

Exercise 2: Making a valid TEI Document

James Cummings

28 January 2014

1 Exercise 2: Making a Valid TEI Document

1.1 Learning Outcomes

When you successfully complete the first part of this exercise you should be able to:

- discern the elements and attributes needed for a minimum valid TEI XML file
- associate a TEI XML file with a schema
- use the TEI namespace
- create a minimal TEI header and text body
- check for both validity and well-formedness

1.2 Summary

This part of this exercise will walk you through creating a TEI XML file and inserting the work you did previously into it. You'll learn about the required aspects of the `<teiHeader>` and the basic structure of a TEI file. We will also add some phrase-level markup if we have time. This is really quite a short and simple exercise which will hopefully be fairly easy!

1.3 Starting A New XML File

Let's start a new XML file by following each of these steps:

1. Load up the oXygen XML Editor if it isn't already loaded by using the Start Menu (in Windows), or double-clicking the icon on the desktop.
2. Once the editor has fully loaded from the File menu select New and expand the New Document section so you can select XML Document. Doing this should open up a blank document with an XML Declaration added.

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 are 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.

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

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

</TEI>
```

2. 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?
3. 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.
4. If it is not showing that it is invalid, ask and we'll show you how to associate your document with the schema at: http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_all.rng

1.5 Adding a `<teiHeader>`

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

1. 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 at this point should look like:

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

</TEI>
```

2. Notice that your file still has an angry red square rather than a happy green square. This is because there are some required elements missing, even though you've added some of the markup.
3. First, add a title like "Attorney-General's Information about Libellous Songs in Sheinton" by adding this text between the starting and closing `<title>` tags. There are other elements which are allowed here in `<titleStmt>` such as `<editor>` (J. Alan B. Somerset), have a look at what is allowed (by typing '`<`'. You could also 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 this isn't strictly necessary for this exercise.
4. Then add a paragraph `<p>` inside the `<publicationStmt>` with some text to record what this file is for, perhaps something like "An exercise for learning TEI."

5. Inside `sourceDesc` we should add a `<p>` with some text like: "REED transcription of London, Public Record Office, STAC 8/20/8."
6. Your file should now look something like:

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Attorney-General's Information about Libellous Songs in Sheinton</title>
      <editor>J. Alan B. Somerset</editor>
    </titleStmt>
    <publicationStmt>
      <p>An exercise for learning TEI</p>
    </publicationStmt>
    <sourceDesc>
      <p>REED transcription of London, Public Record Office, STAC 8/20/8.</p>
    </sourceDesc>
  </fileDesc>
</teiHeader>

</TEI>
```

7. Notice that even though this is a complete `<teiHeader>` with very little information but technically all the required aspects, our file as a whole isn't valid yet.

1.6 Adding some `<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 the `<text>` element by putting in the file we made earlier. To do this:

1. Add a couple of blank lines after the closing `</teiHeader>` just for some space.
2. 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. Note that this is from the 'Document' menu! Select the file you saved earlier if you finished the first exercise. If you didn't then there is a file called 'exercise-01.xml' in the 'spoilers' directory which has the fully completed first exercise.
3. As soon as you've added this, oXygen will still think our file is invalid because we have got that XML declaration before the `<text>` element. Delete this redundant XML declaration!
4. 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.

1.7 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.

1.8 Self-Assessment

Check that 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? (If you are unsure, look them up!)
- Why do you think these elements and attributes are required in TEI XML?

1.9 Further Reading

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

- If you are finished early 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?
- 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>.