

Creating a basic TEI P5 XML file

James Cummings – @jamescummings

2015-03-09

1 Exercise 1: Creating an XML File

1.1 Summary

This exercise will walk you through creating a TEI P5 XML document in the oXygen editor and introduce a variety of ways to mark this document up. You will first start a new TEI P5 XML file, then insert some unmarked up text into the editor, and then mark up the stanzas or line-groups (lg) and lines (l). You will learn to check that your document is well-formed or not.

Note: read each step before you do it.

1.2 Starting A New TEI P5 XML File

Let's start a new TEI P5 XML file by following the following steps:

- Load up the oXygen XML Editor if it isn't already loaded by using the Windows Start Menu, or double-clicking the icon on the desktop. Note: It is the 'Editor' you want, not 'Author'.
- If you get prompted for a license key, copy and paste the license key you have been given.
- Once the editor has fully loaded you may wish to close any additional popup windows, and optionally the 'views' on the left and right-hand side.
- From the 'File' menu select 'New' and double click on 'New Document', scroll down to 'TEI P5', double-click on that and 'All'. This should open up a new TEI P5 XML document.

1.3 Filling in Required Information

Some information is required in every TEI P5 XML file.

- Title: Call your edition "My TEI Edition" (or similar)
- Publication statement: Put the name of the workshop here, or your own statement of why this is being created.
- Source description: Here state which text(s) you are using for this exercise. You should choose from between the following texts (though you could use your own if you really want) or do more than one of them as separate divisions.
 - Demain, dès l'aube... (Victor Hugo)
 - Nähe des Geliebten (Johann Wolfgang von Goethe)
 - A red, red, rose (Robert Burns)

1.4 Creating a Division

Let's assume we're going to be creating a collection of poems each in its own division. Delete any existing markup inside the `<body>` element and create a division using the `<div>` element. This is a generic division or section element.

- On the line below the XML declaration type: `<div>`
- Notice what happens when you type the final '>'. oXygen is trying to help you and inserts in the closing `</div>` tag. This is because it knows the rules of XML, and knows that if you type an opening `<div>` you are required to have a closing `</div>` sooner or later.

1 EXERCISE 1: CREATING AN XML FILE

- We haven't said what *type* of division this is, which means it is just a generic division. Let's categorise it as 'poem' by adding a *type* attribute. Move the cursor back until you just after the letter 'v' in the opening tag. Press space, and then start to type: **type="** and notice what happens. oXygen tries to help you because it knows you are editing a TEI document.
- Also note what happens when you type the quotation mark. oXygen is again trying to help you by putting the closing quotation mark, because it knows that attribute values must always be quoted.
- In between the quotation marks type **poem** to categorise our division as being verse.
- Move back until you are directly in between the opening **<div>** and closing **</div>**. Press 'enter' a couple times to give yourself some space inside the element.

1.5 Inserting Some Text

It would waste a lot of time if we asked you to type the whole poem in, so we've done that for you.

- Make sure your cursor is in-between the opening **<div>** and the closing **</div>**
- Highlight the text of the poem in your web browser with the mouse and press control and 'c' to copy. (Depending on your operating system, but you really should know how to copy and paste text.)
- Go back to the oXygen window and ensuring your cursor is between the opening and closing "div" tags, then press control and 'v' to paste (again, depending on your operating system)
- The start of your **<body>** element should look like:

```
<text>
  <body>
    <div type="poem">
      [Your pasted poem text here]

    </div>
  </body>
</text>
```

1.6 Encoding the Heading (using 'Surround with Tags')

The first line of these files has the title of the poem. This is a heading, so the TEI **<head>** element should be used to mark this. To mark this do the following:

- Highlight the text of the title with the mouse or keyboard.
- Either press control and 'e' (or command and 'e') as a shortcut key, or right-click and under 'Refactoring' select 'Surround with Tags'. A box should pop up and you can type **head** into it and press enter. Notice how oXygen helps you by listing the elements as you type 'head' and again by putting the opening tag before what you had highlighted and the close tag afterwards.

1.7 Marking Stanzas (using both 'Surround with Tags' and 'Split Element')

Let's mark the stanzas that appear doing the following steps:

- Highlight the first stanza as a whole.
- Using control-e, or the menus, as you did above, mark this stanza as an `<lg>` element.
- Add a type attribute with a value of 'stanza' to the `<lg>` element so it looks like: `<lg type="stanza">`.
- This portion of your document should now look like:

```
<text>
  <body>
    <div type="poem">
      <head>[Author's Name]</head>
      <lg type="stanza"> [text of first stanza]
      </lg>

      [More text of the poem]

    </div>
  </body>
</text>
```

- But if we have a number of stanzas, marking each one of them seems a lot of work, but there is a (possibly) easier way.
- Highlight the entire rest of the poem, from after the end of the `</lg>` to just before the author's name, and then surround all of it in an `<lg>` element (by pressing control-e)
- Of course it is silly to have the *entire* rest of the poem marked as a single line-group, but go and add a type attribute with "stanza" as the value to the opening tag. Obviously this is just not true, but it allows us to use a particular shortcut for marking large amounts of text.
- If you move the cursor to just before the start of each stanza, and press alt-shift-d (or select Refactoring -> Split Element from the right-click menu), oXygen should split the `<lg>` element, ending it here and starting it just before where there cursor is located.
- Do not worry about the opening and closing tags not being precisely in the right place. We will 'format and indent' at the end of the exercise.
- Do this for other stanzas that are not yet marked.

1.8 Marking Lines

We've marked all the stanzas but we've not marked the lines.

- Highlight the first line in the first stanza, press control-e to surround with a tag, and type 'l' as the element name. (`<l>` is the line element, meaning a line of metrical verse).
- It might be a bit painful to mark up each and every line this way, you could try using the split-element technique above, but there is another shortcut to try as well. Highlight the second line and press control-/ and notice that oXygen has wrapped the line in a `<l>` element. The reason for this is that control-/ is the 'surround with the last element I surrounded something with' shortcut key.
- Using this technique, quickly mark all the remaining lines.

1.9 The Author's Signature

Clearly this isn't a manuscript copy we're working from, but let's pretend that the name at the bottom of the poem is actually an authorial signature.

- Highlight the name, press control and 'e' for surround-with-element, and mark it up as `<signed>`.

1.10 Format and Indent

Our poem is marked up, but some of the markup might be a bit messy.

- Make sure that your file is 'well-formed'. You'll be able to tell it is well-formed because oXygen will have a *happy green square* in the upper right-hand corner. If it is an *angry red square*, you better find the problem (where a red bar on the right-hand side is) and correct the mistake! (Note: if you are red/green colour blind, ask me and I'll show you how to change this to colours you can differentiate between!)
- Now let's format and indent our file. This tidies up some of the whitespace and indents elements based on their place in the hierarchy. Either select the 'Format and Indent' icon from the toolbar (it looks like some indented lines), or go to the menus: 'Document' -> 'Source' -> 'Format and Indent'.
- Formatting and indenting your markup is not necessary, it could all be on one big long line, but it makes it much easier for other people to read.

1.11 Saving Your Work

Let's save our work:

- Is your work well-formed? Do you have a happy green square or an angry red one?
- From the 'File' menu select 'Save' or click on the Save icon (looks like an old-style 3.5" disk)
- Save the file using a name of your choice.

1.12 Self-Assessment

Check if you understand some of the core principles of this exercise by answering the following questions to yourself:

- How do you start a new TEI P5 XML document in oXygen?
- What information is mandatory in a `<teiHeader>`?
- What is a well-formed document?
- How do I 'Surround with tag' and repeat that action quickly?
- Why might using the 'Split element' approach be useful?
- What is the advantage of formatting and indenting your markup?

1.13 Next?

If you are finished early you may wish to add divisions for the other poems or browse through the TEI Guidelines online at <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index-toc.html>. In particular you might want to look at the Elements appendix for reference pages for individual elements. Consider looking up all the elements you've used in this file to see how they are defined.