

# DE Specs Working Group Meetings

Klaus Thoden

## 1 Meeting on September 19

### 1.1 DE Specs 0.1

Wolfgang presented version 0.1 of the DE Specs. They were then discussed in detail concerning structure and contents.

#### 1.1.1 Structure

There should be a logical order in the specifications. The most frequent features should occur first.

Examples should contain a picture and the text that is going to be typed (using another font).

There should also be an appendix which contains one transcribed page, the tags that were used, a list of ligatures and how they are going to be resolved and a list of characters that should be typed in directly.

The specifications should have a modular character so that instructions could be easily added or removed for special books or languages.

#### 1.1.2 Contents

**Columns** An illustration will show how columns are to be typed. They should be numbered from left to right.

**Hyphens** Instances of all sorts of hyphens should be included in the manual so that they are recognized correctly.

**Notes** Marginal notes are to be typed in the line they appear closest to and it should be stated on which side of the body they occur.

Footnotes are divided into two parts, the mark in the body and the text at the bottom of the page.

**Figures** The figure tag should be assigned to all sorts of images, be it illustrations, smaller ornaments or pictures at the beginning of a chapter.

There will be clear instructions in the specifications where to insert the tag depending on where the image is on the physical page.

**Help and gap tags** If there are characters that are unrecognizable due to a physical defect but might still be considered as readable by an expert, the `<help>` tag should be written.

The `<gap>` tag is used to denote totally unreadable parts of the text.

**Italics and small caps** Words in italics should be surrounded by an underscore. Whole paragraphs in italics will be marked up by an argument in the paragraph tag.

Text in small caps is to be surrounded by opening and closing tags.

## 2 Meeting on September 23

### 2.1 DE Specs 0.2

Version 0.2 of the DE Specs was presented. Contents and structure were again discussed. The pictures in the examples should include a mark pointing directly to the item in question.

#### 2.1.1 New issues

**Handwriting** A tag should be added where a handwritten text or figure occurs so that these instances can be found more easily.

**Quotations** Two structures are to be distinguished: inline quotations and block quotations. Paragraphs that only contain quotation are tagged as a block quotation.

**Tables** Tables should be typed in lines with special field separator tags (e. g. #). The fact that distinguishes tables from columns is that they do not contain running text.

**Footers** Footers, if recognized, are assigned the same tags as headings.

#### 2.1.2 Typing conventions

Most of the characters found in the books should be typed in directly. Most of the ligatures are going to be resolved into the basic letters with a tag that there was a ligature.

If a character cannot be typed directly it should be marked as an unknown character with an ID. The digitized text should then also contain a list of these unknown characters.

Greek should be entered in Unicode.

## 2.2 First introduction to XML Schema

There are two documents that will serve as starting points for the schema. They can be accessed via the wiki. The *Dublin Core* standard will be used for the metadata.

Other issues will be discussed after delivering the DE Specs 1.0.

## **2.3 Next Steps**

### **2.3.1 Milestones**

Version 1.0 will contain only specifications for texts written using the latin or greek alphabet. Later versions will also cover Chinese and fraktur texts.

### **2.3.2 Work sample**

Before all books are sent to China, a small work sample will be made so that the specifications can be modified.