

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

Tunis, Tunisia
29 Oct. - 4 Nov. 2006

**Electronic Versions of ICD-10 –
Files, Formats, Functions**
Michael Schopen

Abstract

Beginning with version 2003 ICD-10 has become available electronically on the DIMDI webserver in an area that is only accessible for the WHO-FIC Collaborating Centres.

However, access statistics for this area indicate that these files are only used by a few people.

The paper will present the different files and formats that are available and will provide some examples for their application.

Content

Abstract	1
Introduction	3
Volume I – Tabular List	3
Volume II – Instruction Manual	7
Volume III – Alphabetical Index	7
References	8
Appendix	9

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.

Introduction

Beginning with ICD-10 2003 an electronic version of this classification has become available on the website of the WHO-FIC Electronic Tools Committee. A second edition of the classification was printed and an eBook became available, both based on these files. For the maintenance of this electronic version the SGML-based tool developed at DIMDI was used (1, 2). The files for this electronic version have been updated annually and can be downloaded at

<http://www.dimdi.de/dynamic/en/klassi/koop/who/closed/index.htm>

A special password was given to the WHO-FIC Collaborating Centres to protect WHO copyright, as a licence is needed for commercial use of these files.

Download statistics show that the files are accessed regularly, but questions DIMDI has been asked concerning these files indicate that the knowledge about the files and their structure is rather low. Furthermore, extensive documentation is not yet available.

This paper aims at improving the knowledge about these files and their possible use and shall serve as basic documentation.

Volume I – Tabular List

The entire Tabular List (including the report about the revision conference, the morphology codes, and the nomenclature regulations) is available in the three formats: Microsoft Word[®], Adobe PDF[®] and Windows text.

Files in **Microsoft Word[®] format** are prepared mainly for book production, but they are also helpful for maintenance purposes, e.g. to discuss possible changes until they are consolidated in "track changes" mode. These files render the original book layout for ICD-10. There is almost no structural information available in these files, however, the classification hierarchy can be taken from Word[®] paragraph styles as follows:

Paragraph Style	Classification Hierarchy
Heading level 2	ICD chapter
Heading level 3	ICD block, 1 st level
Heading level 4	ICD block, 2 nd level
Heading level 5	ICD block, 3 rd level
Heading level 6	ICD three-character category
Heading level 7	ICD four-character subcategory

Files in **Adobe PDF® format** are generated from these Word® files and can be used if the classification is to be distributed in a format that cannot be changed.

Files in **Microsoft Windows text format** are available for traditional computer processing of the whole Tabular List. Their line length is restricted to 80 characters, they are formatted with spaces, carriage return/line feed, and empty lines and are enhanced by two control characters in the first two positions of each line. LZ marks empty lines, on all other lines the first character marks the classification hierarchy, the second character the classification contents. For the classification hierarchy the following control characters are used in the first column of a line:

Control Character in Position 1	Classification Hierarchy
0	ICD chapter
1	ICD block 1 st level
2	ICD block 2 nd level
3	ICD block 3 rd level
4	ICD three-character category
5	ICD four-character subcategory
X	Start of a subclassification
V	Four-character code of a subclassification
F	Five-character code of a subclassification
L	Empty line for rendering purposes (not for longer texts)
P	Start of a new paragraph
<space>	Continuation of the previous line

For the classification content the following control characters are used in the second columns of a line:

Control Character in Position 2	Classification Content
*	List of asterisk codes at the beginning of a chapter
G	List of blocks at the beginning of a chapter
B	"Use additional code"
E	Exclusion notes
H	Notes
I	Inclusion notes
R	Explanations
S	Reference to a subclassification
T	Code and title
Z	Empty line for rendering purposes (not for longer texts)
<space>	Continuation of the previous line

The following example shows the use of these control characters:

```

0T Chapter III
0T Diseases of the blood and blood-forming organs and certain disorders
  involving the immune mechanism
0T (D50-D89)
LZ
0E Excludes: autoimmune disease (systemic) NOS (M35.9)
0E          certain conditions originating in the perinatal period (P00-P96)
LZ
0G This chapter contains the following blocks:
0G D50-D53  Nutritional anaemias
0G D55-D59  Haemolytic anaemias
0G D60-D64  Aplastic and other anaemias
LZ
0* Asterisk categories for this chapter are provided as follows:
0* D63*     Anaemia in chronic diseases classified elsewhere
LZ
1T Nutritional anaemias
1T (D50-D53)
LZ
4T D50      Iron deficiency anaemia
4I          Includes: anaemia:
4I              . asiderotic
4I              . hypochromic
LZ
5T D50.0    Iron deficiency anaemia secondary to blood loss (chronic)
5I          Posthaemorrhagic anaemia (chronic)
5E          Excludes: acute posthaemorrhagic anaemia (D62)
5E          congenital anaemia from fetal blood loss (P61.3)
    
```

The text files have even been used by software companies to load the entire four-character classification into a relational database.

The four-character classification is available in two additional formats: HTML and CSV.

The version in **HTML format** is split into some 300 files at block level to achieve sufficient response time in a web browser. It comes with a search option for ICD codes at the three character level. The classification hierarchy can be accessed from a navigation menu which allows browsing through the chapters and blocks of the hierarchy. References to ICD codes in exclusion notes are converted to HTML links so that they can easily be followed in a web browser.

HTML files can be used as a stand-alone version on a PC ("poor man's eBook"), in a local area network (e.g. in a hospital) or via the Internet. As formatting makes use of a CSS style sheet, colours and styles can easily be modified to integrate this version into existing web sites.

The HTML files are used "as they are" on the DIMDI web server. As an example for a successful integration into another environment the WHO website can be visited at <http://www3.who.int/icd/currentversion/fr-icd.htm>

The HTML files are not yet fully accessible as they make use of JavaScript and HTML frames. This shall be overcome in a later version.

The ICD-10 **meta data files** are in CSV format (character separated values) and contain codes and titles at all levels of the classification hierarchy. Inclusion or exclusion notes, annotations and subclassifications are not included. The meta data files can easily be loaded into a relational database or into spreadsheets.

Meta data files are available for the four-character classification and for the five special tabulation lists. ICD codes are always listed without dagger, as the dagger can be added to most non-asterisk codes if it makes sense in medical context and as the dagger codes in the Tabular List are not exhaustive. Codes are given in three formats: with dot and asterisk, with dot and without asterisk, without dot and asterisk as all these formats are in use in different countries.

Meta data are mainly used in databases, e.g. to check the validity of ICD codes in medical records or for reimbursement purposes.

The Tabular List is stored in three database tables for the chapters, blocks and codes. These three tables are linked by primary keys and foreign keys so that the hierarchy of the classification can be represented in a database system. Furthermore, the 5 special tabulation lists are included as database tables and linked to the ICD codes. Thus it is possible to tabulate mortality or morbidity data both according to the classification hierarchy and according to the special tabulation lists.

Meta data files could be enhanced by different edits, e.g. for patient gender or age. Such edits have been added for the German version. It is suggested to agree on

international edits for these purposes and add them to the meta data files of the WHO version so that all countries use the same edits.

The appendix to this paper gives an extensive description of the meta data files and provides SQL statement to set up a relational database.

Volume II – Instruction Manual

The Instruction Manual is available in Microsoft Word® and in Adobe PDF® format. The outline of the Instruction Manual is represented using Microsoft Word® paragraph styles (i.e. heading level).

Volume III – Alphabetical Index

The Alphabetical Index comes in three formats: in Microsoft Word® and Adobe PDF® and in Microsoft Windows text. Again, Microsoft Word® format uses paragraph styles to represent the different levels of the index entries. Heading Level 1 is used for the lead terms, heading level 2 for terms preceded by a single dash, heading level 3 for terms preceded by two dashes, etc. The same convention is used in tables, i.e. for neoplasms and drugs.

The text files use fixed columns for the ICD codes, which are followed by the text entries of the index. The files are restricted to 120 characters per line with continuation lines for longer texts. "See" and "see also" references have been attached to the texts.

Each line is preceded by two control characters, the first one lists the section of the index (K = diseases, N = neoplasms, U = external causes, C = drugs and chemical substances, <space> = continuation of the previous line), the second one lists the level of indentation (0 = lead term, 1 = one dash, 2 = two dashes, etc., <space> = continuation of the previous line). The ICD codes are listed with dagger and asterisk. Up to three codes are provided per entry in the following sequence: primary code, optional secondary code, and optional morphology code. For the table of neoplasms and the table of drugs and chemicals up to five codes are provided in the sequence of the book.

The following example shows the file structure and the use of the control characters:

K0	Abnormal, abnormality, abnormalities - see also Anomaly
K1E87.4	- acid-base balance (mixed)
K2	- - fetus - see Distress, fetal
K1R77.0	- albumin
K1R77.2	- alphafetoprotein
K1O41.9	- amnion, amniotic fluid
K2P02.9	- - affecting fetus or newborn

This text file can be used to load the entire index into a relational database. It is even possible to repeat the higher level texts during transformation, so that full texts can be stored in the database. The previous example can be transformed into the

following format (codes have been put to the end of a line for easier reading):

Abnormal, abnormality, abnormalities - see also Anomaly
Abnormal, abnormality, abnormalities: acid-base balance (mixed) E87.4
Abnormal, abnormality, abnormalities: acid-base balance (mixed): fetus - see Distress, fetal
Abnormal, abnormality, abnormalities: albumin R77.0
Abnormal, abnormality, abnormalities: alphafetoprotein R77.2
Abnormal, abnormality, abnormalities: amnion, amniotic fluid O41.9
Abnormal, abnormality, abnormalities: amnion, amniotic fluid: affecting fetus or newborn P02.9

Conclusion

The ICD-10 is available in the restricted area in multiple formats. Soon another format will be added: The tabular list in ClaML as this was agreed to be the WHO-FIC standard exchange format for classifications.

With the use of these files synergies in maintenance can be utilised. E.g. for the meta data shared use of age limits or gender information can help to standardise plausibility checks for mortality and morbidity data internationally.

References

1. Schopen M: Electronic Publishing of the German Language Edition of ICD-10. Annual Meeting of the Heads of the WHO Collaborating Centres for the Classification of Diseases 1998 in Paris. WHO/GPE/ICD/C/98.24
2. Schopen M: Electronic Publishing of the Alphabetical Index to ICD-10. Annual Meeting of the Heads of the WHO Collaborating Centres for the Classification of Diseases 1999 in Cardiff. WHO/GPE/ICD/C/99.33

Appendix

1. Detailed description of ICD meta data files

CHAPTERS.TXT: ICD-10 chapters	
Field 1	chapter number, 2 characters
Field 2	chapter title, variable length
BLOCKS.TXT: ICD-10 blocks (lowest level block for chapters with nested blocks)	
Field 1	first category of block, 3 characters
Field 2	last category of block, 3 characters
Field 3	chapter number, 2 characters
Field 4	block title, variable length
MORBL.TXT: Morbidity Tabulation List	
Field 1	code
Field 2	title
MORTL1_1.TXT: Blocks of 1st Mortality List	
Field 1	block code
Field 2	block title
MORTL1_2.TXT: Codes and titles of 1st Mortality List	
Field 1	code
Field 2	block code
Field 3	title
MORTL2.TXT: Codes and titles of 2nd Mortality List	
Field 1	code
Field 2	title
MORTL3_1.TXT: Blocks of 3rd Mortality List	
Field 1	block code
Field 2	block title
MORTL3_2.TXT: Codes and titles of 3rd Mortality List	
Field 1	code
Field 2	block code
Field 3	title
MORTL4.TXT: Codes and titles of 4th Mortality List	
Field 1	code
Field 2	title

CODES.TXT: Codes and titles of categories and subcategories	
Field 1	level of hierarchy, 1 character
	3 = category
	4 = subcategory
Field 2	location in classification tree, 1 character
	T = leaf node (used for coding)
	N = non-leaf node (not used for coding)
Field 3	type of subcategory, 1 character
	X = explicitly listed (pre-combined)
	S = by subclassification (post-combined)
Field 4	chapter number, 2 characters
Field 5	first category of block, 3 characters
Field 6	code without dagger
Field 7	code without asterisk
Field 8	code without dot, dagger, or asterisk
Field 9	title
Field 10	reference to 1 st Mortality Tabulation List
Field 11	reference to 2 nd Mortality Tabulation List
Field 12	reference to 3 rd Mortality Tabulation List
Field 13	reference to 4 th Mortality Tabulation List
Field 14	reference to Morbidity Tabulation List

2. How to set up a relational database

The following SQL statements can be used to set up a relational database from these files:

```
CREATE TABLE Chapters (
    ChapNo TEXT(2) CONSTRAINT ChapX PRIMARY KEY,
    ChapTi TEXT(255)
);
```

```
CREATE TABLE Blocks (
    BStart TEXT(3) CONSTRAINT BIX PRIMARY KEY,
    BEnd TEXT(3),
    ChapNo TEXT(3) CONSTRAINT ChBIX REFERENCES Chapters,
    BITi TEXT(255)
);
```

```
CREATE TABLE MorbList (
    MorbLCode TEXT(5) CONSTRAINT MorbLCodeX PRIMARY KEY,
    MorbLTi TEXT(255)
);
```

```
CREATE TABLE MortList1Blocks (  
    MortL1BICode TEXT(5) CONSTRAINT MortL1BICodeX PRIMARY KEY,  
    MortL1BITi TEXT(255)  
);  
  
CREATE TABLE MortList1 (  
    MortL1Code TEXT(5) CONSTRAINT MortL1CodeX PRIMARY KEY,  
    MortL1BICode TEXT(5) CONSTRAINT MortL1MortL1BI REFERENCES MortList1Blocks,  
    MortL1Ti TEXT(255)  
);  
  
CREATE TABLE MortList2 (  
    MortL2Code TEXT(5) CONSTRAINT MortL2CodeX PRIMARY KEY,  
    MortL2Ti TEXT(255)  
);  
  
CREATE TABLE MortList3Blocks (  
    MortL3BICode TEXT(5) CONSTRAINT MortL3BICodeX PRIMARY KEY,  
    MortL3BITi TEXT(255)  
);  
  
CREATE TABLE MortList3 (  
    MortL3Code TEXT(5) CONSTRAINT MortL3CodeX PRIMARY KEY,  
    MortL3BICode TEXT(5) CONSTRAINT MortL3MortL3BI REFERENCES MortList3Blocks,  
    MortL3Ti TEXT(255)  
);  
  
CREATE TABLE MortList4 (  
    MortL4Code TEXT(5) CONSTRAINT MortL4CodeX PRIMARY KEY,  
    MortL4Ti TEXT(255)  
);  
  
CREATE TABLE Codes (  
    Level TEXT(1),  
    Location TEXT(1),  
    Type TEXT(1),  
    ChapNo TEXT(2) CONSTRAINT CodesChapters REFERENCES Chapters,  
    BISTart TEXT(3) CONSTRAINT CodesBlocks REFERENCES Blocks,  
    Code TEXT(7) CONSTRAINT CodeX PRIMARY KEY,  
    NormCode TEXT(6) CONSTRAINT NormCodeX UNIQUE,  
    CodeWithoutDot TEXT(5) CONSTRAINT CodeWithoutDotX UNIQUE,  
    Title TEXT(255),  
    MortL1Code TEXT(5) CONSTRAINT CodesMortL1 REFERENCES MortList1,  
    MortL2Code TEXT(5) CONSTRAINT CodesMortL2 REFERENCES MortList2,  
    MortL3Code TEXT(5) CONSTRAINT CodesMortL3 REFERENCES MortList3,  
    MortL4Code TEXT(5) CONSTRAINT CodesMortL4 REFERENCES MortList4,  
    MorbLCode TEXT(5) CONSTRAINT CodesMortB REFERENCES MorbList  
);
```

Afterwards, the data files can be loaded into these tables in the following sequence:

1. CHAPTERS.TXT --> Chapters
2. BLOCKS.TXT --> Blocks
3. MORBL.TXT --> MorbList
4. MORTL1_1.TXT --> MortList1
5. MORTL1_2.TXT --> MortList1Blocks
6. MORTL2.TXT --> MortList2
7. MORTL3_1.TXT --> MortList3
8. MORTL3_2.TXT --> MortList3Blocks
9. MORTL4.TXT --> MortList4
10. CODES.TXT --> Codes

All SQL statements have been tested with MS-Access. As the implementation of the SQL standard may vary from system to system, changes to these SQL statements may be necessary for other database systems.

Dr. Michael Schopen
WHO-FIC Collaborating Centre for the German Language
German Institute for Medical Documentation and Information (DIMDI)
Waisenhausgasse 36-38A
50676 Köln
Germany
Email: michael.schopen@dimdi.de
Phone: +49 221 4724 297
Fax: +49 221 4724 444
<http://www.dimdi.de>