

**MEETING OF WHO COLLABORATING CENTRES
FOR THE FAMILY OF INTERNATIONAL CLASSIFICATIONS**

Tunis, Tunisia
29 Oct. - 4 Nov. 2006

How to Migrate to ClaML?

Stefanie Weber, Can Celik, Susanne Bröenhorst, Michael Schopen

Abstract

At the last WHO-FIC-Meeting in Tokyo, Japan, it was agreed to use the XML-schema based on ClaML (Classification Markup Language) as a standard exchange format for WHO-FIC classifications. As well, new maintenance tools for Classifications should use ClaML as underlying data format. Existing tools should integrate an import and export format to ClaML to enable international comparison of clinical modifications.

There are many different ways established to maintain the WHO-FIC classifications and therefore different approaches are needed to migrate from the former format to ClaML or to establish an interface between the former format and ClaML.

This paper shows a couple of examples on how to migrate to ClaML and the benefits for doing so.

This document is not issued to the general public, and all rights are reserved by the World Health Organization (WHO). The document may not be reviewed, abstracted, quoted, reproduced or translated, in part or in whole, without the prior written permission of WHO. No part of this document may be stored in a retrieval system or transmitted in any form or by any means - electronic, mechanical or other - without the prior written permission of WHO.

The views expressed in documents by named authors are solely the responsibility of those authors.

Content

Abstract 1

Status quo 3

Obstacles 3

 Level of granularity 4

 Operating maintenance systems 4

Crossover 4

Benefits 6

 International support 6

 Comparability 6

 Use of the Maintenance and Publication Tool 6

Discussion 6

References 7

Status quo

Through presentations in the Electronic Tools Committee over the last years it became apparent that the need for an electronic maintenance of the classification arose in many countries. A questionnaire was sent out through the Electronic Tools Committee to assess the ways the ICD-10 was handled in electronic tools.

Some countries already had existing electronic versions and their maintenance in place. In a paper of Stahl et al. [1] in 2000 in Brazil two ways of classification maintenance were compared: The Australian way of maintaining the ICD-10-AM in a Microsoft Access® database versus the German way of maintaining the ICD-10 in an SGML-structured file. Whilst pointing out the differences the paper emphasized the importance of a flexible electronic way of maintaining the classifications. The main difference between both approaches was seen in the granularity: Beyond the code level the SGML-based approach provided considerably more information than the database approach, e.g. for the inclusion or exclusion notes.

In the 2001 meeting it was reported [2] that the US analyzed their classification maintenance and with the knowledge of the first paper chose neither Access® nor SGML for classification maintenance but rather an enterprise version of an SQL database system.

First reports on the ICD-10 in XML were given in 2003: In one paper the analysis of an XML-version of the ICD-10 available to WHO was presented [3]. Unfortunately this version of the ICD-10 was insufficient at that time. At the same meeting Hölzer et al. [4] pointed out the benefits of maintaining several language versions of the ICD-10 in one XML-file. They considered a previous version of ClaML and developed an XML structure to represent the ICD. They worked on the three digit code level with subordinated XML documents representing the finer details below the three digit level.

In a survey on electronic tools for the ICD-10 [5] conducted through the Electronic Tools Committee a couple of countries or Collaboration Centres replied to have electronic tools for the ICD-10 in place, some of them for mortality coding, others for morbidity use and some for maintenance of the classification itself.

These publications show that there is a lot of interest in electronic maintenance of the ICD-10 but there are different approaches in place up to date. These approaches mirror the different national requirements on classification maintenance and the historic developments in the different countries.

At the last meeting in Tokyo it was decided to use a redefined ClaML based schema as the standard exchange format for WHO classifications and to ideally even use ClaML in maintenance [6].

Obstacles

Of course such a recommendation holds its obstacles:

Level of granularity

Unfortunately not only the formats for the classification maintenance vary throughout the countries. Due to the different perspectives of classification maintenance countries chose different levels of granularity for the maintenance and different content as well. Whereas e.g. in Australia the special morbidity coding rules are maintained in the same system as the classification itself in Germany they are completely separate as they are handled through a different organization.

This and other differences resulted in varying structure of the classification within the files. It is now important to find the highest common denominator for each country's classification files and ClaML.

Operating maintenance systems

As stated above many countries do have a customized maintenance in operation that does fit their special demands. These tools most likely do not base on ClaML, most of them probably not even on XML. If they do want to switch their maintenance towards ClaML these countries would have to change their maintenance significantly which of course involves resources and time, both of which are scarce in most countries for such topics.

Crossover

For some countries the crossover to ClaML might be easy as they do maintain their classification already on a very granular level. In Germany the SGML files hold the classification in a very detailed way:

```
<SUBINH><I2><EINTRAG><EINSP>
<KOPF>Diabetic:</KOPF>
<LISTE><BMEHRSP SPZAHL="2"><LBSP RKLAMM="RRE">
<BATOM><TXT TYP="">Acidosis</TXT></BATOM>
<BATOM><TXT TYP="">Ketoacidosis</TXT></BATOM>
</LBSP>
<RBSP RKLAMM="RNEIN" LKLAMM="LNEIN">
<ATOM><TXT TYP=""> without mention of coma </TXT></ATOM>
</RBSP></BMEHRSP></LISTE>
</EINSP></EINTRAG></I2></SUBINH>
```

Example 1: Granularity of ICD-10 in SGML from DIMDI

Therefore it is easy for DIMDI to convert the ICD-10 to ClaML. As DIMDI maintains the original WHO-version for WHO Headquarters with its SGML-based maintenance tool, it will be easy to provide an English version of the ICD-10-WHO in ClaML as well. The French version of the ICD-10-WHO was received and corrected in SGML over the last years and is now available, too.

For others it might be possible to switch most of their content to ClaML automatically and to manually correct the rest of the classification towards ClaML. E.g. the fragmentation of the inclusion and exclusion notes in ClaML might pose a problem if they are not already maintained in a structured way.

```
<Rubric kind="inclusion" id="E04-0000385"><Label xml:lang="en">  
  <Fragment col="1" type="listhead">Diabetic:</Fragment>  
  <Fragment col="1" type="listitem">acidosis</Fragment>  
  <Fragment col="2" type="item">without mention of coma</Fragment>  
</Label></Rubric>  
<Rubric kind="inclusion" id="E04-0000386"><Label xml:lang="en">  
  <Fragment col="1" type="listhead">Diabetic:</Fragment>  
  <Fragment col="1" type="listitem">ketoacidosis</Fragment>  
  <Fragment col="2" type="item">without mention of coma</Fragment>  
</Label></Rubric>
```

Example 2: Sophisticated fragmentation of an inclusion note table in ClaML

This problem can be overcome if a country decides to switch its maintenance to ClaML as it is a one time only problem for these countries.

But it was as well recommended at the Tokyo Meeting that countries who are not switching their maintenance to ClaML are encouraged to implement import and export routines to their maintenance which handle ClaML. For these routines manually intervention with each export or import are a resource intense option and cannot be asked for on a regular basis.

Countries, who do not maintain their classifications automatically or are using a simple text format (like Microsoft word), might as well wait for the Maintenance and Publication Tool developed through DIMDI and WHO [7]. This tool will handle the classification in ClaML and will provide full support for classification maintenance and publication.

WHO is also planning to develop a tool called Translator's Tool for those who are maintaining a translation of the standard ICD-10. This tool will be simpler to use than the maintenance tool since the basic structure of the ClaML document will be exactly the same as the original English version. The tool will guide the user in the preparation of the translated version by displaying the structure and the original text. It will also ensure that the structure of the translation is exactly the same as the original. Once the translation entered using the tool, the Publication Tools mentioned above could be used in a similar fashion.

Benefits

International support

As WHO will maintain the ICD-10-WHO in ClaML, countries who maintain their classification in ClaML as well will be able to integrate updates and changes to the ICD-10 easily when provided. If a country is using a different language as provided through WHO at least the structure of the updates can be imported easily. This will avoid mistakes in national classifications and reduce the resources needed to implement the updates.

Comparability

Many countries have started to modify the ICD-10 of WHO towards national modifications. Examples are ICD-10-AM (Australia), ICD-10-TM (Thailand), ICD-10-GM (Germany). A comparison of the modifications in an electronic format is almost impossible at the time being. To enable such a comparison, which of course will not obviate intellectual interpretation of the findings, it is necessary to define a format which the comparison will be handled in.

ClaML handles multiple versions of a classification in one file and is therefore well suited to be used as the format for electronic comparisons. A file containing all clinical modifications at once can be the basis for the ICD-10-XM [8], which could serve as a basis for the next revision of the ICD.

But even for parts of the classification that have been identified to pose a problem in many countries in its recent subdivision the comparison in an electronic format could help to further identify the problematic codes and to serve as a basis for an update proposal.

Use of the Maintenance and Publication Tool

As stated above DIMDI and WHO are developing a classification maintenance and publication tool for the WHO-FIC classifications. This tool is based on ClaML. Most likely this tool will be provided through WHO in the next one or two years. Switching to ClaML will enable Collaborating Centres to start classification maintenance with the new tool right away and even to integrate older versions for generation of crosswalk tables with the tool.

Additionally with the new Publication Tool each collaborating centre will be able to produce a variety of formats for publication: HTML, ASCII, PDF, Crosswalk tables,...

Discussion

Beyond doubt there is some benefit for each country to switch to ClaML. International work can be progressed and the revision of the ICD technically supported. The benefit multiplies if more countries switch to ClaML or at least implement an export to ClaML of their own classification.

Still, there might be countries where the resources needed for migrating to ClaML or for implementing an export exceed the benefits in the first place. As some of the benefits can primarily be seen on an international level and will only affect the respective country on the long run resistance of a migration can be possible in these countries. But, without a ClaML version of almost all clinical modifications a comparison towards an ICD-10-XM is only half-heartedly and will not suffice as a basis for the next revision.

Therefore it should be discussed what the sweetener for these countries could be to kick off a migration or export routine to ClaML.

One option could be to define a "basic conformity" level of ClaML for the ICD-10 exchange format that suffices for international comparison but not for maintenance. It could be in coarser granularity than the ClaML used for maintenance and publication but should have all the structural elements needed for the comparison (e.g. a ClaML without the fragmentation of the inclusion and exclusion note or even a ClaML file with only the codes and the preferred terms for a start). Of course this would avoid the effects of the other benefits, like the use of the maintenance tool for the respective country as well.

For other classifications like the ICF, which has been developed in an electronic environment, the above problems will be easier to solve as there are hardly any elements in the ICF that do impose complicated XML structures to represent layout requirements.

For further revisions of the ICD hopefully the electronic maintenance will be considered whilst the development in order that easy handling and comparison in electronic format will not be complicated.

References

- [1] Electronic Maintenance of Clinical Classifications: Comparing Two Approaches. Stahl C, Walker SM, Garrett C, Truran D, Roberts R, Schopen M; WHO/GPE/ICD/C/00.33
- [2] Developing a Database Version of the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Clinical Modifications (ICD-10-CM). Berglund DJ, Fisher LJ; WHO/GPE/CAS/C/01.65
- [3] Report on the XML Version of ICD-10; Schopen M; WHO/HFS/CAS/C/03.108
- [4] XML Representation of Hierarchical Classification Systems; Hoelzer S, Schweiger RK, Liu R, Rudolf D, Rieger J, Dudeck J; WHO/HFS/CAS/C/03.95
- [5] <http://www.dimdi.de/static/en/klassi/koop/who/etc/index.html>
- [6] Executive Summary of the 2005 WHO FIC Meeting, Section 6c; <http://www.who.int/classifications/network/Executive%20Summary%20Tokyo%202005.pdf>
- [7] Maintenance and Publication Tool for WHO-FIC Classifications – Proceedings in development; Stefanie Weber, Susanne Bröenhorst, Can Celik, Tarek Ahmed, Michael Schopen; submitted for WHO-FIC Meeting 2006 in Tunis
- [8] ICD-10-XM; Schopen M, Roberts R, Üstün TB; WHOFIC/04.073