Serveur Terminologique

Recherche préliminaire en rapport avec le développement d'un serveur belge de terminologie dans le secteur de la santé.

Ref. Ctr-n09-03-serveur terminologique-MIM, financé par le SPF Santé Publique.

Partie 3/3 : Opérationnalisation

Version : 1.1

31 octobre 2010

Coordinateur-rédacteur : Philippe Vandenbergh

Comité rédactionnel :

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

Editeur. : MIM

Index

Introduction

This document is the prolongation of the second part of the preliminary study for the development of a Belgian terminology server in the healthcare sector, the “structured argumentation brief” or “argumentary”.

According to the methodology of the brief, a bottom-up approach was used to identify the needs related to terminology of the “users” of information technologies in the Healthcare sector.
One should consider users in a broad perspective; e.g. users are clinicians, personnel of the Ministry of Health, INAMI/RIZIV/LIKIV, but also the patient as participative citizen.

The domains extend from typical clinical applications as Electronic Health Records, used by medical doctors, nurses and other paramedical professionals, Personal Health Record, clinical and operational decision support systems, to billing registrations, epidemiology studies, business intelligence appliances.

The documentation of the clinical aspects is by far the most complex because of the exhaustivity of the concepts and the multiplicity of “not computer literate” stakeholders. The highest granularity of data is required for the documentation of clinical concepts while on the other hand a high level of usability – a fast, intuitive and single reusable capture of concepts – is mandatory for acceptance of the registration of data by care professionals.
Some registered data have an administrative accent, institution, care professional and patient identification, role and contact information but most of the data are related to the health status documentation (reason of encounter, observation, diagnosis, invalidity, medication, clinical biology, radiology protocol integration, diagnosis and findings) and the actions taken (procedures, prescription of medications or orders).
Some data registrations (eg CareNet, DRG’s, epidemiology, eCare registration suite,…) are compulsory, other not.
Anyway, the continuity of the registration must be guaranteed.
The multi-language requirements and the need to collaborate with other countries must be borne in mind.

So, semantic interoperability is a matter of people, information technology and comprehensive knowledge exchange.
We learned from the former part of the deliverable there is a need and a business case.

In this part we explore several aspects of operationalisation of terminology services in Belgium. It is an attempt of description of a vision, strategy for a Belgian Terminology Centre including terminology management aspects - the setup and the maintenance of the terminology – and the terminology services. Aspects of operational management, requirements for the organisation of these services and functionalities of internal applications and services toward third parties are presented. Roadmaps and scenarii are proposed as well as next steps.

This document should be considered as a work document to build upon.

Chapter 1 - The Terminology Centre

Introduction

The Belgian Terminology Centre (TC) is the organisation governing the development and the maintenance of the Reference Vocabulary and Reference Terminology to be used in the context of the Belgian Healthcare system. It is the unique authentic validated source for terminologies in the sector. In analogy with the CBIP/BCFI (www.bcfi.be), the reference centre provides to end-users, directly or indirectly access to a validated reference terminology set.

The reference centre provides services
- directly to the users through a portal application allowing to brows the Reference Vocabulary and Reference Terminology and all the source terminologies supported by the TC or
- indirectly through a computer application accessing all or a subset of the terminologies provided by the TC at run time through internet or through integration of files provided by the TC to the application provider. The later is similar to the today practice in the Electronic Medical Record for the drug reference data base of the CBIP/BCFI which is incorporated into the end-user application.

The end user shouldn't notice he moved from his application environment - the software supporting his activities - to another, in casu the terminology services.

A Terminology Centre in Healthcare, a challenge

The domain of healthcare is one of the most knowledge intensive with a high level of complexity and heterogeneity of the concepts, but also the diversity of the stakeholders, from experts in fields to the patient citizen which have different information needs, different information processing capacities and different "linguo argot".

It is pretty clear that a multilingual list of terms, with a source term and a target term, with no additional information is not enough. A spreadsheet with x columns will not provide the services needed at user level, end user of clinical applications or terminology management systems.

The collection, selection, maintenance, of terminologies must be managed in order to maintain continuity and sustainability.
The complexity and exhaustivity of the domain terminology require multidisciplinary teams with a leading, vision and strategy and information technology instruments to support their work.

Goals of the Terminology Centre

At high level, the main goal of the terminology services provided by the TC is to facilitate the interoperability between end-users and computer applications and between computer applications in order to support the collaboration between the care providers. (see also fig. 1 of the "Avant-Propos of part 2 of the deliverable) Finally the terminologies must ensure the capacity to automate and improve the quality, security and efficiency of the care processes trough semantic interoperability.

The terminology centre will support services as

- medical content management (indexing and querying for medical information)
- medical knowledge management (computable knowledge bases)
- clinical decision support for quality and safety management
- operational decision support for a better business process management
- natural language support for the capture and representation of information in an adequate language (language, choice of word related to the end-user) and qualitative way. At end-user human level, terminologies must help natural language support in order to facilitate the use of computer technologies by humans, care professionals and the patient/citizen and to improve the quality of registration of (clinical) data.

Terminology Centre Comp

The Terminology Centre creates and maintains a reference data set for all the terminologies used in the healthcare sector today and tomorrow. It is the pivot common data structure.

Based on a high level consensus for the building of the Belgian terminology, the terminology centre must promote and enforce the alignment of vocabularies, classifications and terminologies and its use in Belgium in a scenario of continuity and sustainable working .

Based on the requirement of the argumentary part of the deliverable, the Terminology Centre should develop an instrument for :

1. managing healthcare concepts
2. managing external references to concepts
3. managing the internal representation of concepts
4. managing the data and the mapping of concepts
5. managing the mapping of concepts to vocabularies accessible repository of multi language medical conceptual knowledge mapped with the natural languages vocabularies
6. managing the data and functionality required to map concept entities to and from existing coding and classification schemes
7. providing dynamically generated local nomenclatures or 'coding schemes'
8. providing alert and disaster recovery procedure
9. managing the distribution of the updates of the concepts and associated terminologies
10. managing the links with the national and international stakeholders (INAMI/RIZIV, Ministry of Health, academic, epSOS…)

11. managing the education and continuous training of the share- and stakeholders
12. supporting R&D in the field of terminologies, especially ontologies and Natural Language Processing

Concepts encompass clinical but also medical administrative aspects such as billing, registration items (INAMI/RIZIV; eCare registration,…).

The Terminology Centre is not a clinical data collector. The terminology server is not supposed to collect patient data. It records and managed clinical information models at knowledge level, not at patient level, nor at activity of the care providers level.

As described in the second deliverable, several terminologies are used in the care sector. No one is able to embrace all the needs of all the different actors. The objectives of several of them are diametrically opposed to each other. Actors use different terminologies.

In the first deliverable we learned that even though a lot of medical concepts have a common understanding by the different healthcare actors these concepts are often identified by different terms and codes from different terminologies. For the same domain of interest the users work in different settings with different vocabularies, terminologies or classifications. The different terminologies are not (or not systematically) mapped. A fortiori there is no description of the relation, as identical, parent child, narrower or broader medical concept…

An exercise of mapping between the existing terminologies and a controlled vocabulary showed us the need for accepting all these terminologies and the need for the implementation of a sound methodology and adequate management. The concepts, representing the knowledge, within the terminologies but also linguistic aspects will change in time. This dynamic process will last forever since science is in perpetual evolution.

There is no panacea universalis; the progress toward semantic interoperability can be achieved only through the mapping of these resources to a reference vocabulary.
In order to cover all the functional requirements, a “Layered Terminological Architecture” is proposed. (Fig 1)

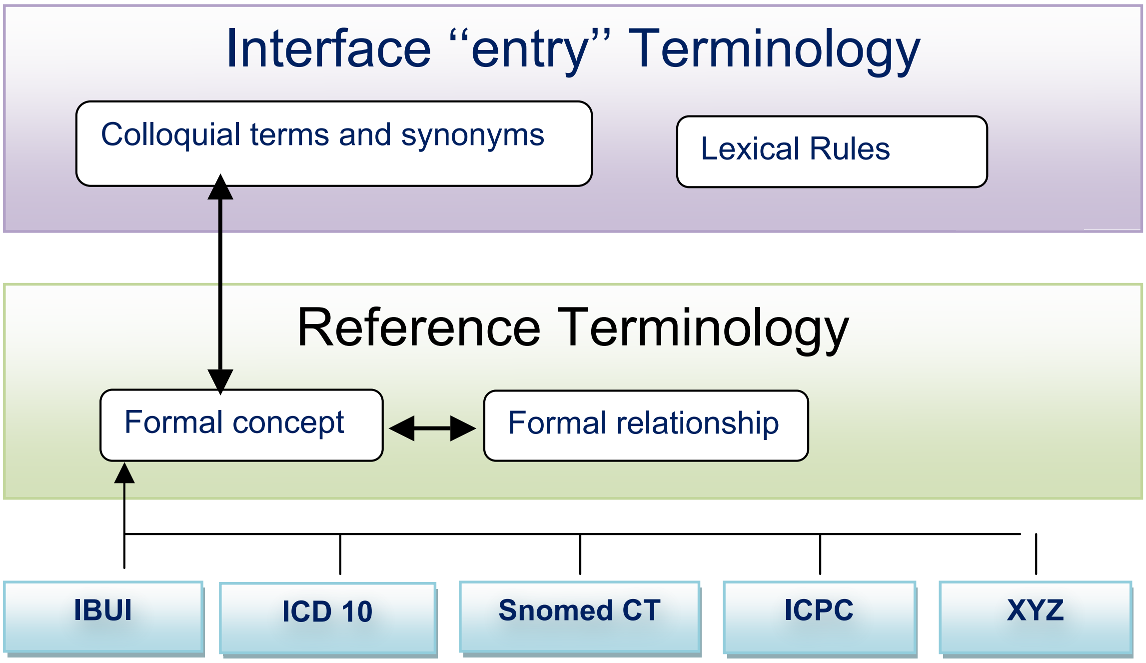


Figure 1 Conceptual model of layered terminology architecture.

The Reference Terminology contains all concepts at knowledge level, with context. The formal relationship between the concepts using a semantic grammar allows the representation of the knowledge and the biomedical logic of nosological entities. A biomedical ontology is created where all what is biomedical, disease, symptoms and signs, structures and their functions,... is described. The Foundational Model of Anatomy (FMA) is an example of a biomedical ontology describing the body structures[1].

The example of the word "sterilisation" illustrates the need to provide a context. The word alone has several meanings depending of the context the word is used (sterilisation of individual, of surgical instrument, of milk or other beverages…). Contexts describe the environment in which a concept is used; it is an attribute to the concept. We can define several contexts, some are applicable for "all" concepts and others are specific for concept categories. Examples of contexts are subject relationship context (eg who the finding applies to, user role, a user type), temporal context (eg when the finding applied), finding context (whether the finding applied) and associated finding context (what finding applies). A context restrains and enforces precision of a concept. This is crucial for the exact comprehension/interpretation of the data registered (alignment between the idea concept in the brain of the end-user and the "code") but also for the representation of the data, the associated linguistic aspects. This is essential for Natural Language support for example.

Domain ontologies exist that describe:

- occurrences, occurrence ontology, describing how it occurred (when, who reported it?, what were the circumstances, etc?...)
- processes, process ontology, describing within which activity something happens (who, when, within what activity, framework, care plan, …)
- a Reference Information Model (RIM) or how things are organized for end-users (active problems, encounters, observations, prescriptions, templates en queries,…)

One of the main issues for representing information in Electronic Health Records is the link between terminology models and structural models. There is a grey area with gaps and even worse, overlap that is a semantic killer[2]. See also annex 3.

The Interface Terminology contains all terms for the correct capture of concepts and their representation in a human friendly way as close as possible of the human natural language. Natural language support can be seen from the linguistic perspective or from the knowledge, concept, perspective. The first is called Natural Language Processing, the second Natural Language Generation. In both cases, the capture of information is provided in an adequate language (language, choice of word related to the end-user) and in a qualitative way. Of course here also contextual information is needed in order to provide a correct concept proposition and formulation. For the representation of the information – this is the use case depicted in the "avant-propos" of the argumentation brief, of Alice reads data registered by Bob - where Alice has another profile/role than Bob and uses another language, other terms to express the "same" concept, the user interface has to produce words understandable by the receiver of the information and of adapted granularity.

---

[1] http://sig.biostr.washington.edu/projects/fm/AboutFM.html

[2] Terminology Binding Requirements and Principles; D. Markwell, NHS Connecting for Health; 2008. See also annex 3.

Colloquial terms should be considered in the broad meaning of the word and applied as such. Colloquial should be understood as informal- as a citizen speaks - as well as specialized in the context of a care professional, eg a scientific jargon.
Lexical rules are the grammar in the language, the rules for the composition of sentences, which is important to express and understand concepts, words in an unambiguous way … The example of "metastasis of" quoted in part 1 B of the deliverable illustrates the need to define rules in order to guarantee alignment in understandings. A metastasis in the colon and metastasis originated from the colon are different cases!
In the next paragraphs the main aspects of terminology management are identified. Finally a roadmap proposal for a Belgian terminology work up will be presented.

In annex 1, Fundamental Requirements of Medical Terminologies are summarized. It is the result of the aggregation of the information found in the literature. In annex 2, practical principles of terminology binding are cited.

Chapter 3 - The Terminology Server

The Terminology Server must support two application domains, terminology services and terminology management.

Terminology Services

The Terminology Services (TS) supports operational systems at run time with dynamic querying, interpretation and encapsulation of natural language expressions, codes, and references. The TS will help medical applications in their interactions with end-users.

Terminology Management System

The Terminology Management System (TMS) supports knowledge engineers and clinical analysts in the development and maintenance of the terminological system : Terms, concepts, codes, lexical layer semantic model, NLP.

Standards and normative references

Some standards and relationship with a short definition is listed here. It is not the intention to share an exhaustive list. Definition of terminologies was done in the argumentary.

- Ontology is the study of what there is. *Ontology* (sense informatics) is a logical model of the meanings of the entities about which information is to be expressed for use in computers.
  Formal ontologies are theories that attempt to give precise mathematical formulations of the properties and relations of certain entities (Stanford Encyclopedia of Philosophy), in respect with ontological principles, such as the use of well-defined, unambiguous, and non-idiosyncratic types and relations *[Schulz2006]*.

  Formal ontologies can be expressed in two formats:

  - Natural Language
    "Every hepatitis is an inflammatory disease that is located in some liver"
    "Every inflammatory disease that is located in some liver is hepatitis"
  - Logic
    $\forall x\text{: } instanceOf(x, Hepatitis) \Leftrightarrow instanceOf(x, Inflammation) \wedge$
    $\exists y\text{: } instanceOf(y, Liver) \wedge hasLocation(x,y)$

  Logic is computable: it supports machine inferences but it only scales up if it has a very limited expressivity

Terminologies - Entities of Language	**Formal Ontologies -** Entity Types
Describe: meaning of human language units	Describe: entities of reality as they generically are independent of human language
"Concepts": aggregate (quasi) - synonymous terms	"Types": represent the generic properties of world entities
Relations: informal, elastic Associations between Concepts ……	Relations: rigid, exactly defined, quantified relationships between particulars
Description pattern: Concept1Relation Concept2	Description pattern: for all instance of Type1: there is some…
E.g.: UMLS, Terminologia Anatomica	E.g.: Foundational Model of Anatomy, Gene Ontology, OpenGALEN,

- The Web Ontology Language, OWL, is a semantic markup language for publishing and sharing ontologies on the World Wide Web. OWL is developed as a vocabulary extension of RDF (the Resource Description Framework) and is derived from the DAML+OIL Web Ontology Language.
  The OWL language provides two specific subsets. OWL Lite was designed for easy implementation and to provide users with a functional subset that will get them started in the use of OWL. OWL DL (where DL stands for "Description Logic") was designed to support the existing Description Logic business segment and to provide a language subset that has desirable computational properties for reasoning systems. The complete OWL language (called OWL Full to distinguish it from the subsets) relaxes some of the constraints on OWL DL so as to make available features which may be of use to many database and knowledge representation systems, but which violate the constraints of Description Logic "reasoners". http://www.w3.org

- W3C Simple Knowledge Organization System, SKOS, is a specification that defines types and a common syntax for the representation of knowledge collections such as thesauri, taxonomies, classification schemes and subject heading systems. SKOS does not specify behavior.
-
- The Object Management Group, OMG, is a not-for-profit, open-membership computer industry specifications consortium; the members define and maintain the MOF specification that they publish. Software providers of every kind build modeling tools that manipulate models in MOF-compliant format - export, import, store, transform, generate code, and so on. OMG doesn't provide any of the software - they provide only the specifications that make software products interoperate. http://www.omg.org/

- Model Driven Architecture, OMG's MDA is the industry-standard architecture of OMG. Based on MOF-enabled transformations, the MDA unifies every step of the development of an application or integrated suite from its start as a Platform- Independent Model (PIM) of the application's business functionality and behavior, through one or more Platform-Specific Models (PSMs), to generated code and a deployable application. The PIM remains stable as technology evolves, extending and thereby maximizing software ROI. Portability and interoperability are built into the architecture. MDA relies on the MOF to integrate the modeling steps that start a development or integration project with the coding that follows. You can read about the details on our MDA Specifications Page, which starts by describing the importance of MOF to MDA and continues with references to the additional OMG standards that complete the set. Model portability is so natural that many MDA code generators do not include their own modeling capability - they require users to use their preferred modeler out of a list of compatible candidates!

- Meta-Object Facility MOF is an extensible model driven integration framework for defining, manipulating and integrating metadata and data in a platform independent manner. MOF-based standards are in use for integrating tools, applications and data.
- Unified Modelling Language: A specification defining a graphical language for visualizing, specifying, constructing, and documenting the artifacts of distributed object systems.
- Ontology Definition Metamodel, ODM, specifies the Object Constraint Language (OCL), a formal language used to describe expressions on UML models. These expressions typically specify invariant conditions that must hold for the system being modeled or queries over objects described in a model. Note that when the OCL expressions are evaluated, they do not have side effects (i.e., their evaluation cannot alter the state of the corresponding executing system). OCL expressions can be used to specify operations / actions that, when executed, do alter the state of the system. UML modelers can use OCL to specify application-specific constraints in their models and to specify queries on the UML model, which are completely programming language independent.

Chapter 4 - Terminology Services

The Terminology Services (TS) supports operational systems at run time with dynamic querying, interpretation and encapsulation of natural language expressions, codes, and references. The TS will help medical applications in their interactions with end-users.

The roles of the Terminology Services

1. Shared reference terminology service
2. Improve the usability of clinical applications through an interface* terminology service
3. Facilitate the development of the clinical applications through reuse of the content and logical resources of the terminology service

> **The major premise of a clinical terminology server is that it is used by clinicians to enter patient observations, findings, and events. This is imposing a terminology design allowing to expose to the user an interface as near as possible to his natural clinical*

Terminology File standardisation:

Because terminologies are shared, the exchange of terminology file must respects standards. The Terminological Markup Framework is an ISO 16642 standard[3] and is suggested as the one to adopt and apply as common practice for terminology file exchange within the Belgian terminology centre as well as when applicable in all implementation and usage in Belgium.

The types of clinical applications to be supported by a Terminology Service

1. Public health applications	Conversion amongst coding and classification schemes for data consolidation and aggregation to allow intelligent querying for research and epidemiological analysis.
2. Medical Records ePrescription Applications	User friendly interfaces for clinical information systems for quality data input
3. Clinical decision support systems	Semantic of the EHR data and knowledge bases for unambiguous inference and rules
4. Medical Content Applications	Indexing, Bibliographic retrieval

[3] http://www.loria.fr/projets/TMF/: http://www.iso.org/iso/catalogue_detail.htm?csnumber=32347

High level functional requirements

In the perspective of exploitation of the services[4]:

(1) support efficient and user-friendly data entry and query formulation;
(2) record and archive clinical knowledge information;
(3) support sharing and reuse of clinical knowledge information
(4) infer and suggest knowledge according to decision support algorithms;
(5) support the terminology maintenance;
(6) support natural language output

The TS must support both semasiological and omasiological lexicological approaches.

> *Semasiology departs from a word and asks what it means, or what concepts the word refers to. An Onomasiology departs from a concept and asks for its names. An onomasiological question is, e.g., "what are the* ***names for*** *an infectious disease of the lungs (answers: pneumonia while a semasiological question is, e.g., "what is the* ***meaning of*** *the term pneumonia?" (answer: "infectious disease of the lungs")*

Use Cases

An external application is *connecting to* the Terminology Server to perform a series of *requests:*

1.Lexical (Semasiological) Query

The Terminology Server is presented with a lexical entity (term). The Terminology Server checks if his vocabulary contains the term

1.1 the term exists : it returns the reference concept witch consists of
  a) the reference code
  b) the list of synonyms per language with the preferred term from the reference vocabulary
  c) the hierarchies where it classifies in the Reference Ontology
  d) the list of equivalent and of similar[5] codes with the corresponding terms from the mapped vocabularies, terminologies , classifications
  e) the hierarchies where it classifies mapped vocabularies, terminologies , classifications
  f) the list of similar codes with the corresponding terms from the mapped vocabularies, terminologies , classifications

[4] Adapted from (A. Rector, Chute, Spackman)
[5] The exact mapping or translation is not always possible because the differences in concepts representations in the target system. In many situations the best that can be done is to provide the application with the information on the potential matches It is then up to the external application program to decide how to deal with this information according to its own particular requirements.

In case of homonymy it returns the user a list of homonym definitions with a method to choose the right concept.
In case of an expression composed of several terms the TS returns the list of reference concepts in the order of the most relevance.

1.2 the term do not exists : generate and classify the new concept entity which is submitted as candidate in the approval workflow

2.Semantic (Onomasiological) query

The Terminology Server is presented with an expression of several terms, which is the definition of a concept. The Terminology Server checks if the expression is legal (composed in accordance with the Reference Semantic Model).

1.1 This is a legal expression: it check for redundancy and returns

a) the reference code
b) a short definition and complete definition (per language)
c) the list of synonyms per language with the preferred term from the reference vocabulary
d) the hierarchies where it classifies in the Reference Ontology (the general concepts that are subsuming it and the more specialized concept entities it subsumes)
e) the list of equivalent and of similar[6] codes with the corresponding terms from the mapped vocabularies, terminologies , classifications
f) the hierarchies where it classifies mapped vocabularies, terminologies , classifications
g) the list of similar codes with the corresponding terms from the mapped vocabularies, terminologies , classifications
h) semantic metadata about the concept entity : the semantic type and the legal relations ("attributes, qualifiers")

1.2 If it is not a legal expression: it executes a lexical query on expression (Use Case 1.1)

Mapping using the Terminology Server is a two-stage process — first map an expression into the Reference Model and then map it into the target external representation.

This allows the TS to answer the following type of questions:

a) What are the external expressions for this concept entity in a particular external system? What is the preferred term for this concept entity in that system?
b) What are the natural language expressions for this concept in a particular language? What is the preferred form for a particular 'clinical linguistic group'?
c) Are these two concept entities derived from two different external representations the same? If not, how do they differ?
d) Find all of the expressions in a given external representation that corresponds to children of this concept entity, i.e. all of the codes that this concept entity subsumes.

[6] The exact mapping or translation is not always possible because the differences in concepts representations in the target system. In many situations the best that can be done is to provide the application with the information on the potential matches It is then up to the external application program to decide how to deal with this information according to its own particular requirements.

This allows the Terminology Server to compensate for the deficiencies in the organisation of external coding systems. For example, forms of heart disease are found in at least five different chapters of ICD-9.

Chapter 5 - Terminology Management System

The Terminology Management System is the other application domain in the TC.

The TMS supports knowledge engineers, clinical analysts in the development and maintenance of the terminological system that includes the semantic model as a reference model, the concepts, the codes and the lexical layer.

Objectives

The Terminology Management System should support the management of the terminology lifecycle

- the management of external references to concept entities,
- the management of the internal representation of concept entities,
- the management of the data and functionalities required to map concept entities to vocabularies accessible repository of multi language medical conceptual, knowledge mapped with the natural languages vocabularies
- the management of the data and functionalities required to map concept entities to and from existing coding and classification schemes
- the management of dynamically generated local nomenclatures or 'coding schemes'
- the management of translation
- the management of process automation through reasoning, including advanced queries and rules and automated workflow
- the change management, conflict detection and resolution
- distributed collaborative authoring
    - Role based authoring rights
    - Approval workflow
    - Quality control

Layered Structure

In order to achieve a progressive development and deployment of the service we have to implement a modular service orientated system architecture in 3 phases. The phases are described in the chapter "Roadmap".

Phase 1 : Reference Vocabulary Management System

Phase 2 : Reference Terminology Management System

Phase 3 : Interface Terminology System and NLP Module

1. Reference Vocabulary Management System

The controlled vocabulary will define the unique identifier, the name, the preferred term and the synonyms, the homonyms. In practice, this phase is a quality check of the merge result of the 3BT /UZBrussels terminology set. 15.000 concepts are described.

The Reference Vocabulary Management System must a minima support following set of use cases:

1. Import external vocabularies , classifications, terminologies (ICD10 , IBUI, …)
2. Display query and browse external (source) vocabularies , classifications, terminologies
3. Display query and browse the Reference Vocabulary
4. Select a source term in the external vocabularies, classifications, terminologies and execute following action:
   Acquire it as new reference term in the Reference Vocabulary
5. Select a source term in the external vocabularies, classifications, terminologies and a term in the Reference Vocabulary and execute following action (Example in the fig 2) :
   Link the external term to the internal term under 3 mapping types: 1.identical, 2.broader as 3.narrower as.
6. Declare a term as preferred

Other use cases are:
1. Edit the Reference Vocabulary terms: name, preferred or synonym, definition
2. Translate Reference Vocabulary Terms

The Lexical capacities of the Reference Vocabulary Management System

1.Word Normalization

The user can enter lexical variants of words that may not match their corresponding representations in a terminology. Common variations include gender, tense, case, pluralization, possessive inflection, or punctuation. Such variations can impair the matching between entered words and phrases with target terminology entries.
The TMS should have a word normalization module to improve the relevance of the query return

2.Word Completion

The TMS will return a list of terms based on the word completion of the submitted string.
Ex: the external applications submits a query with "pneumoc." The TMS returns a list containing pneumoccoc, pneumocyst

3.Spelling Correction

While word normalization and lexical matching offer substantial flexibility and advantage, there will always be a spelling variability. The TMS should have a Spell Checker. The spell checker can return the list of correctly spelled terms for the user to chose from.

4.Composed Term Completion

The TMS returns the list of composed concepts containing the submitted term.
For example, seeking the concept Turner's syndrome a user might enter simply "Turner," expecting the system to complete the full term by adding the "syndrome" ending.

Limitations of a Vocabulary Management System

Diabetic polyneuropathy feet
Diabetic polyneuropathy leggs

Diabetic polyneuropathy feet legs
Diabetic polyneuropathy legs feet

Diabetic polyneuropathy Parenthesis
Diabetic polyneuropathy Dysesthesis

Diabetic polyneuropathy Paresthesis Dysesthesis
Diabetic polyneuropathy Dysesthesis Paresthesis

Diabetic polyneuropathy feet Paresthesis
Diabetic polyneuropathy Paresthesis feet

Diabetic polyneuropathy feet Dysesthesis
Diabetic polyneuropathy Dysesthesis feet

Diabetic polyneuropathy legs Paresthesis
Diabetic polyneuropathy Paresthesis legs

Diabetic polyneuropathy leggs Dysesthesis
Diabetic polyneuropathy Dysesthesis legs

Diabetic polyneuropathy feet legs Paresthesis
Diabetic polyneuropathy legs feet Paresthesis
Diabetic polyneuropathy Paresthesis legs feet
Diabetic polyneuropathy Paresthesis feet legs

Diabetic polyneuropathy feet legs Dysesthesis
Diabetic polyneuropathy legs feet Dysesthesis
Diabetic polyneuropathy Dysesthesis feet legs
Diabetic polyneuropathy Dysesthesis legs feet

Diabetic polyneuropathy feet legs Paresthesias Dysesthesis
Diabetic polyneuropathy legs feet Paresthesias Dysesthesis

Diabetic polyneuropathy feet legs Paresthesis Dysesthesis
Diabetic polyneuropathy legs feet Paresthesis Dysesthesis

Diabetic polyneuropathy feet legs Dysesthesis Paresthesis
Diabetic polyneuropathy Dysesthesis Paresthesis feet legs

Diabetic polyneuropathy legs feet Dysesthesis Paresthesis
Diabetic polyneuropathy Dysesthesis Paresthesis legs feet

Figure 2: The flat list. The concepts in red are "synonyms" of the leading concept.

Note that this use case illustrates also the limits of this approach. Insulin-dependent diabetes mellitus is a composed concept but could be accepted since it is a frequent disease and a popular term. This is semantically inadmissible unless post decomposition is done and the composed concept recognizable as such. The Reference Terminology Management System will decompose the concept in single concepts and define the relations between them. This is a phase 2 topic.
Let's look in figure 2 at another example. It illustrate the difference between what we find in a terminology as 3BT or the merge and a reference semantic terminology

2. Reference Terminology Management System

The Reference Terminology Management System is used to enrich the controlled vocabulary through a reference semantic model. The Semantic Reference Model defines the categories of the medical domain, their relations and semantic constraints.

The Reference Terminology Management System will extend the functionality of the reference vocabulary management system through the – incremental – acquisition of semantically valid new terms and by assigning them to the corresponding categories of the reference semantic model (ontology).
It is important to start with a clean desk and to choose a valid methodology and ontology description.

The generic model of the Semantic Reference Model

Concept

- Unique ID
- Hierarchy, category and attributes. For each categories a set of attributes are defined.
- Concept name
    - Language name
        - Scientific name
        - Popular name
        - Short definition
        - full definition
        - Preferred name
- Semantic formula, eg
    - Has laterality
    - Has cardinality
    - Has part
    - Has evolution
    - …
- rules in order to define (describe) the domain and range
- Mapped external concept
    - identical to
    - similar with
    - more generic than
    - more specific than

As illustration the SNOMED CT components of their model is presented in the figures 3 and 4. Snomed CT has 18 built-in upper level hierarchies. For each hierarchy a set of attributes is defined.
Relationship between concepts are defined using semantic formula. See also annex 4.

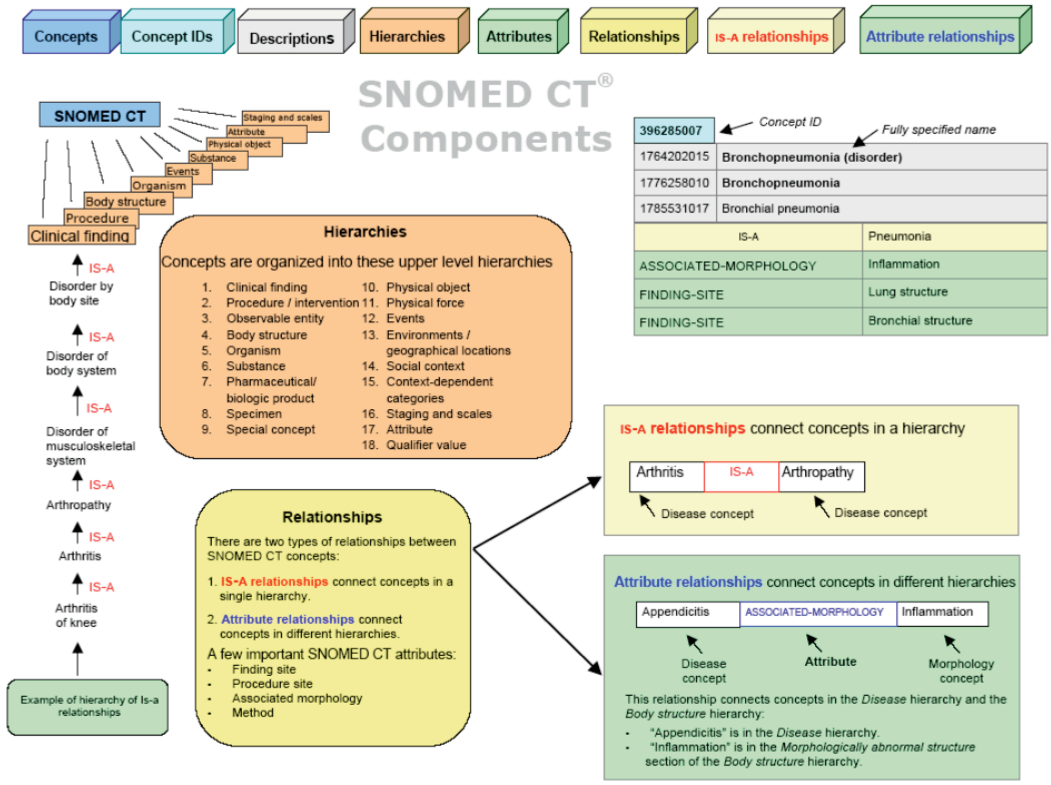


Fig 3 SNOMED CT terminology components of a bio medical ontology.

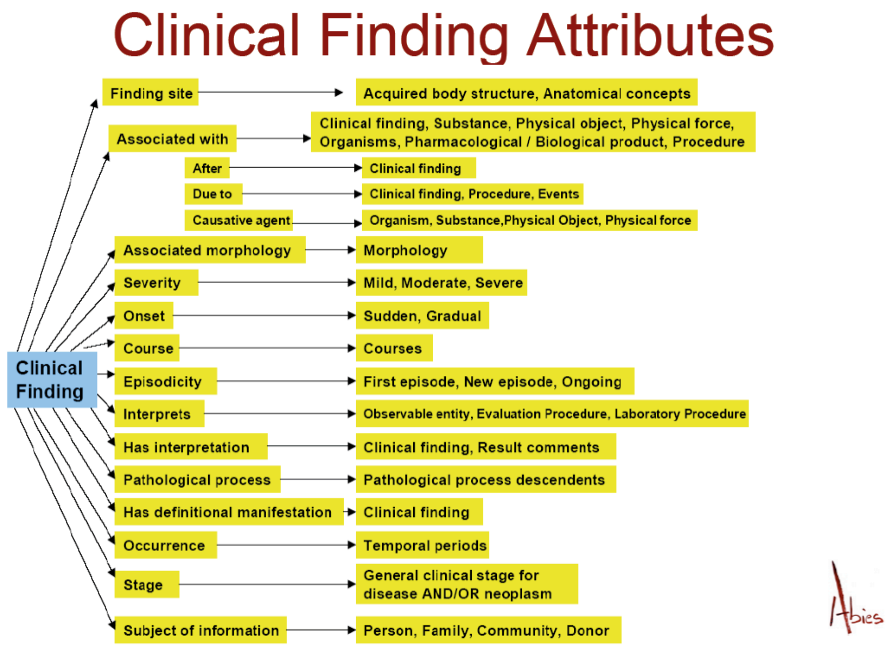


Fig 4 Snomed CT attributes of the category clinical findings

In figure 5 an example from the Foundational Model of Anatomy[7] (Open Source biomedical ontology) illustrates hierarchy and the links between entities with limited semantic relations.

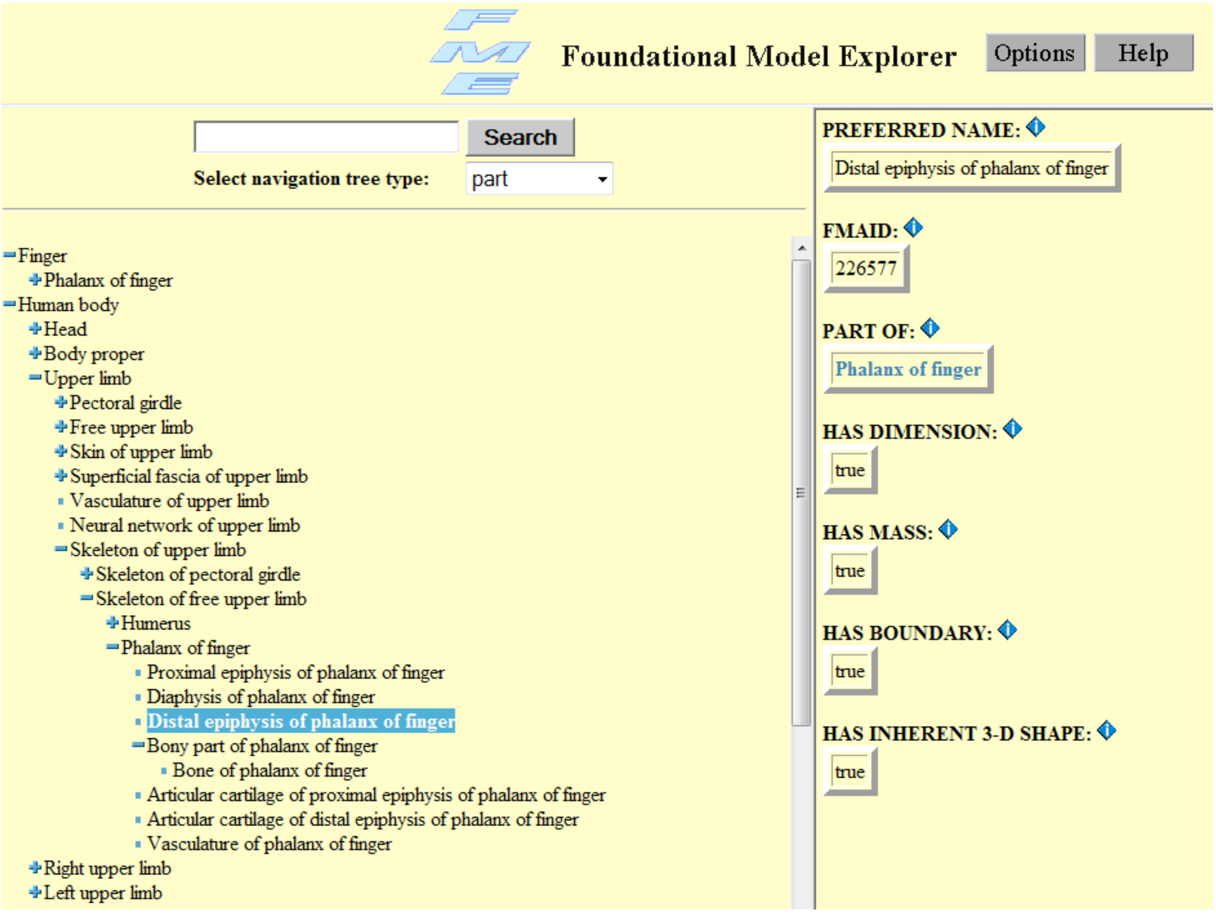


Fig 5 Foundational Model of Anatomy, browsing through a hierarchy view.

If we consider the same concepts illustrated in figure 2 in the limitation of a **Vocabulary Management System** and use a reference terminology we have for the same example in the former section:

Diabetic polyneuropathy

has localisation (body part list)

feet

legs

has symptoms (list neurological symptoms)

paresthesias

dysesthesias

[7] http://sig.biostr.washington.edu/projects/fm/AboutFM.html

In this simple example our Formula Generator creates : **36 entities (= concepts names)** out of which :

14 concepts

22 synonyms

9 time 1 synonym for 1 concept

2 time 3 synonyms for 1 concept

1 time 7 synonyms for 1 concept

One will fairly suggest diabetic neuropathy is a composed concept. This means that representing all the possibilities in one list will augment dramatically the numbers of lines (if there is no reference terminology behind the scene) but will not affect seriously a managed reference terminology.
The impact of the reference terminology is directly observable at interface level as we see in the next section. Long picking list, a usability killer, is history.

3. Interface Terminology System

Clinical interface terminology was defined as a systematic collection of health care – related phrases (terms) that supports clinicians' entry of patient - related information into computer programs, such as clinical ‘‘note capture’’ and decision support tools. Interface terminologies also facilitate display of computer-stored patient information to clinician-users as simple human readable text. The ‘‘interface’’ of interface terminologies (which have also been called colloquial terminologies, application terminologies and entry terminologies) links health care providers' own free text patient descriptors to structured, coded internal data elements used by specific clinical computer programs *[Rosenbloom2006].*

Interface terminology supports the creation of new concepts using two methods: pre-coordination and post-coordination. With ***pre-coordination*** (also called enumeration), developers model relevant levels of detail in the terminology with distinct concepts, typically derived from real-world, unconstrained usage by clinicians; by contrast with ***post-coordination***, complex concepts of differing levels of detail are composed from quasi-independent axes that contain more fundamental concepts (called ‘‘atomic’’ or ‘‘kernel’’ concepts).

The studies by Chute et al., Campbell et al., and Humphreys et al. all demonstrated that existing terminologies allowing post-coordination were better able to represent phrases and concepts extracted from clinical documents than existing pre-coordinated terminologies. Because users of such terminologies can both access existing concepts and dynamically compose new concepts as needed, such terminologies would be expected to have greater domain coverage than those that only allow users to access existing concepts. However, investigators have demonstrated three limitations of post-coordination in clinical terminologies:

(1) difficulty in restricting composition to medically meaningful concepts;
(2) create unrecognized duplicate concept representations;
(3) inefficiency with respect to composing complex concepts from simpler concepts.

Rector et al. and Rassinoux et al. *[Rassinoux 1997]* have separately pointed out that post-coordination may be used to generate meaningless concepts by combining two or more meaningful concepts. For example, a user could combine the concepts ‘‘chest’’ and ‘‘pain’’ and then add the concepts ‘‘radiating to’’ and ‘‘ankle’’ to create the composite ‘‘chest pain radiating to the ankle,'' which makes little sense clinically. Working together, Horrocks and Rector proposed a solution in 1995 called sanctioning.

Sanctioning allows developers to create rules for potential concept combinations requiring the composition to be medically sensible. Horrocks added that sanctioning can be permissive (i.e., no compositions allowed unless permitted) or restrictive (i.e., all compositions allowed unless restricted). In all cases, rules for sanctioning must be added on a concept-by-concept basis, increasing the effort required for terminology development.

Concept duplication occurs when a single concept is represented more than one time in a terminology or when there are multiple independent ways to use the terminology to represent a concept. Concept duplication can reduce the accuracy of information retrieval if it goes unrecognized.

In a usability study, McKnight et al. explored the competing tensions between greater content coverage achieved by terminologies that allow post-coordination and the enhanced ease of use resulting from pre-coordination. The study suggests that composing complex concepts from simpler concepts as part of standard documentation processes may be inefficient for the general practice of most health care providers.

Rector has suggested that tension between clinical usability and meticulous knowledge representation may result from a fundamental conflict between the needs of humans and those of computer programs that use terminologies. According to Rector's view, human users require flexible, expressive terminologies that model common colloquial phrases, while computer programs are generally designed to process formally defined concepts having rigidly defined interrelationships. This echoes the statement by Rassinoux and colleagues that pre-coordination and post-coordination may serve complementary roles.

Interface terminology usability correlates with the presence of attributes that enhance efficiency of term selection and composition:

(1) presence of relevant assertional medical knowledge;
(2) adequacy of synonymy;
(3) a balance between pre-coordination and post-coordination;
(4) mapping to terminologies having formal concept representations.

Improving and evolving interface terminologies require evaluation metrics such as adequacy of attributes, degree of synonym coverage, quantity and quality of relevant assertional knowledge and degree of compositional balance.

Summarized, let's consider a reference semantic terminology model with formal description of all concepts and relations that is available for software vendors. Graphic user interfaces can be generated in a diversity of "lay out" without jeopardizing the meaning of the concepts. Each vendor can propose their own ™ user interface according to the "look and feel" of their application, in harmony with the other screens, which is a step towards a better usability for the end-users.
In figure 6 a template from 2 fictive software's illustrates how interface terminology build upon a reference terminology model can represent semantically identical data in a distinctive way.

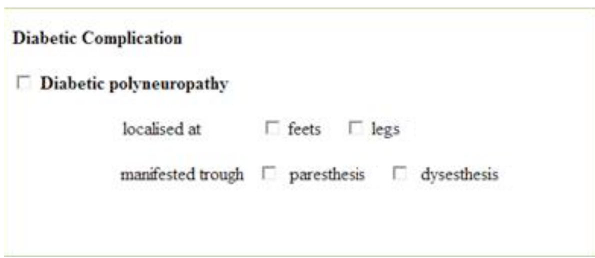
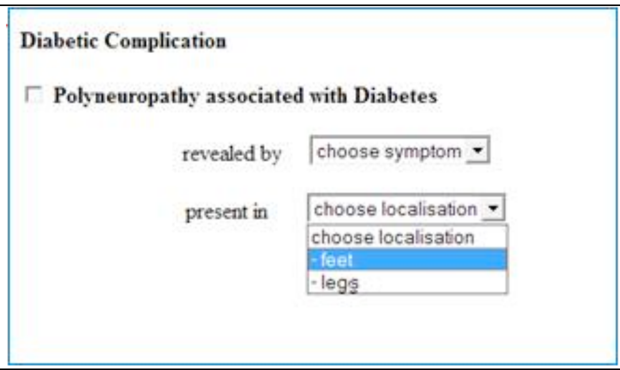
Software A	Software B
	

Figure 6

On the other hand, once a semantically formal terminology is acquired, it can be used to facilitate the Natural Language Processing approach. The success of this approach depends on the richness of the knowledge base the Natural Language Processing tools can rely on. If there is insufficient knowledge background, including contexts, the system will produce long list of terms from where the end-user must pick up the most adequate.
The problem is the exhaustively of the medical concepts and the importance to obtain the correct granularity when decision support will be invoked. Also the production of extended lists is against the good practice rules for usability of computer applications.
The NLP Module should add the natural language processing rules for usability facilitation.

Distributed Collaborative authoring

For all the 3 phases we must apply best practice knowledge engineering methodologies. These methodologies are based on roles and workflows with “build in” control mechanisms in order to allow a maximum flexibility while ensuring the coherence and quality of the vocabularies and terminologies to be developed

Specific information technology tools must support following roles and processes.

End-User

- Profile: healthcare actors
- Process: Every (end-) user can submit a new category, relation, concept, term or translation. The new entity or translation gets a candidate status until validated by the Approval Committee.

Professional experts

- Profile: Professional experts are care professionals, eg physicians, nurses, with expertise in their specific field, cardiologists, orthopaedic surgeons, nurses, physiotherapists, general practitioners, propaedeutic specialists…
- Process: They assist the terminologists in the selection and validation of concepts and terms submitted in the fields of their expertise.

Terminologist

- Profile: The terminologists are physicians trained to the methodology and the software tool for vocabulary and terminology acquisition and management. As a physician he/she has a good knowledge of medical science domain. Besides this knowledge he/she should have expertise in data management and terminology. A chief terminologist, experienced, leads the team.
- Process: The terminologists are validating the concept and term submissions. A chief Terminologist makes the final approval for quality assurance.

  In phase 2 the terminologists are assisting the knowledge engineer in defining the semantic model (ontology) underlying the Reference Terminology. They are analyzing the change requests affecting the ontology level: categories and relations and validate the change proposal. Together with the knowledge engineer they are designing the mapping schemes to external vocabularies and classifications.

Linguists

- Profile: Linguists are specialists in linguistics. These specialists of the "words" are the human communication experts, experts in natural language.
- Process: Linguists will collaborate in the phase 3 when NLP algorithms must be added to the Reference Terminology.

Knowledge engineer

- Profile: Knowledge engineers are software experts specialised in knowledge modelling and processing.
- Process: The knowledge engineer together with the terminologists will define the semantic model (ontology) underlying the Reference Terminology. They are analyzing the change requests affecting ontology level: categories and relations and validate the change proposal. Together with the terminologists, they design the mapping schemes to external vocabularies and classifications.

Approval committee and procedure

- Profile: The approval Committee is composed by 3 terminologists and 3 knowledge engineers.
- Process: Validation procedure must apply through independent vote. At least 2 terminologists and 2 knowledge engineers should be aligned in order to accept the new or the change request for categories and relations. They validate the versioning of the SRM.

Application developers

- Profile: software architect and software developer.

- Process: the software architect gathers the requirements of the users and professional experts, analyses them with the terminologists, designs the solutions with the knowledge engineer and models the solution to be implemented by the developer. He is managing the development, testing and deployment.

Workflow

Workflow at set up

At setup the set of 15000 concepts will be considered at an individual basis, concept per concept. The actions at terminology management level are described later in the chapter “Roadmap”.
Terminology management is composed of several disciplines for different functions. In the first phase, the quality control of the merge result, terminologist and knowledge engineer will check at the semantic level the validity of the concept and the expressions. Also professional experts will validate the concepts at knowledge level, meaning scientific coherence of the concepts. The linguist will control the lexical aspects. In the second phase, terminologists and knowledge engineer will work on the modelling of the terminology reference model or participate in the studies for the acquisition of a terminology (model).
An approval committee must validate the final proposal before a concept as “data set” is published to the terminology server and distributed. In case of rejection, the concept and the denial motivation will be review by the terminology management team in order to assess the motivation and when applicable to propose a corrected version to the approval committee.

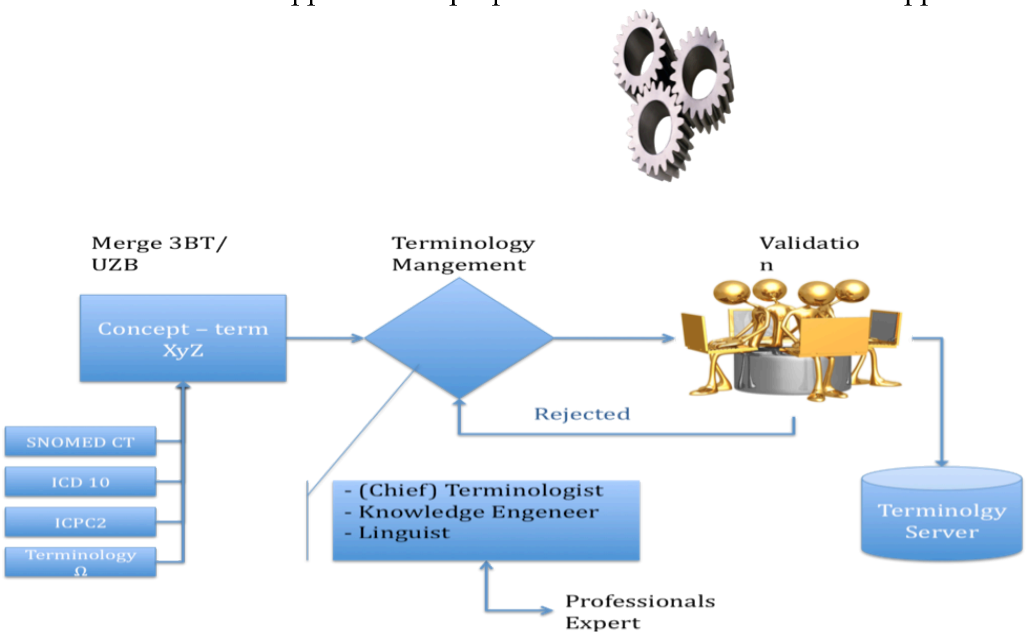


Figure 7.

Workflow in maintenance mode

The workflow as describe above is valid for new proposals of concepts, terms, synonyms and preferred terms but also for suggestion of modification, update or notification of error.

This workflow opens the possibilities to send requests or remarks to all users of software in the healthcare sector.
The proposal is treated by several people and can be rejected, corrected or accepted. Validation happens at each node, a final validation with feed back is present.

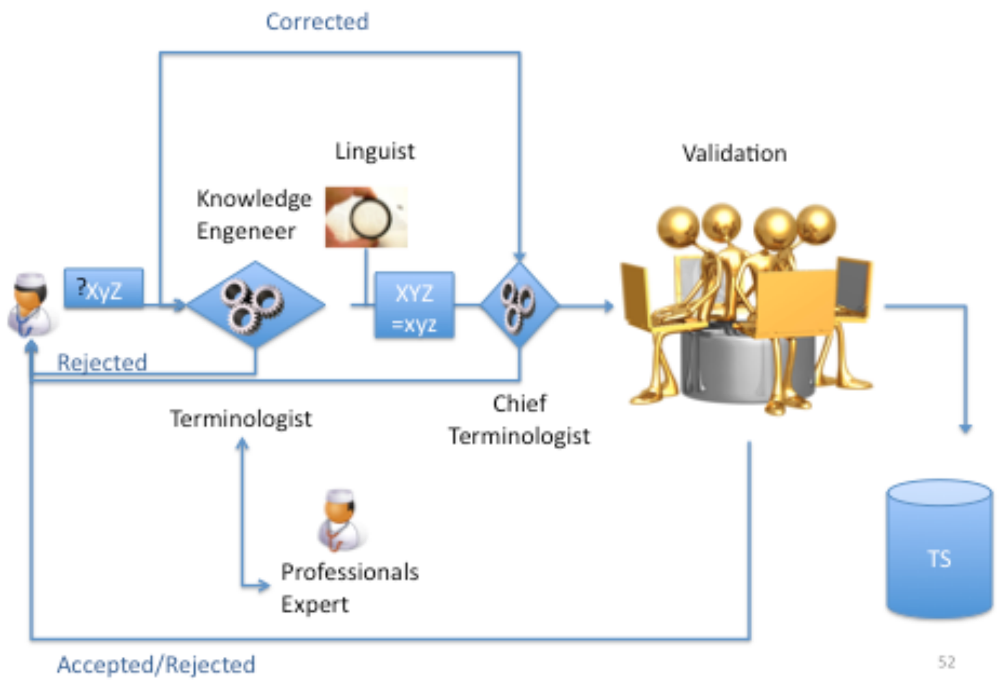


Figure 8. The figure illustrates the workflow of a use case in maintenance mode when an end-user presents a term for integration in the Belgian terminology server.

Distribution

All the users are informed through a portal of the requests and can participate at pools proposed by the analysts in order to understand the community preferences.

Tools for the management of terminologies

In order to support the work of the end-users and all the people concerned by the terminology management processes, information dedicated tools are mandatory.

The final Terminology Management System must allow:

- Browsing
    - Search terms
    - View terms and all related items. Items are
        - hierarchy, categories and attributes.
        - concept name and language name, popular name, preferred name, scientific name including synonyms

        - small and a full definition
    - select different hierarchies, categories

- Querying
    - query builder interface based on the semantic reference model

- Multiple views Display
    - parallel display for mapping and translation
    - list of external identifier for each concept

- Acquisition

- Classification
    - list of categories;
    - concept categories which participate in specified attribute (relationship);
    - List of attributes and relationships for a given concept;
    - Enumeration of concepts corresponding to a specified attribute value;
    - List of concepts names by category and language and their associated metadata (relationships);
    - Composed Concept Management;
    - Represent concepts as coordinated terms or composite entries;
    - Reasonner appliance capbilities.

- Internal management:
    - Log and track changes
    - Back up
    - User management

Tool for phase 1 - Reference Vocabulary Management System

As we will use a 3 phase progressive development the first module will be the Reference Vocabulary Management System.

In the use case illustrate in figure 9, ICD 10 is the source terminology where it was browsed for insulin-dependent diabetes mellitus. In this case a flat list is presented, in other use case an hierachical list could be presented. On the right side the Belgian Reference Vocabulary present 3 concepts with preferred terms and synonyms. A fourth term is added in the preferred vocabulary with a synonym and a definition. According to the SKOS[8] semantic a relationship toward other concept is defined. The concept of “insulin-dependent diabetes mellitus” is not considered as a new concept but linked to the internal term; it is considered broader than diabetes type 1 since also some type 2 diabetes patients are insulin-dependent.
The translation of the term is possible in French and English in the example.

[8] Simple Knowledge Organization System ; http://www.w3.org/TR/skos-reference/#L895

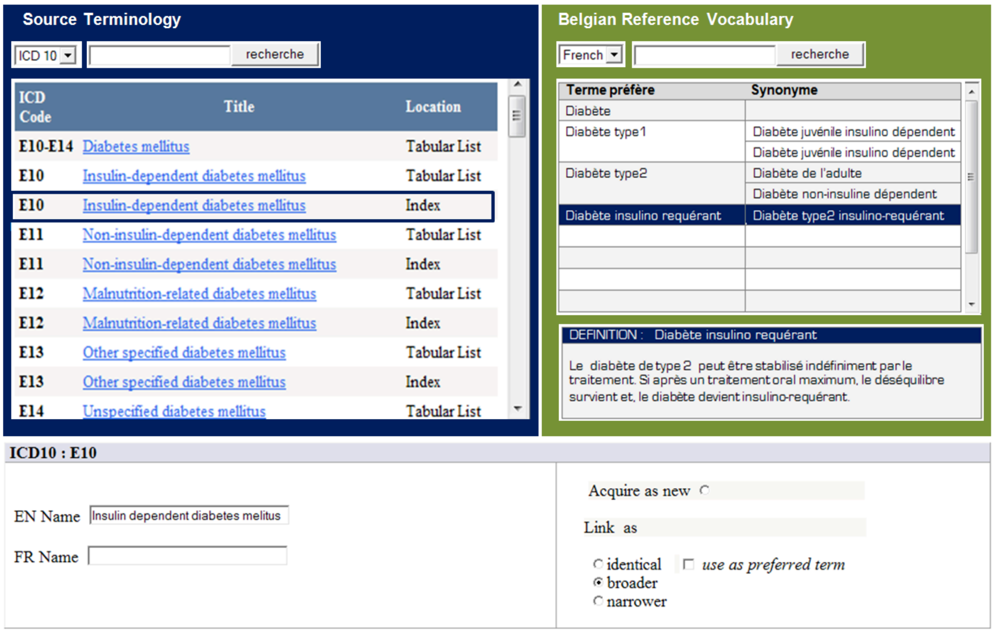


Figure 9. This is a simplistic illustration of a Vocabulary Management System.

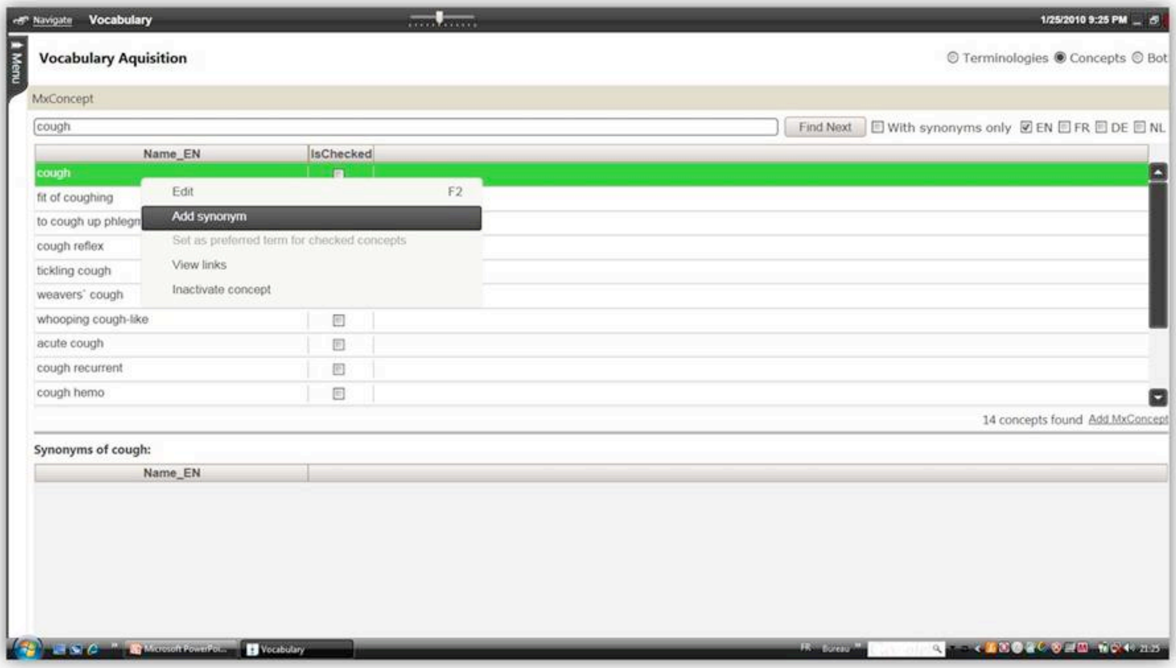


Figure 10. Another illustration of a vocabulary management system.

Examples of tools for Vocabulary Management

Microsoft: Excel
Apelon: Distributed Terminology System ; http://www.apelon.com/Products/DTS/tabid/97/Default.aspx
Medicognos www.medicognos.com
TopQuadrant: TopBraid Enterprise Vocabulary Net; http://www.topquadrant.com/solutions/ent_vocab_net.html

Tool for Phase 2 – the Reference Terminology

In the Phase 2 a formal ontology: classes, relations, domain and range constraints, semantic definitions and rules. A semantic language must be chosen, OWL DL is presented. The Vocabulary Management System will be integrated into a Terminology Management System. In the Terminology Management System we will

- identify composed concepts, decompose these in primitive concepts and properly declare the concepts through a semantic formula. These formulas are described in the compositional grammar. The declaration of the formula should be OWL compliant. This is a crucial moment in the development of the terminology server. We know SNOMED CT did the exercise but because of inconsistencies in the relations, 30% of their composed concepts cannot be exploited by computer applications.
- classify properly and define the inferences in the set of managed concepts;
- create/use semantic mapping rules with other implemented terminologies

The terms of the Reference Vocabulary will be classified under the corresponding classes, semantically defined and then automatically classified by the inference engines (reasoners).

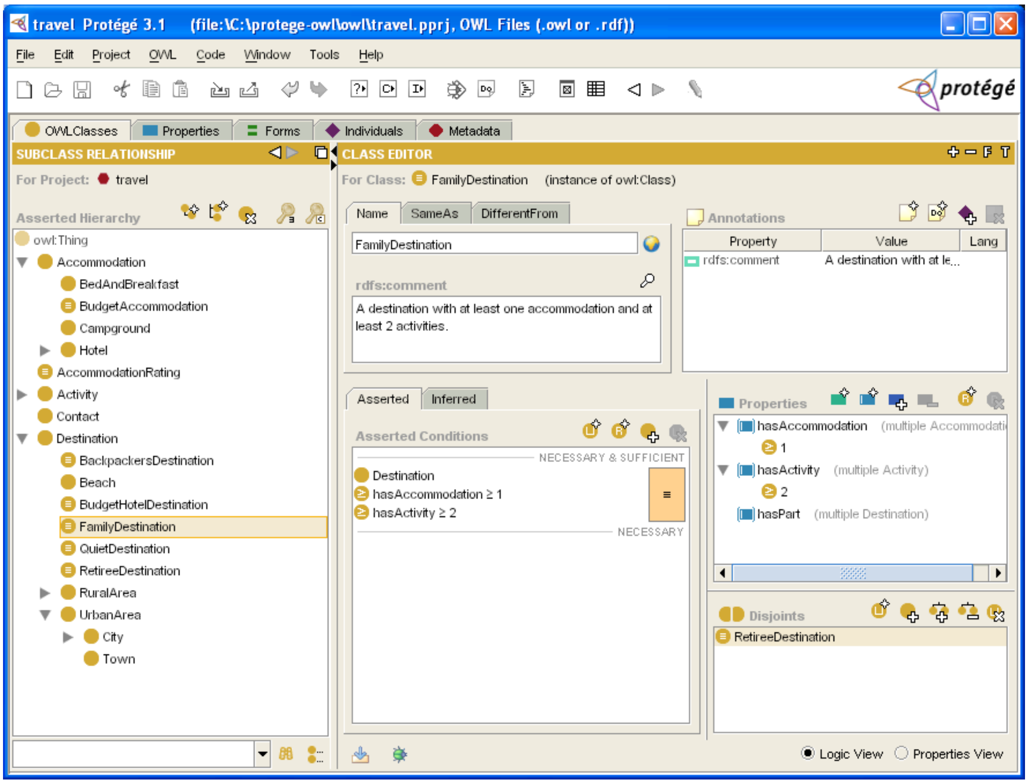


Figure 11A

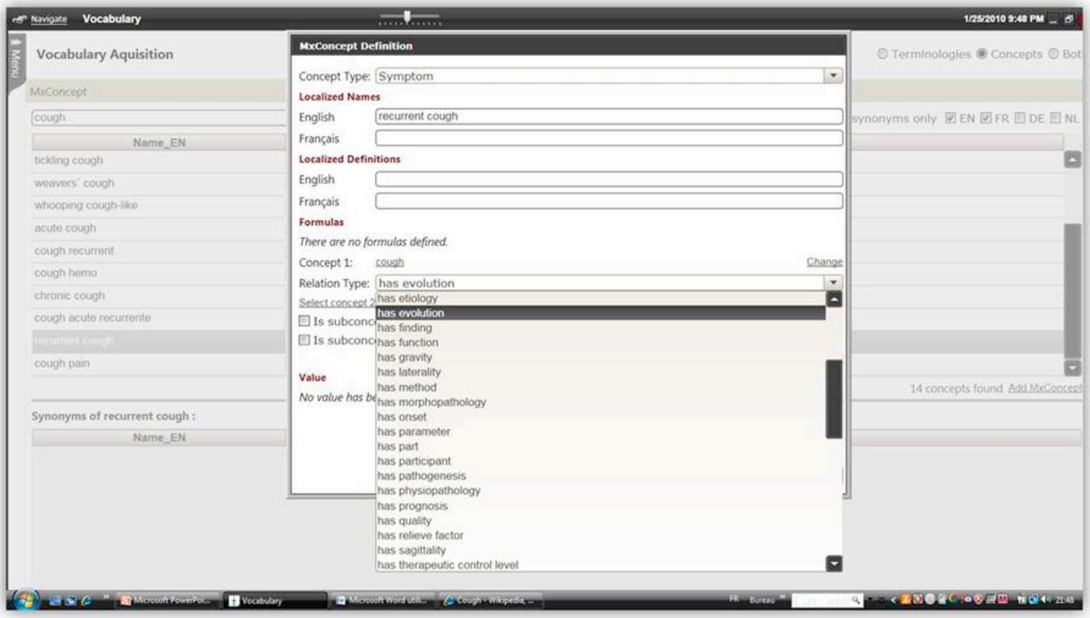


Figure 11B. The screenshot illustrates the use of a semantic formula to define a symptom, cough. In the case recurrent cough means cough has evolution recurrency.

Examples of tools for Terminology Management

Apelon: Distributed Terminology System ;
http://www.apelon.com/Products/DTS/tabid/97/Default.aspx
Carecom: http://www.carecom.dk
Clinclue: http://www.cliniclue.com
Collibra: www.collibra.com
International Institute for the Safety of Medicines (ii4sm): http://www.ii4sm.com
Medicognos www.medicognos.com
Mondeca : http://www.mondeca.com/index.php/en
SNOMED CT workbench:
Protégé: http://protege.stanford.edu/
TopQuadrant: TopBraid Suite; http://www.topquadrant.com/products/TB_Suite.html

Tool for Phase 3 – The Interface Terminology

In the Natural Language Generation approach, the software vendors are free to use their interface. The only but mandatory condition is the respect of the reference terminology and the standards of use.
For the Natural Language processing, several tools are available. Pragmatically the tool should support the Belgian national languages, Dutch, French and German. In that case Nuance is probably the first vendor to approach.

Representation of concepts.
Post coordination is Representation of a clinical idea in an Expression using a combination of two or more *Concepts* where Expression is a collection of references to one or more *Concepts* used to express an instance of a clinical idea.

Example of tools for Natural Language Processing

Medlee: http://www.nlpapplications.com/index.html
Nuance: http://www.nuance.com/for-healthcare/index.htm

Chapter 6 - Roadmap

Terminology Server

For obvious efficiency reasons a national initiative should be prioritized.

The final aim is to achieve a terminological service capable to guarantee the semantic interoperability in the healthcare and public health domain.
Since we have to take into account that such a model with a minimal reference terminology set is not available today, an alternative is proposed as step to. In order to reach tangible results in the shortest possible time while minimizing the cost and the risks we should start from the Merge result 3BT and UZBrussels CMV - a set of circa 15000 concepts - and augment stepwise the quality and enrich progressively according to the demands and the needs from the users.

Why start from the Merge 3BT-UZB?

- the data set is limited offering a manageable start;
- parts are used by some end-users today and is not abstract for the end-users;
- 3BT is integrated in some medical applications;
- it is a free of charge hybrid terminology avoiding all issues on intellectual property and copyright as well as dependencies and vendor lock in;
- it is cross-mapped with other terminologies and with a controlled medical vocabulary validated by care actors in a hospital setting.

Nevertheless the effort will be consequent. We have to keep in mind that the review of the existing data set must be done at different level:

- conceptual,
- linguistic with a vocabulary - meaning words – scientific or popular and preferred (?) in French and Dutch
- the mapping with operational terminologies.

The execution according to the guidelines and recommendations for a solid terminology, at conceptual and linguistic level and the management of all these jobs force us to limit the number of concepts to treat during the set up period of the Terminology Centre. Therefore a limitation to 15.000 concepts – the onomasiological use case – en 5000 words – the semasiological use case – has reach a consensus within the group of experts. This “ref set” or reference set is a start with reachable objectives. Of course, once the learning curve is behind and the adapted information technology tools are available, the terminology set will expand gradually.

As we stated before a 3 phases approach could fulfil this strategy.
The first phase is the quality check and cleaning phase in order to provide a consolidated terminology set, even flat files are produced.
In the second phase we will define the relations between the concepts and classify them in disjoint taxonomies to obtain a reference terminology. Ideally both phases should be realized in parallel.

In the third phase software vendors or the Belgian authorities will develop user-friendly interface terminologies based on the semantic of the reference terminology or provide NLP tools.

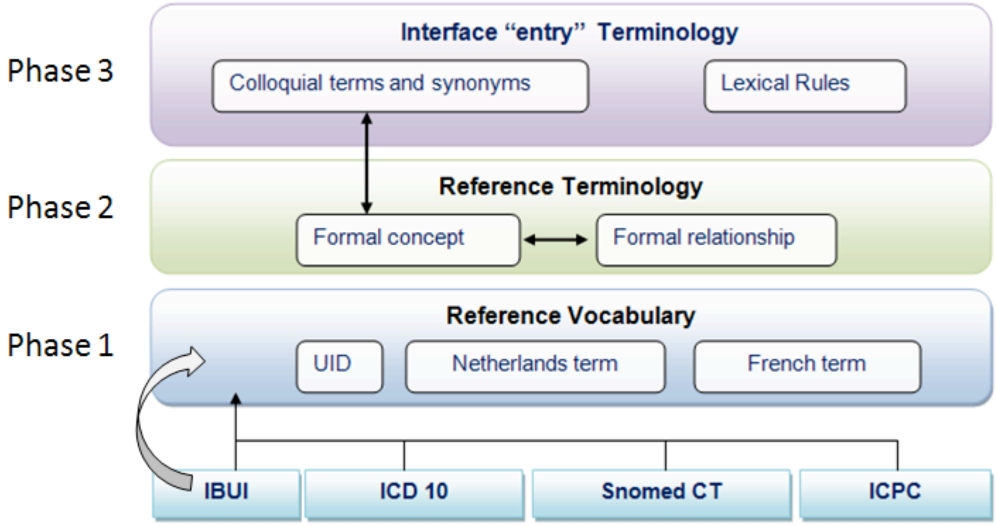


Figure 12. Adapted Conceptual model. It illustrate the suggested workflow, phase 1 consolidation and validation of the reference set, phase 3 the building of the reference terminology which is an ontology and phase 3 the user interface.

A parallel engagement for the first 2 phases and even the 3rd phases is possible. It could be reasonable to focus on quick wins and concentrate the effort in the consolidation of the existing tables and consolidated the concepts.

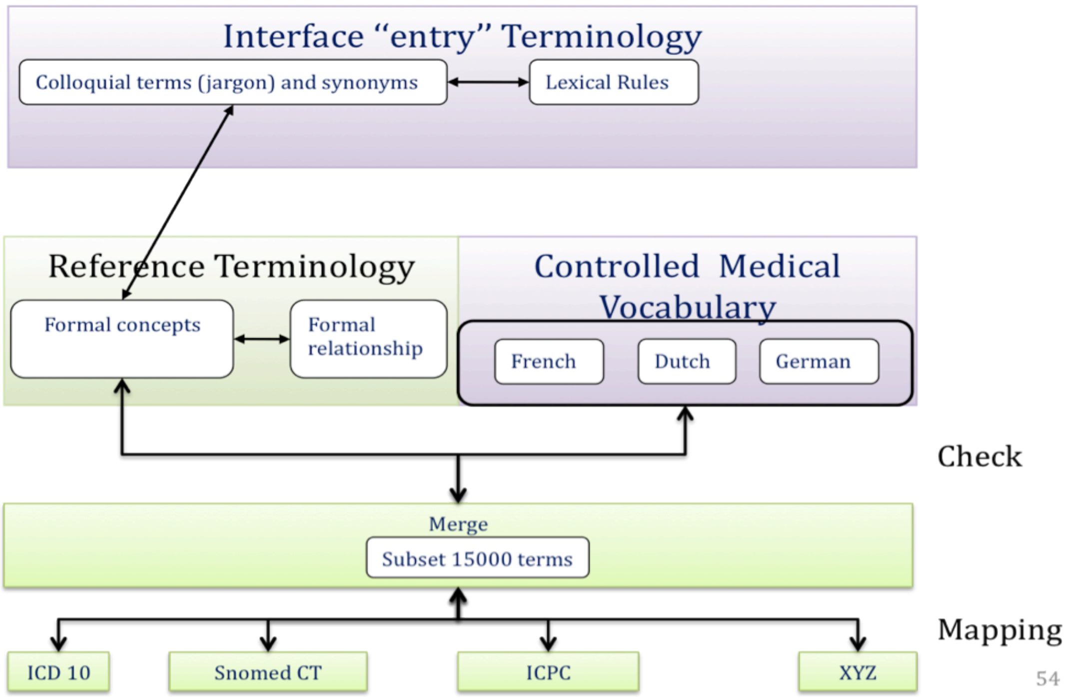


Figure 13. Adapted Conceptual Model with a parallel approach of consolidation and validation of the merge result and the building of a sound reference terminology.

This architecture must provide a progressive implementation starting from the existing vocabularies, classifications and terminologies.

The Phases

Phase 1

This is the phase where the merge result will be controlled, corrected and finally validated. This step is a "sine qua non" passage to the reference terminology.

The review exercise must use a rigourous methodology. The concepts has to be reviewed one by one, in a line per line fashion in order to:

- normalize the concept definition, description that should be unambiguous per concept and standardized
- identify and correct/exclude identical or "twin" concepts in order to respect the unicity principle, search engines (queries) must help the identification of the "twins"
- manage vocabulary acquisition and validate the preferred terms as used at the VUB
- manage synonymy
- manage the vocabulary terms in all languages supported with respect to normalisation, completion of the single and the composed terms and the term spelling
- manage homonymy
- validate the mapping with other terminologies (SNOMED CT, ICD 10, ICD 9 CM, ICPC2,…)

A Reference Vocabulary Management System as described earlier is used to extract from the merge a core set of terms into the Reference Vocabulary. In the same time the reference terms are mapped with standard classifications. (see example for ICD 10 in the fig 9)

For the extension of the content of the core set we will methodologically select the terms largely used and those identified as having the fastest and most significant Return On Investment:

- Largely used (high sensitivity):
  - medical data with administrative emphasis such as gender
  - primary care

- Niche (high specificity):
  - Where multi-disciplinarily is already formalised, eg Medication and prescription, Oncology, Care pathway DM and chronic kidney failure, duty reports
  - Drug information, MMP, a subset of 1830 terms
  - Existing registries (eCare suite, Vaccinet,…) (see also part 2 of the deliverable)
  - epSOS
  - A call for proposal can be launched or a list of proposal be proposed. Each initiative should be objectively evaluated using rules described in a procedure.

- Evaluation and inventory of potential sources to use for the build of the Belgian Reference Identification of missing terms and term validation workflow procedure.
    - Sources are LOCAS, ICPC, SNOMED, WHO-ART,
    - proprietary dictionaries of Electronic Medical Records (Sosoeme, Corilus suite and Health One),
    - vocabularies from non medical professions, nurses (ICPN, NIC, NOC, NANDA), physiotherapists and other paramedical professions.

- Evaluation and validation of the Mapping requirements

    - Main internal sources to map with are:

        - IBUI/3BT
        - INAMI/RIZIV nomenclature
        - Current ICD uses for DRG's, CareNet
        - Albert II

    - Main external Sources to map with are:
        - SNOMED CT
        - ICD10
        - ICD O
        - ICD 9 CM
        - ICPC2
        - LOINC
        - epSOS catalog
        - NIC/NOC/NANDA

The methodology of enrichment of the data set at operational level is described in the workflow.
The Reference Vocabulary Management System will be used and presented to the Information Systems provider to align their vocabularies with the Reference Vocabulary.

For mapping the similar but not identical terms we propose the SKOS semantic: "*broader as", "narrower as", "identical"* witch allows to palliate the concept granularity problem through a parent-child (taxonomical) classification.
This is a first step in classification/hierarchisation of concepts and forces the terminologists to define a rough hierarchy. As a result of this phase a higher level of coherence in the dataset of concepts with vocabularies is reached. It prepares the second phase of the tasks, namely the integration of the concepts and the terms into a workbench for classification / hierarchisation, relationship definition and description, …

Phase 2

In the Phase 2 a formal ontology will be defined: classes, relations, domain and range constraints, semantic definitions and rules. For example SNOMED CT has 18 categories. In each category attributes are defined.

The use the OWL DL language is proposed.
In this phase, the content of the Vocabulary Management System will be integrated into a Terminology Management System.

Method:

- Identify composed concepts, decompose these in primitive concepts and properly declared through a semantic formula. These formulas are described in the compositional grammar. The declaration of the formula should be OWL compliant. We know SNOMED CT did the exercise but because of inconsistencies in the relations, 30% of their composed concepts cannot be exploited by computer applications
- classify properly and define the inferences in the set of managed concepts
- semantic mapping rules with other implemented terminologies

The terms of the Reference Vocabulary will be classified under the corresponding classes, semantically defined and then automatically classified by the inference engines (reasoners).

Once a semantically formal terminology is acquired, it can be used to facilitate the Natural Language Support approach. The success of this approach depends on the richness of the knowledge base the Natural Language Processing tools can rely on. If there is insufficient knowledge background, including contexts, the system will produce long list of terms from where the end user must pick up the most adequate.
The problem is the exhaustivity of the medical concepts, the explosion of composed concepts. The problem will and the importance to obtain the correct granularity when decision support will be invoked. Also the production of extended lists is against the good practice rules for usability of computer applications.

As we mentioned earlier, phase 1 and 2 should not happen in sequential order. The second phase can be initiated before completion of the first one. Ideally the terminology model and the methodology be defined in advance in order to avoid step backs and work repetition.

<u>Phase 3</u>

The way of realisation of the third phase should be studied since several approaches are possible.
However, providers of software must respect the reference terminology and the semantic when they build their user terminology. They can use templates or other graphic supportive tools.
The best solution is the post composition since it is not possible to present or inject all possible expression in the GUI.

Organisational aspects: To Do's

- Creation of the Terminology Centre, virtual or real: a visit in existing centres for terminology in Europe. Examples are UK (http://www.connectingforhealth.nhs.uk/), Sweden, Denmark,… NICTIZ, the Netherlands, has been mandated to manage studies on and to implement SNOMED-CT in 3 uses cases (see part 2 of the deliverables).

- Look up in the industry world, at companies with a sound experience in agile dynamic data processing based on knowledge could be a source for inspiration. The organisation of workshops with experts from companies such as "Collibra SA/NV" (www.collibracom) or "i.know" (www.iknow.be), CETIC (www.cetic.be) among others, will be fruitful. Even more, a partnership with one of them could be considered in order to boost internal knowledge and enhance the processes of creation of the terminology centre.

- Structure and people
    - Organigram, : profile definition, roles and prerogatives of the centre and the members
    - Statute and legal entity
    - Positioning of the terminology server towards the different healthcare institution of the Belgian authorities: Ministry of Health, INAMI/RIZIV, eHealth platform,…
    - Identification of experts in the professionals care fields and data management and identification of working groups and work modality
    - International links and representation of Belgium:
        - Representation of Belgian Terminology Reference Center
        - epSOS compatibility
        - liaison with WICC
        - liaison with SNOMED
    - Education and training

- Set up of sectoral working groups and identification of validated information sources
    - Hospitals (FOD/SPF, RIZIV-INAMI)
        - Terms diagnoses, procedures
    - Primary Care (WIV/ISP, RIZIV-INAMI)
        - Terms diagnoses, procedures, observations
    - Nursing (Nl: NVKVV (ISV), Fr: SIXI)
        - Terms diagnoses, ADL, social factors, procedures, observations
        - Needed access to data EMD in function of type of nurse and type of care element
        - Guidelines orders in function of care element
    - Physiotherapy
    - Medication, devices and medical products
    - Anatomopathology, cancer
    - Laboratory
        - Loinc
    - Accidents, safety (FOD/SPF)

- Market study of product for terminology management:
    - Actions are:
        - Definition of Mandatory requirements
        - Definition of optional requirements
        - Identify issues to be discussed (governance model, legal aspects,…)
        - Define Evaluation criteria
        - Methodology of selection procedure
        - Define Normative references and relationship to standards (non exhaustive list)
            - TMF for file exchange
            - W3C SKOS,
            - OWL 2 semantic profiles, (using OWL DL)
            - Object Management Group Model Driven Architecture,
            - Ontology Definition Metamodel, Meta-Object Facility, Unified Modelling Language, OCL
            - …
    - Inventorisation of Tools:
      Tools for authoring of terminologies, vocabularies and biomedical concepts should be privileged above others. Of course a journey through all the tools provides insight at managerial level and helps in the writing of the strategy and roadmap.

        - ReTam (https://retamacc.smals-mvm.be/retam/); developed by SMALS.
        - Apelon**:** http://www.apelon.com/Products/tabid/62/Default.aspx
        - Carecom: http://www.carecom.dk
        - Cliniclue: http://www.cliniclue.com
        - International Institute for the Safety of Medicines (ii4sm): http://www.ii4sm.com
        - Medicognos www.medicognos.com
        - Medlee: http://www.nlpapplications.com/index.html
        - Mondeca : http://www.mondeca.com/index.php/en
        - Nuance: http://www.nuance.com/for-healthcare/index.htm
        - Protégé: http://protege.stanford.edu
        - SNOMED workbench (made by Apelon)
        - TopQuadrant: http://www.topquadrant.com/index.html

For each case an ROI should be calculated.

- Selection of test fields for the validation of terminology services. Suggested during Seminop meetings were: GP EHR, medical problem list in Hospitals, drug prescription and reimbursement forms, duty mailer for GP's, code finder in soft for DRG coding.

Epilogue

The document is still a draft and must be considered as “pierre à casser”.

Many aspects must be completed. Sometimes the market is still immature and the document will complete the argumentary by providing insights in complicated matter and emancipate the stakeholders. For other aspects such as the problem addressed by the reference information model and the reference terminology model, the current status of understandings and realisation reaches R&D level, for other items the dream level is the only one achieved yet. Another objective of this document is to illustrate possibilities, to identify paths leading to the setup of a terminology server in Belgium. The suggested path in 3 phases is pragmatic and achievable.

ANNEX

Annex 1

The Fundamental Requirements of Medical Terminologies

(adapted after J. Cimino , A. Rector, C. Chute , P.Elkin , D.Markwell)

Concept Orientation	each concept (in the vocabulary) has a single meaning
Concept Permanence	the meaning of a concept, once created, is inviolate. Its preferred name may evolve, and it may be flagged inactive but its meaning must remain.
Nonsemantic Concept Identifier	unique identifiers for the concepts which are free of hierarchical or other implicit meaning semantic connotations Problems with hierarchical identifiers : once assigned a code, a concept can never be reclassified without breaking the hierarchical coding scheme and if a concept belongs in more than one location in the hierarchy a single hierarchical identifier is no longer possible
Polyhierarchy	There seems to be almost universal agreement that controlled medical vocabularies should have hierarchical arrangements. This is helpful for locating concepts through "tree walking"), grouping similar concepts, and inferring meaning
Formal Definitions	Computable definitions as opposed to narrative text variety, such as those found in a dictionary Usually, these definitions are expressed as some collection of relationships to other concepts in the vocabulary. For example, the concept "Pneumococcal Pneumonia" can be defined with a hierarchical ("is a") link to the concept "Pneumonia" and a "caused by" link to the concept "Streptococcus pneumoniae". If "Pneumonia" has a "site" relationship with the concept "Lung", then "Pneumococcal Pneumonia" will inherit this relationship as well. This information can be expressed in a number of ways, including frame-based and description logic semantic networks.
Domain Coverage	Is the degree to which a terminology contains all the relevant concepts of its domain of interest? To enlarge their domain coverage, terminologies are using two pre-coordination (also called enumeration) and post-coordination.
Pre-coordination	Is a method for creating new concepts. Developers model relevant levels of detail in the terminology with distinct concepts, typically derived from real-world, unconstrained usage by clinicians; ‘‘chest pain,’’ ‘‘substernal chest pain,’’ and ‘‘crushing substernal chest pain’’ each exist as individual concepts, with unique terms and synonyms
Post-coordination	Is a method for creating new concepts complex concepts of differing levels of detail are composed from quasi-independent axes that contain

	more atomic concepts For example, with a post-coordinated terminology, a user can dynamically create the concept ‘‘chest pain’’ by combining the anatomic concept ‘‘chest’’ and the semiologic concept ‘‘pain.’’ The user can introduce further detail by selecting new concepts from additional axes, such as ‘‘substernal’’ and ‘‘crushing nature.’’ Although studies demonstrated that existing terminologies allowing post-coordination were better able to represent phrases and concepts extracted from clinical documents than existing pre-coordinated terminologies they have also intrinsic limitations like (1) difficulty in restricting composition to medically meaningful concepts, (2) creating unrecognized duplicate concept representations; Actually these limitations are to a large extent surmountable (numerous investigators developed methods that address them by standard logical formalisms).
Recognize Redundancy Concept Duplication	Redundancy is the condition in which the same information can be stated in several different ways. Synonymy (term level redundancy) is a type of redundancy which is desirable. By contrast the code level redundancy must be avoided. As such redundancy is inevitable in distributed authoring systems allowing term composition the terminology must provide a mechanism by which it can recognize redundancy and render it transparent . While duplication may occur in terminologies regardless of whether they permit post-coordination, post-coordination enables a greater opportunity for duplication to occur. In SNOMED the concept for appendicitis can de modeled through at least four paths. A computer program trying to identify cases of appendicitis from a SNOMED–encoded data set would need to search for all possible ways that the concept could have been expressed. Although SNOMED is using algorithms (description logical) and equivalency tables to solve the problem, the risk for duplication is still difficult to asses.
Meaningful composition	Post-coordination can generate clinically meaningless concepts by combining two or more meaningful concepts. For example, a user could combine the concepts ‘chest’’ and ‘‘pain’’ and then add the concepts ‘‘radiating to’’ and ‘‘ankle’’ to create the composite ‘‘chest pain radiating to the ankle,’’ Build in rules and algorithms can manage this problem to a reasonable extent.
Multiple Granularities	Depending on the purpose a terminology usually implicitly, there is always a preconception of the level of granularity at which the concepts must be expressed. For example, the concepts associated with a diabetic patient might be (with increasingly finer granularity): "Diabetes Mellitus", "Type II Diabetes Mellitus", and "Insulin-Dependent Type II Diabetes Mellitus" As there always be a debate over exactly what constitutes an atomic component, e.g. whether Colon Cancer should be decomposed into the atomic elements of Colon and Cancer, or itself constitutes an atomic notion the TS should support multiple levels of granularity for term composition As the real life terminological needs of the user are variable with respect to the granularity of the terminology we need to introduce a way allow a flexible choice from different level of deepness of the hierarchies. The tests in usability laboratory are showing that users find very helpful the display of terms from the semantic neighbourhood (parent, child and sibling terms) helpful, and will often select such recommended terms.

Reject "Not Elsewhere Classified"	Since no vocabulary can guarantee domain completeness all of the time, it is tempting to include a catch-all term which can be used to encode information that is not represented by other existing terms. The problem with such terms is that they a definition based on exclusion - that is, the definition can only be based on knowledge of the rest of concepts in the vocabulary. Not only is this awkward, but as the vocabulary evolves the meaning of NEC concepts will change in uncontrollable ways.
Representing Context	One the hardest difficulty today with using standard controlled terminologies is that they are created independent of the specific contexts in which it is to be used. This leads to semantic gaps and overlapping when concepts are recorded in some specific context, for example, in an electronic patient record. The experience of “terminology binding” with patient record information model (HL 7, openEHR) imposed the evidence that the best way to avoid the semantic “chaos” (A Rector) is too align the two semantic models (Markwell Report for the NHS , IHTSDO Snomed openEHR Foundation 2010 announcement) That means for the terminology to contain also a formal context representation.
Terminology Usability	Studies have proven that user interfaces for composing complex concepts from simpler concepts may generate important usability problems in the daily practice (McKnight et al) Even rigorous terminology development methodologies (Cimino, Chute Elkin, Rector) et al. do not guaranteed to create terminologies that are easily and directly usable by health care providers during routine clinical tasks. The tension between clinical usability and meticulous knowledge representation results from a fundamental conflict between the needs of humans and those of computer programs that use terminologies. Human users require flexible, expressive terminologies that model common colloquial phrases, while computer programs need to process formally defined concepts. (A.Rector) Recognizing the need to balance terminology domain coverage with clinical usability, Spackman et al., 89 Rector,65 and Chute et al.9 have all suggested that terminology developers limit their scope from creating a single monolithic terminology that meets all users’ needs to building terminologies designed for at last 3 specific categories: 1. classifying clinical data for administrative purposes such as billing and public health 2. clinical concepts and their interrelationships for computer storage, retrieval, manipulation, and analysis. 3. supporting efficient documentation of clinical findings into medical records.

Annex 2

Terminology Binding Requirements and Principles

D. Markwell
NHS Connecting for Health, 2008

Practical principles of terminology binding:

- Understandability
- Reproducibility
- Usefulness
- Re-usability and common patterns
- Transform-ability and normal forms
- Tractability
- Practicability
- Scalability
- Limiting arbitrary variation
- Responsive participating standards

Annex 3

Markwell's report for NHS, Connecting for Health, in 2008. Terminology Binding Requirements and Principles

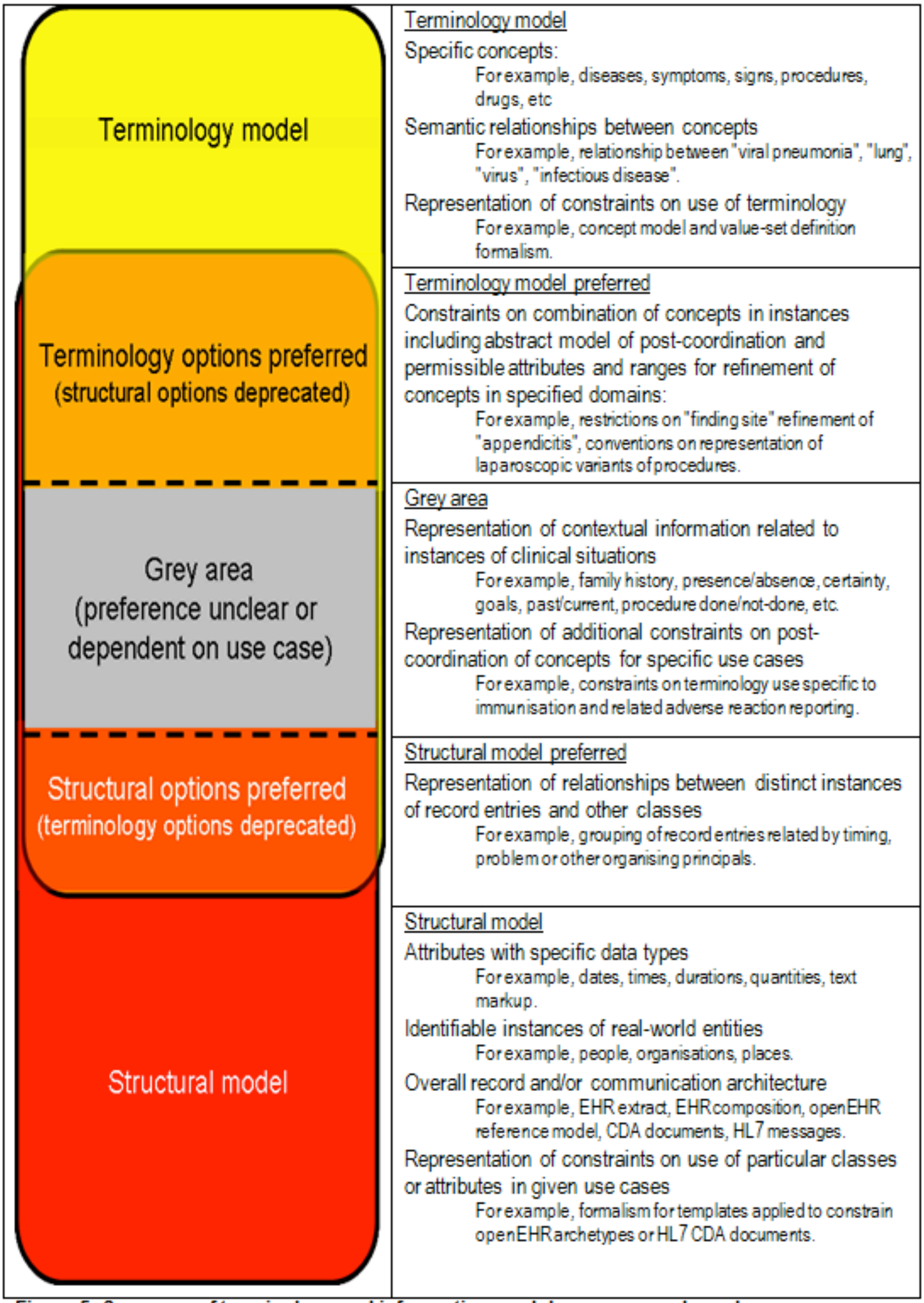


Figure 5. Summary of terminology and information model coverage and overlaps

Annex 4

Concepts

Over 300,000 active concepts with formal logic-based definitions are organized into top level hierarchies:

Clinical Finding
- Finding (*Swelling of arm*)
- Disease (*Pneumonia*)

Procedure (*Biopsy of lung*)
Observable entity (*Tumor stage*)
Body structure (*Structure of thyroid*)
- Morphologically abnormal structure (*Granuloma*)

Organism (*DNA virus*)
Substance (*Gastric acid*)
Pharmaceutical/biologic product (*Tamoxifen*)
Specimen (*Urine specimen*)

Qualifier value (*Right*)
Record artifact (*Death certificate*)
Physical object (*Suture needle*)
Physical force (*Friction*)
Events (*Flash flood*)
Environments/geographical locations (*Intensive care unit*)
Social context (*Organ donor*)
Context-dependent categories (*No nausea*)
Staging and scales (*Barthel index*)
Linkage concept
- Link assertion (*Has etiology*)
- Atributes (*Finding site*)

Special concept (*Inactive concept*)

Descriptions

Contains more than 1 million active English language descriptions for flexibility in expressing clinical concepts

Relationships

Over 900,000 defining relationships enable consistency of data retrieval and analysis

Attributes

Clinical Finding
Finding Site
Associated With
- After
- Causative Agent
- Due To

Associated Morphology
Severity
Onset
Course
Episodicity
Interprets
Has Interpretation
Pathological Process
Has Definitional Manifestation
Occurrence
Stage
Finding Method
Finding Informer

Body Structure
Laterality

Procedure
Procedure Site
- Direct
- Indirect

Procedure Device
- Direct Device
- Indirect Device
- Using
- Access Instrument

Procedure Morphology
- Direct Morphology
- Indirect Morphology

Method
Direct Substance
Access
Approach
Priority
Has Focus
Has Intent
Recipient Category
Revision Status
Route of administration

Measurement Procedure
Component
Measurement Method
Has Specimen
Time Aspect
Property
Scale Type

Context
Associated Finding
Associated Procedure
Finding Context
Procedure Context
Subject Relationship Context
Temporal Context

Specimen
Specimen Procedure
Specimen Source Topography
Specimen Source Morphology
Specimen Substance
Specimen Source Identity

Pharmaceutical / biologic product
Has Active Ingredient
Has Dose Form

Bibliography

[EC Report 2006] **Impact of ICT on Patient Safety and Risk Management in Healthcare** - 2006 Report commissioned by European Commission Directorate-General Information Society and Media, ICT for Health

[Dogac2007] Dogac A, Eichelberg M, Smith B, DeMoor G, Rossi Mori A, Mentzas G, Vitvar T, Wein B, Ceusters W. **EU funded roadmap project for interoperability of eHealth systems leading to recommendations for actions and to preparatory actions at the European level** (Action plan of the eHealth Communication COM 356). Ride, 2007.

[EC Report 2008] **Ontologies and Terminologies, Background, Findings and Recommendations** SemanticHealth Report - commissioned by European Commission Directorate-General Information Society and Media, ICT for Health

[Walker] 2005 **The Value Of Health Care Information Exchange And Interoperability** Jan Walker, Eric Pan, Douglas Johnston, Julia Adler-Milstein, David W. Bates, and Blackford Middleton MarketWatch2005

[Rector] **Clinical terminology: why is it so hard ?** Rector AL. *Methods Inf Med. 1999*

[Schulz2007] **SNOMED CT's Ontological Commitment** Stefan Schulz, Ronald Cornet **;** Nature Precedings 2009

[Markwell 2008] Terminology Binding Requirements and Principles 2008 Repport for NHR Connetcting for health D. Markwell

[Zhang 2010] **Large-scale, Exhaustive Lattice-based Structural Auditing of SNOMED CT** GQ Zhang , PhD and Olivier Bodenreider , AMIA Annual Symposium 2010

[Rosenbloom 2006] **Interface Terminologies: Facilitating Direct Entry of Clinical Data into Electronic Health Record Systems** S. Trent Rosenbloom, MD, MPH, Randolph A. Miller, MD, Kevin B. Johnson, MD, Peter L. Elkin, MD, and Steven H. Brown, MD

[Ball2003] **Leveraging IT to Improve Patient Safety**. Marion J. Ball, Ed.D, David E. Garets, Thomas J. Handler, M.D. International Medical Informatics Association (IMIA) Yearbook 2003

[HIMSS2007] **EMR Sophistication Correlates to Hospital Quality Data**. Comparing EMR Adoption to Care Outcomes, HIMSS 2007

[Crosson2007] **Electronic Medical Records and Diabetes Quality of Care: Results From a Sample of Family Medicine Practices**. Jesse C. Crosson, PhD1 et all ANNALS OF FAMILY MEDICINE MAY/JUNE 2007

[Johnston2004] **Patient Safety in the Physician's Office - Assessing the Value of Ambulatory CPOE**. Douglas Johnston, M.T. S. Eric Pan, M.D., M.Sc. Janice Wa l k e r, R.N., M.B.A. David W. Bates, M.D., M.Sc. Bl a c k f o rd Middleton, M.D., M.P.H., M.Sc. Center for Information Technology Leadership , April 2004

[Kaushal2006] **Return on Investment for a Computerized Physician Order Entry System**. Kaushal R, Jha AK, Franz C, Glaser J, Shetty KD, Jaggi T, Middleton B, Kuperman GJ, Khorasani R, Tanasijevic M, Bates DW. J Am Med Inform Assoc. 2006 Feb 24

[Teich2005] **Clinical Decision Support in Electronic Prescribing: Recommendations and an Action Plan Report of the Joint Clinical Decision Support Workgroup**. Teich J, Osheroff J, Pifer E, Sittig D, Jenders R. The CDS expert review panel; JAMIA 2005

[Niland2006] **An Informatics Blueprint for Healthcare Quality Information Systems** Running head: Healthcare quality informatics blueprint.

[Bates2003] **Ten Commandments for Effective Clinical Decision Support: Making the Practice of Evidence-based Medicine a Reality**. Bates D, Kuperman G, Wang S, Gandhi T, Kittler A, Volk L, Spurr C, Khorasani R, Tanasijevic M, Middleton B. JAMIA 2003.

epSOS; WP 3.5