



WORLD HEALTH ORGANIZATION

WHO/HFS/CAS/C/03.97
Distr.: LIMITED
ENGLISH ONLY

MEETING OF WHO COLLABORATING CENTRES FOR THE FAMILY OF INTERNATIONAL CLASSIFICATIONS

Cologne, Germany
19-25 October 2003

Title: Towards an Electronic Version of ICD-10 - A Status Report

Authors: Michael Schopen

Purpose: for information

Recommendations: -

Abstract:

Work on an electronic version of ICD-10 started in February 2003. It was based on the results of an evaluation of an XML version of ICD-10 that became available to WHO. The XML version seemed to be not suitable for an electronic version of ICD-10. Thus, the files of ICD-10 available from WHO on floppy disk were converted to SGML. All updates to be implemented by 2003 were added. Using the DIMDI maintenance system the following output formats were produced: RTF, PDF, HTML, ASCII and metadata for import into databases. A PDF browser and files for book publishing became available. The paper gives detailed background information and an outline of the future work.

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.

Towards an Electronic Version of ICD-10 A Status Report

Michael Schopen

In November 2002 an XML version of ICD-10 became available and was evaluated by members of the Electronic Tools Committee. Details of this evaluation are available in a separate report (1). This XML version seemed to be not suitable for an electronic version of ICD-10 due to structural weaknesses and insufficient data quality. Thus, we recommended an approach based on the German maintenance system for ICD-10 (2).

Results and available files

1. We started with the ASCII files that were available from WHO on floppy disks. All parts of Volume I and Volume III were converted to SGML. A few parts were missing and had to be recorded again, as they were even not available at WHO. All updates to be implemented by 2003 were added based on the cumulative list of the Update Reference Committee (3). Special mark-up was used so that updated can be identified later.
2. The WordPerfect files of Volume II were converted to RTF and formatted with style sheets. This took considerably longer than expected, as these files were in rather bad shape. Afterwards, updates were added using the MS-Word® feature "Track Changes". Furthermore, certain parts of Volume II were revised by WHO as they were no longer current.
3. The DIMDI maintenance system uses special style sheets to convert SGML to RTF. These style sheets had to be modified considerably to meet WHO layout requirements for Volumes I and III. Word makros had to be programmed to generate the running headers in Volume III.
4. After several months of work on these files, the formatting routines were good enough to produce RTF/PDF files that could be used for book publishing. The PDF files were sent to WHO in July 2003.
5. Starting with these files a contractor (Dr Zaiß, Department of Medical Informatics, University of Freiburg) prepared a PDF browser. It was sent to WHO in August 2003.
6. Using the DIMDI maintenance system the SGML files of the Tabular List were converted to HTML and made accessible on the DIMDI webserver at

http://www.dimdi.de/dynamic/en/klassi/diagnosen/icd10/x_htmlengl2003/fr-icd.htm
User code and password are required and available from schopen@dimdi.de

7. In the next step ASCII files formatted with blanks and new lines and minimal structural information were generated from the SGML files.
8. Finally meta data were generated from the SGML files. They contain database tables for the chapters and blocks and a separate table for the codes and titles. These tables reflect the hierarchy of ICD-10. Furthermore, the codes are linked to the five Special Tabulation Lists to allow data aggregation for statistics.

Detailed notes on data formats

RTF / PDF

Rich Text Format (RTF) was generated directly from SGML. PDF was generated from RTF by Acrobat software. There is one version where the Tabular List, the Instruction Manual and the Alphabetical Index are linked by ICD-10 codes. This version was used to build the PDF browser. It has bookmarks for chapters, blocks and three-character categories. A special PDF index file is available to speed up searches and to support boolean operators. There is also a PDF version without hyperlinks. This version is to be distributed free of charge over the internet (read-only version) as decided at the Brisbane meeting of the WHO-FIC Collaborating Centres. The version with hyperlinks is to be sold by WHO to generate more resources for classification work.

HTML

The HTML version is available for the Tabular List only and is split into small files based on ICD-10 blocks to ensure acceptable download times over networks. It makes use of HTML frames to provide navigational support and to add headers and footers. It can be searched for ICD-10 codes. The layout comes quite close to the book. In the version currently available, headers and footers are the files used on the DIMDI webserver. However, this version can be customized by exchanging the files for header and footer (header.htm, footer.htm), by editing a style-sheet (style.css) and by modifying the HTML frameset (fr-icd.htm). Thus it is possible to offer access to the Tabular List over the internet or over local area networks.

ASCII

The ASCII files are formatted with white space and new lines and their layout comes quite close to the book. Two characters precede each line to provide minimal structural information, one digit for the hierarchical level of the classification and a letter to indicate whether it is basic information (code and title), a note, an inclusion or exclusion note etc. An ASCII file of the Instruction Manual is also available.

The ASCII files of the Alphabetical Index have a different format. Again, each line is preceded by two characters, one indicating the part of the index (e.g. diseases, abortion, neoplasms) and another one indicating the level of indentation. These two characters are followed by five columns for the ICD-10 codes (primary code, asterisk code, morphology code, neoplasms codes etc.), which in turn are followed by the text.

Metadata

Metadata are stored as character separated values (CSV). They are intended for import into relational database systems. They contain codes and titles from the Tabular List. Furthermore, the blocks and chapters from the Tabular List and the codes of the Special Tabulation Lists are linked to the ICD codes. Thus, ICD codes can be statistically aggregated according to blocks and chapters from the Tabular List and according to the five Special Tabulation Lists.

Further work to done:

Database version

A database version with the complete Tabular List and Alphabetical Index is to be generated from the SGML files. This database version should follow the Australian database schema for ICD-10-AM which has a clear and straight-forward structure.

“XMLification” and Maintenance System

The SGML structure is to be “XMLified”. This will include a change from the currently used set of tags in German language to an English language set of tags and a completely new set of programs for the transformation to the target formats with documentation in English.

We consider that to be a larger project. As the support for SGML is gradually breaking away and as XML is a format that is widely accepted, DIMDI plans a redesign of its ICD-10 maintenance system with a change from SGML to XML. This project is planned for 2004 and 2005. Our aim is to set up a system that cannot only handle the German ICD-10 but also other languages.

French Version

In a first step, the French Version will be treated as the English version. However, authorized updates for the French version are not yet available. In a second step the French version should be integrated into the new XML-based maintenance system like the English Version.

References:

1. SCHOPEN Michael: Report on the XML version of ICD-10. Annual Meeting of the WHO Collaborating Centres for the Family of International Classifications 2003 in Cologne. [WHO/HFS/CAS/C/03.]
2. SCHOPEN Michael: Electronic Publishing of the German Language Edition of ICD-10. Meeting of Heads of WHO Collaborating Centres for the Classification of Diseases. Paris 1998. [WHO/GPE/ICD/C/98.24]
3. http://www2.fhs.usyd.edu.au/ncch//WHO%20URC/official_WHO_updates_combined%201996%20to%202002.pdf

Dr. med. Michael Schopen
WHO Collaborating Centre for the
Family of International Classifications
German Institute of Medical
Documentation and Information (DIMDI)

Waisenhausgasse 36 – 38 a
D-50676 Köln

Telephone: +49 221 4724-325
Telefax: +49 221 4724-444
e-mail: schopen@dimdi.de
www: <http://www.dimdi.de>