WHO-FIC NETWORK MEETING

Tokyo, Japan
16-22 October 2005

Maintaining Medical Classifications in XML – ClaML redefined for use with WHO-FIC classifications

Stefanie Weber, Egbert van der Haring; Susanne Bröenhorst, Michael Schopen, Pieter Zanstra

Abstract: Maintaining Medical Classifications in XML – ClaML redefined for use with WHO-FIC classifications

The Schema described in this paper is designed to contain all the information needed to maintain and publish medical classifications such as ICD-10. It was derived from an early version of ClaML (CEN/TS 14463) and after a detailed process of adjustment of this standard towards ClaML and vice versa, ClaML (and therefore this schema as well) is expected to become an ISO-standard.

Primarily this schema was created to match the needs of official WHO medical classifications such as ICD-10. Still, we do think that with this schema it will be possible to maintain a broad variety of medical classifications.

This document will give some explanation of the elements of the schema and examples on how the schema will be used. The examples mostly show the usage of the schema for ICD-10.

Content

ClaML as an XML schema

The Schema described below is designed to contain all the information needed to maintain and publish medical classifications such as ICD-10. It was derived from an early version of ClaML (Classification Markup Language, [1]) developed by Egbert J. van der Haring* and Pieter E. Zanstra* (*University Medical Centre, Nijmegen, the Netherlands). After a detailed process of adjustment of this standard towards ClaML and vice versa, ClaML (and therefore this schema as well) is expected to become an ISO-standard.

Primarily this schema was created to match the needs of official WHO medical classifications such as ICD-10. Still, we do think that with this schema it will be possible to maintain a broad variety of medical classifications.

This document will give some explanation of the elements of the schema and examples on how the schema will be used. These examples are mostly concerning ICD-10.

Some of the graphs are derived from an automated documentation of XML-Spy, an editor used by DIMDI to maintain the schema which can be obtained for free in a home version [2].

The XML-schema and files with examples for the use of ICD-10 in XML according to the schema can be obtained from the DIMDI Internet sites [3].

Root Element: CodingScheme

The Root element of the schema is the element *CodingScheme*. It has a couple of children as shown in figure 1.

CodingSchemeId defines one or more unique identifiers assigned by different issuing authorities to the classification. It uses the attributes *authority* to define the issuing authority (for example HL7) and the attribute *uid* for the unique identifier (2.16.840.1.113883.6.3 in case of ICD-10 in HL7).

Authors gives an outline of the authors that were involved in this version of the classification. Each author is given a unique id.

Editions can be used if the xml-file contains more than one edition of a classification like the GM-, AM- or WHO-edition for ICD-10. Like the *Authors* each edition is given an id that can be referenced from other elements in the xml-file.

Name, *Version* and *Title* are simple type elements that can contain a string of text. The *Date* is of the type date and specifies the date of the Version. For example if the ICD-10 (*Name*) called International Classification of Diseases (*Title*) is published in its *Version* 2004 on the 15th of August 2003 this would be the date to be contained by the element *Date*.

The element *ClassKinds* of *CodingScheme* has one or more child elements *Kind* with the attribute

name and the element *Display*. With this element an outline of the existing kinds of Classes e.g. chapter can be given. There should be one *Kind* element for each existing value, which is stored in the attribute *name*.

The element *RubricKinds* of *CodingScheme* also has one or more child elements *Kind*. As with *ClassKinds* with this element an outline of the existing kinds of rubrics can be given. There should be one *Kind* element for each existing value, which is stored in the attribute *name*. The child element *Display* can be used for a text to be displayed in the printed version of the classification. For example a *Kind* "inclusion" could represent an inclusion note of the ICD-10 and have the information "incl.:" in the element *Display* to be added in the print version to the beginning of every inclusion note in the classification.

The element *Modifier* is used in many elements of the schema and is referenced to by *ModifiedBy* and *ExcludeModifier* (see Class). *Modifier* is a subclassification within the classification used for specific concepts. In the element *ModifierClass*, which is also a child of *CodingScheme* each element represents a part of the subclassification grouped by *Modifier*.

If for example a section has a list of locations in which a diagnosis can apply, a *Modifier* is used to introduce this list and each element of the list is contained in a *ModifierClass*. The *Modifier* of the list is referenced to as *SuperClass* in the *ModifierClass*. The category – i.e. the Code – can then reference the *Modifier* and will get all the information of its *ModifierClasses* as well (example 1).

Class

Each concept of a classification is represented by an element Class, which is child of *CodingScheme* (figure 2). What type of concept it represents can be defined by its attribute *kind*. A *Class* can be of the *kind* chapter (e.g. H00-H59), of the *kind* section (e.g. H00-H06) or of the *kind* category (see example 2), which then represents a single code (e.g. H00).

The use of *Class* is defined by the attribute *use*, which can be a "dagger" or "asterix" for the ICD-10 or "optional" for the german ICD-10.

The attribute *edition* allows the maintenance of two or more editions in one xml-file to avoid redundancy. It references an edition from the *Editions* tag. If a code is the same in two editions it can be kept in the xml-file just once without the edition attribute and those codes that differ in two editions are kept twice with different edition attributes. This attribute should not be used to define the language of the rubric text. Although ClaML supports multiple languages using the standard ISO attribute *xml:lang*, at DIMDI we decided to use one language per xml-file. The DIMDI maintenance tool will be able to access more than one xml-file consecutively to enable comparisons or changes in two languages in one session.

If there are changes made in the current version of a classification the element *Class* can have the attribute *changed* switched to true, otherwise it would be false. Optionally the changes and reasons for the changes can be recorded in the tag *History*, which includes attributes to record the *author*

that made the change, the *edition* of the classification, and the *date* of the change. The xml-files produced by the DIMDI maintenance tool will not contain the history tag as the history of the DIMDI maintenance tool will be handled in a separate database or file.

The element *Symbol* contains the symbol (code) of the *class*. For the ICD-10 this could be "H00-H59" if the *Class* is a chapter or "H00" if the *class* is a category.

Superclass is an element of *Class* that has the information on the superordinated classes. If the *Class* is category and its content is H00 the *Superclass* would be H00-H06 which again would have the *Superclass* H00-H59.

Each *Class* can be modified by a *Modifier*. The reference to the *Modifier* is stored in the element *ModifiedBy*.

The *ModifiedBy* tag is inherited by the descendants of a class. In the example (see example 3), this means that the modifier *MD1* also modifies the descendants of *M40-M54*. Sometimes a modifier is only applicable to a limited subset of the descendants; ClaML defines the tag *ExcludeModifier* to exclude the use of a modifier at certain subclasses and their descendants. For example, it is useful to repeat the modifier at the direct children of *M40-M54*, but it is not useful to see the modifier at the class *M50*.

It is also possible that only some ModifierClasses of an assigned modifier may be used with a certain concept. In this case the element *ValidModifierClass* can be used to restrict the ModifierClasses. For example in ICD-10 the code M07.0 is modified by the modifier S13M00 with the ModifierClasses 0-9 but only the ModifierClasses 0,4,7,9 are valid for this code (see example 4)

Each *Class* does have zero or more elements *Meta* which contain the Metadata of the *Class*. The element *Meta* has the attributes *name* and *value*. For example in ICD-10 the *Meta* element can be the following:

<meta name="used for mortality coding" value="true">

<meta name="upper age limit" value="1year">

The main information of a class is contained in the element *rubric* as described below.

Rubric

The element *Rubric* is the element containing most of the information of the classification. It is a child of *Class, Modifier* and *ModifierClass*. (figure 3)

The attributes of *Rubric* are *xml:lang, kind, edition, changed* and *ID*. The attribute *kind* gives the information on the kind of Rubric that is in use. Its values for ICD-10 are: *text, preferred, preferredLong, preferredShort, footnote, inclusion, exclusion, note, definition, title, introduction* and *coding-hint*. As mentioned above these kinds of *Rubric* are described in *RubricKinds*, where

additional global information is stored that can be added to each Rubric entry by the processing tools.

The attributes *edition* and *changed* of *Rubric* are used the same as for the element *Class*.

If an entry is fragmented into a list or columns, its content can be displayed in the element *Fragment* of *Rubric*. *Fragment* has the attribute *use* (as dagger, asterix etc. in ICD-10), the required attribute *type* with the possible values *item*, *listhead* and *listitem* and the optional attribute *col* which is a positive integer value for the column number if the text of the superordinate rubric is to be formatted as a table (example 2).

The element *Rubric* can hold a reference to an element in the same classification or an external reference to another classification. Therefore the element *Reference* has the attributes *bracket* (if the reference should be printed in brackets, *use* (as dagger, asterix etc. in the ICD-10), the *name* of the reference and the *scheme* that is referenced to. The attribute scheme specifies another classification (like ICD-O) and is optional.

Formatting of parts of the text of a rubric can be achieved with the tags *Italics*, *Bold*, *Subscript* and *Superscript*.

The tag *Include* in a *Rubric* can be used to include rubrics, which are defined at another place in the classification.

Formatting information in Element Para

If a text should be printed in a special structured way this information can be put into the element *Para* (figure 4). *Para* does use the element *Reference* to refer to other parts of the classification as well as the formatting elements (as described in *Rubric*). The element *List* is used to hold information on text that is structured as a list and each list item is marked by a special symbol (like a dot or a minus or with a sequentially incremented label).

The element *table* of *Para* is derived from an official table DTD from Docbook. The DTD was transferred to a schema and then added in Para to enable an easy conversion of the XML-files to Docbook for easy printing. This table-DTD is described in “XML Exchange Table Model Document Type Definition” [4].

References

[1] CEN/TS 14463 “Health informatics – A syntax to represent the content of medical classification systems (ClaML)”
[2] http://www.altova.com/xmlspy
[3] http://www.dimdi.de/static/en/klassi/koop/who/etc/index.html (will be available in October)
[4] http://www.oasis-open.org/specs/tm9901.html

Technical Annex

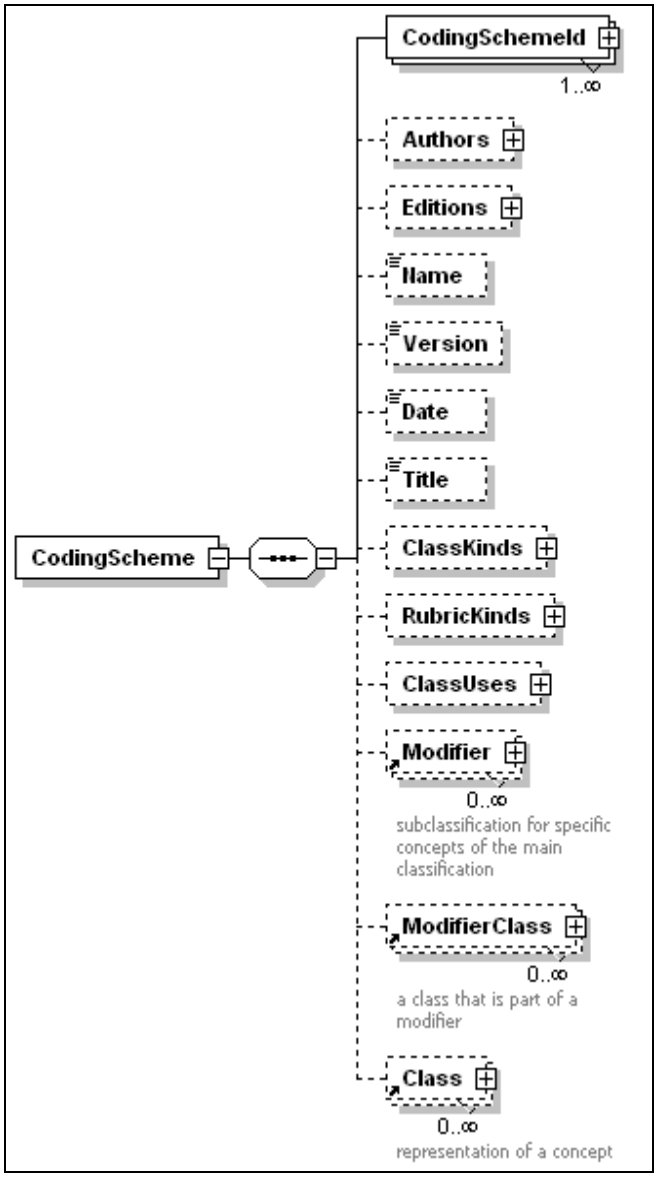

Figure 1 The element <CodingScheme>

```
<Modifier>
        <Symbol>S04E10#4</Symbol>
        <Rubric xml:lang="en" kind="text">
                <Para>The following fourth-character subdivisions are for use with
categories
E10-E14:</Para>
        </Rubric>
</Modifier>
<ModifierClass>
        <Symbol>.0</Symbol>
        <SuperClass>S04E10#4</SuperClass>
        <Rubric xml:lang="en" kind="preferred" id="id-1">With coma</Rubric>
        <Rubric xml:lang="en" kind="inclusion" id="id-381">
                <Fragment col="1" type="listhead">Diabetic:</Fragment>
                <Fragment col="1" type="listitem">coma with or without
ketoacidosis</Fragment>
        </Rubric>
        <Rubric xml:lang="en" kind="inclusion" id="id-382">
                <Fragment col="1" type="listhead">Diabetic:</Fragment>
                <Fragment col="1" type="listitem">hyperosmolar coma</Fragment>
        </Rubric>
        <Rubric xml:lang="en" kind="inclusion" id="id-383">
                <Fragment col="1" type="listhead">Diabetic:</Fragment>
                <Fragment col="1" type="listitem">hypoglycaemic coma</Fragment>
        </Rubric>
        <Rubric xml:lang="en" kind="inclusion" id="id-384">
                <Fragment col="1" type="item">Hyperglycaemic coma NOS</Fragment>
        </Rubric>

</ModifierClass>
```

Example 1: Use of <Modifier> and <ModifierClass>

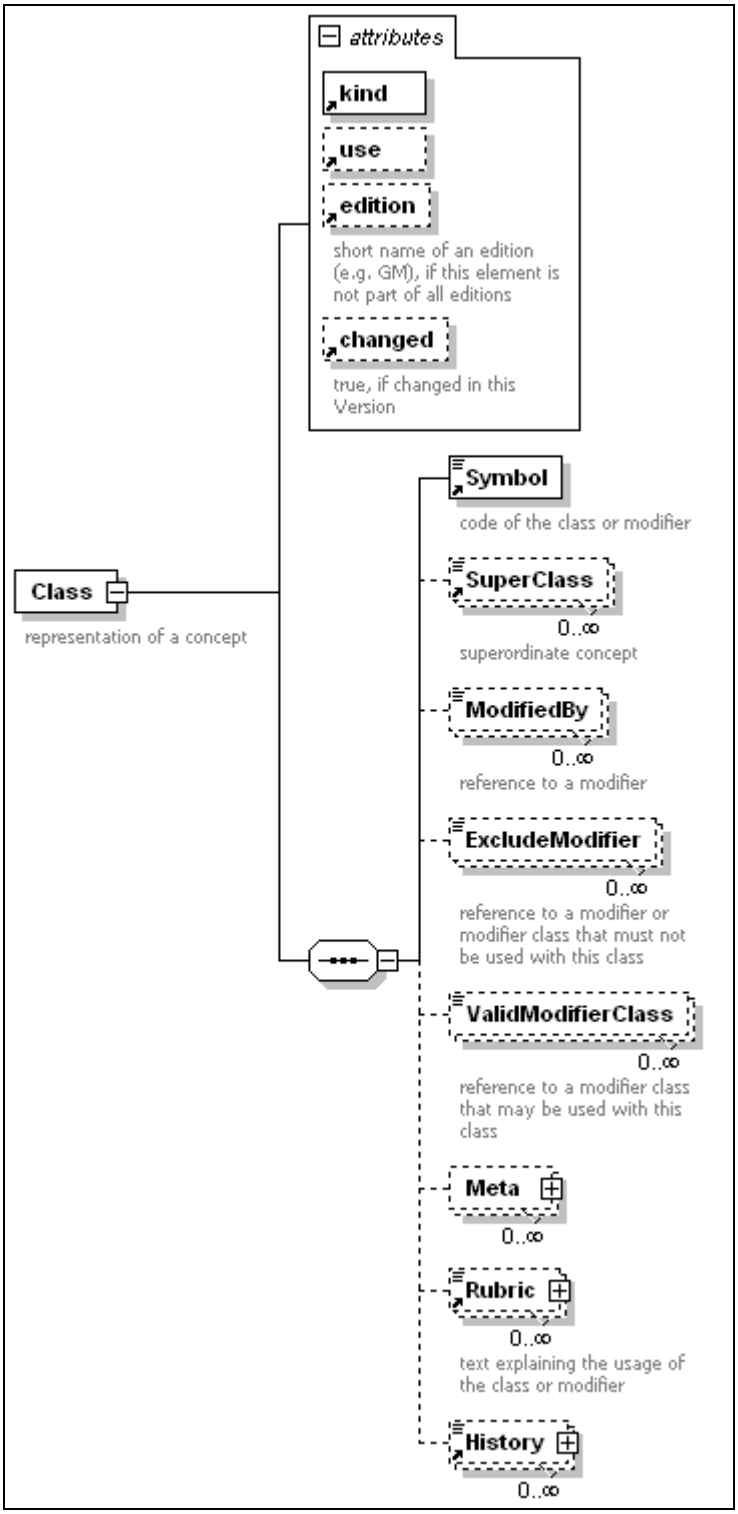
attributes
kind
use
edition
short name of an edition (e.g. GM), if this element is not part of all editions
changed
true, if changed in this Version
Class
representation of a concept
Symbol
code of the class or modifier
SuperClass
0..∞
superordinate concept
ModifiedBy
0..∞
reference to a modifier
ExcludeModifier
0..∞
reference to a modifier or modifier class that must not be used with this class
ValidModifierClass
0..∞
reference to a modifier class that may be used with this class
Meta
0..∞
Rubric
0..∞
text explaining the usage of the class or modifier
History
0..∞

Figure 2: The Element <Class>

```
<Class kind="category">
        <Symbol>A59.0</Symbol>
        <SuperClass>A59</SuperClass>
        <Rubric xml:lang="en" kind="preferred" id="id-315">Urogenital
trichomoniasis</Rubric>
        <Rubric xml:lang="en" kind="inclusion" id="id-1298">
                <Fragment col="1" type="item">Leukorrhoea (vaginalis)</Fragment>
                <Fragment col="2" type="item">due to Trichomonas
(vaginalis)</Fragment>
        </Rubric>
        <Rubric xml:lang="en" kind="inclusion" id="id-1299">
                <Fragment use="dagger" col="1" type="item">Prostatitis</Fragment>
                <Reference code=”N51.0” bracket="true" use="aster"/>
                <Fragment col="2" type="item">due to Trichomonas
(vaginalis)</Fragment>
        </Rubric>
```

Example 2: Use of <Class kind> and <Fragment>

```
<Class kind="section">
        <Symbol>M40-M54</Symbol>
        <SuperClass>M00-M99</SuperClass>
        <ModifiedBy>MD1</ModifiedBy>
        <Rubric xml:lang="en" kind="preferred" id="id-294">Dorsopathies</Rubric>
</Class>
<Class kind="section">
        <Symbol>M50-M54</Symbol>
        <SuperClass>M40-M54</SuperClass>
        <Rubric xml:lang="en" kind="preferred" id="id-360">Other
dorsopathies</Rubric>
        <Rubric xml:lang="en" kind="exclusion">
                <Fragment col="1" type="item">current injury - see injury of spine by
body region</Fragment>
        </Rubric>
        <Rubric xml:lang="en" kind="exclusion">
                <Fragment col="1" type="item"> discitis NOS</Fragment>
                <Reference code=”M46.4” bracket="true">M46.4</Reference>
        </Rubric>
</Class>
<Class kind="category">
        <Symbol>M50</Symbol>
        <SuperClass>M50-M54</SuperClass>
        <ExcludeModifier>MD1</ExcludeModifier>
        <Rubric xml:lang="en" kind="preferred" id="id-361">Cervical disc
disorders</Rubric>
        <Rubric xml:lang="en" kind="inclusion" id="id-865">
                <Fragment col="1" type="item">cervical disc disorders with
cervicalgia</Fragment>
        </Rubric>
        <Rubric xml:lang="en" kind="inclusion" id="id-866">
                <Fragment col="1" type="item">cervicothoracic disc
disorders</Fragment>
        </Rubric>
</Class>
```

Example 3: Use of <ModifiedBy> and <ExcludeModifier>

```
<Class use="aster" kind="category">
        <Symbol>M07</Symbol>
        <SuperClass>M05-M14</SuperClass>
        <ModifiedBy>S13M00</ModifiedBy>
        <Rubric xml:lang="en" kind="preferred" id="id-446">Arthritis psoriatica und
Arthritiden bei gastrointestinalen Grundkrankheiten</Rubric>
        <Rubric xml:lang="en" kind="text">
                <Reference code="S13M00" bracket="false"/>[Schlüsselnummer der
Lokalisation siehe am Kapitelanfang]
        </Rubric>
        <Rubric xml:lang="en" kind="exclusion">
                <Fragment col="1" type="item">Juvenile Arthritis psoriatica und juvenile
Arthritiden bei gastrointestinalen Grundkrankheiten</Fragment>
                <Reference code=”M09.-“ bracket="true" use="aster">M09.-</Reference>
        </Rubric>
</Class>

<Class use="aster" kind="category">
        <Symbol>M07.0</Symbol>
        <SuperClass>M07</SuperClass>
        <ValidModifierClass>0</ValidModifierClass>
        <ValidModifierClass>4</ValidModifierClass>
        <ValidModifierClass>7</ValidModifierClass>
        <ValidModifierClass>9</ValidModifierClass>
        <Rubric xml:lang="en" kind="preferred" id="id-447">Distale interphalangeale
Arthritis psoriatica<Reference code="L40.5" bracket="true" use="dagger"/>
        </Rubric>
</Class>
```

Example 4: Use of <ModifiedBy> and <ValidModifierClass>

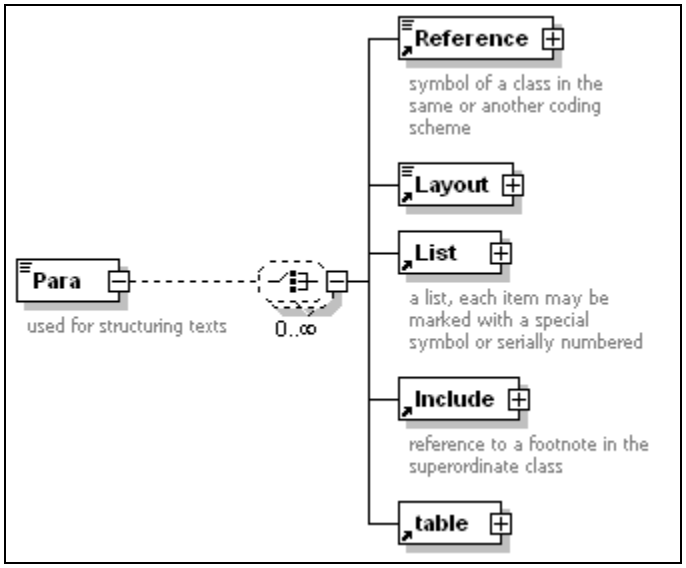

Figure 3: The Element <para>