-Proc/Suppl/
17 D2R Berlin University Linked Data Server: http://www4.wiwiss.fu-berlin.de/diseasome/
18 Jena: http://jena.sourceforge.net/
19 Sesame: http://jena.sourceforge.net/
20 Virtuoso: http://virtuoso.openlinksw.com/

4 Discussion

In this section, we discuss several points that were particularly salients during the preparation of the report.

4.1 First Dilemma: Onomasiological Versus Semasiological approaches to Multilingual Reference Terminologies

Linking to external concept-based (onomasiological) resources (preserve unambiguous relationships). The mapping to international classification is easier with an onomasiological approach because the mapping is made at the level of the concepts. As a result, the different lexical representations of the concept, in the different languages inherit the mappings to the classifications. There was the opportunity to see the reference terminology as a subset of a very elaborate concept-oriented resource (SNOMED-CT) inheriting its sophisticated ontological structure and the facility to achieve international semantic interoperability.

Multilingualism in itself is easier to manage with the conceptual approach.

- The meaning of each lexical representation is inherited from the definition which is found at conceptual level. This is valid both within a language and between languages. Synonyms are found for each language, below the concept. Preferred terms are directly and easily found for a given concept in an onomasiological resource, whereas in a lexical, semasiological resource, it requires information both about the status of the lexical entry – here “preferred term” - and the intended “sense” pertaining to that lexical entry.
- Homonyms exist within a language and between languages (“car” in english and french and within french). They will appear on the final nodes of the structure with a conceptual approach, whereas they would appear on top of a lexical system structure.

The alternative approach, lexically-oriented makes it easier to deal with homonyms, to perform natural language processing and integrate the diversity of the languages. All these desirable features can be dealt with creating a bridge between the multilingual reference terminology and possibly uni-lingual lexical resources.

4.2 Second Dilemma: SKOS as an alternative to TMF

SKOS[21] has been developed during the same period as ISO standard TMF, in the frame of the Semantic Web Initiative of the W3C. The goal is to use existing semantic standards – RDFS and OWL - to easily publish thesaurus, controlled vocabularies, taxonomies, lexicons, glossaries on the web. Hence the name “Simple Knowledge Organization System”.

SKOS is successfully used for the CiSMeF[22] project by the CHU de Rouen to integrate multiple health-related terminologies[Grosjean, 2011]. In particular SKOS files are used to generate all the user interface on the side of the web client.

SKOS has not been developed for terminologies and finds its origins in standards for monolingual and multilingual thesauri [ISO2788, 1986], [ISO5964, 1985]. Even though the editors of SKOS specify that it should not be seen as a limitation of the scope of the standard, not specifically geared for linguistic

21 http://www.w3.org/standards/techs/skos

22 CiSMef: http://pts.chu-rouen.fr/

terminologies .

- SKOS is following a conceptual approach, similar but not fully aligned with TMF, which is the international ISO standard for computer applications in terminology.
- The relationships and labels found in SKOS are not aligned with the international ISO standard for Data Categories (ISO 12620). SKOS only provides a loose and small subset of them.
- Multilingualism in SKOS is only dealt with using a set of labels and language codes (example taken from SKOS Reference - http://www.w3.org/TR/2009/REC-skos-reference-20090818/):

```
MyClass> rdf:type owl:Class ;
      skos:prefLabel "animals"@en ;
      skos:altLabel "fauna"@en ;
      skos:hiddenLabel "aminals"@en ;
      skos:prefLabel "animaux"@fr ;
      skos:altLabel "faune"@fr .
```

- - SKOS does not differentiate the linguistic information from the ontological information, making it not convenient for linking to external ontologies and linguistic resources [Cimiano et. al., 2010].

As a result, even though SKOS provides the desirable feature of being already integrated to the Semantic Web stack, its downsides, mainly in terms of terminological features, do not make it the chosen framework to realize the reference terminology.

4.3 Third Dilemma: Relationship with the end-user terminologies

There are vocabulary differences between Specialists and GP. There is even more differences between the patients and the doctors. The vocabulary being used evolves over time and space and many local expressions exist to designate the same diseases, body parts. In order to cope with this heterogeneity in a (semi)automated manner, in fact to perform the task of the doctor being the interface between the medical wor(l)d and the lay person wor(l)ds, natural language processing is necessary. Even though the Data Categories found in ISOcat.org for TMF terminologies exist, the effort to maintain an evolving multilingual terminology containing all the linguistic intricacies and the local lexical variants for a concept would be daunting.

That is why a specific, lexical and often uni-lingual approach, could be more appropriate for end-user terminologies. In order to deal with multilingualism a link between the reference terminology and these external resources must be assured.We provide a dual mechanism to do so:

- The concept in the reference terminology is linked to the sense part of a lexical resource. This mechanism preserves the conceptual integrity.
- The alignment of the lexical representation of the concept in a specific language with the corresponding lemma in a lexicon of that language.

Our proposal is to use the ISO 24613 LMF Meta-Model [ISO 24613, 2008], [Romary, 2010], for that purpose because it is conceived to deal with linguistic intricacies, uses the same set of linguistic isoCAT Data Categories. LMF provides guidelines on how to link its entries with TMF and other concept based representation systems. Finally the meta-model contains a mechanism to deal with

multiple senses.

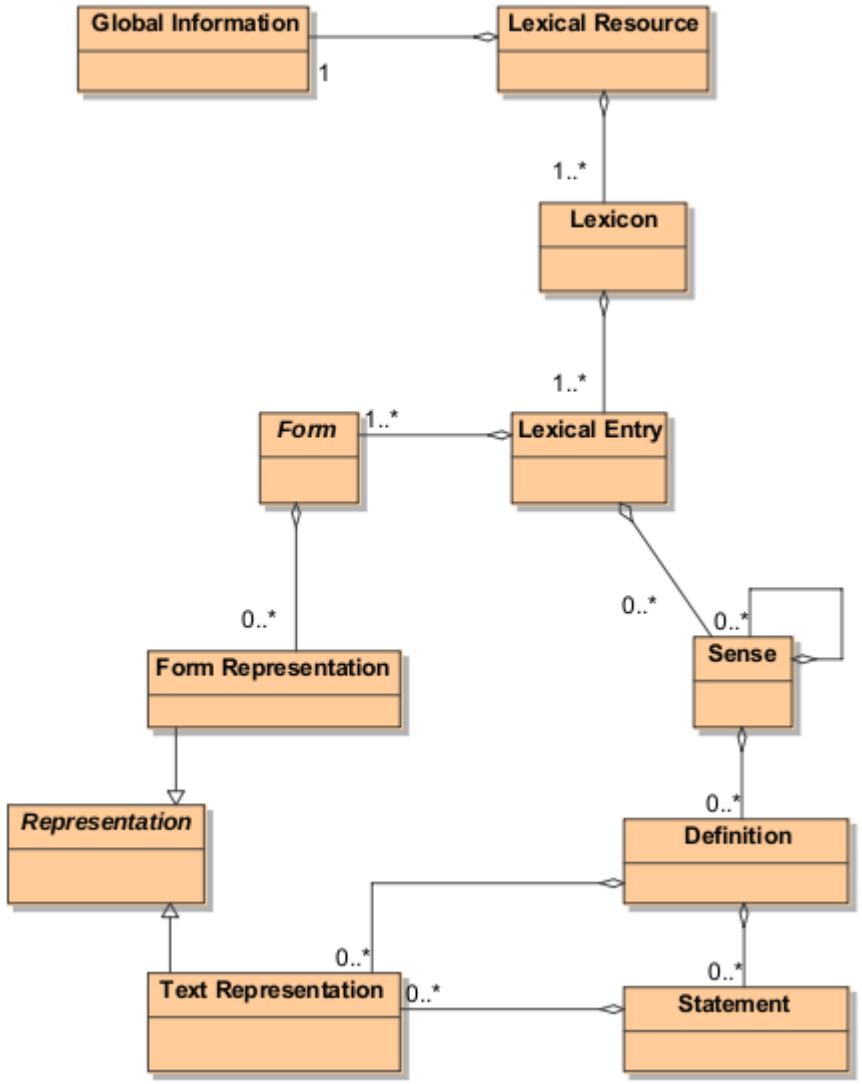

Illustration 5: LMF Core Package, ISO TC 37/SC 4 N453

4.4 Additional considerations for choice

4.4.1 Required investment in software developments

No extended search for software providers was made but current offer is small and the company that developed MyTerm is currently not reachable. The new company by the same owners is using the same base, but the software is only used as a back-end. On the other hand, the ATILF implementation for the Termsciences.org project is running and integrating a lot of sources into its core, as well as providing web services to exploit the data through different ways. Finally, being built on XML, there is plenty of

free Open Source software that can serve as a back-end to the server.

4.4.2 Preserve room for evolution & anticipate on the growth and specificness of specialized linguistic resources

We assumed that creating a central terminology that would contain all the information at once is a bad design choice. The proposed design for the terminology is simple and preserves room for evolution while still allowing linking to external, more precise and exploitation-oriented sources. The linguistic resources will evolve their independent way, and attempts to integrate everything in one central resource will fail at some point.

4.4.3 Opportunity to create a pioneer resource, strengthened by scrutiny of external users

If the resource is quickly set up and published in an open manner – especially in linked open data, it will gain adopters early by being one of the first to do so. As a result the quality of the resource will be improved by scrutiny of external users and the reference terminology server will benefit from it.

4.4.4 Taking into account the Terminological authoring work flow

The terminological authoring work-flow requires a series of step in the adoption of a term, that is for the moment left to proper authoring tools. Though, Data Categories exist that could be evaluated to fulfill this need. /elementWorkingStatus/ is one of these and is already included in the current TML.

5 Conclusion

In this methodological report for the Belgian eHEALTH platform, a proposal is made for the structure of a national reference terminology for health care professionals, integrated with several international classifications and the SNOMED nomenclature, as well as with end-user terminologies for patients.
In this proposal, approaches are described to harness the size and assure the maintainability of such a reference terminology, by assuring external mapping and linking to autonomously evolving nomenclatures, classifications, thesauri, lexicons and glossaries. terminologies, in a multilingual and multidisciplinary approach.
Publication of this resource as Linked Data will assure connectivity of the reference terminology with the rapidly evolving collection of resources in the semantic web.
Results are presented as a data model, compliant with the Terminological Markup Framework Meta-model (TMF); a Data Category Registry, compliant with ISOCAT; and a Terminological Markup Language in XML, based on the relevant ISO standards.
These building blocks can be used for the actual construction of a state-of-the-art reference terminology for Belgium, serving multilingual, multidisciplinary communication between health care professionals, using semantically interoperable information systems.

6 Bibliography

[3BT, 2006] 3BT Belgian Bilingual Biclassified Terminology. Gent & ULB universities 2006.

[Anonymous, 2010] Anonymous. EU eHealth Interoperability Roadmap. Calliope Network. December 2010.

[Bodenreider, 2004] Bodenreider, O., The unified medical language system (UMLS): integrating biomedical terminology, Nucleic acids research, S.-267, 2004

[Bodenreider, 2008] Bodenreider, O., Biomedical ontologies in action: role in knowledge management, data integration and decision support, Yearb Med Inform, S.79, 2008

[Bizer et al., 2010] Bizer, C. and Heath, T. and Berners-Lee, T., Linked data-the story so far, sbc, S.9, 2010

[Cardillo et al., 2008] Cardillo E, Eccher C, Serafini L, Tamilin A. Logical analysis of mappings between medical classification systems. AIMSA 2008: 311-321

[Cardillo et al., 2009a] Cardillo E. A Lexical-Ontological Resource for Consumer Heathcare. International Semantic Web Conference 2009: 949-956

[Cardillo et al., 2009b] Cardillo E, Serafini L, Tamilin A. A Hybrid Methodology for Consumer-Oriented Healthcare Knowledge Acquisition. K4CARE 2009: 38-49

[Cardillo, 2011] Cardillo E. A lexi-ontological resource for consumer healtcare. The Italian Consumer Medical Vocabulary. [Doctoral Thesis]. Fondatione Bruno Kessler. April 2011.

[Cimiano et. al., 2010] Cimiano, P. and Buitelaar, P. and McCrae, J. and Sintek, M., LexInfo: A declarative model for the lexicon-ontology interface, Web Semantics: Science, Services and Agents on the World Wide Web, 2010

[Couto et al., 2001] Couto FM, Silva MJ. Disjunctive shared information between ontology concepts: application to Gene Ontology. J Biomed Semant 2001;2:5

[Chute et al., 1999] Chute CG, Elkin PL, Sherertz DD, Tuttle MS. Desiterata for a Clinical Terminology Server. Proc AMIA Symp. 1999: 42–46.

[Dean et al., 2004] Dean M., Schreiber G., Bechhofer S., Van Harmelen F., Hendler J., Horrocks I., McGuinness D. L., Patel-Schneider P. F., Stein L. A., OWL web ontology language reference. w3c recommendation , (http://www.w3.org/tr/owl-ref/), 10 février 2004.

[Gillam et al., 2007] Gillam, L. and Garside, D. and Cox, C., Developments in Language Codes standards, Data Structures for Linguistic Resources and Applications. Proc. of GLDV, S.11-13, 2007

[Grosjean, 2011] Grosjean, J; Merabti, T; Dahamna, B; Kergourlay, I; Thirion B; Soualmia LF & Darmoni, SJ. Health Multi-Terminology Portal: a semantics added-value for patient safety. Patient Safety Informatics - Adverse Drug Events, Human Factors and IT Tools for Patient Medication Safety , Studies in Health Technology and Informatics, Volume 166, Pages 129-138, 2011. PSIP.

[Horrocks, 2007] Horrocks, I., Semantic web: the story so far, Proceedings of the 2007 international cross-disciplinary workshop on Web accessibility (W4A), S.120-125, 2007

[ICPC-2-R, 2005] ICPC-2-R - Wonca (World Organisation of Family Doctors) ICPC-2-R, International Classification of Primary Care (revised 2nd Ed). OUP. 2005

[ICD-10, 2010] ICD-10 -International Statistical Classification of Diseases and Related Health Problems. 10th Revision. Version for 2007.Tabular List of inclusions and four-character subcategories- http://apps.who.int/classifications/apps/icd/icd10online/

[Isaac et al., 2009] Isaac, A. and Summers, E., Skos simple knowledge organization system primer, Displays, 2009

[ISO 16642, 2003] ISO 16642, Computer applications in terminology - Terminological markup framework (TMF), 2003.

[ISO 12620, 1999] ISO 12620:1999, Computer applications in terminology – Data categories, 1999.

[ISO 24613, 2008] ISO 24613:2008, Language resource management — Lexical markup framework (LMF)

[ISO 11179-3] ISO/IEC 11179-3, Information Technology -- Metadata registries (MDR) – Part 3: Registry metamodel and basic attributes

[ISO5964, 1985] ISO 5964:1985 Documentation - Guidelines for the establishment and development of multilingual thesauri. First edition. ISO TC 46/SC 9, 1985.

[ISO2788, 1986] ISO 2788:1986 Documentation - Guidelines for the establishment and development of monolingual thesauri. Second edition. ISO TC 46/SC 9, 1986.

[Jamoulle and Roland, 1993] « LOCAS-CISP, logiciel de codage et d'acquisition de synonymes pour la Classification Internationale des Soins Primaires » (Jamoulle M, Roland M Fédération des Maisons Médicales, Bruxelles), 1993.

[Roland et al., 2000] Roland, M. and Jamoulle, M. and Dendeau, B. LOCAS 2, logiciel de codage et d'acquisition de synonyme basé sur la CISP version 2. Care Editions asbl, 2000.(available on http://trix.docpatient.net)

[Khayari et al., 2006] Khayari, Majid and Schneider, Stéphane and Kramer, Isabelle and Romary, Laurent, Unification of multi-lingual scientific terminological resources using the ISO 16642 standard. The TermSciences initiative., 2006

[Kemps-Snijders et al., 2009] Kemps-Snijders, M. and Windhouwer, M. and Wittenburg, P. and Wright, S. E., ISOcat: remodelling metadata for language resources, International Journal of Metadata, Semantics and Ontologies, S.261-276, 2009

[Lindberg et al., 1993] Lindberg, D.A. and Humphreys, B.L. and McCray, A.T., The unified medical language system, Methods of information in medicine, 1993

[Lipcomb, 2000] Lipscomb, C. E., Medical subject headings (MeSH), Bulletin of the Medical Library Association, S.265, 2000

[Okkes et al., 1999] Proc AMIA Symp. 1999: 42–46. Okkes IM, Jamoulle M, Lambert H, Bentzen N. ICPC-2-E: the electronic version of ICPC-2. Differences from the printed version and the consequences. Fam Pract 2000;17:101-107.

[Okkes et al., 2002] Okkes IM, Becker HW, Bernstein RM, Lamberts H. The March 2002 update of the electronic version of the ICPC-2. A step forward to the use of ICD10 as a nomenclature

and a terminology for ICPC-2. Fam Practice 2002;19:543-317.

[Picth et al., 2001] Picht H. and K.-D. Schmitz (Eds.), 2001. Terminologie und Wissensordnung, Ausgewählte Schriften aus dem Gesamtwerk von Eugen Wüster, TermNet Publisher.

[Romary, 2006] Romary, Laurent and Kramer, Isabelle and Alt, Susanne and Roumier, Joseph, Gestion de données terminologiques : principes, modèles, méthodes, S.13, 2006

[Romary, 2010] Romary, Laurent, Standardization of the formal representation of lexical information for NLP, 2010

[Rosenbloom et al., 2006] Rosenbloom ST; Miller RA, Johnson KB, Elkin PI, Brown SH. Interface terminologies: Facilitating direct entry of clinical data into electronic health record systems. J Am Med Inform Assoc 2006;13:277-288.

[Rosenbloom et al., 2008] Rosenbloom ST, Miller RA, Johnson KB, Elkin PI, Brown SH. A model for evaluating interface terminologies. J am Med Inform Assoc 2008;15:65-76.

[Tuderache et al., 2010] Tuderache T, Faloconer S, Nyulas C, Joy NF, Musen MA. Will semantic web technologies work for the Development of ICD-11 ? International Semantic Web Conference (2) 2010: 257-272.

[Zeng et al., 2006] Zeng QT, Tse T. Exploring and developing Consumer Health Vocabularies. J Am Med Inf Assoc 2006;13:24)29.

[Zeng et al., 2007] Zeng QT, Tse T. Divita G, Keselman A, Crowell J, Browne AC, Goryachev S, Ngo LM. Term Identification Methods for Consumer Health Vocabulary Development. J Medical Internet Research 2007;9:e4.

7 ANNEXES

7.1 RelaxNG XML schema for theTerminological Modeling Language

Corresponds to file “refterm-v1.1.rng”

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- Reference Terminology RelaxNG v1.1 - 20111003 -->
<!-- adapted from RelaxNG niveau DataCollection TermSciences V2 - le 11/05/2007 -->
<grammar xmlns="http://relaxng.org/ns/structure/1.0">
  <start>
    <ref name="termDatCol_element"/>
  </start>
  <define name="attLang">
    <optional>
      <attribute name="xml:lang">
        <choice>
          <value>nl</value>
          <value>fr</value>
          <value>en</value>
          <value>de</value>
          <value>it</value>
        </choice>
      </attribute>
    </optional>
  </define>
  <define name="AttId">
    <optional>
      <attribute name="xml:id"/>
    </optional>
  </define>
  <define name="AttTarget">
    <optional>
      <attribute name="target"/>
    </optional>
  </define>
  <!-- Data Categories definition -->
  <define name="classificationCode_DC">
    <attribute name="type">
      <value>classificationCode</value>
```

```
  </attribute>
  <text/>
</define>
<define name="conceptType_DC">
  <attribute name="type">
    <value>conceptType</value>
  </attribute>
  <text/>
</define>
<define name="definition_DC">
  <attribute name="type">
    <value>definition</value>
  </attribute>
  <text/>
</define>
<define name="note_DC">
  <attribute name="type">
    <value>note</value>
  </attribute>
  <text/>
</define>
<define name="originatingDatabaseName_DC">
  <attribute name="type">
    <value>originatingDatabaseName</value>
  </attribute>
  <text/>
</define>
<define name="originatingInstitution_DC">
  <attribute name="type">
    <value>originatingInstitution</value>
  </attribute>
  <text/>
</define>
<define name="elementWorkingStatus_DC">
  <attribute name="type">
    <value>elementWorkingStatus</value>
  </attribute>
  <text/>
</define>
<define name="entrySource_DC">
```

```
  <attribute name="type">
    <value>entrySource</value>
  </attribute>
  <text/>
</define>
<define name="geographicalUsage_DC">
  <attribute name="type">
    <value>geographicalUsage</value>
  </attribute>
  <text/>
</define>
 <define name="lemmaIdentifier_DC">
  <attribute name="type">
    <value>lemmaIdentifier</value>
  </attribute>
  <text/>
</define>
<define name="normativeAuthorization_DC">
  <attribute name="type">
    <value>normativeAuthorization</value>
  </attribute>
  <choice>
    <value>admittedTerm</value>
    <value>deprecatedTerm</value>
    <value>legalTerm</value>
    <value>preferredTerm</value>
    <value>standardizedTerm</value>
    <value>supersededTerm</value>
  </choice>
</define>
<define name="source_DC">
  <attribute name="type">
    <value>source</value>
  </attribute>
  <text/>
</define>
<define name="subjectField_DC">
  <attribute name="type">
    <value>subjectField</value>
  </attribute>
```

```
  <text/>
</define>
<define name="domainExpert_DC">
  <attribute name="type">
    <value>domainExpert</value>
  </attribute>
  <text/>
</define>
<define name="term_DC">
  <attribute name="type">
    <value>term</value>
  </attribute>
  <text/>
</define>
<define name="termProvenance_DC">
  <attribute name="type">
    <value>termProvenance</value>
  </attribute>
  <text/>
</define>
<define name="languageIdentifier_DC">
  <attribute name="type">
    <value>languageIdentifier</value>
  </attribute>
</define>
<define name="termDatCol_DC">
  <attribute name="type">
    <value>terminologicalDataCollection</value>
  </attribute>
</define>
<!-- end of DC -->
<!-- Structure -->
<define name="terminologicalEntry_DC">
  <attribute name="type">
    <value>terminologicalEntry</value>
  </attribute>
</define>
<define name="languageSection_DC">
  <attribute name="type">
    <value>languageSection</value>
```

```
  </attribute>
</define>
<define name="termSection_DC">
  <attribute name="type">
    <value>termSection</value>
  </attribute>
</define>
<define name="comInf_DC">
  <attribute name="type">
    <value>complementaryInformation</value>
  </attribute>
</define>
<define name="gloInf_DC">
  <attribute name="type">
    <value>globalInformation</value>
  </attribute>
</define>
<define name="termComponentSection_DC">
  <attribute name="type">
    <value>termComponentSection</value>
  </attribute>
</define>
<!-- end of Levels -->
<!-- Term Data Collection Level -->
<define name="termDatCol_element">
  <element name="struct">
    <ref name="termDatCol_DC"/>
    <zeroOrMore>
      <ref name="termDatCol_element_struct"/>
    </zeroOrMore>
    <ref name="termDatCol_element_DCures"/>
  </element>
</define>
<define name="termDatCol_element_DCures">
  <element name="feat">
    <choice>
      <ref name="note_DC"/>
      <ref name="originatingDatabaseName_DC"/>
      <ref name="originatingInstitution_DC"/>
      <ref name="source_DC"/>
```

```
      <ref name="subjectField_DC"/>
    </choice>
  </element>
</define>
<define name="termDatCol_element_struct">
  <choice>
    <ref name="gloInf_element"/>
    <ref name="comInf_element"/>
    <ref name="termEnt_element"/>
  </choice>
</define>
<!-- Global Information Level -->
<define name="gloInf_element">
  <element name="struct">
    <ref name="gloInf_DC"/>
    <ref name="gloInf_element_DCures"/>
  </element>
</define>
<define name="gloInf_element_DCures">
  <element name="feat">
    <choice>
      <ref name="note_DC"/>
      <ref name="originatingDatabaseName_DC"/>
      <ref name="originatingInstitution_DC"/>
      <ref name="source_DC"/>
    </choice>
  </element>
</define>
<!-- Complementary Information Level -->
<define name="comInf_element">
  <element name="struct">
    <ref name="comInf_DC"/>
    <ref name="comInf_element_DCures"/>
  </element>
</define>
<define name="comInf_element_DCures">
  <element name="feat">
    <choice>
      <ref name="note_DC"/>
      <ref name="originatingDatabaseName_DC"/>
```

```
      <ref name="originatingInstitution_DC"/>
      <ref name="source_DC"/>
    </choice>
  </element>
</define>
<!-- Term Entry Level -->
<define name="termEnt_element">
  <element name="struct">
    <ref name="terminologicalEntry_DC"/>
    <ref name="AttId"/>
    <interleave>
      <zeroOrMore>
        <ref name="termEnt_brack"/>
      </zeroOrMore>
      <zeroOrMore>
        <ref name="termEnt_element_DCures"/>
      </zeroOrMore>
      <oneOrMore>
        <ref name="termEnt_element_struct"/>
      </oneOrMore>
    </interleave>
  </element>
</define>
<define name="termEnt_brack">
  <element name="brack">
    <oneOrMore>
      <ref name="termEnt_element_DCures"/>
    </oneOrMore>
  </element>
</define>
<define name="termEnt_element_DCures">
  <element name="feat">
    <choice>
     <ref name="conceptType_DC"/>
      <ref name="elementWorkingStatus_DC"/>
      <ref name="subjectField_DC"/>
      <ref name="source_DC"/>
      <ref name="originatingDatabaseName_DC"/>
      <ref name="originatingInstitution_DC"/>
      <ref name="domainExpert_DC"/>
```

```
      <ref name="entrySource_DC"/>
      <ref name="termProvenance_DC"/>
      <group>
        <ref name="definition_DC"/>
        <ref name="attLang"/>
        <ref name="AttId"/>
      </group>
      <group>
        <ref name="note_DC"/>
        <ref name="attLang"/>
        <ref name="AttId"/>
      </group>
      <ref name="classificationCode_DC"/>
    </choice>
  </element>
</define>
<define name="termEnt_element_struct">
  <ref name="langSec_element"/>
</define>

<!-- Language Section Level -->
<define name="langSec_element">
  <element name="struct">
    <ref name="languageSection_DC"/>
    <ref name="AttId"/>
    <interleave>
     <zeroOrMore>
        <ref name="langSec_brack"/>
      </zeroOrMore>
      <oneOrMore>
        <ref name="langSec_element_DCures"/>
      </oneOrMore>
      <oneOrMore>
        <ref name="langSec_element_struct"/>
      </oneOrMore>
    </interleave>
  </element>
</define>
<define name="langSec_brack">
       <element name="brack">
```

```
            <interleave>
                  <zeroOrMore>
                        <ref name="langSec_brack"/>
                  </zeroOrMore>
                  <oneOrMore>
                        <ref name="langSec_element_DCures"/>
                  </oneOrMore>
            </interleave>
      </element>
</define>
<define name="langSec_element_DCures">
  <element name="feat">
    <choice>
      <ref name="domainExpert_DC"/>
      <ref name="elementWorkingStatus_DC"/>
      <ref name="originatingDatabaseName_DC"/>
      <ref name="originatingInstitution_DC"/>
      <ref name="source_DC"/>
      <group>
        <ref name="definition_DC"/>
        <ref name="attLang"/>
        <ref name="AttId"/>
      </group>
      <group>
        <ref name="note_DC"/>
      </group>
      <group>
        <ref name="languageIdentifier_DC"/>
        <choice>
          <value>nl</value>
          <value>fr</value>
          <value>en</value>
          <value>de</value>
          <value>it</value>
        </choice>
      </group>
    </choice>
  </element>
</define>
<define name="langSec_element_struct">
```

```
  <ref name="termSec_element"/>
</define>
<!-- Term Section Level -->
<define name="termSec_element">
  <element name="struct">
    <ref name="termSection_DC"/>
    <ref name="AttId"/>
    <interleave>
      <zeroOrMore>
        <ref name="termSec_brack"/>
      </zeroOrMore>
      <oneOrMore>
        <ref name="termSec_element_DCures"/>
      </oneOrMore>
      <zeroOrMore>
        <ref name="termSec_element_structs"/>
      </zeroOrMore>
    </interleave>
  </element>
</define>
<define name="termSec_brack">
  <element name="brack">
    <interleave>
      <zeroOrMore>
        <ref name="termSec_brack"/>
      </zeroOrMore>
      <oneOrMore>
        <ref name="termSec_element_DCures"/>
      </oneOrMore>
    </interleave>
  </element>
</define>
<define name="termSec_element_DCures">
  <element name="feat">
    <choice>
      <ref name="classificationCode_DC"/>
      <ref name="term_DC"/>
      <ref name="subjectField_DC"/>
      <ref name="domainExpert_DC"/>
      <ref name="elementWorkingStatus_DC"/>
```

```
      <ref name="entrySource_DC"/>
      <ref name="geographicalUsage_DC"/>
     <ref name="lemmaIdentifier_DC"/>
      <ref name="normativeAuthorization_DC"/>
      <ref name="originatingDatabaseName_DC"/>
      <ref name="originatingInstitution_DC"/>
      <ref name="source_DC"/>
      <ref name="termProvenance_DC"/>
      <group>
        <ref name="note_DC"/>
        <ref name="attLang"/>
      </group>
    </choice>
  </element>
</define>
<define name="termSec_element_structs">
  <zeroOrMore>
    <ref name="termComSec_element"/>
  </zeroOrMore>
</define>
<!-- Term Component Section Level -->
<define name="termComSec_element">
  <element name="struct">
    <ref name="termComponentSection_DC"/>
    <ref name="AttId"/>
    <interleave>
      <zeroOrMore>
        <ref name="termComSec_brack"/>
      </zeroOrMore>
      <oneOrMore>
        <ref name="termComSec_element_DCures"/>
      </oneOrMore>
    </interleave>
  </element>
</define>
<define name="termComSec_brack">
  <element name="brack">
    <interleave>
      <zeroOrMore>
        <ref name="termComSec_brack"/>
```

```
        </zeroOrMore>
        <oneOrMore>
          <ref name="termComSec_element_DCures"/>
        </oneOrMore>
      </interleave>
    </element>
  </define>
  <define name="termComSec_element_DCures">
    <element name="feat">
      <choice>
        <ref name="classificationCode_DC"/>
        <group>
          <ref name="definition_DC"/>
          <ref name="attLang"/>
          <ref name="AttId"/>
        </group>
        <ref name="elementWorkingStatus_DC"/>
        <group>
          <ref name="note_DC"/>
          <ref name="attLang"/>
        </group>
        <ref name="source_DC"/>
      </choice>
    </element>
  </define>
</grammar>
```

7.2 XML example

Corresponds to file “5465-mi-v0.94.xml”

```
<?xml version="1.0" encoding="UTF-8"?>
<struct type="terminologicalDataCollection">
  <struct type="terminologicalEntry" xml:id="TE.5465">
    <brack>
      <feat type="classificationCode">22298006</feat>
      <feat type="originatingInstitution">IHSTDO</feat>
      <feat type="originatingDatabaseName">SNOMED-CT</feat>
    </brack>
    <brack>
      <feat type="classificationCode">I21.9</feat>
      <feat type="originatingInstitution">WHO</feat>
      <feat type="originatingDatabaseName">ICD10</feat>
    </brack>
    <brack>
      <feat type="classificationCode">10039912</feat>
      <feat type="originatingDatabaseName">3BT</feat>
      <feat type="originatingInstitution">Ministère de la Santé Belge</feat>
    </brack>
    <brack>
      <feat type="classificationCode">K75</feat>
      <feat type="originatingInstitution">WICC</feat>
      <feat type="originatingDatabaseName">ICPC2</feat>
    </brack>
    <brack>
      <feat type="source">Dorland, 27th ed</feat>
      <feat type="definition" xml:id="TE.5465.DE.1" xml:lang="en">Gross necrosis of the
myocardium, as a result of interruption of the blood supply to the area.</feat>
    </brack>
    <brack>
      <feat type="domainExpert">Michael Card</feat>
      <feat type="originatingInstitution">University Hospital</feat>
    </brack>
    <feat type="conceptType">Diseases</feat>
    <feat type="elementWorkingStatus">pending</feat>
    <feat type="subjectField">cardiology</feat>
    <feat type="subjectField">general Practitioner</feat>
    <feat type="termProvenance">existing classification</feat>
```

```
    <!-- Language Section -->
    <struct type="languageSection">
      <feat type="languageIdentifier">en</feat>
      <brack>
       <feat type="domainExpert">Franck English</feat>
       <brack>
              <feat type="originatingInstitution">University Hospital of Leeds</feat>
       </brack>
      </brack>
      <feat type="elementWorkingStatus">pending</feat>
      <!-- Term Section -->
      <struct type="termSection" xml:id="TE.5465.TS.1">
        <feat type="term">myocardial infarction</feat>
        <feat type="normativeAuthorization">standardizedTerm</feat>
       <brack>
              <feat type="domainExpert">Steve Franck</feat>
              <feat type="originatingInstitution">University Hospital of Leeds</feat>
       </brack>
       <feat type="elementWorkingStatus">pending</feat>
       <feat type="termProvenance">existing classification</feat>
       <brack>
              <feat type="lemmaIdentifier">x84</feat>
              <feat type="originatingDatabaseName">English Lexicon</feat>
              <feat type="originatingInstitution">University of Leeds</feat>
       </brack>
      </struct>
    </struct>

    <!-- Language Section -->
    <struct type="languageSection">
      <feat type="languageIdentifier">nl</feat>
      <brack>
        <feat type="definition" xml:id="TE.5465.DE.2" xml:lang="nl">Een hartinfarct, is het
afsterven van een deel van de hartspier door onderbreking van de bloedtoevoer ervan door de
kransslagaderen.</feat>
        <feat type="source">http://nl.wikipedia.org/wiki/Hartinfarct</feat>
      </brack>

      <!-- Term Section -->
      <struct type="termSection" xml:id="TE.5465.TS.2">
```

```
        <feat type="term">myocardinfarct</feat>
        <feat type="normativeAuthorization">preferredTerm</feat>
       <brack>
             <feat type="domainExpert">Elma Jansen</feat>
             <feat type="originatingInstitution">University Hospital of Ghent</feat>
       </brack>
       <feat type="elementWorkingStatus">pending</feat>
       <feat type="termProvenance">existing classification</feat>
       <feat type="geographicalUsage">nl_be</feat>
       <brack>
             <feat type="lemmaIdentifier">x42</feat>
             <feat type="originatingDatabaseName">Dutch Lexicon</feat>
             <feat type="originatingInstitution">University of Ghent</feat>
       </brack>
      </struct>
    </struct>

    <!-- Language Section -->
    <struct type="languageSection">
      <feat type="languageIdentifier">fr</feat>
      <brack>
        <feat type="definition" xml:id="TE.5465.DE.3" xml:lang="fr">L'infarctus du myocarde
(IDM étant une abréviation courante) est une nécrose (mort de cellules) d'une partie du
muscle cardiaque.</feat>
        <feat type="source">http://fr.wikipedia.org/wiki/Infarctus_du_myocarde</feat>
     </brack>

      <!-- Term Section -->
      <struct type="termSection" xml:id="TE.5465.TS.3">
        <feat type="term">infarctus cardiaque</feat>
        <feat type="normativeAuthorization">preferredTerm</feat>
        <brack>
             <feat type="domainExpert">Amandine Desart</feat>
             <feat type="originatingInstitution">University Hospital of Brussels</feat>
       </brack>
       <feat type="elementWorkingStatus">pending</feat>
       <feat type="termProvenance">existing classification</feat>
       <feat type="geographicalUsage">fr_be</feat>
       <brack>
             <feat type="lemmaIdentifier">a21</feat>
```

```
            <feat type="originatingDatabaseName">French Lexicon</feat>
            <feat type="originatingInstitution">University of Liège</feat>
       </brack>
       </struct>

       <!-- Term Section -->
       <struct type="termSection" xml:id="TE.5465.TS.4">
         <feat type="term">infarctus du myocarde</feat>
         <feat type="normativeAuthorization">admittedTerm</feat>
       <brack>
            <feat type="domainExpert">Amandine Desart</feat>
            <feat type="originatingInstitution">University Hospital of Brussels</feat>
       </brack>
       <feat type="elementWorkingStatus">pending</feat>
       <feat type="termProvenance">existing classification</feat>
       <feat type="geographicalUsage">fr_ca</feat>
       <brack>
            <feat type="lemmaIdentifier">a23</feat>
            <feat type="originatingDatabaseName">French Lexicon</feat>
            <feat type="originatingInstitution">University of Liège</feat>
       </brack>
      </struct>
    </struct>
  </struct>
  <feat type="note">this is an example of a terminological entry</feat>
</struct>
```

7.3 Management of a Reference Terminology as a controlled vocabulary

1. Criteria to select concepts into the reference terminology
   - Corresponds with the sense of a frequently used word or paraphrase
   - Has a perfect (or almost perfect) match in a SNOMED concept (if not, a request for concept creation is to be sent to IHTSDO
   - Can be categorized to a typology of concepts
2. Procedure to add a concept into the reference terminology
   1. Decide on whether to accept or reject the concept for the reference terminology
   2. Classify the concept into typology
   3. Find the right Snomed-ct Concept
   4. Explore level of abstraction, precision of match; use snomed-ct browser and snomed-ct rule
   5. Import the SNOMED number, fully specified name and clinical label of the concept
   6. Translate the SNOMED label literally (acc. Snomed-ct rules)
   7. Choose the technical term for local use
      - On the basis of term/knowledge extraction projects
      - Advised by clinical experts
      - Based on term component analysis or post coordination
3. Optional Actions
   - If possible add a link to end-user terminology system for lay language (popular terms)
     - On the basis of term/knowledge extraction
     - By creation of phrase or neologism
   - Choose the type of term and link to relevant classifications
4. Validate at all relevant levels

7.4 Types of Health Care Professionals in Belgium

Primary care
General Practitioner/ Family Physician / Primary Care Physician
Certified medical specialists
Anesthesiologist
Surgeon
Neuro-surgeon
Plastic surgeon
Gynecologist
Ophtalmologist
Urologist
Orthopedic surgeon
Dermatologist
Internist
Pneumologist
Gastro-enterologist
Pediatrician
Cardiologist
Neuro-psychiatrist
Rheumatologist
Physiotherapist (medical)
Pathologist
Radiologist
Radiotherapist
Nuclear Medicine Physician
Dentist
Stomatologist

Medical Specialists not officially certified
Multidisciplinary specialists : e.g. oncologist
Combination of disciplines : e.g. pediatric pneumologist
Sub-specialties : e.g. proctologist
Uncertified specialists : Nephrologist, clinical pharmacologist, emergency physician
Pharmacy
Community pharmacist
Clinical Pharmacist
Hospital Pharmacist
Allied Health Personnel
Nurses
Physiotherapist (non-medical)
Psychologist
Social assistant
Dietician
Health care manager

7.5 CROSS TABLE : CONCEPT TYPE / CLASSIFICATION

(Snomed-ct used as pivot in most cells, hence not mentioned).

Note: non-exhaustive list.

CONCEPT TYPE \ CLASSIFICATION	Primary Care	Specialists	Allied health care Personnel	Health Care Management
Reason for Encounters	ICPC		NANDA	
Signs	ICPC		NANDA	
Fear of	ICPC			
Functional Status	Wonca, ICF	ICF		
Symptoms	ICPC		NANDA	
Diseases	ICPC, ICD	ICD		DRG
Causes of death	ICD	ICD		
Risk Factors			NANDA	
Social/Ethnological/Sociological: Problems	ICPC			
Disabilities	ICF	ICF	NANDA, ICF	
Procedures	ICPC, CPT	ICHI		ICHI
Laboratory examinations	Loinc	Loinc		
Physical Diagnostic/Examination				
Imaging	Loinc	Loinc		
Anatomical concept	Mesh	MeSH, FMA	FMA	
Physiological process	MeSH	MeSH		
HealthCare Agent/ Institution				
Devices	MeSH	MeSH		
Therapeutic procedures	ICD9-CM	ICD9-CM		
Drugs	ATC, MedDRA, NDC	ATC, NDC		
Prevention terms	MeSH	MeSH		

Illustration 6: Cross Table : Concept Type / Classification

7.6 Alignment between the Belgian Merge Project Excel file and the proposed Terminological Markup Language (TML)

An alignment is available in a separate spreadsheet: merge-excel-tml-correspondance-v3-20111028.ods

Legend:	**TMF Model**	**Corresponding level**	TE	TE
proposed additions		**Corresponding DC**	/classificationCode/ (value = Snomed-CT code)	/ elementWorkingStatus/ (value=pending)
workflow management (including new columns)	**Excel Model**		**Snomed CT number**	**Snomed ok**
			274663001	
			274665008	
			43724002	
			38880002	
			274640006	
			7520000	
			386661006	

TE	TE	TS in LS(lang=en)	TS in LS(lang=en)	TS in LS(lang=en)	TS in LS(lang=en)
/definition/	/ elementWorkingStatus/	/term/, /normativeAuthorization/ (value=standardizedTerm)	/ elementWorkingStatus/	/term/, /normativeAuthorization/ (value=preferredTerm)	/ elementWorkingStatus/
definition	**definition ok**	**FULLYSPECIFIEDNAME**	**FULLYSPECIFIEDNAME ok**	**EN_Clinical_Label**	**EN_Clinical_Label ok**
not foreseen in the current application but required for future evolutions		Acute pain (finding)		Acute pain	
		Chronic intractable pain (finding)		Chronic intractable pain	
		Chill (finding)		Chill	
		Rigor (finding)		Rigor	
		Fever with chills (finding)		Fever with chills	
		Pyrexia of unknown origin (finding)		Pyrexia of unknown origin	
		Fever (finding)		Fever	

TS in LS (lang=en) /term/, /normativeAuthorization/ (value=admittedTerm), /geographicalUsage/(value=UK)	TS in LS(lang=en) / elementWorkingStatus/	TS in LS(lang=nl) /term/, /normativeAuthorization/ (value=standardizedTerm)	TS in LS(lang=nl) / elementWorkingStatus/	TS in LS(lang=nl) /term/, /normativeAuthorization/ (value=preferredTerm)
EN_UK	**EN_UK ok**	**NL_Literal_translation**	**NL_literal_translation ok**	**NL_Clinical_Label**
				acute pijn chronische onbehandelbare pijn rillingen rigor koude rillingen met koorts koorts van onbekende oorsprong koorts

TS in LS(lang=nl) / elementWorkingStatus/	TS in LS (lang=nl) /lemmaIdentifier/, /provenance/ (value=3BT)	TS in LS (lang=nl) /term/, /normativeAuthorization/ (value=admittedTerm), /geographicalUsage/(value=BE)	TS in LS(lang=nl) /elementWorkingStatus/	TS in LS (lang=fr) /term/, /normativeAuthorization/ (value=StandardizedTerm)
NL_Clinical_Label ok	**IBUI**	**NL_BE**	**NL_BE ok**	**FR_Literal_Translation**
	10112164 10121983 10087203 10087183 10049408 10049426 10049305			

TS in LS(lang=fr) / elementWorkingStatus/	TS in LS (lang=fr) /term/, /normativeAuthorization/ (value=preferredTerm)	TS in LS(lang=fr) / elementWorkingStatus/	TS in LS (lang=fr) /lemmaIdentifier/, /provenance/ (value=3BT)	TS in LS (lang=fr) /term/, /normativeAuthorization/ (value=admittedTerm), /geographicalUsage/ (value=BE)
FR_Literal_Translation ok	**FR_Clinical_Label**	**FR_Clinical_Label ok**	**IBUI**	**FR_BE**
	douleur aiguë douleur chronique irréductible frissons raideurs fièvre avec frissons fièvre de cause inconnue fièvre			

TS in LS(lang=fr) / elementWorkingStatus/	TS in LS (lang=fr) /term/, /normativeAuthorization/ (value=admittedTerm), /geographicalUsage/ (value=CA)	TS in LS(lang=fr) / elementWorkingStatus/	TS in LS (lang=de) /term/, /normativeAuthorization/ (value=StandardizedTerm)	TS in LS(lang=de) / elementWorkingStatus/
FR_BE ok	**FR_CA (FR_Canada)**	**FR_CA (FR_Canada) ok**	**DE_Literal_Translation**	**DE_Literal_Translation ok**

TS in LS (lang=de) /term/, /normativeAuthorization/ (value=preferredTerm)	TS in LS(lang=de) / elementWorkingStatus/	TS in LS (lang=de) /term/, /normativeAuthorization/ (value=admittedTerm), /geographicalUsage/ (value=BE)	TS in LS(lang=de) / elementWorkingStatus/	TE / elementWorkingStatus/ (value=pending)
DE_Clinical_Label	**DE_Clinical_Label ok**	**DE_BE**	**DE_BE ok**	**Concept OK?**
not foreseen in the current application but required for future evolutions				

TE / classificationCode/ (value = icpc2 code)	TE / elementWorkingStatus/ (value=pending)	TE /classificationCode/ (value = ICD10-CM code)	TE /elementWorkingStatus/ (value=pending)	TE /note/
ICPC_2_Code	**ICPC_2_Code ok**	**ICD10_CM_Code**	**ICD10_CM_Code ok**	**Specific**
A01		R52.0		
A01		R52.1		
A02		R68.8		
A02		R68.8		
A03		R50.0		
A03		R50.9		
A03		R50.9		