# Using the GRAIL language for Classification Management

P.E. Zanstra[1], E.J. van der Haring[1] , F. Flier[2], J.E. Rogers[3] , W.D. Solomon[3]
[1]*Dept. Medical Informatics, Epidemiology and Statistics, University of Nijmegen, NL*
[2]*WCC, National Centre for Medical Terminology, Zoetermeer, NL*
[3]*Medical Informatics Group; University of Manchester UK*
*http://www.ehm.kun.nl/mi/galen or email:Galen@mie.kun.nl*

**Abstract**. This paper describes a novel approach in classification management where a formal model of medical semantics is being used for manipulations on existing classification systems. The paper addresses the issue of semi-automatically making specialist classifications that are compatible with the source classification. The examples in this paper are from a limited domain. At the time of the presentation results will be shown of the present modelling work within the GALEN-In-Use project. The model will then contain several thousands of medical procedures from four different classification centres.

## 1. Introduction

In medicine, standard classifications, such as ICD and ICPM, are typically intended for a wide range of users and settings. Consequently they are not adapted to the needs of any particular user or the circumstances of any particular setting. To make standard classifications more attractive to users, they often have to be adapted in some way. Adaptation may involve selection of relevant classes, refinement of the classification by adding relevant subclasses, and rearrangement of classes in a more appealing order. Rearrangement, in particular, has to be done with care, to avoid breaching the compatibility with the standard classification. To date, production of compatible adaptations is mainly done manually, by means of a word processor, which can be a very time-consuming and thus costly task. The last version of ICD has taken 15 years to come to deployment. The translation and local adaptation takes an additional number of years, and a lot of effort. Consequently new ways are being sought to support this presently mainly manual craftsmanship.
Within the GALEN-In-Use-project a Classification Manager (ClaM) is being developed to assist Classification centres with their task of producing compatible adaptations and reliable conversions between different classifications. In addition the ClaM will support the translation of classifications.
In the third year of the GALEN-In-Use project a major validation of this approach will take place by nine Classification Centres in nine different countries

## 2. The Core Reference Model

*GALEN* aims to build a compositional generative model for medical terminology. This model comprises a well defined ontology of atomic medical entities with rules to combine these entities such that all and only sensible medical expressions can be generated. This means that with a relatively small model in principle billions of medically sensible expressions about patients can be made. Ultimately it is the intention that GALEN covers all of medicine. The language in which this knowledge is represented is called GRAIL, the GALEN Representation And Integration Language. The model of this knowledge is called the Common Reference Model, or CRM for short. The present version of the CRM comprises some 7000 elementary entities with about 15000 links. It is expected that to cover a general layer of medicine some 25000 elementary entities are required. [1,2,3]

## 3. Classification Schemes

The ClaM stores a classification scheme in a hierarchy consisting of *classes*. Each class consists of a *code* and one or more rubrics. A rubric can be further specified by its *kind* (e.g. *preferred*, *includes*, *excludes*) and a *language* (e.g. *Dutch*, *English*). For example the ICD-10 class A18.61 could have the rubrics:
Dutch preferred “Tuberculeuze otitis media”
Dutch synonym “Tuberculeuze middenoorontsteking”
Dutch note “Not sure whether the Dutch synonym is OK”
English preferred “Tuberculosis of ear”

A class in one classification scheme can reference a class in another classification scheme. This can, for example, be used to link a class in an existing classification scheme to a specialist classification scheme. Similarly, a class in a classification scheme can reference a concept in the CRM. The latter references are called *mappings* to differentiate them from references between classes in classification schemes. Both *references* and *mappings* are specified by their *kind*, e.g. *isEquivalentTo*, *isBroaderThen*, etcetera.

The ClaM supports the following functionality:

- *Activities involving a single classification scheme*
  - de novo creation
  - creating, modifying, moving and/or deleting classes
- *Activities involving multiple classification schemes*
  - creating and maintaining relationships between different classification schemes and/or different versions of one classification scheme
  - copying classes from one classification scheme to another
- *Activities involving a classification scheme and the GALEN CRM*
  - mapping classes of a classification scheme to concepts in the CRM
  - making selections of existing classification schemes knowledge in the CRM
  - extending existing classification schemes using knowledge in the CRM
  - rearranging existing classification schemes using knowledge in the CRM

## 4. Operations on a Classification Scheme

Medical specialists have specific requirements for the terminologies they use. They find the existing systems mostly not sufficiently detailed, or having a lack of precision. They

also do not want to be bothered by terminology outside their own field of specialty, at least they do not want all the high detail outside their own domain. Making specialist classifications is tedious and therefor expensive, as it mostly can not be performed solely on the basis of the classifications own structure, nor by simple string matching procedures. Experience at WCC has shown that such approaches yield only 60% of the target specialist classification scheme (5). The remaining parts are gathered by hard work.
Below we will outline a number of operations that are performed using the CRM as a backbone for manipulations. The examples here are meant to be illustrative for the kind of operations that are possible with this formalism. In the presentation we will show real examples of manipulations on the classifications of medical procedures. At the time of this writing four classification centres are working on the analysis of those classifications.

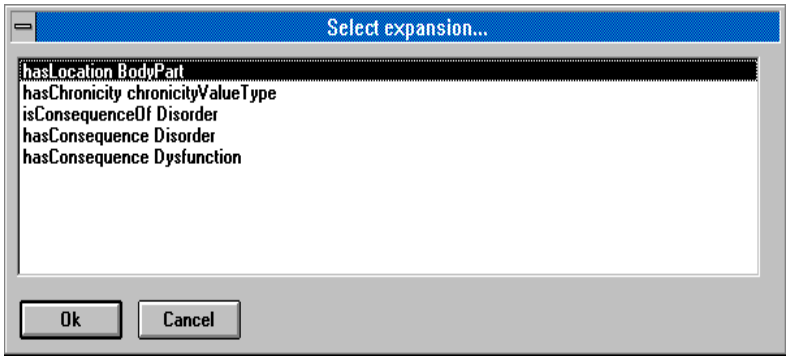


figure 1. The sensible statements for the refinement of the concept *Cellulitis of the external ear* (ICD-10) according to the GALEN CRM.

## 1 Refinement of a Classification Scheme

The ClaM can automatically create sub-classes for a class in a classification scheme based upon the concept in the CRM to which the class is mapped, and its sensible particularisation's in the CRM. For example suppose the class *H60.1 Cellulitis of the external ear* (ICD10) is mapped to the concept *(Cellulitis which < hasLocation ExternalEar >)* in the CRM. You can then select a statement from the list of sensible statements for the concept (figure 1) which will be used to create refined sub-classes for *H60.1*.
The entry *hasChronicity chronicityValueType* for example means that it is sensible for the concept *(Cellulitis which < hasLocation ExternalEar >)* to be refined using criteria of the form *hasChronicity chronicityValueType.* Upon selection of *hasChronicity chronicityValueType* the ClaM will generate the corresponding refinements e.g.,

- (Cellulitis which < hasLocation ExternalEar hasChronicity chronic >)
- (Cellulitis which < hasLocation ExternalEar hasChronicity acute >)
- (Cellulitis which < hasLocation ExternalEar hasChronicity subacute >)

In addition, the ClaM will generate subclasses below *H60.1 Cellulitis of the external ear* and map them to the new concepts. The rubrics of these new classes will be generated by the ClaM's natural language generator. This will result in these classes:

- H60.1.0 Chronic cellulitis of the external ear
- H60.1.1 Acute cellulitis of the external ear
- H60.1.2 Subacute cellulitis of the external ear

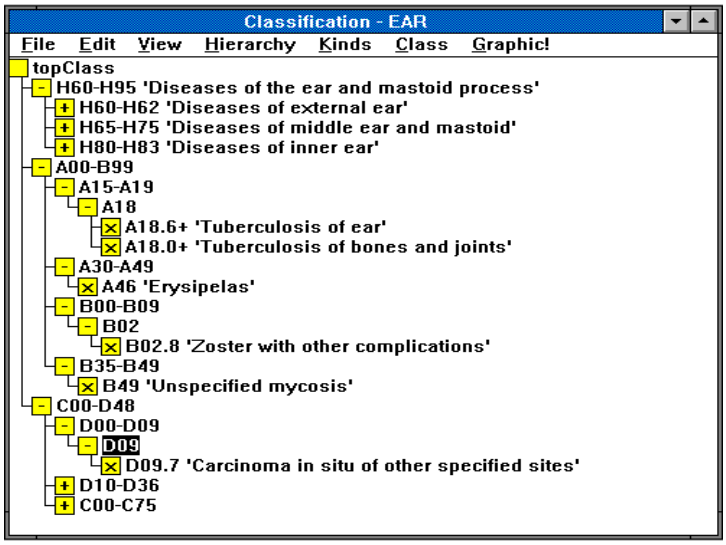


figure 2. En example of a selection from ICD10 based upon (Disorder which hasLocation ear). In the hierarchy an 'x' in the square box means this is a leaf node. Squares having a '+' mean that they have children underneath, which may be shown with a single click of the mouse on that box.

## 2 Selection from a Classification Scheme

The ClaM can select a number of classes from an existing classification scheme and copy them into a new classification scheme. An example of making a selection is to copy all the classes in the ICD10 that involve a disorder of the ear (figure 2) to a new classification. The ClaM collects all the descendants of e.g., *(Disorder which < hasLocation Ear >)* from the CRM, this will all be disorders that involve the ear or parts of the ear. Then the ClaM looks up all the classes in the source classification that map to any of these concepts. These classes and their ancestors will be copied to the target classification. The ClaM will also create the mappings between the classes in the new classification scheme and the CRM. You can modify the new classification as required, e.g., by adding or removing classes and rubrics.

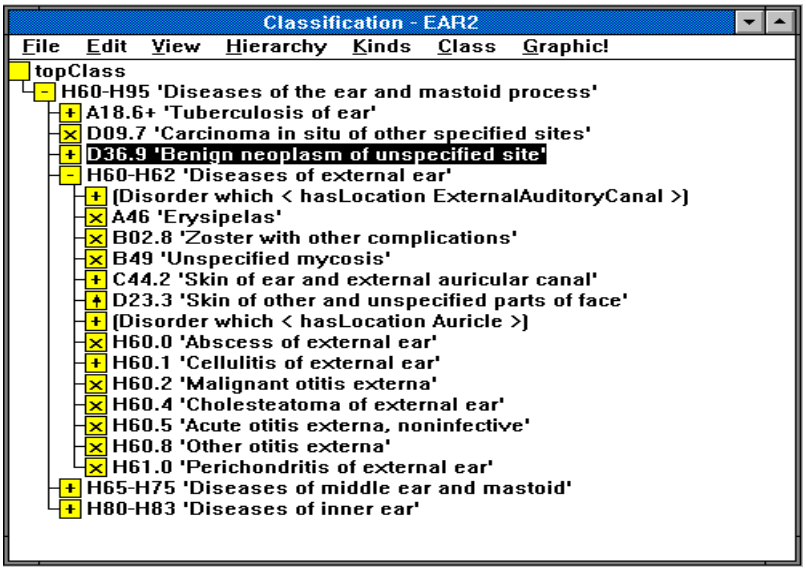


figure 3. An example of a re-arranged part of the ICD10 based upon the concept hierarchy below 'Disorder which hasLocation Ear'. Where there is no rubric in the original classification ClaM fills in the Grail concept (e.g.(Disorder which <hasLocation Auricle)). The next version will generate a natural language string for that GRAIL concept

## 3 ReArranging a Classification Scheme

This is similar to the previous operation, the main difference is that in *Selection from a classification scheme* the hierarchical structure of the original classification scheme is preserved in the new classification scheme. When the ClaM rearranges a classification scheme, the hierarchical structure of the new classification scheme reflects the structure in the CRM. For example, you could rearrange all the classes of the ICD10 that involve disorders of the ear. In ICD10 you will find such classes in several chapters. Again, the ClaM first collects all the descendants of *(Disorder which < hasLocation Ear >)*. The hierarchical structure of these concepts is copied to the new classification scheme. Then the ClaM looks up the classes in the source classification that map to these concepts, and copies their codes and rubrics to the new classification scheme.
In the example all the classes involving disorders of the ear from the ICD10 have been re-arranged on the basis of the subsumption hierarchy in the CRM. Note that this is a multi-axial hierarchy. It contains a hierarchy by type of disorder (A18.6+; D09.7; D36.9) , and a hierarchy by topography (H60-H62; H65-H75; H80-H83).

## 5. Conclusions and Future Developments

The results given in this paper suggest that the GALEN CRM can be used for selection and refinement of specialist classification schemes. It must however be realised that the present CRM is only covering a small portion of medicine. Also there are only small sections of systems like ICD and SNOMED being mapped to the CRM. Therefor it is too early now to draw final conclusions. Present work on medical procedures in the GALEN-In-Use project should however give the answers of the usefulness of the GALEN approach for classification management purposes.
Future developments in the ClaM are that it will support the generation and analysis of natural language expressions. Natural language generation will take place in the present project period, analysis is pending on separate funding. Later this year the ClaM will be integrated with existing GRAIL based natural language generators. In the second quarter of 1997 we will assess the first results of language generation with the individual centres. By the end of 1997 natural language generation is planned for Dutch, English, Finnish, French, German, and Italian. This will allow the ClaM to automatically translate a classification scheme from one language into another.

## 6. References:


1. Rector, A. (1994). Compositional models of medical concepts: towards re-usable application-independent medical terminologies. in Knowledge and Decisions in Health Telematics P. Barahona and J. Christensen (ed.). IOS Press. 133-142.

2. Rector, A., W. Solomon, W. Nowlan; T. Rush; W. Claassen; and P.E. Zanstra (1995). A Terminology Server for Medical Language and Medical Information Systems. Methods of Information in Medicine, Vol. 34, 147-157

3. Rector, A. and W.A. Nowlan (1993). The GALEN Representation and Integration Language (GRAIL) Kernel, Version 1. The GALEN Consortium for the EC AIM Programme. (Available from Medical Informatics Group, University of Manchester).

4. Anonymous (1995) Draft WCC Methodology for the development of specialist classifications.(in Dutch)

5. Anonymous(1993) Draft WCC standard classification on behalf of Ear, Nose and Throat Medicine, adaptation of the ICD-10 (in Dutch)