

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

Tunis, Tunisia  
29 Oct. - 4 Nov. 2006

---

**Maintenance and Publication Tool for WHOFIC  
Classifications –  
Proceedings in development**

Stefanie Weber, Susanne Bröenhorst, Can Celik, Tarek Ahmed, Michael Schopen

---

**Abstract**

*Since the last WHO-FIC meeting, where the functional specifications of the planned maintenance and publication tool were discussed, the technical specification of the tool was finalized and the programming has begun.*

*As described in the functional specifications the toolset will be a standalone tool which will hopefully run under the most common operating systems.*

*The database containing the classification will be an XML database (Berkeley DB XML) which will be accessed by the maintenance tool and the publication tool via an interface.*

*As a basis for the tools the Eclipse Rich Client Project will be used with the features realized as plug-ins, e.g. the WYSIWYG editor or the history functionalities.*

*Documentation and programming will be in English.*

*For the discussion of sharing resources and assuring WHO needs towards this software a meeting was held in Cologne, Germany and WHO agreed on programming certain parts of the system. This cooperation will guarantee the adjustment of the toolset to WHO requirements and therefore enables possible worldwide use of the software.*

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

Technical Specification and Realized Components ..... 3

    Database ..... 3

    Tool Framework ..... 3

    Programming language ..... 3

    Editor ..... 3

    Plausibility checks ..... 5

    History ..... 6

    Output Generation ..... 6

    Crosswalks Generation ..... 7

Cooperation with WHO Headquarters ..... 7

Problems ..... 8

Outlook ..... 8

References ..... 8

---

## Technical Specification and Realized Components

Since last years WHO-FIC Network Meeting in Tokyo lots of steps have been taken towards realization of the Classification Maintenance and Publication Tool. In DIMDI and WHO headquarters programming has begun. The first step towards realization was the development of the technical specification for the toolset which was finished in early spring. Most of the components described below are still work in progress and might be altered after presentation and publication of this paper.

### Database

As ClaML - the structure for the markup of the classification - is an XML schema several XML databases were evaluated under the following aspects:

- Is the database available free of charge?
- Is the database widespread and common in use?
- Will the database be supported in the long run?
- Is the database fast and easy to handle with large quantities of data?

It was decided to choose the Berkeley XML DB as it does fit all the above needs and performed well in tests [1].

### Tool Framework

After evaluation of some tool frameworks the Eclipse platform was chosen to be used both as framework for the tools set and as development environment [2]. It provides a flexible base with preexisting plug-ins that can be used for the new toolset.

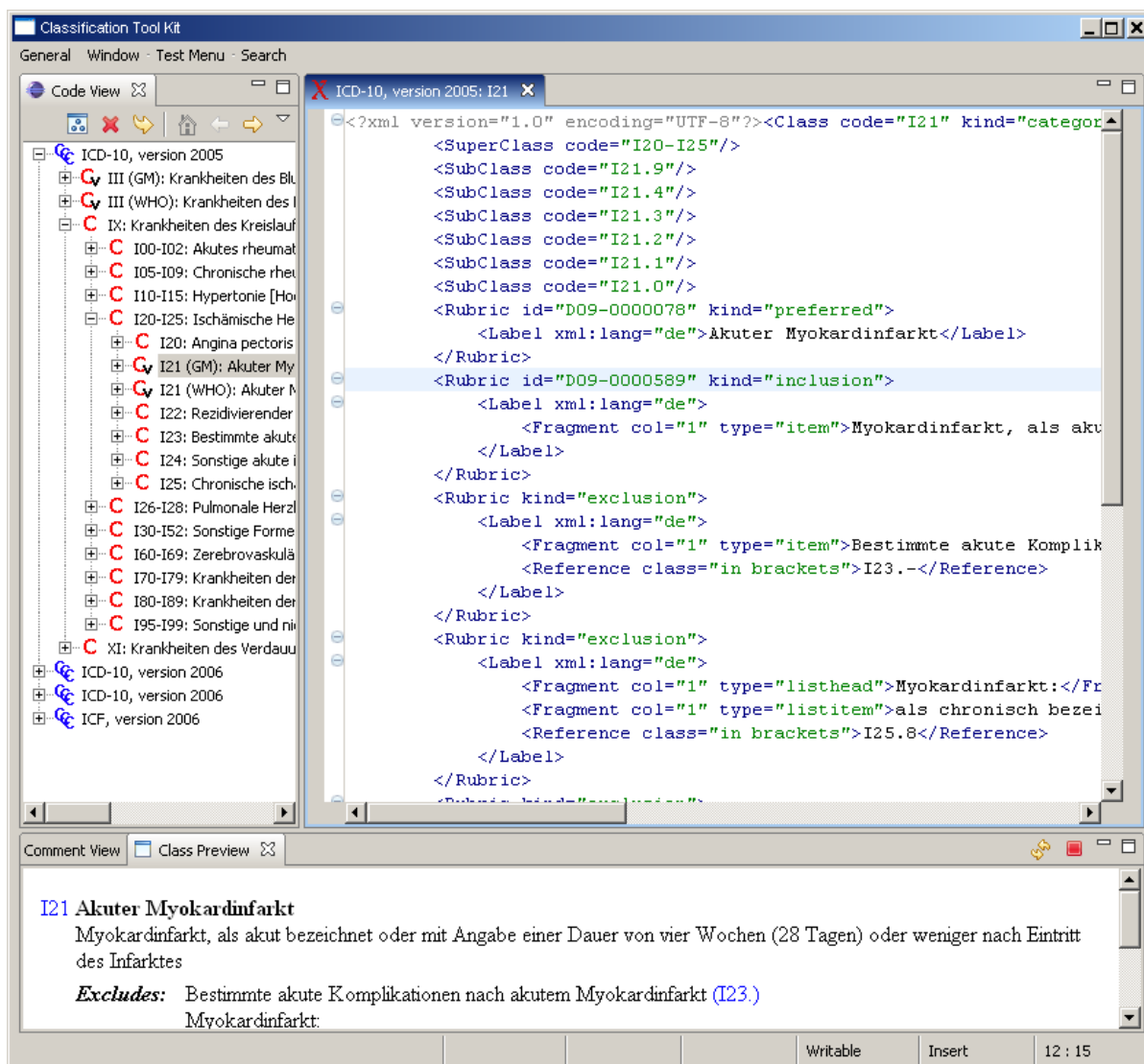
The Eclipse Community is spread worldwide and therefore discussion forums and other helpdesks are easy to access.

### Programming language

As there are already lots of very good plug-ins for the Eclipse platform in Java and this is a common and widely used programming language, Java was chosen to be the programming language for the toolset. It is easy to use and known by almost every computer specialist worldwide.

### Editor

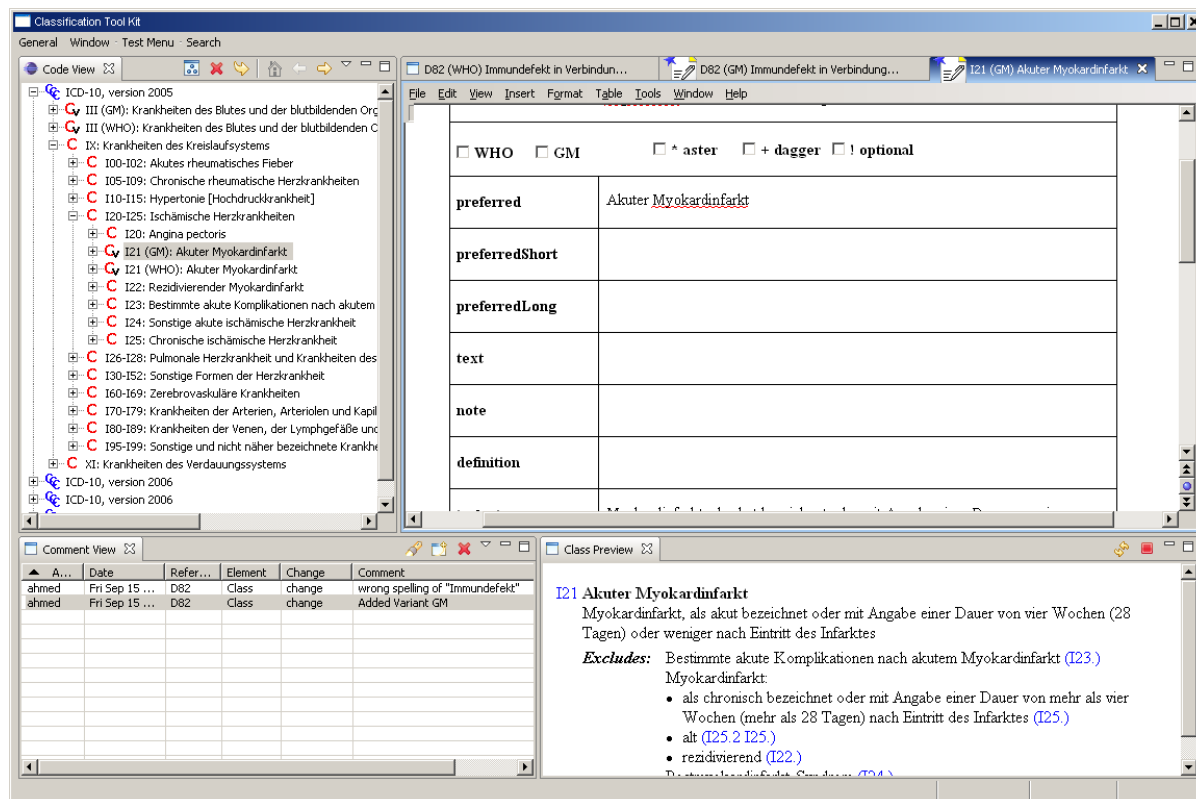
Implementing an XML editor in the toolset was realized to give the specialist easy access to the classification with using the other features of the tool (history, plausibility checks).



Screenshot 1: The XML editor in the Eclipse environment (right upper window)

Still the main user of the Maintenance Tool will be a classification specialist who cannot be expected to have high level XML knowledge. Therefore the text editor will be the key feature of the new Maintenance Tool and needed to be chosen very carefully. It has to be able to edit classification elements in a precisely defined way. The classification elements have to be presented to the user of the tool according to the printed layout of the classification or at least in such a intuitive way that the user will recognize the semantic structure of the classification and will edit it in a proper way.

After ample research the open office writer showed all the needed features, was free of charge and editable. Its basis is an XML document structure which gives the possibility to easily transform ClAML to the open office XML and backwards.

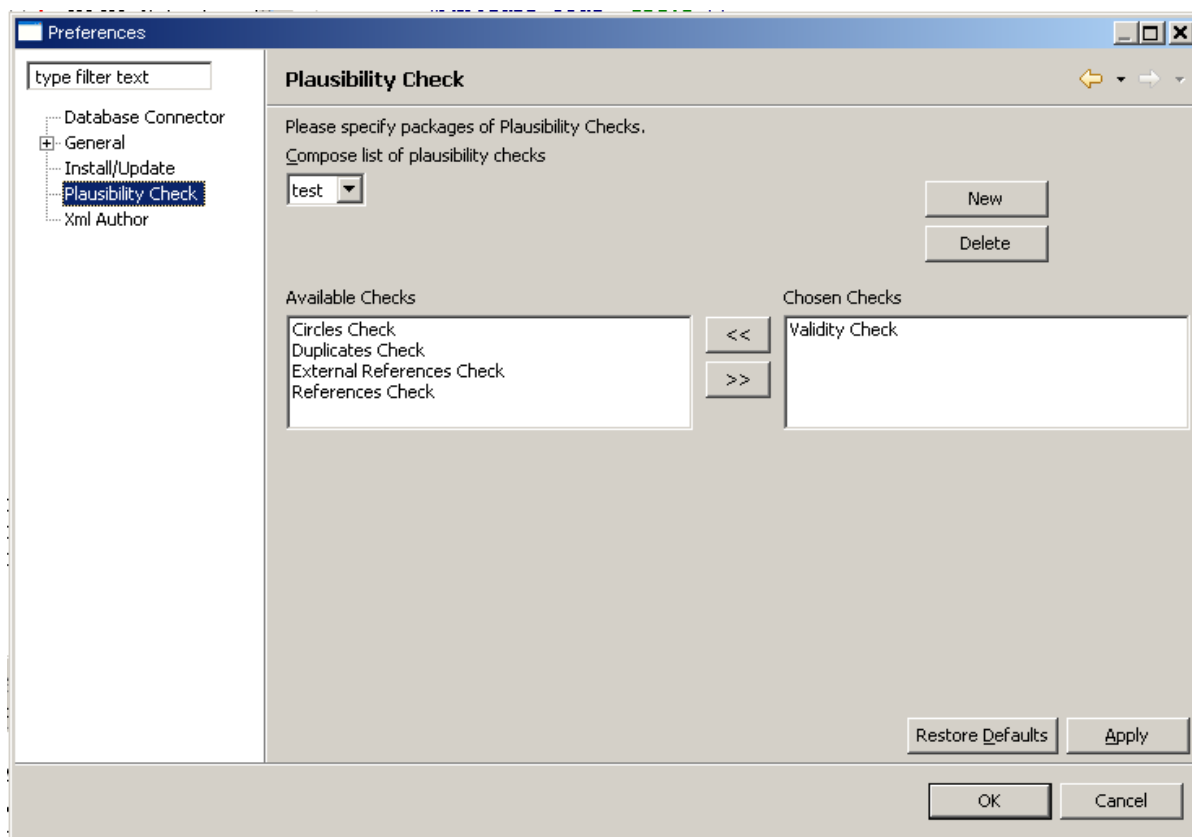


Screenshot 2: Prototype of the text editor (right upper window) The left upper window shows the tree view of two variants of the ICD-10. In the lower window the comments from the history are shown and the lower right window shows an HTML preview of the selected category. The text editor window is still in an early stage of development and will be designed to meet usability requirements later.

## Plausibility checks

There are lots of aspects in a ClAML based classification that can be checked automatically, starting from simple validity of the XML markup over spell checks to cycles in the class hierarchy. We decided to make the toolset easily extendable by implementing each type of plausibility check as a plug-in. Thus it is possible for third parties to write their own checks without the need of rebuilding the platform. So far the following checks are implemented:

- Well-formedness and validity of XML code
- existence of superclass for each class
- check for duplicate codes
- cycles in the class hierarchy
- cross-check of internal references
- cross-check of external references



Screenshot 3: Plausibility check pick list

We are working on more checks right now. Some checks are specific to the type of classification (ICD, OPS, ICF, ...). A corresponding interface is planned.

## History

In the needs specification it was requested that an automated documentation of changes was needed to create a history of the classification maintenance.

This was implemented as a plug-in in Eclipse and works automatically without bothering the user. Still, the user has the option to create or alter entries in the history table or to search and browse the history information if needed.

## Output Generation

Through the toolset publication of the classifications should be easy, flexible and possible at any time in an update cycle of the classification.

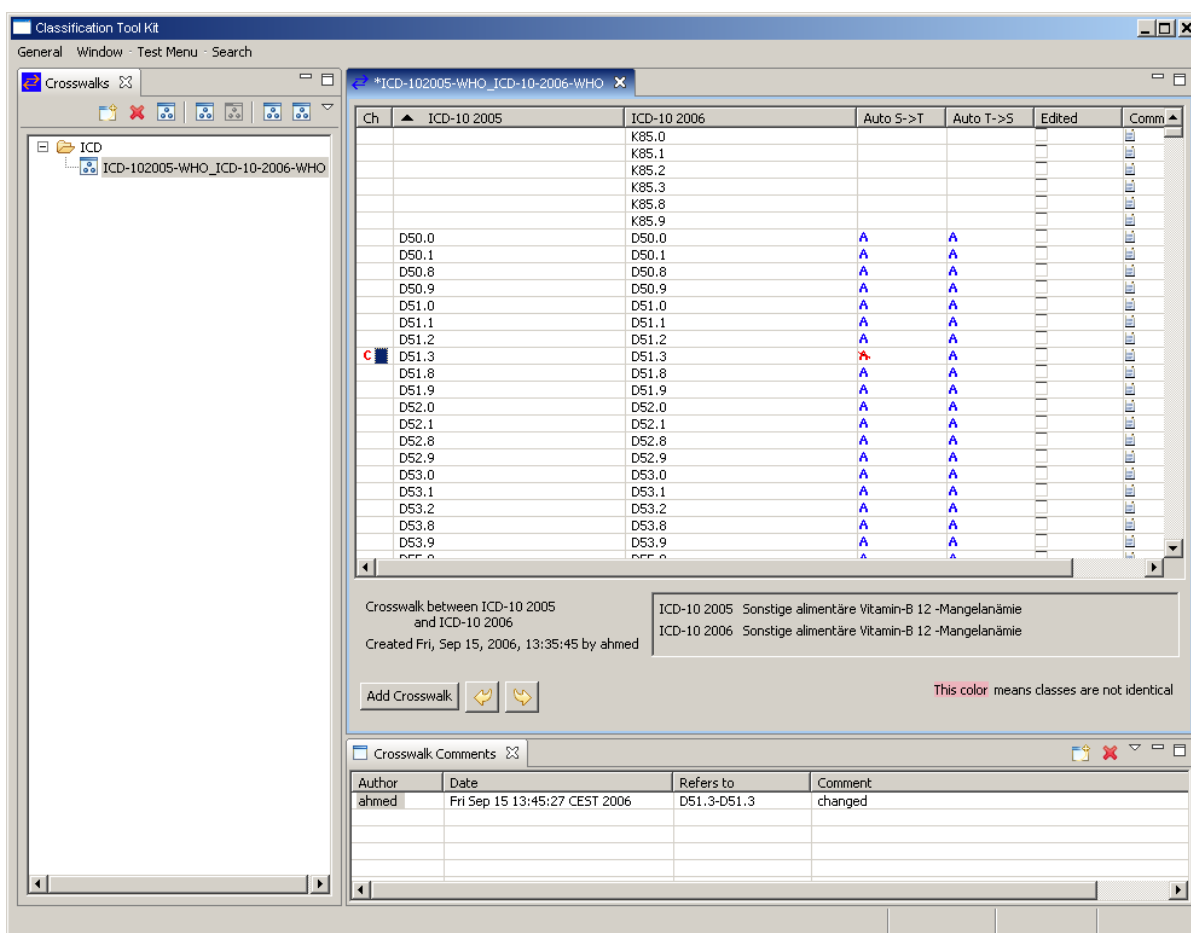
Publication as a printed format (rtf or pdf) will be solved through the open office plug-in. This tool already provides the needed features for this output.

HTML and ASCII output will be created through transformation via XSL-T directly from the XML database.

### Crosswalks Generation

To generate crosswalks between different versions of one classification a plug-in was programmed. Of course such automated crosswalks can only be considered to serve as a starting point for intellectual work. But for classifications of the size of the ICD-10 automated help is essential. Crosswalks generation supports 1:1, 1:n, and n:m relations and the possibility to specify best matches for automatic mapping into both directions.

The algorithm may have to be altered for every classification according to its structure. Therefore we started with the most widely used classification ICD-10 and implemented a crosswalks generator.



Screenshot 4: The automatically created crosswalks can be altered by the user and a documentation of its alterations will be saved with the option to add a comment.

### Cooperation with WHO Headquarters

In a meeting in Cologne in early spring of this year the agreed cooperation of WHO and DIMDI on the realization of the Maintenance and Publication Tool was discussed. As most of the components have to be developed in Eclipse itself it seemed a good

---

way to have them programmed in DIMDI only. Therefore WHO agreed to take on the development of the style sheets and routines for most of the different output formats (HTML, ASCII, etc.) and some transformation routines to ClaML for the existing SGML files.

Work has begun in WHO Headquarters as well and already some results were integrated in the Eclipse platform (HTML transformation).

## Problems

Some of the needs on classification maintenance as requested in the needs specification have already caused problems in realization. A main "troublemaker" is the maintenance of two variants of the same classification in one file. As this reduces redundancy in maintenance and will be helpful for WHO to create an ICD-10-XM if needed it is a feature that cannot be abandoned. Still, the realization did cause lots of extra effort for the IT specialists – and still does. For example even simple things as the realization of the tree structure of a classification aggravates a lot if it needs to show the variants of a classification in one tree.

Furthermore there are problems to be expected in the conversion from ClaML to ODT (Open Document) and above all from ODT to ClaML. This remains to be seen though, as we have only just begun working on this aspect.

Due to the goal of having the tools set run on several platforms (Windows, Linux, Macintosh) and our heavy use of external software (Berkeley DB XML, Open Office) some extra efforts seem likely here too.

## Outlook

As shown above the realization of the tool has progressed considerably. Still, some key features need further work to be used in routine application.

DIMDI is planning to have an extensive testing phase toward the end of 2006 and will, after fixing the problems found in testing, start using the tool for routine maintenance next spring. It is very likely that the new tool will be used in parallel with the existing SGML-based tools for one production cycle to make sure that both tools provide comparable files. Especially the output features should be ready early in the updating cycle of the German classifications which runs from March until October. Already software developers and other users of the classifications have been informed that DIMDI will maintain and publish its classifications in ClaML in the near future and responses were quite positive.

If development and testing of the new tools set are progressing further according to the given timeline, DIMDI is quite optimistic to show a finalized first version by the WHO-FIC Network Meeting in 2007.

## References

[1] <http://www.sleepycat.com/products/bdbxml.html>

---

[2] <http://www.eclipse.org/>

[3] <http://java.sun.com/>

Dr. Stefanie Weber

WHO-FIC Collaborating Centre for the German Language

German Institute for Medical Documentation and Information (DIMDI)

Waisenhausgasse 36-38A

50676 Köln

Germany

Email: [stefanie.weber@dimdi.de](mailto:stefanie.weber@dimdi.de)

Phone: +49 221 4724 485

Fax: +49 221 4724 444

<http://www.dimdi.de>