An infrastructure for quality assessment in intensive care. Prognostics models and terminological systems

de Keizer, N.F.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 7

Understanding terminological systems I; Terminology and typology

N.F. de Keizer, A. Abu-Hanna, J.H.M. Zwetsloot-Schonk

Based on
Understanding Terminological Systems I; Terminology and Typology.
Accepted for publication in Methods of Information in Medicine.
Abstract
Terminological systems form an important research issue within the field of medical informatics. For the precise understanding of existing terminological systems a referential framework is needed that provides a uniform terminology and a typology of terminological systems themselves. In this chapter a uniform terminology is proposed by combining relevant fundamental notions and definitions used by standardization organizations such as CEN and ISO into perspective, and interrelating them in order to arrive at a useful typology of terminological systems. The resulting typology is illustrated by applying it to five general well-known existing medical terminological systems and one existing terminological system for intensive care.

7.1 Introduction
Clinicians are increasingly aware of the benefits of electronically stored medical data and the use of these data for many purposes, such as direct patient care, research, costs, management support and education. The use of medical data stored in computer-based patient records (CPRs) or Patient Data Management Systems (PDMSs) has increased the need for structured and controlled data entry and data representation [1]. The usual way to cope with this need is to enforce the use of standard terms derived from a dedicated medical terminological system. Consequently, terminological systems form a very important issue of research within the field of medical informatics. In this thesis the term “terminological system” is used as an all-embracing term for the notions “classification”, “thesaurus”, “vocabulary”, “nomenclature” and “ontology”, which are further described in section 7.3.

Before an existing terminological system can be used or a new system can be developed, good understanding of the terminological system is essential. In the context of the NICE project in which we aimed to develop a terminological system for intensive care diagnoses, we met considerable difficulty in the analysis of the literature on (existing) terminological systems. This is caused by a great heterogeneity and indistinctness in the terminology used to describe the terminological systems, and by an incomplete description of the structure and characteristics of the various systems.

Some efforts have already been made by organizations for standardization such as the International Standards Organization (ISO) and the Comité Européen de Normalisation (CEN), which have developed standards for medical informatics [2]. In 1990 the CEN, Technical Committee 251 (CEN/TC251) was set up for standardization in medical informatics. One of its working groups covers the activities on health care terminology, semantics and knowledge bases. Its prestandard “Medical informatics - Categorical structures of systems of concepts - Mr:del for representation of semantics” [3] contains the names and definitions of concepts in the field. This CEN prestandard uses many standards already defined by ISO [4, 5] and is intended to provide a comprehensive basis for systematic understanding, development and use of computer-based terminological systems in health care. The standards developed by these organizations are freely available, however, the support they provide for the understanding of the proposed terminological systems in order to assess their merits is minimal because they only enumerate dry definitions of notions. We feel an overall accessible framework for the understanding of terminological systems, in which definitions are placed in perspective and put in practice, is still lacking.
This chapter describes the first part of such a referential framework in which definitions have been arranged in coherence and put into perspective, resulting in a typology of terminological systems. To illustrate the first part of the referential framework, we apply it to five existing terminological systems with a general medical domain and one with an intensive care domain intended to record diagnoses for evaluative research. The second part of the new framework, described in chapter 8, provides a basis for a formal and thereby unambiguous description of the structure of a terminological system. In chapter 8 we also describe our positive experience with terminological system formalization in terms of (1) understanding existing terminological systems; (2) recognizing patterns in the structure of the systems and (3) developing a new terminological system (see chapter 9).

The goal of chapter 7 and 8 is to provide a referential framework, by which terminological systems can be characterized, understood and compared, and to describe our experience with the use of this framework. We hope that this framework provides assistance to those interested in understanding and applying terminological systems.

In section 7.2 the building blocks for conceptualization of terminological systems: object, concept, term, code, and the different types of relations between concepts, are described. The typology of terminological systems is described in section 7.3. In section 7.4 we applied the typology described in this framework to five general known medical terminological systems and one terminological system with an intensive care domain. Section 7.5 is the summary and discussion of this chapter.

7.2 Terminology and definitions

Before we discuss terminological systems, some basic elements have to be explained: objects, concepts and designations, which form the so-called semiotic triangle or meaning triangle (see Figure 7.1) [3, 6]. Reality can be conceived as consisting of objects (things), such as “heart valve” or more abstract things such as “pain”. We use characteristics of objects to form cognitive constructs, called concepts, which are units of thoughts. Although objects and concepts are different notions, in the rest of this chapter we only talk about concepts, because terminological systems mainly comprise of general concepts, e.g. diseases, that are used to describe instances of patient’s diseases (i.e. objects) recorded in patient’s records.

To be able to communicate about concepts we use linguistic labels, called terms, to designate a concept. It is complex to designate concepts fully and it can be overwhelming because there are many (informal) languages and one language can use many different terms for the same concept. Definitions are (for example formal) statements of the meaning of a concept and can be used to cope with this language problem and can be used to further clarify the meaning of concepts. Codes (letters, numerals or a combination thereof) can be used to designate concepts for their representation in a computerized system. Agreements about using the same code for referring to corresponding concepts facilitate sharing them among different parties even if language differences hamper simple communication.
7.2.1 Concepts and their characteristics
Concepts can be described by their characteristics. Two types of characteristics can be distinguished: intrinsic and extrinsic characteristics. Intrinsic characteristics describe the concept itself, e.g. colour ("red"), size ("5 cm\(^2\)") and shape ("ellipse") describe the concept "ulcer". Extrinsic characteristics are based on relationships between concepts, e.g. an extrinsic characteristic of the concept "ulcer" can be localization ("duodenum"). Relationships are treated in the section 7.2.2. If intrinsic or extrinsic characteristics are considered essential, this implies that the characteristics are part of the definition of the concept. For example a "duodenum ulcer" is always located in the duodenum so "location is duodenum" is an essential characteristic of the concept "duodenum ulcer". These essential characteristics are called necessary conditions in the definition.

7.2.2 Relationships between concepts
According to ISO and CEN standards [3-5], concepts can be related to each other by hierarchical and non-hierarchical relationships. In a hierarchical relationship an order is expressed between (at least) two concepts: a superordinate concept (e.g. heart valve) and a subordinate concept (e.g. mitral valve). Hierarchical relationships can be generic or partitive.
In the past, generic relationships were called *logical* relationships in ISO standards. A generic relationship, the “Is_a” relation, is a relationship between a genus (superordinate concept in a generic relationship) and a species (subordinated concept in a generic relationship) where the *intension* of the genus is contained and extended in the *intension* of the species. The intension of a concept is the set of uniquely describing characteristics including relationships of that concept, e.g. the set {health problem with anatomical localization “liver”, dysfunction “infection” and etiology “virus”} constitute an intensional definition of the concept “viral hepatitis”. Hepatitis B (species) is a “viral hepatitis” (genus) and therefore it has the same intensional definition but extended with: etiology is “hepatitis B virus”. According to Flier et al. [7] there are three types of Is_a relationships each representing other types of knowledge about the relation between subordinate and superordinate concepts. This division of Is_a relationships falls outside the scope of this thesis.

Other types of hierarchical relationships are *partitive* relationships in which the superordinate concept denotes an object which represents the whole, and the subordinate concepts refer to its parts, e.g. a heart valve is *part of* the heart.

In the past, the ISO standard described *ontological* relationships, which nowadays are split up in partitive and non-hierarchical relations. Non-hierarchical relations describe a wide range of relationships between concepts such as spatial, temporal, causal or any arbitrary one, for example the relationship between an operative procedure and the organ to be operated on. In theory there is an unlimited number of relationships between concepts. However, in practice only a small portion of relationships, those determined by human’s knowledge and experience, is widely useful. Therefore, ontological relations are called *philological* relations (distinguishing concepts and coherence between concepts as being determined by cultural history, i.e. subject to the constraints of language, art and science) by W.M. Hirs [8].

Figure 7.2 shows the above described types of relationships between concepts.

---

**Figure 7.2** Types of concept relations. The current ISO and CEN standards are represented above the dotted line, former ISO/CEN standards and W.M. Hirs’ relation types are represented below the dotted line.
7.2.3 Definitions of concepts
Hierarchical relations can be used to order concepts, and non-hierarchical relations between concepts represent (non-hierarchical) characteristics of concepts. Hence relationships (and characteristics) between concepts can be used to structure and define concepts in a system. An intensional definition is a definition based on the intension of the concept (see section 7.2.2). A *definition per genus et differentium* is the intensional definition of the genus (if it has one) extended with extra characteristics of the species. Instead of describing a concept intensionally, an *extensional definition* can be used, i.e. the set of all specific concepts (species) of a superordinate concept (genus) is enumerated. For example “Granulocytes” can be intensionally defined by “Leucocytes with abundant granules in the cytoplasm” or extensionally by the enumeration of “Neutrophil”, “Eosinophil” and “Basophil”, all the species of the genus “Granulocytes”.

7.2.4 Codes
Conventional terminological systems in the medical domain are commonly provided with a coding scheme. A *coding scheme* is a collection of rules for assigning codes to concepts. Codes can be used to refer to concepts and are usually stored in the computer-based patient record. In general we can distinguish two types of codes: *non-significant codes* and *significant codes*. Non-significant codes are context free, meaning that the code’s value is not related to the meaning of the concept. Examples of non-significant codes are *sequential codes* (a number or string taken sequentially from an ordered set, e.g. 101, 102, 103,..) and *random codes* (one of a set of possible unarranged code values). The advantage of using non-significant codes is that concepts can be altered in the structure of the terminological system without consequence for the codes, and that new concepts can easily be added to a terminological system, contrary to the case in which significant codes are used. Significant codes are related to the characteristics of the concept and its place in the terminological system. Examples of significant codes are *mnemonic codes* (codes containing one or more characters related to a name describing the concept to aid the user to memorize the meaning of the code e.g. A900=Anatomical component “lung”); *abbreviation based codes* (short representation generated directly from a concept’s name, e.g. ENT for ear-nose-throat); *hierarchical codes* (codes of which the relationship between a genus and a species is recognized, for example viral pneumonia is coded by 480 and pneumonia caused by the adenovirus is coded by 480.1); *juxtaposition codes* (composite codes with segments, representing independent characteristics, placed next to each other, e.g. a code for pneumonia A900.P100 (A900= anatomical component “lung”, P100=process “inflammation”)); *combination codes* (composite codes with segments, representing dependent and often hierarchically related characteristics, placed next to each other). Any code type can be used to compose juxtaposition or combination codes. Significant codes usually use fixed number of positions and a fixed set of characters per position. A fixed number of positions of a code restricts the level of detail (the depth of the hierarchy) in the system, a fixed number of characters per position restricts the breadth of the hierarchy.
7.3 Typology of terminological systems
A terminological system relates concepts, of a particular domain, among themselves and provides their terms and possibly their definitions and codes. In literature, terms such as “terminology”, “thesaurus”, “vocabulary”, “nomenclature” and “classification” are often confused. This section describes a typology of terminological systems based on literature and existing standards, including definitions and relationships between different types of terminological systems. This typology can be additionally used to existing typologies such as described by Rossi Mori [2].

A **terminology** is a list of terms referring to (all) concepts in a defined particular domain. A **thesaurus** is a terminology, in which terms are ordered e.g. alphabetically or otherwise, and in which concepts can possibly be described by more than one (synonymous) term. When a concept in a terminology or thesaurus is accompanied by a definition, it is called a **vocabulary** or **glossary**. A **nomenclature** is a system of terms composed according to pre-established composition rules or the set of rules itself for composing new complex concepts. **Classification** is an arrangement of objects or concepts (by the is_member_of relation) based on their essential characteristics into groups of concepts, called classes. A **taxonomy** is an arrangement of classes according to the is_a relationship from the subordinate class to the superordinate class. The CEN standards [3], however, does not distinguish between “classification” and “taxonomy”. Therefore, we use the term “classification” in this chapter loosely to also include taxonomy. A **nosology** is a classification of diseases. A terminology, thesaurus, vocabulary, nomenclature or classification is called a **coding system** when the system uses codes for designating concepts. Figure 7.3 is a model of different types of terminological systems and their relations. Each terminological system is at least a terminology with possibly additional characteristics, e.g. it is also a vocabulary when the system include definitions for the concepts described by the terms of the terminology. Sometimes the notion of ontology is used as a synonym for different types of terminological systems. An ontology is a (formal) specification of concepts, relations and functions in a domain [9] and hence focus on concepts. Concepts are important in terminological systems but they also focus on the terminology itself. An ontology is usually used to model consensus in understanding a domain between different partners.

7.4 Typology of existing terminological systems
To further illustrate the definitions of sections 7.2 and 7.3, this section describes the historical background, the types of the terminological systems and the coding schemes used in the terminological systems ICD-9-CM /ICD-10, NHS Clinical Terms, SNOMED, UMLS, GALEN and ICNARC Coding Method (ICM). We have chosen these six terminological systems for their wide diagnostic use or, in the case of ICM, special domain of intensive care diagnoses as this research is actuated by the Dutch national intensive care evaluation (NICE) project, which requires a terminological system for classifying intensive care diagnoses [10]. In Table 7.1 we summarize the types of each terminological system and coding schema. One • represents that the type is partially applicable. Further explanation about the types of each terminological system can be found in sections 7.4.1 to 7.4.6.
Figure 7.3 Model of relations between terminology, thesaurus, vocabulary, nomenclature, classification and coding system. Each terminological system is either a terminology or includes a terminology as part of a thesaurus, classification, vocabulary, nomenclature or coding system.
Table 7.1 Types of well-known terminological systems or systems with an intensive care domain.

<table>
<thead>
<tr>
<th>Type</th>
<th>ICD</th>
<th>NHS Clinical Terms</th>
<th>SNOMED</th>
<th>ICM</th>
<th>UMLS</th>
<th>GALEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terminology</td>
<td>••</td>
<td>••</td>
<td>••</td>
<td>••</td>
<td>••</td>
<td>••</td>
</tr>
<tr>
<td>Thesaurus</td>
<td>••</td>
<td>••</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Nomenclature</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Classification</td>
<td>••</td>
<td>••</td>
<td>••</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Ontology</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Coding schema</td>
<td>Significant</td>
<td>Non-significant</td>
<td>Significant</td>
<td>Significant</td>
<td>Non-significant</td>
<td>Non-significant</td>
</tr>
</tbody>
</table>

7.4.1 ICD-9-CM / ICD-10
The International Classification of Diseases (ICD) is perhaps the best known terminological system used in medicine, originally intended for statistical abstraction of disease incidence and mortality. The first “ICD” was developed in 1893, and through a hundred years it evolved into the ICD-10 [11], which was published in 1993. The predecessor of ICD-10, the ICD-9, is used for mortality registration. To be able to use this classification in a clinical setting, the ICD-9-cm (clinical modification of ICD-9) [12] was developed for morbidity registration and is currently in use in most countries. Some countries already use ICD-10 now, but mostly for mortality registration only. The first versions of the ICD were maintained by the Statistical International Institute but for the last decades it has been under the responsibility of the World Health Organization.

The ICD is a classification (recall that we do not distinguish here between classification and taxonomy) of generic related diagnostic terms represented in 17 (ICD-9-CM) or 21 (ICD-10) “chapters”, mainly arranged according to anatomical system or etiology. The ICD-10 also contains an alphabetical index, which makes it also a thesaurus. The ICD-9-CM further contains two extra classifications, the V-list for factors influencing health status and contacts with health services, and the E-list for external causes of injury and poisoning. It also contains a number of appendices, one of which is a very limited nomenclature that further specifies the morphology of neoplasms of chapter 2 (this explains one “•” in Table 7.1). In the ICD-10 the V- and E-list have been added to the chapters and the nervous system chapter of the ICD-9-CM is split into three different chapters (diseases of the nervous system, diseases of the eye and adnexa, diseases of the ear and mastoid process).

As shown in Figure 7.4, the codes of the ICD-9-CM and ICD-10 are hierarchical and in the ICD-10 they are also mnemonic e.g. all terms in the chapter “Diseases of the digestive system” start with a “K”.
In the ICD-10 some terms have two alternative codes: one marked with a dagger (†) belonging to the etiology, and the other marked with an asterisk (*) belonging to the localization or manifestations. For example “meningococcal meningitis” has two codes: A39.0† in the chapter “Certain infectious and parasitic diseases” (etiology) and G01* in the chapter “Disease of the nervous system” (localization). In general dagger codes are used for mortality registration, asterisk codes are used for morbidity registration.

### 7.4.2 NHS Clinical Terms

The NHS Clinical Terms, formerly called Read Clinical Classification or Read Codes [13, 14], was developed in the 1980s for automated description of clinical and administrative data to enable medical record summarization in General Practice. The first version used four character alphanumeric codes determining the position of a concept in the hierarchy, so this version is known as the Four-Byte Set. Since 1990 the system (version 1) has tried to cover the entire field of health care, incorporating more detail and synonyms. In version 2 mapping to other classifications such as OPCS4, ICD-9/ICD-10, the British National Formulary and the Anatomic and Therapeutic Chemical Classification Index (ATC) was realized to meet the needs of hospitals. In version 3 (and 3.1), developed during 1992-1995, important adaptations were realized to facilitate daily care practice instead of the limited use in medical record summarization. The UK National Health Service Center for Coding and Classification adopted the system as a standard and takes care of the maintenance of the system. Over 2000 health care professionals were involved in the development and quality assurance of this version.

The NHS Clinical Terms forms a classification of generic related medical concepts representing the domain of: diseases, occupations, history/symptoms, examinations/signs, diagnostic procedures, laboratory procedures, radiology/diagnostic imaging, preventive procedures, surgical procedures, other therapeutic procedures, administration, drugs/apparatuses, health status measurement. Each concept has a preferred term and some synonyms (if applicable) which are ordered hierarchically, qualifying it as a thesaurus. Since version 3.1 it is also a nomenclature in which rules are given to modify terms in a controlled way. For each concept the accompanying characteristics (qualifiers) and allowable values are defined in a template table. These qualifiers are also used to (partially) define concepts in the system, giving it a vocabulary character.

Until version 2, the codes were significant, representing the relative place of the concept in the hierarchy. Concepts could be assigned by more than one code when they are placed in more than one part of the classification. From version 3 on each concept has a unique non-significant code existing of characters and numbers. Codes for very detailed concepts (e.g. severe respiratory insufficiency due to head trauma) will be composed (combination codes) by qualifying the basic NHS Clinical Term code with codes of other concepts.
7.4.3 SNOMED

In 1975 the College of American Pathologists published the Systematized Nomenclature of Medicine (SNOMED) to provide terms for a broad range of clinical domains. A number of revisions resulted in SNOMED International [15, 16] which is intended to be incorporated into a computer-based patient record.

The SNOMED International structure consists of eleven modules, also called axes or dimensions, which can be conceived as distinct classifications. The eleven modules contain concepts hierarchically related to each other and concern the following domains: Topography, Morphology, Function, Living organisms, Chemicals & drugs & biological products, Physical agents, Occupations, Social context, Disease and Diagnosis, Procedures, General modifiers. Almost all ICD-9-CM terms and codes can be found in the “Disease and Diagnosis” module. By linking concepts of the various modules one can compose new complex medical concepts. The discrete terms together with the information about the cross-reference relations between the different axes provide (at least a partial) definition of each concept, giving SNOMED its vocabulary character. Although SNOMED is an acronym including the term nomenclature, it is questionable whether SNOMED is a nomenclature that can generate sensible compositions. It is a nomenclature in the sense that new concepts can be composed, but rules which define which concepts may be sensibly linked to compose new concepts are missing. Due to the lack of suitable composition rules it is possible to represent the same clinical concept by different composites and even to arrange clinically irrelevant composites, e.g. a fractured left lung encoded by T-28500 (Left lung); M12200 (Fracture,open). SNOMED-RT [17], which is under development, is claimed to address these problems. Codes in SNOMED are significant. Within a module they are hierarchical and mnemonic, e.g. topographical codes start with a “T”. Newly composed concepts get juxtaposed codes by joining the codes for the associated terms from the different modules.

7.4.4 ICNARC Coding Method

The ICNARC Coding Method (ICM) was developed in 1995 by the UK Intensive Care National Audit and Research Centre (ICNARC) [18]. ICM supports the collection of reasons for admission in Intensive Care which is part of the data collection used in the UK national audit of intensive care.

ICM is a terminology without generic relations between end-terms (diagnostic conditions). The system has a so-called “hierarchical five tier structure”: type (surgical/non-surgical); body system; anatomical site; physiological/pathological process; and diagnostic condition. The higher tier prescribes the allowable values on the lower tier, for example the selected body system prescribes which anatomical sites can be chosen. ICM is a vocabulary in the sense that each diagnostic condition is (at least partially) defined by the type - body system - anatomical site - physiological/pathological process path.

The ICM codes are significant. At each of the five tiers in the hierarchy, a selection is made from a list of options and a code (number) is returned. The code is not unique at all tiers, but it is unique in combination with the preceding tiers. The final unique code is the result of the five selections. For example, admission for ICU following surgery for a gastric tumor is coded as: surgical (type="T."), gastrointestinal (body system="3."), stomach (anatomical site="3."), tumor/malignancy (process="39."), tumor (condition="1."). Figure 5.1 shows that conditions can be reached by different paths (different combination of type - system - site - process) and therefore can have different codes. These are linked in a conversion table.
7.4.5 UMLS

In 1987 the National Library of Medicine developed the Unified Medical Language System (UMLS) [6, 19]. The goal of the UMLS is to facilitate the retrieval and integration of information from multiple biomedical information sources such as patient record systems and bibliographic databases.

The UMLS consists of four knowledge sources: the Metathesaurus, the Semantic network, the Specialist lexicon and the Information Sources Map. The Metathesaurus and the Semantic network are the most relevant in the context of this paper. As the name indicates, the Metathesaurus is a thesaurus in which concepts are linked to (synonymous and preferred) terms which are alphabetically ordered. The Metathesaurus is manually enriched with hierarchical relations between concepts from established terminological systems such as ICD-9-CM, Read Codes, SNOMED and MeSH which give it characteristics of a classification. As each concept has an attribute “definition”, the metathesaurus is also a vocabulary. The Semantic Network provides information about concepts (semantic categories) and their relationships but only at a high aggregation level (the reason for one • in table 1). All concepts represented in the Metathesaurus are categorized into semantic categories of the Semantic Network which regulates (together with the Metathesaurus) the mapping of concepts from different sources to each other. Although in theory semantic relations defined in a semantic network could aid a user to make new composites, in practice the UMLS does not support the composition of new concepts.

Each concept and term in the Metathesaurus has a unique UMLS code. Although these codes are somewhat mnemonic (concept codes start with a “C” and term codes with a “L”) they are not significant in the sense that the numbers next to the starting character are non-significant.

7.4.6 GALEN

Growing out of earlier work in PEN&PAD, the Generalized Architecture for Languages Encyclopaedias and Nomenclatures in Medicine (GALEN) project started in 1992 [20-22]. The project was funded by the European Commission and is a collaboration of European universities, agencies and vendors. The goal of the Generalized Architecture for Languages Encyclopaedias and Nomenclatures in Medicine (GALEN) project is to formally describe and model the medical domain by which the interchangeability of electronic medical data of different data sources can be supported. The “Terminology Server”, the implementation of GALEN's goal, integrates three modules: the Concept Module, the Multilingual Module and the Code Conversion module.

The Concept Module implements a formal language, the GALEN Representation and Integration Language (GRAIL) which is used to represent concepts and their characteristics and relations in the Concept Reference (CORE) model. The CORE model is an ontology, which comprises in 1998 the definitions of approximately 5000 concepts (e.g. bones, organs, and fractures) and 1000 relations (e.g. HaveLocation, HaveComplexity). The GRAIL formalism allows developers of terminologies to create models containing these concepts and relationships, and to derive new concepts which are valid compositions of existing ones. The composition rules included in GRAIL make the system a nomenclature, the definition rules make it a vocabulary.

The Multilingual Module manages the mapping of concepts to synonym and preferred terms (thesaurus) and hence information entered in one language can in principle be displayed in another. The Multilingual Module contains terminology and grammar rules which make it
possible to generate phrases for complex compositions. The mapping of concepts to and from existing coding systems is managed by the Code Conversion Module. Each concept in the CORE model has a unique non-significant code. Different applications can communicate with the terminology server. By this direct communication it is possible to generate random unique numbers for new composites which are saved at the server.

7.5 Summary and discussion
Historically, medical data was coded mainly for (retrospective) statistical, epidemiological and administrative purposes. Nowadays with the electronical availability of medical data, the importance of these data in daily care practice, and especially the importance of these data in interdisciplinary communication and for clinical research, has increased. This shift in use of medical data implies new requirements on terminological systems concerning for example the level of detail and the structure of the terminological systems and hence it is important to (re)study this.

Although a major goal of terminological systems is the standardization of terminology to improve communication, the notions used in the literature to describe terminological systems themselves are not uniform, which makes it hard to communicate their underlying ideas. A good understanding of terminological systems is essential before one can assess whether an existing terminological system is appropriate for use in certain circumstances or when one has to develop a new system. Therefore a referential framework for understanding terminological systems is needed. Such a framework includes at least two components. First a terminology and typology of terminological systems and second a uniform (formal) representation of the structure of the terminological system. Existing standards such as ISO and CEN [3, 4] only describe the first part. Moreover, this is restricted to a rather dry enumeration of definitions about notions in the field. Therefore this chapter describes the first part of a framework for understanding terminological systems and summarizes the notions and definitions used by standard organization such as ISO and CEN, but enriched with interrelations between these notions, including a typology of terminological systems. This typology is illustrated by applying it to six medical terminological systems including diagnoses.

This chapter plays a facilitating role for chapter 8 of this thesis. Chapter 8 includes our positive experience with the application of a conceptual and formal representation formalism to describe the structure of terminological systems. We hope these two papers would support researchers to interpret the merits and limitations of existing terminological systems and to build on existing work in the field.

Acknowledgement
We would like to thank Ronald Cornet for his valuable discussions and helpful remarks on an earlier draft of this chapter.
References


