

Getting started

Lou Burnard Consulting

2019-09-14

This tutorial shows you how to create an ELTeC level zero file, and how to complete the metadata required by the TEI Header.

If you haven't done so already, download the working files for this tutorial, unpack them into a folder somewhere on your machine, and then open the ELTeC.xpr project file in oXygen (choose Project->Open project from the menu, or hit CTRL-F2).

1 Getting started

- Open oXygen and click on the New icon (first left on the toolbar); or select File->New from the menu; or type CTRL-N
- The New File dialogue appears. Under the heading **Global templates**, you should see the template for ELTeC-0. Select it. (If it's not there, you need to tweak your installation of oXygen; or you can simply open the template file ELTeC-0.xml from your Work directory).
- Change the filename in the bottom of the window from Untitled.xml or ELTeC-0 to something more appropriate: we suggest using the identifier for the file you're planning to work on. Either way...
- ... the main oXygen editing window opens, with the beginnings of an ELTeC XML file in it. Note the following:
 - At the top of the file there are two purple lines beginning `<?xml-model`.
 - The file contains tags (in blue) and attribute names (in orange) and helpful comments (in green) ... but no text other than some bits of the header.
 - To the right of the main window there is a status bar with an angry red square at the top and some red error flags. These correspond with parts of the text display underlined in red.
 - Underneath the editing window there is an error message corresponding with the first error flag: **value of attribute "xml:id" is invalid; must be an XML name without colons**
- The red flags tell us that although oXygen has done its best some additional information is needed to create a valid header. Let's fix up some of these errors!

2 Making the header valid

- The schema requires a proper value for @xml:id but we have not yet supplied one. Put the cursor between the quote marks and type in the identifier for your novel. A new error message appears! **element "TEI" missing required attribute "xml:lang"**.
- Put the cursor before the > at the end of the <TEI> start-tag and type a space to see the available attributes. *@xml:lang* appears in the list in bold because it's required here. Select it, and then add the language code for your document.
- The next red line is complaining about the content of the <measure> element at line 17, which should contain a number indicating how many words the text contains. If you don't know the wordcount for your text, you can enter any number you like for the moment, but don't forget to count the words in your text and correct this when you've finished! Note that the count is supplied as content of an element, not an attribute value.

2 MAKING THE HEADER VALID

- The next three red lines are all to do with the text classification codes. You need to decide what value to enter for the "authorGender" category (M, F, or U); what size code corresponds with the word count ('short', 'medium', or 'long'); and what time slot code (T1,T2,T3 or T4) is applicable. You have the data: refer to the header documentation (for example on the wiki) for a reminder on what the codes mean.
- To remove the only remaining red line in the header, enter today's date, in the right (YYYY-MM-DD) format as value for the *@when* attribute. Phew! no red marks anywhere in the Header – which is now therefore valid: click the disk icon (or type CTRL-S, or choose File-Save from the menu) to save it for further work.
- This is a good start, but validity is not necessarily the same thing as truth, and we would prefer our document to be true – the title of this work is definitely not **Title**, and your name is not **Your Name Here!** More seriously, we have not yet provided any information at all about the source from which this ELTeC text derives nor about its author. Over to you... Referring to the metadata you accumulated in the first part of this exercise, you should now add the following:
 - the actual title of the novel
 - the author's name and dates in the recommended format
 - an additional `<respStmt>` to show who was responsible for the original data capture of this title (if you know)
 - the date of publication of this digital version (today!)
 - information about the source as described below
- The source description should contain two `<bibl>` elements, one (with *@type digitalSource*) for the digital source from which your version derives, and one (with *@type firstEdition*) for the first print edition of this work. Without a lot of detective work, it's hard to tell whether this text was transcribed from the first edition, or via some intermediate modern edition, but if you are reasonably sure that a modern edition is the actual source, you could supply a third `<bibl>` with information about it.
- Your research should have provided you with enough information to provide a basic bibliographic description of the first edition of the novel, including for example publisher name, place of publication, as well as the original published title and form of the author's name, which might differ from the versions in use today. At the very least, your `<bibl type="firstEdition">` should contain a `<date>` giving the year of first publication.

When you're happy with your header, and it has no redlines, remember to save the file before proceeding to add something to the `<body>` element!