WORLD HEALTH ORGANIZATION

MEETING OF WHO COLLABORATING CENTRES
FOR THE FAMILY OF INTERNATIONAL CLASSIFICATIONS

Cologne, Germany

19-25 October 2003

Title: Applicability of Dutch ICD-10 Electronic tool for publication of ICD-10 updates and control of derived and related classifications

Authors: Huib Ten Napel, WHO-FIC Collaborating Centre in the Netherlands, Bilthoven & Dept of Medical Informatics, University Medical Centre Nijmegen.

Purpose: for information and discussion

Recommendations:

Abstract:

Production and dissemination of ICD-10 updates as a word processor file is a time consuming activity. It is a laborious work for the Centre responsible for the updates, and it is laborious for the end-users, implementing such a paper-based update, especially when this update has to be processed into a database. Control of a derived or related classification is not so much a laborious work, if it is already available as a database. The major question in this case is how to reach consistency and unambiguousness in the use of terminology, so that aggregated data from that classification is comparable with the reference classification.

In order to cope these different problems the Dutch Centre is using an electronic tool, the Classification Manager (ClaM) within the Classification Workbench, for the production of the electronic versions and the XML (ClaML) versions of ICD-10 and ICF. As the tool is developed in close cooperation it offers a number of facilities required by the Centre, such as comparison of versions, a history record of changes made, an export function for several formats, such as a pre-structured pdf, etc.

In our presentation we will explore the possibilities of the electronic tool for update-messages in ClaML, deriving classifications from source files and mapping of related classifications from a national point of view.

Background

Present paper based versions of ICF and ICD-10 have been processed, using different text processing tools, ranging from WP5.1 to Word 2000, sometimes even used by the same person. For updates of ICD-10 and derived classifications, such as derivations of ICIDH and ICF this is not much different. It is a laborious and time-consuming process for everyone involved, with a high risk factor for errors.

These text files form the source for database construction, which again is a laborious work. Not to mention the effort that is has to be made when updates of the classifications are published.

Another time-consuming work is making ´crosswalks´ between related classifications. These crosswalks are in most cases, manual mappings of ´possible´ related classes of two or more classifications. Here again with a high risk for errors and misinterpretations.

At one hand text processing tools are not equipped for development, version control and management of classification. Text processing is about presentation and layout of a text file.

At the other hand tools for database construction are for, depending on the specific tool, defining simple or more complex relations between data. But not equipped with the requirements for classification construction, and certainly not for presentation and layout of classifications.

Until recent our classification work required the use of many different applications. Working with different applications, and with different people working separate on one classification increases the chance of inconsistency and ambiguousness.

The Dutch Centres expectation is that a centralised, coordinated working in one central source file, using a standard structure for classification will improve the consistency and unambiguousness of classification products, and support the complex classification work importantly.

The Dutch electronic tool

To cope with the different challenges for present and future classification work, the Dutch Centre is now using an electronic tool, called Classification Manager (ClaM). ClaM is part of the Classification Workbench (ClaW), a sophisticated tool for formal modelling, analysis and representation of classification schemes, by means of a reference terminology and for building a local reference terminology. For its work the Centre is focussing on the ClaM first, as this is the tool that covers the functionality we require.

First of all there is the central source file. This is an under construction or already completed classification. This central source file can be edited ´de novo´ from scratch, or imported from a pre-structured text file (EBNF-format or any other tagged .txt file).

In ClaM the technical specifications of CEN/TS 14463 (Classification Markup Language) have been implemented, meaning that working in ClaM is in accordance with this CEN Technical Standard.

ClaM stores classification schemes in an electronic form that preserves the internal structure of the classification. It explicitly represents the rubrics, codes, and the hierarchy that make up the structure of the classification. ClaM provides the flexibility that is required to represent a wide variety of classification schemes in a uniform way.

ClaM has several classification manipulation functions. These functions will be summarised and not explained in detail.

- Edit functions such as: replace, move up move down, etc.
- Standard operations that allow to add, edit, find, delete or move a class
- More complex operations to collect, sort or shift children classes
- Add modifiers and assign these modifiers to classes
- Summarize existing classes under a new parent class.

ClaM offers several views on the classification, you can center the classification on a certain class, expand one level, expand a branch, collaps levels and branches, show the classification from top, and spawn the classification. If the classification is multilingual you can change the language as desired.

After a classification has been completed, it can be indexed as desired. The Class and Rubric structure allows inclusion or exclusion of terms on every level. Also other indexes or thesauri can be imported.

The Referrer tool checks the text of the rubrics in the classification for references to other codes in the same classification. When it finds a reference it adds the ClaML <Reference> tags around the code. These references are shown as clickable hotlinks in the rubrics window of ClaM.

The Comparator tool compares the currently opened classification with another classification. For every class in the classification it check if that class is also present in the other classification. The Comparator reports if a class is absent. If the class is present in both classifications, the Comparator checks if they have the same parents and rubrics. At the end, a summary is written.

ClaM also keeps a complete history of changed made in the classification file.

The central .cla file can be exported in different formats. The ClaML format, which is an XML format, an old ClaM format for compatibility with former versions of ClaM, a comma separated, a tab separated format and RTF format. The RTF format can be structured by means of a style sheet. The style sheet can contain the layout aspects of the printed version of the classification.

It is also possible to export the classification on one, two, three or more digit levels, with or without selected rubrics.

So far it may be obvious that the Dutch Centres requirements are met, the functionality of ClaM is partly due to the close cooperation between the developers and the Dutch Centre.

Exploration of possibilities

Update messages

In the Netherlands the ICD-9 CM is used for statistics and most hospitals use it for registration purposes. The ICD-9 CM is also used by specialist groups to define the diagnose part of the DBC (Diagnosis-Treatment Combination, which a kind of Case Mix). The ICD-10 has not yet become obligatory for registration in health care and it is questioned if it will. Still a number of paper based ICD-10´s in a loos-leaf system have been distributed, along with a high number of CD-Roms with ICD-10 as a printable Word-file. For the loos-leaf system the Centre has issued updates including1999.

Database files of ICD-10 have not been issued by the Centre.

Depending on the application, almost every interested party has its own set of requirements for the database. The construction of the databases is the responsibility of the interested parties.

The expectation is that in future there will be less demand for paper based classifications and an increasing demand for electronic versions of the classification. With the formats the Centre is able to produce, the quests for electronic formats can be met. The question is how the updating will take place.

There is the possibility of making the complete new version in several formats available on the Centres website. There is also the possibility of update messages.

As described, ClaM has a Comparator tool with which one version of a classification can be checked against a next version. The result is a summary of differences. This summary could be used as an import to update the present classification. It requires an import and update mechanism and a standard for the update message. The present CEN 'prEN 13609-1, Updating of coding schemes', is a candidate.

Derived Classifications

Most leading bodies of health care workers have adopted WHO classifications like ICF and ICD, and even ICPM as a standard terminology for registration, documentation and exchange of information within their work processes. In most cases the ICF, ICD, and in some cases ICPM are sufficient for registration of data on a broader level of aggregation. Discipline specific, patient documentation requires a higher level of detail than the present classifications offer. The Centre is pleased with the fact that most disciplines try to be compliant to our Dutch versions of WHO Classifications. The question however is how these derived classification schemes, not only fit into the original classification schemes on term level but on structure level as well, as these derived classification schemes are paper based and developed in an unstructured environment. In best cases developed, but sometimes only supervised by experts. It is a hard job to making derived classification schemes

compatibel with the original afterwards. Merging the two classification schemes would be a time consuming and complicated work.

In most cases specialist disciplines need more detail in depth of the existing scheme and only is some cases additions on existing levels. ClaM offers the possibility of working directly in the central source file with the focus on specific classes. This means that new classification schemes can be developed as an extensions to the original, leaving the original structure in tact.

Working this way has the advantage of working in a structured and controlled enviroment. There is more the emphasis of making explicit what needs to be added and the position where it needs to be placed. At the same time the history of changes is documented within ClaM, which makes changes and decisions on changes traceable and controllable. Via the compare tool the additions can be made visible for comment.

The question is how the editorial process should be sructured? Is there one central editing or a parallel editing?

Mapping

There are many efforts made to produce crosswalks between two or more versions of a classificaton scheme and also between two different classification schemes. In publications on the terminological aspects of classification schemes, especially concerning the formalisation of classes, mapping is a hot item. Presumably there is a notion of mapping. The question is, what is a map, and does it exist?

In crosswalks a class in classification scheme A is manually 'mapped' to a class in classification scheme B. What does this actually mean? Are this the same kind of things? Are this just the same terms? Do they have the same referential meaning? If so, what is the meaning and how has this been proven? The problem is that on this level of association it cannot be proven.

Can we use ClaM to make mappings between classes of successive classification schemes, or between similar terms of different classification schemes?

Here the same question arises, what is a map?

ClaM can compare classification schemes, but only on the level of codes and text strings. Not on the level of semantics, referential meaning, etc.

To understand what a class in a classification scheme means we can use other tools within the ClaW, such as SPET decompositions (explicit definitions) to analyse the scheme. We also need a reference model (RM) or add required formalisms to the existing RM, using the GCE and GRAIL. Than we can map the classes to the RM and see where they classify within the RM. Next we have to analyse and decompose classes from other classification schemes, expand them to the RM and see where these classes end up. If they appear next to each other, does that mean there is a map, is it a close fit, and how close?

ClaW contains the tools we need to formalise classification schemes. It offers us the possibilities to explore the formal relations between classes. It also offers us the

possibility to merge classification schemes, and in the best scenario a close fit of two or more classes.

But first we have to understand what mapping is.

We can also just settle for connecting two things we think are the same, but why would we?