Practical Exercise 2:

Martina Scholger *2016-07-04*

1 Exercise 2: Create a TEI Document

1.1 Learning Outcomes

When you successfully completed this exercise you should be able to:

- discern the elements and attributes needed for a minimum valid TEI XML file

- have used the TEI namespace

- create a minimum TEI header and text body

- check for both validity and well-formedness

1.2 Summary

This exercise will walk you through creating a TEI XML document and inserting the work you did previously into it. You'll learn about the minimal required aspects of the `<teiHeader>` and the basic structure of a TEI file.

1.3 Start a New XML File

Follow the same steps you did for the first exercise to start a new blank XML file. Although we could start a file with a TEI P5 template, for this particular exercise that would be cheating!

- 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.

- Once the editor has fully loaded from the 'File' menu select 'New' and under 'New Document' select 'XML Document'. This should open up a blank file with an XML Declaration added.

- An XML Declaration looks like:

```
<?xml version="1.0" encoding="UTF-8"?>
```

1.4 Inserting a `<TEI>` Element

All TEI files start either with a `<TEI>` element or a `<teiCorpus>` element. In most cases you'll want a `<TEI>` element. These elements have a special pseudo-attribute called 'xmlns' that indicates the namespace a set of elements is from. This is inherited by any elements inside it (unless overridden). This is how we can be sure we're talking about, say, a `<title>` element from the TEI rather than any other schema.

- Add a `<TEI>` element and then add it to the TEI namespace (http://www.tei-c.org/ns/1.0) as the value to an *@xmlns* attribute. Maybe add a few blank lines between the starting and closing `<TEI>` tag. Your file should look like:

```
<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">

</TEI>
```

- Notice what happens in oXygen and how it helps you input this. Also notice that your file may now have an angry red square rather than a happy green one! Is your file well-formed? (yes, it is!) Why is this red then?

- If it is red it is because your version of oXygen is prepackaged with all sorts of TEI goodness, and in this case it recognises that files starting with `<TEI>` in the TEI namespace are to be associated automatically with a TEI schema that it has stored. It is complaining that you do not have a `<teiHeader>` in your file because all valid TEI files must have this.

1.5   Adding a `<teiHeader>`

Inside the `<TEI>` element we need to add a `<teiHeader>` element.

- Put the cursor between the starting and closing `<TEI>` element and type in a `<teiHeader>` element. Notice that oXygen provides the closing `</teiHeader>` element. If the correct option is set in oXygen, it understands the TEI schema and knows that certain content is required inside a `<teiHeader>`. It can automatically provide that markup. If not, you'll have to type it in! The resulting file should look like:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
<teiHeader>
 <fileDesc>
  <titleStmt>
   <title>
<!-- you'll need to add a title here -->
   </title>
  </titleStmt>
  <publicationStmt>
<!-- you'll need to add a paragraph here -->
  </publicationStmt>
  <sourceDesc>
<!-- source will go here -->
  </sourceDesc>
 </fileDesc>
</teiHeader>
<!-- <text> and other elements go here -->
</TEI>
```

- Notice that your file still has an angry red square rather than a happy green square. This is because there is still stuff needed even though you've added some markup. First, add a title of something like "My Shakespeare Collection" by adding this text between the starting and closing `<title>` tags. There are other elements which are allowed here in `<titleStmt>` such as `<author>` (William Shakespeare), that you could add but aren't really required for this exercise. You could use the more general `<respStmt>` (with a `<name>` element with your name and a `<resp>` element with something like 'TEI P5 Encoding' in it) to record your own work if you wish, but as with the other embellishments this isn't necessary for this exercise.

- Then add a paragraph `<p>` inside the `<publicationStmt>` with some text to record what this file is for, perhaps something like "Created as part of the TEI Workshop at DHOxSS16"

- Inside `<sourceDesc>` we should add a `<p>` with some text like: "Shakespeare's Sonnets and First Folio."

- Your `<teiHeader>` should now look something like:

```
<teiHeader>
 <fileDesc>
  <titleStmt>
   <title>My Shakespeare Collection</title>
  </titleStmt>
  <publicationStmt>
   <p>Created as part of a TEI Workshop at DHOXSS16</p>
  </publicationStmt>
  <sourceDesc>
   <p>Shakespeare's Sonnets and First Folio.</p>
  </sourceDesc>
 </fileDesc>
</teiHeader>
```

- Notice that even though this is a complete `<teiHeader>` with all the required aspects, our file as a whole isn't valid.

1.6   Add a `<text>`

All TEI files, in addition to a `<teiHeader>` with `<fileDesc>` containing a `<titleStmt>`, `<publicationStmt>`, and `<sourceDesc>`, need to follow the header with at least one of: `<sourceDoc>`, `<facsimile>`, or `<text>`. In our case we're going to add a `<text>` element. To do this:

- Add a couple of blank lines after the closing `</teiHeader>`.

- Insert a `<text>` element and inside that a `<body>` element. (The `<text>` element requires a `<body>` element because if you don't have a text body, what are you encoding?)

- The `<text>` section of the file should look something like:

```
<text>
 <body>
<!--We will put our sonnet here -->
 </body>
</text>
```

1.7   Adding Our Sonnet

This is a good start but we need to put something inside the body. Luckily, we have already encoded a Sonnet in the previous exercise, so we can use that!

- With the cursor in between the opening and closing `<body>` tags go to the 'Document' menu on the menu bar, and select 'File', and 'Insert File'. Select the file you saved earlier if you finished the first exercise. If you didn't manage to finish the first exercise in time, don't worry, there is a completed version at: view-source:http://tei.it.ox.ac.uk/Talks/2016-07-dhoxss/spoilers/ ex01-spoiler.xml?style=raw and view-source:http://tei.it.ox.ac.uk/Talks/2016-07-dhoxss/spoilers/ ex01-spoiler.xml?style=raw which contains the completed first exercise.

- But wait, as soon as you've added this we get a bit of a problem! oXygen will complain that we've got an XML Declaration in the middle of our file. Delete this redundant XML declaration!

- Your document should now be valid and have a happy green square in the upper right-hand corner! If it isn't, try to solve the problem by looking at the error message that is provided.

1.8 Adding another division: Drama, Baby!

Now, we will add Act 1, Scene 1 of the drama 'Henry IV.' to our Shakespeare collection. You will need the following files:

- Drama: http://tei.it.ox.ac.uk/Talks/2016-07-dhoxss/materials/henryIV.txt

- Images: http://tei.it.ox.ac.uk/Talks/2016-07-dhoxss/materials/FF-p48.jpg and http://tei.it.ox.ac.uk/Talks/2016-07-dhoxss/materials/FF-p49.jpg

- After the closing **\<div\>** of the sonnet, add another **\<div\>** with a *@type* attribute with the value 'drama'. To reduce the amount of typing you have to do, there is a transcript of the scene already prepared for you at. With the cursor in between the opening and closing **\<div\>** tags go to http://tei.it.ox.ac.uk/Talks/2016-07-dhoxss/materials/henryIV.txt in your browser highlight the entire drama, copy it and paste it in the oXygen window

- The angry red square will now resurface - don't worry about, you are going to fix the document as the exercise progresses!

- Now, you will need to manually encode this new type of text.

- The number in square brackets ([48]) and further down ([49]) signify the start of a new page in the text. To encode this, use the empty **\<pb /\>** element with an *@n* attribute and the given number.

- Start to encoding the first text line with the **\<head\>** element.

- The number in the curly brackets ({1}) signifies the start of a column in the text. To encode this, use the empty **\<cb /\>** element with an *@n* attribute and the given number.

- The next line contains stage directions which are encoded using the **\<stage\>** element.

- Your division should now look like:

```
<div type="drama">
 <pb n="48"/>
 <head>Actus Primus. Scœna Prima.</head>
 <cb n="1"/>
 <stage>Enter the King, Lord Iohn of Lancaster, Earle of
   Westmerland, with others.</stage>
King.
SO shaken as we are, so wan with care,
Find we a time for frighted peace to pant
And breathe shortwinded accents of new broils

<!-- And so on... -->
</div>
```

- At this point, a sequence of speeches begins which make up the actual scene. Speeches use the **\<sp\>** element and are composed of a 'Speaker', designated with the **\<speaker\>** element, and his or her text which is encoded line by line using the **\<l\>** element. Highlight the first speech (beginning with 'King' and ending with the line 'In forwarding this deare expedience.') but no other elements. Either press control-e as a shortcut key, or right-click an under 'Refactoring' select 'Surrounding with Tags'. A box should pop up and type 'sp' into it and press enter. Notice how oXygen helps you again by putting the opening before what you had highlighted and the close tag afterwards. This should surround the entire first speech with the **\<sp\>** element.

- Use the same 'Surrounding with Tags' function to surround the speaker in the first line with the **<speaker>** element and the following lines with the **<l>** element.

- The surround-with-tags shortcut is very useful. But also remember the split-element shortcut (alt-Shift-D) from the first exercise to split the line element.

- Format and indent the file as you learned in the previous exercise!

- The beginning of our drama **<div>** should therefor look something like:

```xml
<div type="drama">
 <pb n="48"/>
 <head>Actus Primus. Scœna Prima.</head>
 <cb n="1"/>
 <stage>Enter the King, Lord Iohn of Lancaster, Earle of
   Westmerland, with others.</stage>
 <sp>
  <speaker>King.</speaker>
  <l>SO shaken as we are, so wan with care,</l>
  <l>Find we a time for frighted peace to pant</l>
  <l>And breathe shortwinded accents of new broils</l>
<!-- And so on... -->
 </sp>
</div>
```

- Encode the example using the elements mentioned above

- Your document should now again be valid and have a happy green square in the upper right-hand corner! If it isn't, try to solve the problem by looking at the error message that is provided.

1.9   Adding more markup

- To improve the document further, you may want to encode additional information inside the scene, e.g. by rendering the Speakers in italics using the *@rend* attribute with a value of 'italic'.

- Look at the original rendering of the text. Some text passages are centered. You can simply add a second value to the *@rend* attribute with the value 'center'.

- Furthermore you can specify the nature of the stage directions using the *@type* attribute with a value 'entrance' or 'exit'.

- You may wish to hightlight more persons in the spoken lines by using the **<hi>** element and again adding the *@rend* attribute with a value 'italic'.

1.10   The result

Since we are encoding the first scene of the first act we can add the corresponding **<div>**'s to our TEI document since our text passage is actually nested inside scene 1 which is nested in turn in act 1. Highlight the whole text (except the first page break) and add a **<div>** element with the *@type* attribute of the value 'scene' and the *@n* attribute with the value '1', using the surround-with-tag shortcut. Then repeat this step and add another **<div>** element with the *@type* attribute of the value 'act' and the *@n* attribute with the value '1'.

You might have recognized the '\' in the first stage direction, which signals a line-break in the soure document. Replace the '\' with the empty **<lb />** element.

In the end our result should look something like:

```
<div type="drama">
 <pb n="48"/>
 <div type="act" n="1">
  <div type="scene" n="1">
   <head rend="italic center">Actus Primus. Scœna Prima.</head>
   <cb n="1"/>
   <stage rend="italic center"
    type="entrance">Enter the King, Lord Iohn of Lancaster, Earle of
<lb/>Westmerland, with
       others.</stage>
   <sp>
   <speaker rend="italic center">King.</speaker>
   <l>
    <hi rend="decoratedCapital">S</hi>O shaken as we are, so wan with care,</l>
   <l>Find we a time for frighted peace to pant</l>
   <l>And breathe shortwinded accents of new broils</l>
   <l>To be commenc'd in Stronds a-farre remote:</l>
   <l>No more the thirsty entrance of this Soile,</l>
   <l>Shall daub her lippes with her owne childrens blood:</l>
   <l>No more shall trenching Warre channell her fields,</l>
   <l>Nor bruise her Flowrets with the Armed hoofes</l>
   <l>Of hostile paces. Those opposed eyes,</l>
   <l>Which, like the Meteors of a troubled Heauen,</l>
   <l>All of one Nature, of one Substance bred,</l>
   <l>Did lately meet in the intestine shocke,</l>
   <l>And furious cloze of ciuill Butchery,</l>
   <l>Shall now in mutuall well-beseeming rankes</l>
   <l>March all one way, and be no more oppos'd</l>
   <l>Against Acquaintance, Kindred, and Allies.</l>
   <l>The edge of Warre, like an ill-sheathed knife,</l>
   <l>No more shall cut his master. Therefore Friends,</l>
   <l>As farre as to the Sepulcher of Christ,</l>
   <l>Whose Souldier now vnder whose blessed Crosse</l>
   <l>We are impressed and ingag'd to fight,</l>
   <l>Forthwith a power of English shall we leuie,</l>
   <l>Whose armes were moulded in their mothers wombe,</l>
   <l>To chase these pagans in those holy Fields,</l>
   <l>Ouer whose acres walk'd those blessed feete</l>
   <l>Which fourteene hundred yeares ago were nail'd</l>
   <l>For our aduantage on the bitter Crosse.</l>
   <l>But this our purpose is a tweluemonth old,</l>
   <l>And bootlesse 'tis to tell you we will go:</l>
   <l>Therefore we meete not now. Then let me heare</l>
   <l>Of you my gentle Cousin Westmerland,</l>
   <l>What yesternight our Councell did decree,</l>
   <l>In forwarding this deare expedience.</l>
   </sp>
   <sp>
   <speaker rend="italic">West.</speaker>
   <l>My Liege: This haste was hot in question,</l>
   <l>And many limits of the Charge set downe</l>
   <l>But yesternight: when all athwart there came</l>
   <l>A Post from Wales, loaden with heauy Newes;</l>
   <l>Whose worst was, That the Noble <hi rend="italic">Mortimer</hi>,</l>
   <l>Leading the men of Herefordshire to fight</l>
   <l>Against the irregular and wilde <hi rend="italic">Glendower</hi>,</l>
   <l>Was by the rude hands of that Welshman taken,</l>
   <l>And a thousand of his people butchered:</l>
   <cb n="2"/>
   <l>Vpon whose dead corpes there was such misuse,</l>
   <l>Such beastly, shamelesse transformation,</l>
```

```
      <l>By those Welshwomen done, as may not be</l>
      <l>(Without much shame) re-told or spoken of.</l>
    </sp>
<!-- more speeches -->
  </div>
 </div>
</div>
```

If you didn't manage to finish the exercise in time, don't worry, there is a pre-completed version at: view-source:http://tei.it.ox.ac.uk/Talks/2016-07-dhoxss/spoilers/ex02-spoiler.xml?style=raw

1.11 Saving Your Work

Let's save our work:

- Have you formatted and indented your work automatically?

- 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 the name 'exercise02.xml' or another name of your choice.

1.12 Self-Assessment

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

- Which elements and attributes do you need for a minimum valid TEI XML document?

- What three parts of the `<teiHeader>` are required in all TEI conformant documents?

- Where are these elements and attributes allowed?

- What is the function of each element and attribute you've used?

- Why do you think these elements and attributes are required in TEI XML?

2 Next and Further Reading

These exercises should have given you some experience editing XML and making a valid TEI file.

- You may wish to 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 of reference pages for individual elements. Consider looking up all the elements you've used in this file to see how they are defined.

- What other elements are allowed inside the `<text>` element? What would you use them for?

- What other parts of the `<teiHeader>` are there? What are they for?

- You may wish to read the chapters on Default Text Structure http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html or Elements Available to All TEI Documents http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html.