

# Practical Exercise 3: Creating a Better Header

James Cummings

19 September 2013

## 1 Practical Exercise 3: Better Encoding Practice

### 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!

## 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.

## 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/exercise2-finished.xml' and then saving it under a new name where you put your other files.

## 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 dropdown list of possibilities).
2. Underneath the `<title>` add an `<author>` element. The content of this should be 'Wilfred Owen'.
3. Below this add an `<editor>` element with the content of 'Renée van Baalen'. (She transcribed the letter for our teaching purposes.) How does one type in 'é' in oXygen? **Hint:** the 'Edit' menu contains a 'Insert from Character Map' entry.

## 5 IMPROVING THE PUBLICATIONSTMT

---

4. After this add a `<principal>` element to record the person primarily responsible for the project. In this case, use your own name!
5. Below this add a `<meeting>` element with the content of 'Pre-JADH 2013 TEI Workshop'.
6. After that add `<respStmt>` with a `<resp>` inside it saying 'Improved encoding' and a `<name>` with your name.
7. Your `<titleStmt>` should now look something like:

```
<titleStmt>
  <title>Letter to Leslie Gunston</title>
  <author>Wilfred Owen</author>
  <editor>Renée van Baalen</editor>
  <principal>[your name here]</principal>
  <meeting>Pre-JADH 2013 TEI Workshop</meeting>
  <respStmt>
    <resp>Improved 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/ja/html/REF-ELEMENTS.html> given above.

## 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, 'TEI @ Oxford' is the publisher.
3. Below the `<publisher>` add a `<distributor>` containing 'Pre-JADH 2013 TEI Workshop'.
4. After this add an `<authority>` element, to detail under whose authority it is published. In this case let's say it is under your authority, so add your name.
5. Next, inside a `<pubPlace>` element, which itself contains an `<address>` element. Use the current address, or you can use the IT Services, University of Oxford address. So for example include an `<orgName>` (IT Services, University of Oxford), a `<street>` address (13 Banbury Road), a `<settlement>` (Oxford), a `<postCode>` (OX2 6NN), and a `<country>` (United Kingdom).
6. After the `<pubPlace>` element, but still inside the `<publicationStmt>` add a `<date>` element with content of '19 September 2013'. The `<date>` element can have a `@when` attribute to take a standardised YYYY-MM-DD form of the date, add `<date when="2013-09-19">` 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 letter!

- 
- 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.)
  - Your `<publicationStmt>` should now look something like:

```
<publicationStmt>
  <publisher>TEI @ Oxford</publisher>
  <distributor>Pre-JADH 2013 TEI Workshop</distributor>
  <authority>[Your name here]</authority>
  <pubPlace>
    <address>
      <orgName>IT Services, University of Oxford</orgName>
      <street>13 Banbury Road</street>
      <settlement>Oxford</settlement>
      <postCode>OX2 6NN</postCode>
      <country>United Kingdom</country>
    </address>
  </pubPlace>
  <date when="2013-09-19">19 September 2013</date>
  <idno>[Insert an ID number here]</idno>
  <availability>
    <licence
      target="http://creativecommons.org/licenses/by/3.0/">Licensed with a Creative Commons Attribution licence.</licence>
  </availability>
</publicationStmt>
```

## 6 Improving the `<sourceDesc>`

Our `<sourceDesc>` is also fairly limited.

- Delete the entire paragraph that is currently in the `<sourceDesc>` and replace it with a `<biblStruct>`.
- The `<biblStruct>` should have an `<analytic>` with a `<title>` (Letter to Leslie Gunston), and `<author>` (Wilfred Owen).
- The `<biblStruct>` should also have a `<monogr>` for the collection containing:
  - `<title>` (The Wilfred Owen Collection).
  - A `<ref>` (First World War Poetry Digital Archive) containing a `@target` attribute pointing to 'http://www.oucs.ox.ac.uk/ww1lit/collections/document/5243'.
  - An `<imprint>` element containing a `<publisher>` (The First World War Poetry Digital Archive), a `<pubPlace>` (Oxford), and a `<biblScope>` (Two pages) with a `@type` attribute of 'pp', and a `@n` attribute of '2'.
  - Outside the `<monogr>` but inside the `<biblStruct>` add a `<relatedItem>` with a `<bibl>` containing 'The source of this digital resource is a copy from the Harry Ransom Centre.' You could also wrap 'Harry Ransom Centre' in a `<distributor>` element. This is an example of a much less structured bibliographic citation inside a structured one.
  - Your `<sourceDesc>` should now look something like:

```
<sourceDesc>
  <biblStruct>
    <analytic>
      <title>Letter to Leslie Gunston</title>
      <author>Wilfred Owen</author>
    </analytic>
    <monogr>
      <title>The Wilfred Owen Collection</title>
      <ref
        target="http://www.oucs.ox.ac.uk/wllit/collections/document/5243/4769">
First WorldWar Poetry Digital Archive</ref>
      <imprint>
        <publisher>The First World War Poetry Digital Archive</publisher>
        <pubPlace>Oxford</pubPlace>
        <biblScope type="pp" n="2">Two pages</biblScope>
      </imprint>
    </monogr>
    <relatedItem>
      <bibl>The source of this digital resource is a copy from the
      <distributor>Harry Ransom Centre</distributor>.</bibl>
    </relatedItem>
  </biblStruct>
</sourceDesc>
```

## 7 An Even Better <sourceDesc>

That <sourceDesc> is pretty good, but let's make it better by adding a <listPerson>, <listPlace>, <listOrg>, and <msDesc>. But wait, that would take a long time for you to type in, so we've done it for you!

1. Place your cursor after the closing </biblStruct> tag but before the closing </sourceDesc> tag.
2. From the **Document** menu on the menu bar, and select **File**, and **Insert File**. **Note that this is from the 'Document' menu.** The file you want to insert is called 'exercise3-extras.xml'.
3. Inserting this file at this point will add a <listPerson>, <listPlace>, <listOrg>, and <msDesc> elements. Read through each of these to see all the information you have added!
4. Your file should still be valid and have a happy green square. If it is an angry red square, find the problem and solve it!

## 8 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 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. '260 words') or in this case '2 pages'.
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'.

---

## 9 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 it saying something like 'The TEI@Oxford project created these teaching materials.'
3. Next inside the <encodingDesc> add an <editorialDecl> with a <correction> inside that with a paragraph saying something like 'Apparent errors have been marked as <sic> but correct readings not provided'. Mark up <sic> as an element by using <gi> (generic identifier).
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>.
6. Your <encodingDesc> should look something like:

```
<encodingDesc>
  <projectDesc>
    <p>The TEI@Oxford project created teaching materials.</p>
  </projectDesc>
  <editorialDecl>
    <correction>
      <p>Apparent errors have been marked as <gi>sic</gi> but correct readings not
provided.</p>
    </correction>
    <hyphenation>
      <p>Hyphens have been transcribed as they appear.</p>
    </hyphenation>
  </editorialDecl>
</encodingDesc>
```

## 10 Adding a <profileDesc>

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

1. After the closing </encodingDesc> add a <profileDesc>.
2. Inside this add a <creation> with a <placeName> (Craiglockhart) and a <date> (July 1917) perhaps with a @when attribute ('1917-07').
3. In the <profileDesc> next add a <handNotes> with a <handNote> inside it saying something like 'Written in Wilfred Owen's hand'.
4. 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.
5. Next add a <textClass> with a <classCode> with content of '826' and a @scheme attribute of "http://www.oclc.org/dewey/resources/summaries/default.htm". This is the Dewey classification code for 'English Letters'. (I'm not recommending this particular classification scheme in general: The US Library of Congress has some much better ones!)

6. Your `<profileDesc>` should now look something like:

```
<profileDesc>
  <creation>
    <placeName>Craiglockhart</placeName>
    <date when="1917-07">July 1917</date>
  </creation>
  <handNotes>
    <handNote>Written in Wilfred Owen's hand</handNote>
  </handNotes>
  <langUsage>
    <language ident="en">English</language>
  </langUsage>
  <textClass>
    <classCode
      scheme="http://www.oclc.org/dewey/resources/summaries/default.htm">826
    </classCode>
  </textClass>
</profileDesc>
```

## 11 Adding a `<revisionDesc>`

A `<revisionDesc>` gives you a way to record major stages in revision to a document.

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 '2013-02', with a `<persName>` of 'Renée van Baalen' saying that she 'transcribed the Letter to Leslie Gunston document'. You may also wish to mark 'Letter to Leslie Gunston' as a `<title>`.
4. It is standard practice for the most recent `<change>` to be first.
5. Your `<revisionDesc>` should now look something like:

```
<revisionDesc>
  <change when="2013-07-03">
    <persName>[Your name here]</persName> improved the header.</change>
  <change when="2013-02">
    <persName>Ludwig Wittgenstein</persName> transcribed the <title>Letter to
      Leslie Gunston</title> document. </change>
</revisionDesc>
```

## 12 Marking Up Names

If you have time then it might be good to use some of that metadata we imported earlier. We added some additional markup to our header which included a `<listPerson>`, `<listPlace>`, a `<listOrg>`. In addition to the general purpose `<name>` element which can take a `@type` attribute for classification, there are three types of names specifically catered for in the TEI. These are: organizational names (`<orgName>`), personal names (`<persName>`), and place names (`<placeName>`). Occasionally you might want to mark something like 'she', 'I', or 'you' which are not strictly names but references an understood named entity. To do this we use a reference string or `<rs>` element.

1. In the first `<salute>` mark up 'L.' as a `<persName>`.
2. In the first paragraph encode 'Field Club' as an `<orgName>`, and 'Berlitz, Edin.' as an `<orgName>` with a `<placeName>` inside it ('Edin.').
3. In the second paragraph mark up Antaeus, Heracles, Mother Earth, and 'old Herk.' as `<persName>` elements.
4. In the verse encode 'Earth' as a `<persName>` (because it is used anthropomorphically here).
5. In the final division mark up 'Locke's' and 'Swinburne' as a `<persName>` elements.
6. Inside the `<signed>` element mark up 'WEO' as a `<persName>`.
7. There are more names we could mark up, such as the use of the names Leslie Gunston and Wilfred Owen throughout the header, but that is optional.

### 13 Linking Names and Metadata

Having marked all these names, and imported stored metadata about them, it seems a shame not to link the names to this metadata. So let's do that!

1. Go to the `<persName>` you put in the first `<salute>` around 'L.'. Put the cursor immediately after the final 'e' in the opening `<persName>` tag and press space. You should get a drop-down list of attributes, select 'ref', when you do so you should get a drop-down list of `@xml:id` values present in the entire document. Scroll down and select '#LG'. This element now points to the `<person>` element with the `@xml:id` value of 'LG'!
2. This `<salute>` now should look like:

```
<salute>Dear <persName ref="#LG">L.</persName>
</salute>
```

3. The value of `@ref` is a URI, which includes URLs, and in this case a 'fragmentary URL'. It starts with a '#' to let us know it is in the same document. You could also have stored the `<listPerson>` in a separate document, in which case we would put something like 'people.xml#LG', or stored this online somewhere 'http://www.example.com/people.xml#LG'. While it is best if this points to a TEI `<person>` element, it can in fact point to anything which documents the name such as a wikipedia article. (One reason it is better for this to point to a `<person>` element is that inside that you could indeed point to more than one external source of information.)
4. For each `<persName>`, `<placeName>`, and `<orgName>` (for which there is a `<person>`, `<place>` or `<org>` element) go through and add a `@ref` attribute pointing to the correct `@xml:id`.
5. The benefit of an encoder doing all this work is that for each instance of a name someone processing the text could find a standardised form of it, and other metadata, when generating other outputs. (e.g. for help in searching, or displaying this information)

### 14 Referencing Strings

As explained earlier the `<rs>` element can be used to mark things which aren't strictly names in themselves but are understood to reference named entities. For example 'I' and 'you' in this file refer to Wilfred Owen and Leslie Gunston respectively.

- Depending on how much time you have left, mark as many of the instances of 'I' and 'you' as `<rs>` pointing to the appropriate `<person>` element in each case.

### 15 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.

### 16 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>`?
- How does a `<persName>` element in the text point to a `<person>` element in the header?

### 17 More Suggested reading

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