Barriers and challenges of using medical coding systems
Surján, G.

Citation for published version (APA):
Surján, G. (2011). Barriers and challenges of using medical coding systems
Chapter III. THE CULTURAL HISTORY OF MEDICAL CLASSIFICATIONS

Abstract:
This chapter outlines the history of medical classifications in a general cultural context. Classification is a general phenomenon in science and has an outstanding role in the biomedical sciences. Its general principles started to be developed in ancient times, while domain classifications, particularly medical classifications have been constructed from about the 16th-17th century. We demonstrate with several examples that all classifications reflect an underlying theory. The development of the notion of disease during the 17th-19th century essentially influenced disease classifications. Development of classifications currently used in computerised information systems started before the computer era, but computational aspects reshape essentially the whole picture. A new generation of classifications is expected in biomedicine that depends less on human classification effort but uses the power of automated classifiers and reasoners.

1. Introduction
This chapter outlines the history of medical classifications in a general cultural context. While classification and medical terminology is a hot topic of current biomedical informatics, our aim is to show, that nearly all problems we face currently originate from the past. The modern computer era however offers more efficient techniques although the principles of these techniques have been developed through many centuries.

Classification is an essential issue in all scientific activity. Its importance is emphasised by R.A Crowson in his book titled Classification and biology. (Crowson RA, 1970) He argues, that it is often thought that the essence of science is to count or measure things. But before we could do so we have to select what we want to count or measure (and what not). And this distinction presupposes a classification. Indeed, all scientific activity requires a clear scope definition: a distinction between relevant and irrelevant phenomena. This distinction – either made consciously or unconsciously – is at least a dichotomous classification. But this is usually not the last step – even if the aim is not to classify the phenomena of the given domain. This is particularly true in life sciences. If someone – let say – wants to study the alimentary habits of frogs (either in a qualitative or quantitative way), it is necessary to classify the things in the world as frogs and non frogs. But the habit of one particular frog at a particular time is probably not a real scientific issue: science is more about the general rules than about particular phenomena. So we study the habits of a number of frogs. Then we realise that there are many different kinds of frogs, each kind probably having different habits: many different species, but also young and old, male and female ones etc. And now we are in the middle of the sea of the problem of classification:
Which distinctions are relevant and which are not for an actual problem?
Which distinctions are relevant in general?
Which categories are real, which are arbitrary?
Are we able to classify all phenomena correctly – are the categories well defined? Etc.

This is no more the problem of a scope definition: we have to define classes within the scope of our research in order to properly interpret our observations. This is very characteristic to all life sciences, mostly due to the amazing variability of life. Medicine is no exception. Medical classification is not a solved problem, and perhaps never will be totally solved. Beyond the fact, that the rapid development of medicine reshapes classifications from time to time, it points to several philosophical, linguistic and logical problems. Philosophically it is related to the questions about the basic nature and structure of existence, the ontological nature of medical entities etc. The linguistic aspect deals with the naming conventions and the language used to describe medical phenomena, while the logical aspect is related to the problem of reasoning over medical facts.

The goal of this chapter is to show that all problems around medical classification and terminology have historical roots. This history can be seen in a wider and a narrower context. A narrower context would focus on biomedical and health problems, while a wider context includes the development of the theory of classification in general.

While we prefer the wider context, we do not want to provide an exhaustive description of the whole history of classifications and all cultural problems around it. Through selected examples we just want to show, that nearly all the problems we are facing now, already emerged in the past, and that the lessons learnt from this history can help to avoid traps in present and future development. This approach determines the structure of this chapter: first we want to describe the development of a general theory of classifications, pointing to the mentioned philosophical, linguistic and logical aspects. Then we will show how specific domain classifications emerged, particularly in life sciences. The third part of the chapter will go through the history of classifications in the medical domain while the fourth deals with current trends and achievements.

2. Development of a general theory of classifications

2.1. The beginnings

We believe that classification, i.e. identification of discrete entities of the world and grouping them into categories, is an inherent property of human intelligence. In that sense the history of classifications is as long as the history of mankind. Later on, this inherent and often subconscious intellectual activity became subject of scientific investigation. The known history of this conscious investigation, the theory of classifications, started in ancient Greek philosophy. After the initial steps of Plato, Aristotle carried out foundational work by drawing up the principles of categorisation that are still more or less valid. Before his substantial work there was a lack of language necessary to describe the theory of classifications. The fifth book of Aristotle’s Metaphysics (a series of his studies collected by his students) provides a vocabulary of fundamental notions, like principle, substance, quality and quantity, necessary and accidental properties, unity, identity, part and whole, etc. Many of the words e.g. category (κατεγωρία) used today to describe classifications were coined
by Aristotle. Perhaps his most influential work was the invention of the conceptual hierarchy: i.e. the structure that stems from the arrangement of existing things into species and genera (i.e. specific and generic classes). This structure serves as the basis of the definitions of entities by specification of a "genus proximum" and a "differentia specifica". It was recognised rather early that a superior category might have many different, partially overlapping subdivisions. To avoid confusion it is important to find proper 'differentiae' (properties that divide generic categories into non-overlapping subcategories). These differentiating criteria were called fundamentum divisionis (basis of division).

The main goal of ancient philosophers was to make clear how the true nature of certain things can be defined and to find ways to understand the nature of being. None of the known ancient classifications gave an exhaustive, fully comprehensive representation of its domain, since there was no need for any practical use of classifications. The detailed classifications (e.g. the classification of animals developed by Aristotle) served as a test-bed of the theory (Ogle, 2001). In the 3rd century A.D. the Neo-Platonist philosopher, Porphyry wrote an introduction to Aristotle's Categories. This work is known today as Isagoge from the Greek word "Εισαγωγή" that means introduction. This work was discussed and rewritten in Latin by several early medieval thinkers, e.g. Boetius and Peter of Spain. The text is available in English translation by George McDonald Ross (Ross, n.d.) The Isagoge explains the basic notions of Aristotelian classification theory, such as individuals, species and genera. An illustration of this is known through Peter the Spain as "Tree of Porphyry". Its logical structure is shown in (Figure III-1)
This early 'conceptual graph' provides a strictly dichotomous tree structure. Each category has exactly two "differentiae" that lead to two subcategories. The "differentiae" are attribute-pairs based on presence or lack of some property (e.g. rational – irrational, or material – immaterial, etc). Another important feature of the graph is that there is one single highest genus (i.e. the topmost category; the Substance). There are several layers of subordinate genera and the lowest category, called species: Human and Beast, have no further subcategories. Only individuals may belong to species. Except the topmost and the lowest categories all intermediate categories are genera of their subcategories and species of their super-categories:

"What has been said will become clearer if we consider just one category. Substance is in itself a genus, and under it there is body; under body, animated body; under animated body, animal; under animal, rational animal; under rational animal, the human being; and under the human being, Socrates, Plato, and all particular human beings. Of these, substance is so general that it can only be a genus, and the human being is so specific that it can only be a species; whereas body is a species of substance, and the genus of animated body. But animated body is also a species of body, and the genus of animal; and again animal is a
Chapter III – Cultural history of classifications

species of animated body, and the genus of rational animal; and rational animal is a species of animal, and the genus of human being; and the human being is a species of rational animal, but it is not the genus of any sub-division of humanity, so it is only a species; and anything which is immediately predicated of the individual it governs will be only a species, and not a genus. So, just as substance is the most general genus, being the highest, with nothing else above it; so the human being is only a species, and the lowest, or (as we said) the most specific species, since it is a species under which there is no lower species or anything which can be divided into species, but only individuals (such as Socrates, Plato, or this white thing). ." (Ross, n.d.)

These distinctions between types are often used even in modern medical and non-medical classifications. The lowest level categories that have no subcategories just direct instances are called today 'concrete class', while all categories above this that have but indirect instances are called 'abstract classes', while the modern name of the supreme genus is 'top level category'.

The strict dichotomy, i.e. each class must have exactly two direct subclasses with one and only one differentiating criterion, was a Platonic idea; As D. Ross points out Aristotle argued against this strict dichotomy in the 'De partibus animalium' (Ross, 1977).

2.2. The Ars Magna

The middle ages were strongly influenced by ancient Greek philosophers. Great thinkers of the scholastic period tried to synthesize this philosophical tradition with Christianity, but it was not a straightforward discipline. Perhaps the strangest, somewhat odd but obviously exciting example was the Ars Magna written by Ramón Lull (1235-1315). Lull (or Raymond Lully or Raymundus Lullus) lived in a religiously heterogeneous (Jewish, Islamic and Christian) environment, and wondered why the obvious truth of his own religion (Christianity) was not convincing enough for others. He thought that this must be due to the weakness of language, so he tried to invent a system that is able to represent all truth and reason in an obvious, convincing and language independent way.

To achieve this goal, he tried to build up a structure of the ultimate general knowledge, onto which all other more specific sciences are built. Highly influenced by Aristotle, he sought the most general principles, to which all particular principles belong as parts of a whole. To structure this general knowledge, Lullus selected nine letters (from 'B' to 'K'), arranged them onto two disks, labelled with letters 'A' and 'T'. Each letter assigns one absolute, one relative principle, one question, one subject, one virtue and one vice as is shown in Table III-1.
### Table III-1 The Lullian alphabet

The nine letters on disk 'A' refer to the nine absolute principles, the disk indicates them in both in noun and adjective form (goodness and good, wisdom and wise etc.), and all the nine letters are linked by connecting lines to all the others, forming a fully connected graph. (See Figure III-2)

<table>
<thead>
<tr>
<th>General Principles Disk 'A'</th>
<th>Relative Principles Disk 'T'</th>
<th>Questions</th>
<th>Subjects</th>
<th>Virtues</th>
<th>Vices</th>
</tr>
</thead>
<tbody>
<tr>
<td>B Goodness</td>
<td>Difference</td>
<td>Whether</td>
<td>God</td>
<td>Justice</td>
<td>avarice</td>
</tr>
<tr>
<td>C Greatness</td>
<td>concordance</td>
<td>What</td>
<td>angels</td>
<td>prudence</td>
<td>gluttony</td>
</tr>
<tr>
<td>D Duration</td>
<td>contrariety</td>
<td>of what</td>
<td>Heaven</td>
<td>Fortitude</td>
<td>Lust</td>
</tr>
<tr>
<td>E Power</td>
<td>Begin</td>
<td>Why</td>
<td>Man</td>
<td>Temperance</td>
<td>conceit</td>
</tr>
<tr>
<td>F Wisdom</td>
<td>Middle</td>
<td>how much</td>
<td>Imagination</td>
<td>Faith</td>
<td>Acidy</td>
</tr>
<tr>
<td>G Will</td>
<td>End</td>
<td>what quality</td>
<td>Senses</td>
<td>Hope</td>
<td>Envy</td>
</tr>
<tr>
<td>H Virtue</td>
<td>Major</td>
<td>When</td>
<td>Vegetation</td>
<td>charity</td>
<td>Wrath</td>
</tr>
<tr>
<td>J Truth</td>
<td>Equal</td>
<td>Where</td>
<td>Elements</td>
<td>Patience</td>
<td>Lies</td>
</tr>
<tr>
<td>K Glory</td>
<td>Minor</td>
<td>how, with</td>
<td>Instruments</td>
<td>Compassion</td>
<td>inconstancy</td>
</tr>
</tbody>
</table>

### Figure III-2 The absolute principles

The lines connecting each letter to all other letters refer to statements like Goodness is durable, Power is great, Greatness is powerful etc. The letters on the second disk are arranged in a different way, they form three triangles. This disk T depicts the relative principles, as is shown in
Each relative principle has different sorts. E.g. difference may exist between sensual and sensual, between sensual and intellectual, and between two intellectual things. These sorts are also indicated on disk T. Then Lullus creates a table containing all bi-grams (combinations of two letters) without repetition, irrespectively to the order of the letters. (Figure III-4)
The whole 6th part of the Ars Magna deals with that. The letters may refer either to general or relative principles, but to questions, subjects, virtues and vices as well. For instance BF means 36 different statements and one or two questions about each statement ('whether it is' and 'how much'). Let we see some examples:

Statements:
- Goodness is Knowable; Wisdom is Good; Wisdom is Different; Difference is Good; Faith is good; Faith has knowledge; God is wise; God is good; God is different; Avarice has knowledge; Avarice is evil;

Questions:
- Is goodness knowable? How knowable is goodness? How much knowledge does justice have? Is faith good? How wise is God? Is avarice evil?

Lullus slightly modifies the meanings of the letters depending on the subject of the statements or questions, or in case of vices he changes sometimes the propositions to their opposite without any explanation. So God is wise, faith and avarice has knowledge, God and faith are good, but avarice is evil (why not good?). Lullus apparently admired the "combinatorial explosion" (the high number of possible derived meanings of just two letters taken from a set of nine), but he fails to achieve the real goal: to create a tool to discriminate false and true statements. He must sometimes violate his own rules to avoid combinations that are controversial with his belief.

Further on he creates a third disk that consists of three coaxial wheels; each of them contains the nine letters. By rotating the three wheels independently, it is possible to generate all possible tri-grams (729 variations). Lullus extends each trigram using an additional letter 't'. The letters, preceding this 't', refer to the disk A, the following letters refer to disk T. So a trigram, e.g. 'bcd' can be interpreted in different ways: 'bcdt', 'bctd', 'btcd' and 't bcd'. Lullus provides a table that lists all these possibilities. Still, the meaning of these 'codes' are not unambiguous, since the letters may refer to the disks, but also to the questions, subjects, virtues and vices, according the 'alphabet'. Lullus perhaps intentionally allows this 'flexibility'.

"...each science has principles different from those of other sciences, the human intellect requires and seeks one general science with its own general principles in which the principles of all other sciences are contained as particulars of a universal that regulates the principles of other sciences so that the intellect can repose in those sciences by really understanding them and banishing all erroneous opinions. This science helps to establish the principles of all other sciences by clarifying their particular principles in the light of the general principles of this art, to which all particular principles belong as parts of a whole".

(Dambergs, 2003)

This is something like what we call in our days 'top level ontology. In spite of all peculiarities, the ideas of Lullus influenced many thinkers not only in his own age but through many centuries. He had followers and also serious opponents and perhaps parodists as well. Even today many people admire his work. John Sowa mentions him as the inventor of first device for automated reasoning (Sowa, 2000). His work is converted to a software application that allows studying his ideas.
(Abbott & Dambergs, 2003). From our aspect, the following features of this work are important:
Lullus used letters to signify concepts – in the same way as we use codes nowadays. However he used the same letter for several different entities, so his codes were not unique.
He created a combinatorial system that is able to generate a number of complex entities (or statements) very similar to our combinatorial coding systems.
He tried to formally discriminate true and false statements, or formally approve what he thought to be true. This is what we now do with reasoners that are able to make inferences on formally described statements.
In a later work, the *Arbor Scientiae* (The Tree of Science or the Tree of Knowledge) Lullus drew up a categorial structure of all sciences and disciplines of his age. This treelike structure was based on the nine principles described in the Ars Magna. The structure is not a true tree structure in the geometric sense, but a lattice, since some of the branches or sub-trees cross over. (Eco, 1993) This is the problem of poly-hierarchy that is also present in recent medical classifications. (The definition of poly-hierarchy is that it occurs whenever an entity has more than one super-ordinate entities that are hierarchically not related to each other.)
The efforts of Lull reappeared three centuries later in the work of Leibnitz, who himself studied Ars Magna thoroughly.

2.3. *The Real Character*

Further important developments can be observed in the work of Bishop John Wilkins (1614-1672): *An Essay towards a Real Character and a Philosophical Language*, published in London 1668 (Wilkins, 1668).
The 'real character' is a writing system, in which all elements (each single character) have direct meaning, i.e. syntax and semantics have an unambiguous one to one relation. Wilkins was concerned not only about the variety of languages, but also about the alternations of the language over time. E.g. he compared many English versions of the Lord's Prayer as something that should be stable but it is not. After investigating the history of languages and language variations, Wilkins turns to study the history of letters, supposing that all alphabets originate from Hebrew. He states that there are less variations and differences in letters than in languages, because letters are much younger. Anyway, the variety of letters he thought to be an appendix to the curse of Babel. Then he mentions examples in which letters – just for brevity – are used to signify words instead of sounds. (Sort of shorthand writing) Then he continues:
"Besides these, there have been some other proposals and attempts about a real universal character that should not signify words, but things and notions, and consequently might be legible by any nation in their own tongues"
(Wilkins, 1668)
This was the aim why he created his own system. To achieve this goal he thought it necessary to start from a generally accepted systematisation of the world, at least those things that are general, elementary entities. Once there is an unambiguous system of symbols that refer to these elementary things, the combination of the characters will allow representing all other entities. So the second part of his work has this title: *Conteining a regular enumeration and description of all those things and notions*
This second part is a system of categories, that follows Aristotelian principles of categorizations and somewhat resembles the Tree of Porphyry. It starts from 40 top level categories, called genera arranged into a tree structure. These genera are subdivided into "differences" and further into "species" (The word 'difference' refers to 'differentia specifica' but often refers to the subcategory defined by the differentiating criterion, as it is the case here.) In total he had 2030 species. Wilkins tried to fix the number of "differences" for a genus to six (i.e. one category can have at most six subcategories), but he admits that this is sometimes not possible (especially in case of living beings), because they "are of too great variety to be comprehended in so narrow a compass." The species under each "difference" are arranged into pairs "just for better helping memory". The proposed hierarchy is much more complex than a simple three level tree: the top level categories are themselves arranged into a tree and many of the species also have subdivisions. Wilkins apparently tried to enforce an artificial structure on the system but recognised that reality often does not obey this.

The third part of his work is titled 'Concerning natural grammar'. In this part Wilkins tries to set up a general grammar that is independent from each particular language but preserves all features that he thought to be common in all languages. This grammar gives a combinatorial nature to his system, and allows describing many more things than the 2030 entities covered by his categorial structure. This 'grammar' conveys something that we currently call relations. E.g. this allows him to describe butcher as flesh + merchant, carpenter as wood + manufacture and mathematician as quantities + artist.

The fourth part of his work deals with the writing system. Wilkins proposes actually two alphabets; the first is solely for writing, that contains unpronounceable symbols, the second is created for speaking, and contains Latin and Greek letters and diphthongs. It arranges consonants and vowels in such a way that it helps pronunciation. All of the characters are significant: either they signify a category or a grammatical construct. In the fourth chapter of this part Wilkins gives two examples of the usage of his language: he "translates" the Lord's Prayer and the Credo into this language, explaining the translations word by word.

As an appendix, Wilkins included an alphabetical dictionary of approximately 15000 (!) English words, that are either indexed to his "philosophical tables" (i.e. the system of his categories or grammatical entities), "or explained by such Words as are in those tables". Homonyms are indexed to different places according to the different meanings, as the following extract shows:

Abolish.
[Annihilate] AS.I.1.O.
[Destroy] AS.I.4.O.Abate

The letter 'a.' stands for 'active', i.e. the given meaning is a verb (makes something nothing)

By this dictionary Wilkins tries to demonstrate, that his "philosophical language" is really able to express everything that is expressible in English. Beyond the important improvement (compared to Lull) that the codes of Wilkins are unique, he also discovered that codes can express the hierarchy of a system of
Chapter III – Cultural history of classifications

categories. This strategy of code value assignments is used today in e.g. ICD codes, the Read Clinical Classification (version 2) and versions 2 and 3 of SNOMED. This code value assignment has an obvious computational benefit (it is easy to decide if one entity is subsumed by another) but enforces a strict tree structure. Wilkins also drew up classification trees, trying to follow the principles of the Tree of Porphyry, but he could not help violating both the dichotomous principle and the tree structure: the same things appear several times even in the same table (in the same sub-tree). Once the principle of dichotomy is given up, the number of subdivisions of a single category might increase, but the strategy of assigning a single English letter to a subdivision leads to a restriction of the number of subclasses of a category, since in this way it is not possible to represent more than 26 subcategories at a given node, since the English alphabet has no more letters. The principle of mutually unique codes has another drawback: any misspelling or typing error leads immediately to misunderstanding. (Eco, 1993) Natural coding systems, e.g. genetic code and natural language, are usually redundant and hence there is some degree of fault tolerance. (The use of check digits was not invented at that time of course, and would be of little help without using computers.) The system set up by Wilkins was quite large: navigation in the space of 2030 entities in a paper based representation is difficult. And still – as he realised – this system was far from complete. In our modern time we also have to realise, that maintenance problems increase rapidly with the size of classifications, and completeness seems to be never achievable in building medical classifications.

2.4. When to think will mean to calculate

In the 17th century, two extraordinary scientists, Gottfried Wilhelm Leibnitz (1646 - 1716) and Isaac Newton independently invented the infinitesimal calculus (differential and integral calculus). Both of them were polyhistors, working not only in various fields of mathematics but also in several scientific disciplines like philosophy, linguistics, physics and theology. A century before Charles Babbage, the 'father of computers', Leibniz developed a mechanical calculator that was able to multiply and divide. He was heavily influenced by the thoughts of Lullus, and was seeking for a better mathematical foundation of his ideas. A dream of Leibnitz was to eliminate all uncertainty from mathematical proofs, so that true and false statements could be distinguished in a formal way, simply by calculations. "To think will mean to calculate" he claimed. In order to achieve this, he proposed to decompose the whole knowledge into elementary entities and assign prime numbers to them. In other words, he used numeric codes, contrary to the alphabetic codes of Lullus and Wilkins. Complex entities can be coded by multiplication of the codes of their components. Following the idea of Sowa (Sowa, 2000) with some minor modification, let us take the Tree of Porphyry as an example and assign 1 to Substance and the subsequent prime numbers to the differentiae. We get the Table III-2 in this way.
Table III-2 Prime numbers representing differentiae (after Sowa)

<table>
<thead>
<tr>
<th>Classification</th>
<th>Prime Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substance</td>
<td>1</td>
</tr>
<tr>
<td>Material</td>
<td>2</td>
</tr>
<tr>
<td>Immaterial</td>
<td>3</td>
</tr>
<tr>
<td>Animate</td>
<td>5</td>
</tr>
<tr>
<td>Inanimate</td>
<td>7</td>
</tr>
<tr>
<td>Sensitive</td>
<td>11</td>
</tr>
<tr>
<td>Insensitive</td>
<td>13</td>
</tr>
<tr>
<td>Rational</td>
<td>17</td>
</tr>
<tr>
<td>Irrational</td>
<td>19</td>
</tr>
</tbody>
</table>

Then HUMAN as a rational, sensitive, animate, material SUBSTANCE gets the code $2 \times 5 \times 11 \times 17 = 1870$, while LIVING, as animate material SUBSTANCE is $2 \times 5 = 10$, and PLANT as insensitive LIVING is $10 \times 13 = 130$. The benefit of this representation is that the truth of such statements, like "HUMAN is a LIVING being" can be investigated by division of the corresponding code numbers. Since 1870 is divisible by 10, the statement is true. To investigate the statement "HUMAN is a PLANT" we have to divide 1870 by 110; the result is not a natural number, so the statement is false. This representation has the obvious benefit that it makes certain types of syllogism calculable. But, as Sowa claims, still many logical operations, such as implication and negation, cannot be calculated in this way. Leibnitz had further concerns about his own invention. While Wilkins thought that due to the high number of entities, setting up a categorical system is a huge work that requires co-operation of many scientists, Leibnitz realised that not only the number of entities that have to be assigned is infinite, but any smallest piece of world is infinitely complex (Eco, 1993). Therefore he felt that the entirety of prime numbers is insufficient to represent the world. This view is not surprising for somebody who invented the infinitesimal calculus. Modern studies of human thinking do not confirm his concern unanimously. E.g. Sowa argues that all representations in our mind – even the representations of continuums are discrete (Sowa, 1984). If he is right, it is possible to map all concepts to natural numbers or prime numbers – at least theoretically.

The idea of Leibnitz about the formalisation of thinking came back in 19-20th century mathematics in the form of formal languages, and especially in the work of Gödel, who showed the inherent limitations of the dream of Leibnitz.

2.5. "Concept writing": Towards formal languages

One of the most interesting aspects of the *Ars Magna* is that three components: the classification itself, the system of codes to denote categories, and the rules of the used "language" (i.e. the logic and reasoning) were present and developed simultaneously. After Leibnitz, these aspects became separated from each other. The first important steps towards formal languages were made by Gottlob Frege (1848-1925) with his *Begriffsschrift* (concept-writing); but this was only about the syntactic rules that enable the representation of the logical structure of any statement, and not about how the symbols denote concepts nor about how to
categorise of the concepts. The Begriffsschrift contains the following four primitives only:

\[
\begin{align*}
\text{assert } p & \quad \neg p \\
\text{not } p & \quad p \rightarrow q \quad \text{if } p \text{ then } q \\
-\exists x \quad P(x) & \quad \text{for every } x \in P(x)
\end{align*}
\]

These graphical symbols proposed by Frege are not in use today because of the difficulty to map them to spoken natural language. For instance the following symbol (See Figure III-5) stands for the simple statement that some men are tall, but must be read as:
"It is not true that if x is a man then x is not tall." \{S1\}

![Figure III-5 Frege's concept writing](image)

In the recently used syntax of predicate calculus it is:

\[\exists x (\text{tall}(x) \land \text{man}(x))\]

that is to be read as:
"There exist x, where x is tall and x is man." \{S2\}

Note, that \{S1\} and \{S2\} are semantic equivalents.

The syntax evolved through the works of Peirce, Peano, Russell and others. We do not follow here this story in detail, but will return to formal languages when we discuss modern biomedical classifications. We also do not follow the philosophical chain of developing top level categories (the so called top level ontologies) that aim to provide a general framework, under which all classifications can be arranged in a consistent way. A summary of this story is given by Sowa (2000). Instead we turn to domain classifications.

3. **Domain classifications**

Domain classifications do not aim to represent the whole world or the universe of human knowledge, which was an obvious aim e.g. of both Lullus and Wilkins. Rather, domain classifications – as the term itself indicates – focus on a single, well described domain, e.g. plants, chemicals, minerals, diseases etc. Of course, such domain entities already appeared in many early general classifications, mostly as examples. From Aristotle through Lullus up to Wilkins, all philosophers built some domain classifications; biology, later medicine was a favourite test-bed for them. When early philosophers tried their hands on domain classifications, this was done
mainly to test whether their theory – on which the classification was built – suited the entities of the given domain.

A real domain classification however is not only a restriction of the scope; at the same time such a classification aims to be comprehensive. They are designed for some practical use within the given domain, and to achieve this all entities of a domain should be classifiable, even if arbitrary collective categories ("not elsewhere classified") are used. In other words, the difference between top level (general) and domain classifications is not just the difference of their scope, but also the difference in purpose: general classifications aim at the proper understanding of reality, domain classifications aim at some practical use, often for some statistical data collection. This last point is especially significant for medical classifications.

Domain classifications have been developed in a number of different fields. There are classifications both for natural and artificial things; e.g. minerals, clouds, poetry, folk music and transport vehicles. Now let us review some examples from biology, since historically they are the closest relatives of medical classifications.

3.1. Carl von Linné

Perhaps the first pure domain classification that satisfied the criteria mentioned above of which parts are even used today, was the system developed by Carl von Linné (1707-1778) known also as Carolus Linnaeus. He studied medicine at various European universities and was fond of plants. At that time botany was an integral part of medical curricula because of the wide use of medicinal plants. Linnaeus graduated and practised medicine, but his scientific work concentrated on the classification of living organisms, mostly of plants. Linnaeus is often referred to as the father of taxonomy. He arranged plants into a system of hierarchically arranged categories (taxa) following Aristotelian principles. Contrary to Porphyry, Linnaeus restricted the term 'species' to the lowest level categories and 'genus' to the first level above 'species'. (As we mentioned, for Porphyry all intermediate levels are genera of their subcategories and species for their superordinate categories) The levels above genus were called Orders, Classes and Kingdom by Linné. While contemporary biology still uses this structure, the number of levels has been extended, by introducing Phyla, Superclasses, Superorders, Infraorders, Families, Superfamilies and Tribes.

It was also the idea of Linné, that species should be named according to their place in the classification. This system of names he proposed is called the binominal nomenclature as the names consist of two parts: the first name denotes the genus and the second is used for distinguishing the given species from others within the same genus. This fits with the Aristotelian principle of defining things by the 'genus proximum' and the 'differentia specifica'. Linnaeus did not use any codes for his classification, but the proposed names refer to the place of the named thing in the hierarchy of the classes – similarly to many recent medical coding systems, where code values reflect the place of the category in the hierarchy.

From Aristotle up to Linnaeus the systematisation of living organisms was based on the so called scala naturae, (the natural scale), a continuous chain of organisms, with humans on the top and the simplest living organism at the bottom. The ranking of this list of species was based on their complexity, and the different species were not arranged into classes. So this hierarchy was an exclusive hierarchy of
categories, like military ranks. A higher ranked category does not subsume the lower ones. In other words: the *scala naturae* was not a chain of generalisation-specialisation. It was something like a disease classification in which diseases were arranged into a ranked list according to their seriousness.

Contrary to that, the Linnaean system (published in several editions of his *Systema naturae*) was an *inclusive* (subsumption) hierarchy that had two important features that followed Porphyry's principles: (1) The structure he proposed was a strict tree structure. (2) He tried to find a "diagnostic criterion" i.e. a sufficient and necessary condition for the membership of each taxon (category). The hierarchical tree structure he proposed had some weak points: it failed to represent hybrid species that inherit essential properties of two independent species. This is the poly-hierarchy problem that frequently occurs in many classifications and often disturbs developers of classifications even today. Before the age of computers it was a source of technical difficulty, since a poly-hierarchy is hard to represent on paper; nowadays it is a modelling challenge.

Another weakness of the Linnaean taxonomy is that it cannot explain the phenomenon that features characteristic of a genus ("diagnostic criteria") sometimes are missing from certain species (or subordinate genera) of that genus. E.g. fleas are pterygote insects with no wings, whales are mammalians with no hair etc. (Panchen 1992). This is what biologists call *polythetic taxa*.

The essence of the Linnaean "revolution" was the move from a ranked scale to an inclusive hierarchy. The basis of this step was a scientific theory saying that biological species show not only different levels of complexity but show common essential properties and these properties constitute a reasonable basis for classification.

### 3.2. Developmental theory and the classification of species

In 1859 the publication of the 'Origin of species' by Charles Darwin (1809-1882) resulted in an immediate outbreak of a revolution in biology, not only in the sense of rapid development but also in the sense of a heavy fight between supporters and opponents of the theory. While Darwin was a diligent observer who himself collected many facts supporting his theory, most– if not all – of his statements were already stated by different researchers. In the introduction of his work, he lists 34 researchers that contributed to his theory. His ultimate merit was the careful selection not only of the facts but also of the contemporary scientific works that could be arranged as mosaic pieces into a whole and consistent picture.

Darwin's theory served as a new principle of the whole biological classification. The natural diagnostic criterion of a taxon became: having an ancestor in common. But this is not necessarily a visible, intrinsic property: it hardly can be established just by investigating visible features of existing living organisms, but requires knowledge about the phylogenesis. (Nowadays the sequencing of the DNA can essentially contribute to this research, but Darwin could not even dream of it.) Phylogenetic inheritance is different from the inheritance of object properties along ontological hierarchies: descendants of a species may loose essential properties (necessary conditions) of their ancestor (see the problem of polythetic taxa of Linnaeus); different descendants may lose different essential properties, but still belong to the same taxon due to having common phylogenetic roots.
Contemporary classification of species in biology is still based on Darwinian principles; just it is more and more based on the analysis of DNA sequences today. However in many cases biologists are increasingly concerned about the definition of species. The classical definition says that individuals of the same species are able to interbreed producing reproductive descendants. This definition hardly can be defended in case of micro-organisms, like bacteria. On the other hand there are some examples of interbreeding individuals from different species that are able to produce reproductive descendants.

From our perspective three important lessons have to be learned from the history of the biological classification of species.

1) As was shown, developments in biology reshaped the biological classifications. All biological classifications reflected the contemporary theory of life.
2) An important step in the development of biological classification was the introduction of an inclusive hierarchy instead of an exclusive rank of species.
3) The pattern of classification itself — a mere geometrical structure — reflects a basic philosophical view of the world. There are two basic patterns of inclusive hierarchies: logical division and clustering.

Logical division (e.g. the Tree of Porphyry) starts from some unity: i.e. a generic top level category, the "sumnum genus". Then this unity is gradually divided into smaller and smaller subcategories according to strict rules of sufficient and necessary attributes that are inherited by all subsumed classes; until finally we arrive to unique individuals or instances.

Clustering moves in the opposite direction. It starts with the observation of individuals and seeks for groups of individuals that share common properties. On that basis it defines classes, and then — still based on common properties — several classes are merged into superclasses, etc until we arrive at a supreme class that covers all observed individuals of the domain of interest.

In principle these two approaches may lead to the very same categorial systems. At first sight one can imagine a third classification pattern. This is the pragmatic classification. E.g. things that require a certain action could be classified into one category, regardless of their properties. This is important for us, since medical classifications are not so much theory-driven, but pragmatically oriented. But if it is true that certain different things really require the same action, then there must be a theory that explains why this is the case. And this theory will show what is common in those apparently heterogeneous or dissimilar things. Exactly in the same way that the theory of evolution explains what is common in dissimilar species belonging to the same taxon.

4. Development of medical domain classifications

4.1. Ancient time

Undoubtedly early physicians tried to categorise diseases. But no trace of a real ancient medical domain classification was found in the literature. None of the known works of Hippocrates – the most famous ancient physician who lived about two centuries before Aristotle – concentrates on classification. Neither his ancient followers like Galen worked heavily on disease classification.
Chapter III – Cultural history of classifications

The known history of medical domain classifications begins in the 17th century with the classification of diseases. Other medical domain classifications (procedures, drugs, medical devices, etiologic agents, social factors, etc.) were not developed until the 20th century.

4.2. John Graunt and the London Bills of Mortality

During the early 1600s around the great plague epidemic in London, it was decided to collect systematically data of all death cases. Some periodic data collection was started in 1592, but it became continuous about ten years later. The obvious reason was early detection of a plague epidemic outbreak. It was quite a challenging enterprise at that time, since there was not enough medically trained personnel. Moreover there were no practical or theoretical foundations or any guidelines how to do such things. Practically the work was done by so-called searchers, usually women, nominated at each parish. They had to examine all corpses to decide what the person died of. They had no predefined list of diseases; neither were they medically educated. No wonder that some of the recorded causes were really odd. The records they made were collected by each parish and published in print weekly as the London Bills of Mortality for more than a century. (Greenberg, 1997)

John Graunt, (1620-1674), later often referred to as the father of demography, analysed the data published in the Bills. His study entitled Natural and Political Observations (made upon) the Bills of Mortality was first published in 1662. In this work he presented a number of substantial observations based on the analysis of the figures of the Bills. E.g. he tried to estimate the proportion of live-born children dying before the age of six. At that time the Bills of Mortality did not record the age at death of the person. He defined a set of causes that he supposed never occur after the age of six and another set of causes that he supposed to occur under six in half of the cases. With these postulates he was able to calculate the approximate proportion of young children who died. Whether or not such an approach can be justified, the idea, that such data can tell more about a population than what is directly expressed by the numbers – is obviously interesting. This is an early example of the reuse of information: data collected for one specific purpose was used for another. Later investigations revealed that Graunt's estimates were quite good. (World Health Organisation, n.d.) He also observed that the number of deaths due to other reasons than plague also increased during periods of a plague epidemic. This was surprising and conflicted with what everyone would expect. Graunt concluded that deaths due to plague was significantly underreported. Other sources confirm that families of persons died of plague often bribed searchers to report death due to something else to avoid being locked up in their house. By analysing the numbers Graunt was able to estimate that the number of death due to plague was 25% higher than reported. Then he concludes that the true number of plague deaths could not be defined without reporting all diseases. "A true Accompt (=account) of the Plague cannot be kept, without the Accompt of other Diseases" and also that "The ignorance of the Searchers [is] no impediment to the keeping of sufficient, and useful Accompts." A facsimile of the original edition can be found on the web (Stephan, 1996)
These early experiences highlight two important aspects of disease classifications that are relevant even today – needless to say in a very different form. One aspect is the lack of knowledgeable personnel. In that time it concerned the lack of medical knowledge and the low number of trained physicians; today the problem is that in many countries there is a shortage of professionals that are knowledgeable in the theory of classification and medicine (at least in nosology) as well. Nosology comes from the Greek word νοσος (nosos = disease). It means a branch of medicine that deals with the classification of diseases.

The other interesting problem was that financial incentives distorted the data: searchers were paid for not making records of plague. Even today there is a huge literature of the distorting effect of financing incentives, especially with case mix based systems. See (Hsia, Kurshat, Fagan, Tebbut & Kusserow, 1988), (Hsia, Ahern, Ritchie, Moscoe, & Kurshat, 1992) and many others.

We can also learn from Graunt, that under certain conditions, data are reusable, and also that important and reliable observations about a population are possible even if the data are seriously distorted at the individual level.

The London Bills of Mortality practically lacked any theoretical, scientific foundation. Today it is difficult to judge how effective this system was in preventing plague epidemics. We just have to stress that classification in medicine is not a purely academic exercise. The question is not merely how to describe or categorise the reality. Classifications in medicine are tools used for interventions that improve the health of the population.

4.3. Development of the notion of disease in the 18th century

The naïve, layman approach to disease suggests that diseases are entities that can be named. But this was not always thought to be obvious during centuries. Consider that we do not necessarily name situations when our car or computer brakes down. Similarly it is not self evident that various conditions that are considered as illness of our body should have names. During the history of medicine various theories emerged that did not consider diseases as defined or definable entities. The so called 'Galenism' explained all diseases as a disturbance of the balance of the four humours (blood, bile, phlegm, and choler) and the four qualities: heat, cold, wet and dry (corresponding to the four elements: fire, air, water, earth). According to this theory, diseases are not distinct, identifiable "things" but rather some degree of a process that does not display well defined forms. E.g. fever was thought to be a consequence of excess of hot fluids. Just as extreme weather conditions, floods, drought periods, storms usually do not have specific type-names, similarly disease-types in this view do not require names. Miasma theory explained diseases as consequences of some imbalance of the air and these conditions again hardly could be classified into named entities. Such views were strong amongst academically trained physicians through medieval and early modern centuries. A more evidence based theory developed slowly.

In the second half of the 17th century the famous English physician Thomas Sydenham was seriously concerned about the inefficiency of contemporary medicine. He stated that instead of wasting time with studying the basic sciences of medicine (at this time anatomy, physics, chemistry and botany) physicians ought to spend most of their time at the bedside, and make direct observations about the
nature of diseases. He suggested that diseases 'were to be reduced to certain and determinate kinds with the same exactness as we see it done by botanic writers in their treatises of plants!' Sydenham considered diseases as entities having definite properties, in alignment with the naïve approach. As he continues:

Nature, in the production of disease, is uniform and consistent; so much so, that for the same disease in different persons the symptoms are for the most part the same; and the self-same phenomena that you would observe in the sickness of a Socrates you would observe in the sickness of a simpleton.

Cited in (Fischer-Homberger, 1970)

But his view was not unanimously acknowledged by the scientific community at that time. The whole period was characterised by the co-existence of contradictory theories. Besides Galenism, another important theory was contagionism. As early as in 1546 Girolamo Fracastoro published his theory of epidemic diseases that he thought to be caused by small particles that could be transferred from one patient to another. Beyond that, he supposed that these 'spores' are specific to the given type of epidemic disease; hence plague can cause only plague, small pox only small pox etc. But the debate on Galenism, contagionism, miasma theory and some other theories (like iatromechanism, that considered the human body as a mechanical machine) continued for centuries, and even at the time of the plague epidemic in Marseille in 1720 many did not believe that plague was contagious (Delacy, 1999).

Nearly one century after Sydenham, in the midst of the 1700s, – among others – three highly respected physicians tried to follow Sydenham's recommendation to create taxonomies of diseases in the same or similar way as botanist did with plants.

(1) François Bossier de Lacroix, alias Sauvages (1706-1777) first published anonymously his Nouvelles Classes de Maladies in 1731, and about 30 years later a more elaborated version entitled Nosologia Methodica (1763). Its full title was Nosologia Methodica Sistens Morborum Classes, Genera et Species, juxta Sydenhami mentem et Botanicorum ordinem. In this work he grouped 2400 diseases into 315 genera, 44 orders, and 10 classes. (Lesh, 1990)

(2) Carol von Linné (1707-1778), who had good contact with Sauvages, also published a classification called Genera Morborum (1759, 1763) that was strongly founded on Sauvages' work. Linné listed more or less the same diseases, but classified them differently into 325 genera and 11 classes.

(3) Some year later in Scotland, William Cullen published his Synopsis Nosologicae Methodicae (1769). This was an integrative work of the several previous taxonomies (including Sauvages and Linnaeus) and Cullen's own classification. The latter entitled Genera Morborum Præcipua Definita. This work was published several times in Latin, and in English and in a number of European languages, the last time as late as in 1823. We show with the example of Linné's Genera Morborum, how such 'botanical classifications of diseases' were constructed.

An edition of the Genera Morborum from Italy is available electronically (Linneaus, 1776). In this edition he classified 325 diseases into a three level hierarchical structure that has eleven supreme categories. All the diseases are assigned serial
numbers. These code numbers form a continuous series that does not reflect the hierarchy. Each classified condition has a short symptomatic description (informal definition). These descriptions are often cross-referenced: a description of one condition uses another one. E.g. Cholera is defined as a combination of vomiting, diarrhoea and colic. These cross-references are made explicitly by indicating the numbers of referenced entities.

50 Colica  
Intestini dolor umbilicalis cum torminibus

….

183. Vomitus  
Rejectio ingestorum convulsiva

…

186. Cholera  
Vomitus .183. cum diarrhoea .187. colica .50.

187 Diarrhoea  
dejecio faecum liquidarum frequens

So the *Genera Morborum* can be seen as a hierarchical mono-axial cross-referenced coding system of the diseases. It is not a mere tabular list of the diseases known at that time, because all the conditions are defined by symptomatic criteria. (This latter feature is something that we even miss in today’s form of the ICD!) The *Genera Morborum* was merely a scientific enterprise. We are not aware of any public health usage of it. This also seems to be true for the Sauvages’ *Nosologia Methodica*. Their purpose of classification was to improve efficiency of medical treatment by better understanding the nature of diseases:

It was not only hoped that a working therapy would evolve from a systematic nosology of diseases; it was also hoped that such a nosology would facilitate communication between doctors and thus be of didactic use.  
(Fischer-Homberger, 1970)

The latter point became more important. As she continues:

The significance of the communicatory value of nosology is, for instance, clearly stressed by Vincenzo Chiarugi (1759-1820), the nosologist of psychiatry. In the introduction to the systematic part of his work he wrote that considering the prevailing uncertainty and confusion in matters of terminology, it seemed necessary to establish a set of terms with which everyone would associate the same meanings. And Johann Peter Frank (1745-1821), although considering nosological systems as such to be worthless from a scientific point of view, nevertheless conceded that ‘they make medical language accessible to the most diverse nations from pole to pole’.

We have to note, that this was the time when the language of science slowly turned from Latin to national languages. Anyway the goal to make medical texts intelligible regardless of nationality and mother tongue is something what we can read in many recent works on the interoperability of health information systems.
None of these three classifications were maintained further in the following 19th century, which was the century of pathology. Autopsies became more and more common, the dissection of corpses was performed not to study the normal anatomy – since this was already well known at the time – but to study the pathological alterations causing death. Medical observations were not anymore restricted to the surface of the body. This was the time, when Semmelweis made his influential observations about the death of women due to puerperal sepsis that was caused by the fact that physicians and medical students performed autopsies and treated patients without disinfection. Disinfection hand washing improved the development of surgery and this achievement in turn revealed many observations on internal changes of the body under living conditions. Instead of observing external symptoms it became possible to observe the internal pathological alterations that explain the external symptoms and characterise the various diseases. Botanic type classifications that were based on observable superficial properties of diseases were no more the interest of physicians. In the light of pathology it became clear, that many of the ‘diseases’ listed in the 18th century classifications were only symptoms that could be caused by a number of diseases, often rather different ones. At the same time, the medical profession started to specialise, and classification of diseases slowly became an enterprise of public health physicians instead of clinicians. Public health – at that time – was mostly concerned with mortality statistics.

So far we have seen two different and more or less independent threads in the history of disease classifications:

The early mortality statistics required a classification of causes of death. This practical goal was apparently different from the goal of those philosophers who founded the principles of the theory of classifications. It was not about understanding the nature of the existence or the metaphysical status of entities, but to provide a basis for statistical analysis in order to study and improve public health. In that sense, medical (diagnostic) classifications are different from other domain classifications: they are used to provide evidence of need for intervention.

The other thread seems to be a continuation of the work of philosophers; it was a scientific enterprise even in lack of a suitable theory of diseases that could be used as a foundation of the classification. The *Genera Morborum* and the *Systema Naturae* are very similar in the sense that both are based on observable properties that in lack of a proper theory did not lead to a rigorous and consistent classification. Linne’s botanic classification is still respected today, while all similar botanic enterprises of that time and all its contemporary disease classifications have been forgotten. The reason of this perhaps is that Linne was lucky enough to find and select observable properties for plants that later became explained by phylogenetic facts, so many of his taxa are still valid, while other botanic classifications were not so lucky. And nobody was lucky enough to find categories of diseases that could remain defendable in the light of the inventions of the next century.

The two threads gradually converged in the history of ICD, the International Classification of Diseases, but even up to now did not result in a fully satisfactory solution.
4.4. The roots of ICD from the 19th century

From the 19th century on statistics – including health statistics – became more and more official: institutions and organisations were established in various European countries and in the States to collect and publish statistical demographic and health data. One of these institutes was the General Register Office of England and Wales founded in 1837. The head of this office – the Registrar General – was obliged to present a report to "both houses of Parliament by Command of Her Majesty" on births, deaths and marriages. The first report that covered the second half of the year 1837 was published in 1839. (Farr, 1839). This report contained several appendices, one of them was written by William Farr (1807-1883) the first health statistician of the Register Office. This appendix was a letter to the Registrar General, containing many valuable observations on the mortality data presented in the report. From our point of view the most important observation is that Farr realized that the usefulness of mortality statistics strongly depends on an appropriate nosology. He gave a short summary of nosology from Sauvages to Cullen and up to his contemporary nosologists. He demonstrated that there are a number of different nosologies sometimes differing from each other essentially only in the number of classes, but sometimes differing in the principal "fundamentum divisionis". Most often, the main differentiating criterion is the localization of disease, but sometimes it is the severity of disease, the underlying pathological process or even the typical age of onset of the disease. Farr criticised all these classifications and proposals from the point of view of mortality statistics, and pointed out that the Cullen classification that was then in general use in public services became outdated because of the substantial development of pathological anatomy since the time of Cullen. Then he introduced the notion of 'statistical nosology', i.e. a nosology that suits the requirement of population studies and mortality statistics. His principal "fundamentum divisionis" was the 'mode in which diseases affect the population'. In that sense he discriminates three basic classes of maladies: (1) contagious diseases either endemic that prevail in particular localities, or epidemic that spread over countries and (2) sporadic that arise in an isolated manner and originate internally from the body and (3) violent causes of death caused by external agents: injuries or poisoning. (The term 'sporadic' means that the cases appear independently from each other; sporadic diseases are usually but not necessarily rare.) Figure III-6 illustrates the proposed structure (using somewhat modernised terminology)
Chapter III – Cultural history of classifications

Figure III-6 The structure of "statistical nosology" proposed by W. Farr

As the figure shows, he classified sporadic diseases according to their anatomical localisation.

Farr paid attention not only to the structure of classification but also to the terms used to identify the categories. He did not use numbers or any sort of codes to designate a class.

Another concern of Farr was that primary disease causing death often was mistaken for its complications. The problem of selecting one and only one principal cause of death – which is medically not always sensible – will be discussed later when we will describe the coding rules of ICD.

In the 19th century health statistics became an international issue. This fact also emphasized the importance of a standard nomenclature. The first International Statistical Congress (Brussels) in 1853, decided to request Farr and Marc d'Espine to prepare an internationally applicable, uniform classification of causes of death. Even at that time standardisation was a difficult enterprise: at the next Congress (Paris 1855) Farr and d'Espine presented two independent and conflicting lists. These two lists reflected two different views. Farr put emphasis on anatomical localisation, while d'Espine classified diseases according to their "nature", i.e. according to their appearance, observable properties (gout, bleeding, etc). As we already stated, each classification reflects some theory. Farr's classification reflects a theory that says that disease is a structural or functional alteration of a body part. d'Espine's theory put more emphasis on the symptoms. He thought bleeding, for instance, to be an entity independently to the site of bleeding. The Congress adopted a compromise of the two proposals. I do believe that this compromise had serious consequences for the structure of ICD that is present even today: still there is no clear "fundamentum divisionis" in ICD, the uppermost division (into chapters) is still a mixture of anatomical, etiological and pathological division. In the later revisions of ICD, Farr's approach became dominant. But Farr himself was not quite consistent: his basic categorisation consisted of five groups: epidemic,
general (systemic), local, developmental diseases and consequences of violence. (Note that this proposal of Farr was somewhat different from what he described in his first letter to the Registrar General.) Only the third group was really classified according to anatomical manifestation. This classification was adopted and updated several times by the International Statistical Congress, but never became unanimously accepted.

The International Statistical Congress was later succeeded by the International Statistical Institute. This organization established a committee in 1891, led by Jacques Bertillon (1851-1922), charged with the task of preparing a classification of causes of death. We have to emphasize the importance of the fact that from this point on the construction of the classification was not a work of one single expert: the elaboration of the classification was done by a committee, which necessarily represented different views, philosophies and interests, and always led to some compromise.

The Bertillon committee composed a list of causes of death that predominantly reflected the list used in Paris (since Bertillon was the Chief of Statistical Services in the city of Paris), but it also was a synthesis of different national classifications used in England, Switzerland and Germany. Actually the committee created three classifications: a shortlist consisting of 44 entities, a medium list of 99, and a full list of 161 causes. It is not clear from the literature, whether these three lists formed a hierarchy, where the shortlist contained the top level entities, and the others consisted of their subdivisions. It is also possible, that the shorter lists simply disregarded the less important causes or collected them into some "other" garbage categories. In any case this system reflected the fact, that different questions (different problems in health care) require different levels of granularity.

The Bertillon Classification of Causes of Death was accepted by the meeting of the International Statistical Institute in Chicago in 1893. Up to his death in 1922 Bertillon was involved in several revisions of the system that became more and more accepted and widely used in many countries. The need for revision from time to time was obviously motivated by the growing number of improvements in medicine. On the other hand, stability of a statistical classification is crucial in order to obtain comparable data over time (backward compatibility). Following the proposal of the American Public Health Association, the International Statistical Institute decided on decennial revisions.

After the death of Bertillon, the Health Organisation of the League of Nations – the ancestor of the WHO of the UN – had taken more and more interest in classification, (not only in mortality but more and more in morbidity). The work was therefore continued by a so called "Mixed Committee" consisting of representatives both of the Health Organisation and the International Statistical Institute. For a relatively long period classification of causes of death and disease classifications (mostly for hospital statistics) had been developed more or less separately. In 1856, in his sixteenth Annual Letter to the Registrar General, Farr himself proposed to extend his system to diseases that are not fatal but causes public health problems. At the first and second International Conference for the Revision of the Bertillon Classification (1900, 1909) a parallel list for morbidity statistics had been adopted. Non-fatal diseases were represented as subdivisions of certain categories of causes of death. But these early experiments did not result in
an internationally accepted, satisfactory morbidity classification. Therefore a number of national classifications emerged, until finally, the World Health Organisation (WHO), founded after the Second World War created the first proposal of the International Classification of Diseases, Injuries, and Causes of Death, as the sixth revision of the Bertillon classification. The accepted classification was supplemented with an alphabetic list of disease names and a manual that described the coding rules. Even in very common and simple cases more than one code may apply to describe different aspects of that case. Therefore it is necessary to define certain rules that explain how to select the single statistically appropriate "principal diagnosis" code for each case.

One may wonder why one and only one principal diagnosis is allowed for a case, since one person might suffer from several lethal diseases and there are no evident rules to find out which of them actually caused death. The traditional way of thinking in mortality statistics was apparently the following: There is a total number of deaths in a given population in a given period of time. This number can be broken down according to different attributes like age, gender and cause of death. Of course, we expect that the number of male and female death cases should be equal to the total number of death cases. Similarly the sum of the death cases due to all causes is expected to be equal to the total number of death cases. To extend the list of causes with combinations of diseases could be a theoretically possible solution, but a list containing all possible combinations would be too long and unmanageable in paper based systems. This is a good example of the serious difference between clinical and public health views. Today when we more and more often speak about reuse of clinical information, we have to be aware that in many cases there is no easy and straightforward way from clinical records to public health information.

Through the following revisions, ICD became a more and more comprehensive system, while some of the features of the original structure were still preserved. In spite of all the constructional problems the coding system became more and more widely used, not only for the original purposes, but for many others, ranging from clinical data management through resource allocation, financing, reimbursement to health policy decision making.

This expansion of uses generated a need for a much more detailed representation than necessary for health statistics both at an international or national level. Since a fine grained system would not be suitable for statistical purposes, the ninth revision introduced an additional hierarchical level: all statistically relevant entities are represented by a three digit code while finer distinctions can be made as subdivisions, using the fourth (or for some categories even an optional fifth) digit. The former (7th and 8th) versions used basically three digit codes, and only a part of the three digit items were subdivided using a 4th digit character. The so called dagger asterisk notation is also an invention of the 9th edition. This notation is used to cross reference items from different chapters representing the anatomical and pathological categorisation of the same disease. This enables to describe conditions by combined codes, but still keeping the rule that only one principal cause of death (or one principal disease in case of morbidity statistics) is allowed. This issue goes back to the dispute between Farr and d'Espine about the organisation principle. Unfortunately this dagger-asterisk system did not lead to a consistent
combinatorial solution. Daggers are used to mark codes that describe etiology or pathology aspect, while asterisks are used to mark codes that describe the localisation of the disease. But these marks are applied only in a limited number of the cases where ICD allows combining codes. Even in such cases the allowed combinations are strictly defined. E.g. the first chapter of ICD describes infectious diseases. This chapter contains a whole section for tuberculosis from codes A15.0 to A19.9. The entities listed here contain some information about the localisation of the disease and also about the diagnostic method. For instance A15.1 means ’Tuberculosis of lung confirmed by [bacteriological] culture only’. Some of the codes in this section are marked with a dagger. E.g. A17.0 means ’Tuberculous meningitis’ and it is allowed to combine with G01 that stands for ’Meningitis in bacterial diseases classified elsewhere’. Note that in case of A17.0 the specification of the diagnostic method is missing. A17.0 is not allowed to combine with anything else than G01, but this combination says nothing more than A17.0 alone. On the other side, G01 (that is marked with an asterisk) is allowed to be used in combination with a number of codes. Some of them are reasonable combinations (e.g. A69.2 + G01 means Lyme meningitis while A69.2 means Lyme disease only without specification of the localisation), but some are not (A39.0 + G01 means meningococcal meningitis, but A39.0 alone means the same). So instead of offering a freedom for the users to select codes that describe the infective agent (and the diagnostic method if necessary) and combine it with codes describing the localisation, the allowed combinations are ’printed in the stone’ in advance, and carry very few or no information. (These examples are taken from the currently used 10th version, but the problem in the 9th version was the same.)

Around the development of the 9th version it was recognised that important aspects of diseases are not reflected in ICD at all. To overcome this shortcoming, additional classifications (e.g. for functions and disabilities, for morphology of tumours, for medical procedures) were proposed.

The current version of ICD is the tenth revision that preserves the above mentioned hierarchy and realises that health statistics may go further than mortality and morbidity: sometimes there are conditions where a person may require some health service without having an illness. These conditions are also included in ICD-10. (E.g. Pregnancy or being in contact with a patient having infective disease)

The development of ICD is described in a document published by WHO (World Health Organisation, n.d.)

5. Current issues in medical classifications

5.1. Computers step in

Computers reshape our life nearly in all aspects. This is particularly true for medical classifications. A number of constraints arose in the past from the limitation of paper based information management. Enforcing tree structures to avoid multiple inheritance and the need of a single principal cause of death are just examples. But the traditions inherited from the past are strong, and changes come slowly. We are not at the end of this story. Until now, the use of computers reformed medical classifications in the following ways:
a) Classifications increased in number, and extended in scope. From the midst of the 20th century more and more medical domains (procedures, laboratory concepts, social conditions, medical devices, drugs etc.) were classified. Wikipedia currently lists more than 20 medical classifications under the title 'medical classifications'. And this is obviously not an exhaustive list. (We provide a brief of internet resources related to current medical classifications in the appendix.) These classifications often overlap because there is no universally accepted, universally suitable coding system in the medical domain. This is partially a consequence of the lack of a general theory of medicine. However, contemporary classifications are not theory-driven but practically oriented: various classifications have been developed to serve various practical needs.

b) The use of classifications became manifold. While early classifications were used nearly exclusively to create statistical tables (perhaps additionally for educational purposes) in the modern health care systems coded data are often re-used for a number of purposes, e.g. for financing, quality assurance, policy making, medical decision support, document and information retrieval etc.

d) Classifications increased in size. Computers made them more easily manageable, so much larger classifications could emerge. This also means that the granularity of coding systems became more and more fine-grained: nomenclatures emerged. Before the age of computers, very detailed coding systems were useless due to their inconvenience of use.

e) Large combinatorial systems emerged. While a vision of a combinatorial coding system already appeared in the work of Lullus, before the age of computers such systems could not spread and develop because of their inconvenient use in paper based systems.

The idea of a two-axial combinatorial system was raised first at a symposium at the New York Academy of Medicine in 1929 (Chute, 2000). The resulting classification was called Standardized Nomenclature of Diseases, SND for short. The two axes were the topology (i.e. the anatomical localisation of the disease) and the aetiology (in reality it contained pathological and patho-physiological entities) describing the nature of the alteration. Pretty soon, in 1933, the system was extended with operations, resulting in the Standardized Nomenclature of Diseases and Operations (SNDO). In this system diseases can be described as combinations of the codes of the appropriate anatomical site and pathological entity. Since the Operation axis contained only generic categories (like 'removal' or 'incision') actual operations could be coded also by combinations of the code of the anatomical site and the generic operation. The first medical use of a combinatorial system was followed by a number of multi-axial coding systems, since compositionality offered higher flexibility and expressivity than mono-axial systems can ever achieve. New inventions, e.g. newly invented operations can be coded immediately without any modification of the coding system. (Since maintenance of classifications require consensus building in a large community, modifications of medical classifications are rather slow and often seriously behind the rapidly developing medical science).

In the sixties, when computers appeared on the horizon, American pathologists developed the system further according to their needs, creating SNOP, the Systematized Nomenclature of Pathology, that later developed further and extended beyond the need of pathology as SNOMED, Systematized Nomenclature
of Medicine. Subsequent versions have been published in 1979, 1982 and 1994. The history of the development of the three versions highlights the pros and cons of multi-axial approach. When humans create classifications manually, multiple inheritance always causes problems. For the human mind it is much easier to deal with pure tree structures. Unfortunately reality usually does not obey this wish. The multi-axial approach has the promise to get rid of multiple inheritance. The author has published a paper demonstrating that multiple inheritance can always be eliminated by introducing new dimensions (Surján & Balkányi, 1996). It is very clear in the case of ICD for instance, that a disease can be classified according to the affected body part (anatomy), according to the nature of pathological alteration (pathology), and also according to the cause of the disease (aetiology). This is the cause of poly-hierarchy e.g. in case of pneumonia, which is a lung disease, and an inflammation as well. A multi-axial approach can make this situation very clear: represent the lung as a respiratory tract organ, and – in a separate axis – inflammation as a given pathological condition, and pneumonia as a combination of these entities. On the other hand, multi-axial systems have serious limitations. There is a lack of a grammar that could provide clear semantics to the combinations. E.g., the juxtaposition of codes of ‘incision’ and ‘trachea’ together does not necessarily mean ‘tracheotomy’. At least theoretically it could mean 'incision of trachea' but also 'incision by trachea' or even 'trachea of incision'. Of course human knowledge prevents us from such comic misinterpretations, but computers lack the required common sense and medical knowledge. They need an explicit specification of the semantics. As soon as the location of the disease differs from the location of the intervention, what commonly occurs in medical practice, some syntactic rules are required to make clear how to connect pathology and intervention codes to locations. If a given semantic interpretation adheres to an axis, e.g. 'location' means that something exists in the given location. Then there is the problem that one axis can represent one semantic role only. In reality however a certain thing that has a given ontological nature, may play different roles. E.g. in the second version of SNOMED there was an axis for aetiology. This included chemical substances among others, since they can play an etiological role in case of poisoning. Being a ‘chemical substance’ is the ontological property while being a poisoning agent is a role in a disease. One chemical substance that causes poisoning in one case can be a medicine as well as a component of some drug in other situations. In the 3rd version of SNOMED the Etiology axis has been split up into three dimensions, namely physical agents, chemicals, and living organs. This was a shift from the actual role to the ontological nature of the entities. But at the same time the notion of ‘unknown aetiology’ disappeared, since there was not an axis any more where such a notion could be placed. f) Recognition of the above mentioned limitations lead to steps towards a formal representation and formal definitions of the categories, in order to benefit from the potential of reasoners (programs that enables computer to reason automatically). All these changes originate from the essential change in paradigm of using such systems. Before the age of computers, the goal always was to create tabulations for statistical, financial or any other purposes. But the new paradigm, as Rector
Chapter III – Cultural history of classifications

emphasises, is to use codes in ‘patient centred’ software systems (Rector, 1999). As pointed out by Slee and others, earlier classifications were designed for grouping cases to create some output from diagnostic information; while the new coding systems aim at using them for input information into electronic health records. (Slee, Slee & Schmidt, 2005).

During the past several decades the talk moved from ‘classification’ to ‘terminology’ then to ‘concept or knowledge representation’ and finally to ‘ontology’. These were the favourite terms at various scientific events. In spite of this shift of wording, the problem remained the same: how to make medical descriptions semantically computable? Using codes instead of names to identify classes means that we map our semantic units to natural numbers (whether we use numeric or alphanumeric codes they are countable and hence can be mapped to natural numbers). This allows computers to manipulate the information independently from the ambiguity of language. This is an important prerequisite of semantic computing, that makes it possible to use categories of classifications as variables in any formal language. (Formal languages consist of a defined, usually strongly limited number of symbols, a set of syntactic rules, and an unlimited number of free variables). Once it is achieved, it opens the way towards a formal description of the meaning of the categories. The first robust system of formally defined medical entities was the GALEN core model, produced by a project funded by the European Union. (Recto, Nowlan & Glowinski, 1993). Actually GALEN does not use codes, the categories are identified by human readable strings, but the content of a category does not depend on these strings but on their formal definition. GALEN brought three new inventions:

- To represent a new entity it is not necessary to seek its proper place in the whole classification, it is ‘enough’ to define the entity according to the strict formal rules of GALEN
- The formal definitions enable the use of automated reasoners to classify all entities, eliminating human errors
- Users can compose new entities according to their need while the formalism ensures that the new entities will have a clear semantics intelligible for everybody acquainted with the system.

5.2. Co-existence of classifications – the problem of mapping

In the previous section we saw that there are many concurrent, partially overlapping medical classifications, often used in parallel. The reason for this – as we mentioned – is that different classifications serve different tasks. All classifications are abstractions, i.e. details thought to be irrelevant for the given task are rejected. But certain details can be relevant for one task while irrelevant for another. In principle, a very fine-grained representation could carry all information that could be relevant for all conceivable purposes. Then it would be possible to extract the relevant information as required in a given situation. But there is no hope in the close future to develop such single totally satisfying ‘omnipotent’ classification. So we have to learn how to live with the variety of classifications.

One often mentioned way to solve this problem is the ‘mapping’ of various classifications to each other. But it is easy to see, that if two classifications are really different, i.e. they consist of different entities, such a mapping can not be done
without distortion of the information. Let us illustrate this problem with a sheet of paper that is divided into squares on one side and into triangles on the other. (Look at the illustration on Figure III-7) Squares are coded by numbers and triangles by letters. A given point of this sheet can be 'classified': instead of describing its precise co-ordinates we can say that it falls let say, in the square '4' or triangle 'g'. It is very clear, that a mapping from squares to triangles or vice versa is not possible without some distortion: once we do not know the co-ordinates of a point but only the code of the corresponding triangle, we can not tell precisely in which square the point falls. Of course we can create a conversion table using the maximal overlap. E.g. most of the triangle 'e' overlaps square '3'. In some lucky and exceptional cases, the whole triangle overlaps a square, like triangle 'a'. In such exceptional situations mapping is possible in one direction. A mapping of all entities in one direction is possible without distortion only if one system consists merely of subdivisions of the other.

In 1986 the National Library of Medicine (NLM) of the USA started the UMLS project. UMLS stands for Unified Medical Language System. The project strongly built on the experiences with MESH (Medical Subject Headings), a thesaurus used for document retrieval in MEDLINE, the scientific literature database of NLM. (Lindberg, Humphreys, & McCray, 1993)

The background of the project was the recognition that the use of MESH is limited by the fact that the language of users often differs from the language used in the requested documents (McCray, Aronson, Browne, Rindflesch, Razi & Srinivasan, 1993) and also from the terms of MESH. Searchers are more willing to use their own search terms than map their terms to MESH entries. To overcome this barrier it was decided to integrate all relevant thesauri, classifications and terminological systems into UMLS. This forms the first so called knowledge source of UMLS, the Metathesaurus. This is a huge repository of concepts, their preferred names and synonyms, and relations among the concepts. Unfortunately, many of the relations (e.g. 'broader than' or 'otherwise related') have no clear semantics. The second knowledge source of UMLS, called Semantic Network, was created to arrange concepts collected form various sources into an integrated space of broad
Chapter III – Cultural history of classifications

semantic types. Originally, the third knowledge source, the Information Source Map, was created to provide information about the sources that are integrated into UMLS. It was hoped that the UMLS enterprise can help to map categories from one classification to another. Partly due to the lack of explicit semantics (and the consequent inconsistencies) but partly because of the inherent theoretical limitations, this could not be fully achieved. Bodenreider et al. report a 50-65% success rate in mapping (Bodenreider, Nelson, Hole & Chang, 1998).

Of course this does not mean that the whole UMLS project failed. To the contrary, the system was growing through many years, and grows continuously till now. Hundreds of publications report applications of UMLS in various fields from nursing informatics to natural language processing etc. …

5.3. End of the story? – The medical viewpoint

We argued, that all domain classifications reflect an underlying theory of the given domain. From this point of view one may wonder how a disease classification can exist at all, since a scientifically sound general theory of diseases is something that is missing even in modern medicine. A partial answer is that while there is no such a general and defendable theory of diseases, still from the very beginning physicians had some theory, however vague, by which they hoped to explain the causes of diseases. We do not state that in lack of a proper theory classifications cannot exists, but very likely they are not well structured and consistent. Indeed all disease classifications show some inconsistency and arbitrariness. But the other part of the answer is that classifications not only reflect a theory but often are used to improve the theory. If somebody tries to classify entities of an obscure domain this can help in identifying the weak points in the corresponding theory; and often indicates the topics that require further research.

Fischer-Homberger argues that medicine found its theoretical basis in pathology around the 19th century and the importance of nosology decreased from this time on (Fischer-Homberger, 1970) She refers to Sydenham, who first proposed that classification of diseases should follow the way of classification in botany, and at the same time he was sceptical about the usefulness of basic medical sciences, e.g. anatomy. At the time of Sydenham basic sciences could not really help the practical work of physicians. So Sydenham suggested going to the bedside instead of studying anatomy, botanics, and chemistry. Fischer-Homberger points out that the development of pathological anatomy in the 19th century reshaped this picture; in pathological anatomy a science had been found on which medicine could base itself. The sceptical attitude of Sydenham and his followers became unjustified, since from that time on the cause of many diseases had been discovered. According to Fisher-Homberger, this achievement made nosology obsolete and unnecessary. She mentions psychiatry and dermatology as two exceptions, as medical fields where causes of many diseases are still unknown.

This view suggests that nosology is nothing else than a surrogate for sound knowledge. The early nosologists had to follow an empiricist way: their classifications were based on the observable properties of the diseases, pretty in the same way that Linnaeus followed in classification of plants. But recognition of causes of diseases does not mean the end of nosology, like the Darwinian theory of the origin of species did not lead to the end of biological classification. The
development in the 19th century medicine led to the end of the symptomatic
disease classifications only.
We do not believe that medical classifications will be useless and become obsolete.
On the contrary, we expect better classifications in the future. The emergence of
ontologies (here in the sense of formal representations of entities and their
relations) opens the way to get rid of many consistency errors that necessarily
occur when either individual experts or committees construct classifications
without appropriate theoretical foundations. The goal is not to find the right
classes and super-classes any more. The goal is to find the primitives, and to
formally define all other entities based on the primitives. This is a challenging new
way, that has of course certain limitations (whether a formal system is theoretically
and practically decidable is always a question), but obviously new experiences will
be obtained. Formal ontologies can be considered as realisations of the dream of
Leibnitz ('to think will mean to count') in the sense that they make logical
reasoning computable.

5.4. Development of code value assignment
Early classifications sometimes used numbers to identify the entities, sometimes
not. The consistent use of these identifiers instead of the names of the entities is a
relatively late invention. The obvious aim of using codes is to get rid of language
dependency and ambiguity of names. Early health statisticians, like Farr, recognised
that inconsistent use of disease names often leads to statistical errors. Even today it
is often thought, that synonyms and homonyms are the major obstacles for
comparing data and for the interoperability of heterogeneous information systems.
While this is only one – and perhaps not the most serious – aspect of the problem,
use of codes in computerised data processing has well known obvious benefits.
Whenever (either numeric or alphanumeric) codes are to be assigned to the
categories of a classification, the question emerges, how this assignment has to be
done. Linnaeus used a simple sequential numbering in the Genera Morborum. But
once a categorical system is organised, it seems to be reasonable to use code values
that express the structure of the system (hierarchy, combinations). Such a tendency
is present in many current medical coding systems, like ICD and SNOMED
version 3 for instance. This code value assignment strategy has obvious benefit
both for human users and computers. It makes it easy to move up and down in the
hierarchy, and to decide whether a hierarchical relation holds or not between two
entities, etc.
However this assignment strategy often leads to irrational constraints: it makes it
difficult to represent poly-hierarchies; and causes a restriction on the number of
possible subclasses because of the limited number of available characters. To get
rid of this limitation Read codes version 2 uses both numbers and letters and
discriminates upper and lowercase letters to be able to represent up to 62
subclasses for each entity. (Note that Read codes – like the Real Character of
Wilkins – are organised in a way that each single character has actual meaning.)
More complex coding systems developed recently are not intended for manual use.
Once human usage is not the goal, this "expressive" code assignment strategy can
be replaced by other solutions. For computers all strings are meaningless but string
searching can be done very fast. So to find a super-ordinate class can be done
easily not only by manipulating the code (e.g. removing some rightmost characters) but also by looking up in a table that describes the hierarchical relationships. Let us show this with the example how ICD and SNOMED CT represent bacterial and staphylococcal pneumonia. (Table III-3)

<table>
<thead>
<tr>
<th>Entity</th>
<th>ICD code</th>
<th>SNOMED CT concept ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacterial Pneumonia</td>
<td>J15</td>
<td>53084003</td>
</tr>
<tr>
<td>Pneumonia caused by staphylococcus</td>
<td>J15.2</td>
<td>22754005</td>
</tr>
</tbody>
</table>

Table III-3 Representation of two forms of pneumonia in SNOMED and ICD

From the ICD codes it is easy to infer from J15.2 that it is a subordinate category of J15, simply by removing the two last characters (.2) from the code of staphylococcal pneumonia. Such a simple formal inference is not possible with SNOMED identifiers. In case of SNOMED a relational table or some other data source should be used to list all subsumption hierarchies. But computers can process such sources in a fast and effective way, so there is no absolute need for expressing the hierarchy formally by code values. Using a separate data source instead of expressive codes to describe the hierarchy has the following benefits:

- It is easy to represent poly-hierarchies
- There is no limitation on the number of subdivisions for a single category.
- It gives freedom to modify the structure of the system without changing the codes.

For instance, Barry Marshall and Robin Warren (Nobel Laureates in physiology in 2005) discovered the fact, that gastric ulcer is caused by Helicobacter pylori. Earlier it was thought to be a usually psychosomatic disease. Consequently gastric ulcer should be moved from its current place to gastro-intestinal infections. Or when the HIV virus was identified as the etiological agent of AIDS (Acquired Immune Deficiency Syndrome), this condition had to be moved from immunological disorders to viral infections. In case of expressive codes such situations result in the change of code values due to the change of our knowledge but without change of the meaning. If the hierarchy is represented in separate tables, it is enough to correct the corresponding rows while the meaning of the codes remains the same. But even this does not mean that random code assignment is the best choice. The problem of code value assignment can be well demonstrated by the example of ASCII codes that are used to represent letters in computers. ASCII codes are arranged in alphabetic order: the ASCII code of A is 65, B is 66, C is 67 etc. The code of a lowercase letters is equal to the code of corresponding upper case letter + 32. This arrangement has the following benefits:

- It makes it easy to order strings (e.g. names) in alphabetic order, since it follows the order of corresponding code numbers.
- It makes it easy to convert strings from upper case to lower case and vice versa.
- Since 32=2^5 in case of binary digital representation the rightmost five digits are equal for all upper and lowercase letters (e.g. the binary code for 'A' is 0100 0001, for 'a' is 0110 0001). That makes case insensitive processing easy.
Chapter III – Cultural history of classifications

All these things work properly with the English alphabet. As soon as other languages, like French, German, or Hungarian have to be dealt with, an extension of the original ASCII code table should be used and the above mentioned systematic arrangement disappears. In the Hungarian alphabet for instance 'A' is followed by 'Á'. In the so called 852 code page the corresponding codes for 'Á' and 'á' are 181 and 160. So, if there is a need to arrange strings according to the Hungarian alphabet an auxiliary table should be used that describes the order of letters. The same happens when additional tables describe the hierarchy (i.e. a partial order) of classes in case of some current medical classifications.

The history of the development of code value assignment shows two things:

1) If code values convey some additional information beyond the direct meaning of an entity, there is a risk that in case of change of our knowledge some code values should be changed without changing of the meaning. Therefore use of 'mute' codes that merely identify an entity without conveying additional information is recommended, at least in case of generally used systems.

2) Certain value assignment strategies can support effective computing, but this depends on the given task. In local applications this can be a useful technique.

5.5. Lessons from the history

'Historia est magistra vitae' (history is the teacher of life) says an ancient Latin sentence. Understanding that certain problems persisted over centuries may help us to redefine our research targets to avoid everlasting disputes that lead nowhere. Nearly all the problems we face nowadays in the fields of medical terminology, classification, coding systems and knowledge representation already emerged in the past. What we are looking for – currently and throughout the centuries – is a tool that enables us to represent the world, at least the medically relevant phenomena in an unambiguous, comparable, relevant, and – since computers emerged – processible way. The representational techniques used currently and in the past fail to satisfy the entirety of these requirements. I try here to summarise the reasons of failure:

The problem of size: none of the medical classifications and terminological systems created so far was complete. Building large terminologies and classifications is a time and resource consuming enterprise. The necessary consensus building also requires a lot of time, while medicine is changing rapidly.

The problem of representation: before large scale classifications could achieve an acceptable level of comprehensiveness the used representation became obsolete. The first classifications were represented in plain paper based tables, later converted to simple database table structures. Then relational databases, conceptual graph formalisms, various knowledge representations (description logic – DL and first order logic – FOL languages) emerged. Recently the development of web technology seems to be a driving force for standard implementations of these languages; OWL (Web Ontology Language) is the most common example. Not only the representation but medicine itself develops fast: new terms and categories emerge; old ones become obsolete at a speed that is hard to follow.
Structural problem: all approaches in the field provided a structure to arrange the entities. But none of the proposed structures seems reasonable and convincing to everyone. All structures appear to be somewhat arbitrary and accidental. When someone looks at such a structure it is always questionable why this and not another structure was used. The worst case is when the structure represents a compromise of different aspects, resulting in a hotchpotch of ‘fundamenta divisionis’.

The poly-hierarchy problem: as we mentioned, the human mind apparently strives to see tree structures in all hierarchies, while reality is more complex and seems to form lattices instead of trees: nearly always there are several options to find a basis of divisions. Before the age of computers it was simply too hard to represent poly-hierarchies. Computers can easily cope with them. But perhaps human thinking still adheres too much to paper based representations, or human thinking is genuinely bound to tree hierarchies, and we intuitively refuse more complex structures. In principle it is easy to say, that at least at one level of divisions we must not mix more than one discriminating criterion. But apparently the consistent application of this simple rule in a large system often fails. Today in modern ontology engineering it is a modelling challenge whether manual assertion of poly-hierarchy is allowed or not. Usually it is recommended that users should create only a tree structure manually and additional relations should be inferred by automated reasoning.

Use and reuse: it is relatively easy to create a classification that serves one definite goal satisfactorily. We saw historical examples that data collected for one specific purpose under certain circumstances and with certain limitations can be used for a different purpose. But the general rule is that the better the classification fits one purpose the less chance of its reuse for another.

Conflict between representational power and ease of use: coding systems with a finer granularity allow the representation of more details and to express rich semantic content. But increasing the representational power can only be done at the expense of ease of (human) use, and usually also at the expense of increasing coding and classification errors. The problem of coding errors was studied in a number of publications. A review of them was presented by the author (Surján, 1999). Here I would like to show only the historical aspect of validity: The practical use of disease classifications stems from mortality statistics. Individual errors in mortality classification can not directly harm the individual, since the subject of the classification already died. John Graunt showed already in the 17th century, that it is possible to get correct conclusions from statistical data even if there is a high number of individual errors. When the classification was expanded to morbidity statistics, individual harm was still not possible since the data were not used in the treatment process. But this usage of classifications opened the way towards using codes in individual patient management. The inherited potential of errors then became a source of possible danger to individuals.

The importance of the human factors: The story of the London Bills of Mortality has shown that the reliability of classification data strongly depends on human factors, including the knowledge of the person who classifies the cases, and also the financial incentives. Nowadays many automated or semi-automated coding software packages are available, promising that human factors can be eliminated in the future. However none of them have general acceptance. None of them are
perfect but many of them exceed the performance of human coders. Still the general opinion is that we should trust humans more than computers. We have no space here to discuss the reasons for that. But it is very clear that the classifications will become more and more complex, and humans will not be able to manage them, so coding (or more precisely translating medical text to some formal language) should be left to computers in the future. This is true for medical classifications also. There are hot issues today in this field that require some solution. But we can not be sure that these problems ever can be solved.

5.6. Today and tomorrow

More than thirty years after the publication of Fisher-Homberger’s sceptic paper about the end of medical classification (Fischer-Homberger, 1970) we have to see that, in spite of her prediction, the story has not ended. It is true, that clinicians are not strongly interested in large, comprehensive disease classifications. Their attention turns to the details: they classify forms and variations of a disease: types of lymphomas or fractures of a bone etc. I would use the term 'micro-classification' for them, since their scope is rather limited, and they usually consist of a small number of classes. The importance of comprehensive disease classifications shifted from clinical medicine to public health and health information science. And the development did not stop. WHO is preparing the 11th version of ICD (Üstün & al., 2007) and there is a movement to introduce an international standard medical terminology based on SNOMED-CT. Unfortunately neither GALEN nor SNOMED-CT is represented using some standard formal language. Currently the most popular formal language is OWL (the Web Ontology Language) that was designed to be used in World Wide Web technology. There are some experiments to transform GALEN from its original language (GRAIL) to OWL (Héja, Surján, Lukácsy, Pallinger, Gergely, 2007). Top level ontologies are developed to ensure the consistency and inter-operability of various domain ontologies. Certain domains appear to be crucial for nearly all biomedical domains. Anatomy is a good example. In such domains re-usable, foundational or reference ontologies are developing, that could be built in several applications. FMA, the Foundational Model of Anatomy is perhaps the best known example (Rosse, C., Mejino, J.L. Jr., 2003). The rapid development of molecular biology (genomics and proteomics) produces a huge number of newly described entities (gene and protein sequences, cell-components, molecular functions etc.) that are arranged into the fast growing Gene Ontology, GO for short (Ashburner, et al., 2000). The word 'classification' is used perhaps less frequently for these systems. The current buzzword is 'ontology' instead. This reflects two important considerations:

1) It is not enough to create and order classes; categories require formal definitions.

2) The term 'ontology' comes from philosophy, and expresses a sort of philosophical commitment. This means that categories should not be based on human ideas or simply observable properties, but on the substantial properties of the entities.

These statements outline current trends and movements in the field. In any branch of informatics it is hazardous to try to predict the future. In some years dramatic
and totally unexpected changes can happen. And – as we stated above – medical classifications shifted partly to the field of information science. The development of medical science is also speeding up, and it also contributes to the unpredictability of the future of medical classifications. But based on the current trends we can draw a picture, without being sure that this will really happen. According to this picture in the foreseeable future we can not expect one single, robust medical classification, but perhaps we can develop a system of inter-operable classifications. My expectation is to get a generally accepted standard top level ontology that ensures the consistency of all other classifications, and a number of standard foundational ontologies that are used as building component of various application specific ontologies used for different practical purposes. This is illustrated in Figure III-8

Figure III-8 Synergy of ontologies

Whatever will happen in the closer or more distant future, it is obvious that we always have to keep in mind, what happened in the past. This was my intention with this chapter.

Acknowledgements
The author has to express his special thanks to Arie Hasman for his valuable comments, and all members of Health Ontology Forum, a working group of the Medico- biological Section of John von Neumann Computer Society, for their contribution to the fertilising atmosphere in which this work could be carried out.
Chapter III – Cultural history of classifications

References


Chapter III – Cultural history of classifications


Üstün, B., Jakob, R., Çelik, C., Lewalle, P., Kostanjsek, N., Renahan, M., Madden, R., Greenberg. M., Chute, C., Virtanen, M. Hyman, S., Harrison,


Appendix

A short inventory of internet resources related to current medical classifications

ATC/DDD Drug classification http://www.whocc.no/atcddd/
DSM Diagnostic and Statistical Manual of Mental Disorders http://allpsych.com/disorders/dsm.html
HCPCS Health Care Procedure Coding System http://www.cms.hhs.gov/MedHCPCSGenInfo/
ICD International Classification of Diseases
  ICD-O-3 International Classification of Diseases for Oncology, Third Edition
  ICD-10-NA Application of the International Classification of Diseases to Neurology
  ICD-10 for Mental and Behavioural Disorders
  ICD-DA Application of the International Classification of Diseases to Dentistry and Stomatology
  http://www.who.int/classifications/icd/en/
ICD-10-PCS American procedure codes
  http://www.cms.hhs.gov/ICD9ProviderDiagnosticCodes/08_ICD10.asp
ICECI International Classification of External Causes of Injury
ICF International Classification of Functioning, Disability and Health
  http://www.who.int/classifications/icf/en/
ICHI International Classification of Health Interventions -- under development
  http://www.who.int/classifications/ichi/en/
ICPC-2 International Classification of Primary Care
LOINC Logical Observation Identifiers Names and Codes
  http://www.regenstrief.org/medinformatics/loinc/
MedDRA Medical Dictionary for Regulatory Activities
  http://www.meddramsso.com/MSSOWeb/index.htm
MeSH http://www.nlm.nih.gov/mesh
NANDA North American Nursing Diagnosis Association International Taxonomy
Technical aids for persons with disabilities: Classification and terminology ISO9999
  http://www.who.int/classifications/icf/iso9999/en/
SNOMED
  SNOMED International
  SNOMED Clinical Terms http://www.snomed.org/
Chapter III – Cultural history of classifications

TNM Classification of tumour size, lymph node involvement and metastasis
http://www.uicc.org/index.php?id=508

UMLS Unified Medical Language System
http://umlsinfo.nlm.nih.gov/