

Exercise 3: Improving the TEI Header

James Cummings

28 January 2014

1 Exercise 3: Improving the TEI Header

1.1 Learning Outcomes

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

- improve the structure and metadata of a `<teiHeader>`.
- understand the components of a `<fileDesc>` including:
 - `<titleStmt>` for title and intellectual responsibility.
 - `<publicationStmt>` for information about the publication and distribution of the electronic item.
 - `<sourceDesc>` to record metadata about the source document.
- use the `<encodingDesc>` to record the markup used in the file.
- use the `<profileDesc>` to record non-bibliographic aspects of the file.
- record major changes to the file in the `<revisionDesc>`.
- and if you have time, mark up some names!

1.2 Summary

This exercise gives you a chance to significantly improve the header of a TEI XML file and understand its markup and structure. It walks you through improvements to various aspects of the `<teiHeader>` and how to record additional metadata about the electronic file and its sources. If you have time it also gives you a chance to mark up some of the named entities inside it.

1.3 Starting Up

Load up the completed file from the previous exercise. If you did not complete the exercise you can cheat by loading up `'spoilers/exercise-02.xml'` and then saving it under a new name where you put your other files.

1.4 Improving the `<titleStmt>`

As you can hopefully now see, the `<teiHeader>` we created earlier is lacking a lot of information. Let's improve it!

1. Inside the `<fileDesc>` the `<titleStmt>` contains only a `<title>`. What else can `<titleStmt>` contain? (hint: typing `'<'` here will provoke oXygen into providing a drop down list of possibilities).
2. If you haven't already, underneath the `<title>` add an `<editor>` element. The content of this should be `'J. Alan B. Somerset'`.
3. After this add a `<principal>` element to record the person primarily responsible for the electronic version of this file. In this case, use your own name!

4. Below this add a `<meeting>` element with the content of 'TEI Workshop'.
5. Below this add a `<sponsor>` element with the content of 'Records of Early English Drama project'.
6. After that add `<respStmt>` with a `<resp>` inside it saying 'TEI P5 XML Encoding' and a `<name>` with your name.
7. Your `<titleStmt>` should now look something like:

```
<titleStmt>
  <title>Attorney-General's Information about Libellous Songs in Sheinton</title>
  <editor>J. Alan B. Somerset</editor>
  <principal> Your Name Here! </principal>
  <meeting>TEI Workshop</meeting>
  <sponsor>Records of Early English Drama project</sponsor>
  <respStmt>
    <resp>TEI P5 XML Encoding</resp>
    <name> Your Name Here! </name>
  </respStmt>
</titleStmt>
```

If you do not understand what any of these elements are for, make sure to look them up on the TEI-C website at <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/REF-ELEMENTS.html>.

1.5 Improving the `<publicationStmt>`

The `<publicationStmt>` is also fairly limited. It could contain a lot of structured information, but just has a paragraph of prose. Let's replace it:

1. Delete the entire paragraph including the starting and ending `<p>` tags.
2. Inside `<publicationStmt>` add a `<publisher>` element. In this case, 'Records of Early English Drama' is the publisher.
3. Next, inside a `<pubPlace>` element add an `<address>` element. Use the address 'Jackman Humanities Building, University of Toronto, 170 St George St Ste. 810, Toronto, ON, Canada M5R 2M8', but you should break this up. You could either use anonymous `<addrLine>` elements or more specific structural ones. So for example you could include a `<placeName>` element around 'Jackman Humanities Building' perhaps with a `@type` attribute of 'building'. You could follow this with an `<orgName>` of University of Toronto (perhaps with a `@type` of 'institution'), a `<street>` address (170 St George St, Suite 810), a `<settlement>` 'Toronto' (perhaps with a `@type` of 'city'), a `<region>` of 'Ontario' (perhaps with a `@type` of 'province'), `<country>` (United Kingdom), and a `<postCode>` (M5R 2M8). In the end this `<pubPlace>` should look something like>

```
<pubPlace>
  <address>
    <placeName type="building">Jackman Humanities Building</placeName>
    <orgName type="institution">University of Toronto</orgName>
    <street>170 St George St, Suite 810</street>
```

```

    <settlement type="city">Toronto</settlement>
    <region type="province">Ontario</region>
    <country>Canada</country>
    <postCode>M5R 2M8</postCode>
  </address>
</pubPlace>

```

4. Below the <pubPlace> add a <distributor> containing 'TEI Workshop'. This is who is responsible for distributing the electronic file.
5. After this add an <authority> element, to detail under whose authority it is distributed. In this case let's say it is under your authority, so add your name.
6. After the <authority> element add a <date> element with content of '29 January 2014'. The <date> element can have a @when attribute to take a standardised YYYY-MM-DD form of the date, add <date when="2014-01-29"> to the opening tag as well.
7. Add an ID number after this, using <idno>. This should be something like a catalogue number, or a URL at which this document will reside. In this case, make up what you think a sensible ID number would be for your edition of this record!
8. Next add an <availability> statement with a <licence> element containing a description of the licence you would want to distribute this under. We recommend you choose a Creative Commons license with a @target attribute on the licence. (See my example below.)
9. Your <publicationStmt> should now look something like:

```

<publicationStmt>
  <publisher>Records of Early English Drama</publisher>
  <pubPlace>
    <address>
      <placeName role="building">Jackman Humanities Building</placeName>
      <orgName>University of Toronto</orgName>
      <street>170 St George St, Suite 810</street>
      <settlement>Toronto</settlement>
      <region>Ontario</region>
      <country>Canada</country>
      <postCode>M5R 2M8</postCode>
    </address>
  </pubPlace>
  <distributor>TEI Workshop</distributor>
  <authority> Your Name Here! </authority>
  <date when="2014-01-29">29 January 2014</date>
  <idno>[Insert an ID number here]</idno>
  <availability>
    <licence
      target="http://creativecommons.org/licenses/by-nc-nd/3.0/"> Licensed with a Creative Commons At-
tribution Non-Commercial No-Derivatives licence.</licence>
  </availability>
</publicationStmt>

```

1.6 Improving the <sourceDesc>

Our <sourceDesc> is also fairly limited, but for now we're just going to leave it. The best thing to do is replace it with a full <msDesc> (manuscript description) which I'll be talking about next.

1.7 Other components of the <fileDesc>

There are other elements that could appear in your <fileDesc>.

1. Immediately after the closing </titleStmt> tag you could add an <editionStmt> with an <edition> containing a descriptive phrase such as 'First Electronic Edition' for the current edition of the electronic file.
2. Immediately after the closing </editionStmt> you could add an <extent> element with some measure of the size of the text (e.g. '1033 words').
3. Immediately after the closing </publicationStmt> you could add a <notesStmt> with one or more <note> elements inside it. One could contain something saying 'Transcribed for a TEI Workshop'.

1.8 Adding an <encodingDesc>

An <encodingDesc> element will give us a place to document the encoding practices in the document.

1. After the closing </fileDesc> add an <encodingDesc> element.
2. Inside the <encodingDesc> add a <projectDesc> with a <p> inside that containing a description of the REED project. Perhaps something like: "Founded in 1975, Records of Early English Drama (REED) is an international scholarly project that is establishing for the first time the broad context from which the great drama of Shakespeare and his contemporaries grew. Since then REED has worked to locate, transcribe, and edit historical surviving documentary evidence of drama, secular music, and other communal entertainment and ceremony from the Middle Ages until 1642, when the Puritans closed the London theatres."
3. Next inside the <encodingDesc> add an <editorialDecl> with a <correction> inside that with a paragraph saying something like 'Apparent errors have been provided as is and correct readings provided as a note when important.'
4. Also inside the <editorialDecl> add a <hyphenation> with a paragraph saying something like 'Hyphens have been transcribed as they appear'.
5. Look at the other options available to you inside <editorialDecl> and <encodingDesc>. You could use the information it is possible to store here to produce an introductory section on the editorial principles of the text you are editing.
6. Your <encodingDesc> should look something like:

```
<encodingDesc>
  <projectDesc>
    <p> Founded in 1975, Records of Early English Drama (REED) is an international scholarly project
that is establishing for the first time the broad context from which the great drama of Shakespeare and
```

```

his contemporaries grew. Since then REED has worked to locate, transcribe, and edit historical surviving
documentary evidence of drama, secular music, and other communal entertainment and ceremony from the
Middle Ages until 1642, when the Puritans closed the London theatres. </p>
</projectDesc>
<editorialDecl>
  <correction>
    <p>Apparent errors have been provided as is and correct readings provided as a note when impor-
tant.</p>
  </correction>
<hyphenation>
  <p>Hyphens have been transcribed as they appear.</p>
</hyphenation>
</editorialDecl>
</encodingDesc>

```

1.9 Adding a <profileDesc>

A <profileDesc> is a place to store various information concerning the text.

1. After the closing </encodingDesc> add a <profileDesc>.
2. Inside this add a <creation> with a <origPlace> (England, we can't be sure where he wrote it) and an <origDate> (8 November 1614) perhaps with a @when attribute ('1614-11-08'). I've made the text of <creation> make sense by saying it was 'Written in England in 8 November 1614'. <origDate> and <origPlace> are special versions of <date> and <placename> which are intended to hold the canonical origin date and place. Normally, they are used in the <history> section of a manuscript description.
3. Next, add a <langUsage> inside the <profileDesc> with a <language> inside ('English') with an @ident attribute with a value of 'en' for the English language code. You can also add a @usage attribute of '100' to say it is entirely in English.
4. Your <profileDesc> should now look something like:

```

<profileDesc>
  <creation> Written in <origPlace>England</origPlace> in
  <origDate when="1614-11-08">8 November 1614</origDate>.
</creation>
<langUsage>
  <language ident="en" usage="100">English</language>
</langUsage>
</profileDesc>

```

This is a fairly minimal <profileDesc> and there is a lot more we could add, but that is probably good enough for now!

1.10 Adding a <revisionDesc>

A <revisionDesc> gives you a way to record major stages in revision to the electronic file.

1. After the closing </profileDesc> add a <revisionDesc> element.

2. Add two `<change>` elements inside this. On the first one add a `@when` attribute with today's date. Inside the `<change>` add a `<persName>` containing your name, followed by the text 'improved the header'.
3. In the second `<change>` add a `@when` attribute of '2014-01', with a `<persName>` of 'James Cummings' saying that he 'created a marked up transcription based on the REED volume'.
4. It is standard practice for the most recent `<change>` to be first.
5. Your `<revisionDesc>` should now look something like:

```
<revisionDesc>
  <change when="2014-01-29">
    <persName>Your Name Here!</persName> improved the header </change>
  <change when="2014-01">
    <persName>James Cummings</persName> created a marked up transcription based on the REED
    volume </change>
</revisionDesc>
```

While this isn't a detailed `<revisionDesc>` you can understand how these would grow over time and become useful histories of the electronic file and the stages of production and maintenance that it had gone through.

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?
- Have you formatted and indented your work automatically?
- From the 'File' menu select 'Save' or click on the Save icon (looks like an old-style 3.5" disk).
- Or if you prefer use the 'File' then 'Save As' menu item to save the file using the name 'exercise03.xml' or another name of your choice.

1.12 Self-Assessment

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

- What kinds of metadata can you store in a `<titleStmt>`?
- What is a `<publicationStmt>` used for? What can it contain?
- How do you provide details of the source for the file?
- What is the difference between `<bibl>` and `<biblStruct>`?
- What is an `<encodingDesc>` for?
- What order should `<change>` elements be listed in a `<revisionDesc>`?

1.13 Further Reading

- Look up the reference pages for each of the new elements you've used.
- Read some of the chapter on the `teiHeader` <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html>.