

Getting started with oXygen

Lou Burnard Consulting

2014-06-21

1 Introducing oXygen

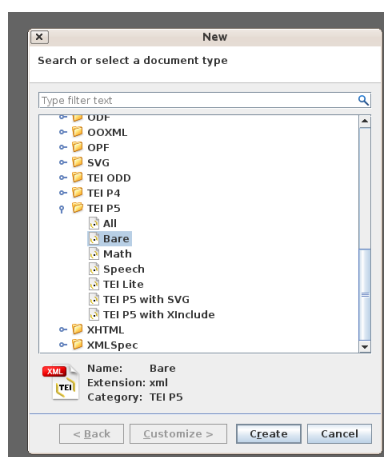
In this first exercise we will use oXygen to :

- create a new XML document
- gradually add markup to the document
- carry out continuous validation of the document
- display the document without tags
- automatically convert a word processor document to TEI

2 A born-digital document

TEI can be used to encode any kind of document, not just to represent previously existing ancient manuscripts or inscriptions. It can also be used for the creation of new documents, starting from scratch....

- Start up oXygen.
- Click on the New icon, top left (or select New from the File menu, or type CTRL-N) to open the New dialogue
- Choose Framework Templates, then TEI P5, then Bare to select a minimal schema.

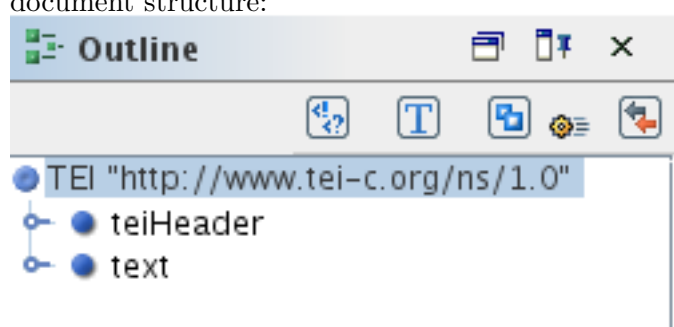


- Click the Create button below. oXygen displays a minimal TEI document for you to complete.

```
Untitled1.xml x
1 <?xml version="1.0" encoding="UTF-8"?>
2 <?oxygen RNGSchema="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_bare.rng" type="xml"?>
3 <TEI xmlns="http://www.tei-c.org/ns/1.0">
4   <teiHeader>
5     <fileDesc>
6       <titleStmnt>
7         <title>Title</title>
8       </titleStmnt>
9       <publicationStmnt>
10        <p>Publication Information</p>
11      </publicationStmnt>
12     <sourceDesc>
13       <p>Information about the source</p>
14     </sourceDesc>
15   </fileDesc>
16 </teiHeader>
17 <text>
18   <body>
19     <p>Some text here</p>
20   </body>
21 </text>
22 </TEI>
```

The TEI bare schema provides only a few tags but still enough for you to make a complete TEI document.

By default, the oxyGen workspace displays several small windows around a big blank space where we will work on our document. On the left, for example, the Outline window shows the document structure:



A TEI document is divided into two parts. The upper part contains metadata, wrapped in the `<teiHeader>` tag. The lower part contains the text, wrapped in the `<text>` tag. Click on the key icon next to each part to expand or contract it, and you will see its components.

You can change the way the document is displayed in the middle window by clicking on the *Text*, *Grid*, or *Author* buttons at the bottom of the screen. We're not frightened of tags, so we will mostly be working in *Text* mode.

First, we will provide three pieces of metadata essential for any new document: its title, its distribution, and its origin. This template provides just bits of text which you need to modify and then complete ..

- Replace the word 'Title' (between the tags `<title>` and `</title>`) by a title for your document. It's up to you what the title is! Something like "Getting started with oXygen" would be fine.
- Replace the two other bits of text by something more appropriate. For example, for publication, you might put 'Unpublished draft' or 'For publication on my blog' ; for `<sourceDesc>`, we suggest something like 'An original work prepared as part of a TEI training course'.

oXygen does its best to keep your document in a valid state. What happens if you make it invalid?

- Delete a few letters within any of the tags : for example, change `<publicationStmnt>` to `<publicat>`.
- oXygen does what it can. Firstly, it changes the corresponding closing tag to match (so it becomes `</publicat>`): consequently, the document remains well-formed. But it is no

longer valid: tags that are not defined in the schema in use are underlined in red; there is an error message at the foot of the screen; and in the right hand sash bar you see several red marks.

- Click on the red marks to see how one validity error can lead to another.
- Click on CTRL-Z (or select Undo from the Edit menu) to return your document to a state of grace ... note how the little red square at top right becomes green when all errors have been corrected.

oXygen only allows you to add tags permitted by the schema you have selected for your document.

- Put the cursor at the end of the 7th line (or the beginning of the 8th), that is, between `</title>` and `</titleSmt>`;
- type `<` and pause a moment ;
- the green square becomes red because you have not yet entered the name of the element you want to insert
- oXygen offers you a drop down list of element names available at this point in the document. Specifically, you can insert an `<author>` element or another `<title>`, or you could close the `<titleSmt>`, or insert an XML comment.
- Press ENTER to accept the first suggestion. oXygen inserts the tags `<author>` and `</author>` into your document.
- The cursor remains inside the opening tag in case you want to specify some attributes for this new element. Since you don't, move the cursor one character to the right using the arrow key, and type your name, since you are the author of this document.

3 Adding content to the document

Now we are ready to begin adding content to your document. It's up to you what content to add and how to tag it!

Firstly, what sort of structural components will your document need? A document containing just one paragraph seems a bit unlikely! Using the TEI bare schema, you can add any of the following:

- titles ... marked up with the `<head>` tag
- chapters ... marked up with the `<div>` tag
- lists ... marked up with the `<list>` tag
- and of course more paragraphs ... marked up with the `<p>` tag

Let's go!

Obviously, you have to create the content yourself. But (as you've already seen) oXygen can usually insert the tags for you. It can suggest tags relevant at a particular point in the document, and it can also be used to add tags to a selected string of text, or to split up existing elements.

- Use the mouse to select just the content of the `<body>`, element, i.e. the string `<p>Some text here</p>` (including the start and end tags).
- Type CTRL-E (or choose XML Refactoring – Surround with tags from the Document menu.

- oXygen offers a drop down list of tags which can legally surround this content at this point in the document. Choose `<div>` and click on **Accept**
- This `<div>` element will contain the first chapter of your document. You might like to add a title for this first chapter. Insert the title in the same way as you inserted an author, but this time select the tag `<head>`. Make sure you put the title in the right place : inside the `<div>`, and before its first `<p>`.
- It would be useful to number the chapters of your document explicitly. The TEI proposes an attribute `@n` for this purpose, which is available on any element, including `<div>`.
- Put the cursor *inside* the `<div>` start-tag, just before the `>`, and type a space. oXygen displays a drop-down menu of the available attributes for this element.
- When `@n` is selected, type RETURN to insert it into the tag. oXygen puts the cursor between the quote marks which it has added so that you can insert a value for the attribute.
- Type (for example) `1`, and your document is valid once more.

Your document should now look something like this :

```
<text>
  <body>
    <div n="1">
      <head>Chapter One</head>
      <p>Some text here....</p>
    </div>
  </body>
</text>
```

Now you need to add some text to this paragraph: preferably something a bit more interesting than ‘some text here’ ! Describe your first reactions to oXygen; share your ideas about the political situation; describe your troubled childhood... it’s up to you.

No doubt you will pretty soon want to start a new paragraph. How do you close this one and start a new one? Just typing blank lines won’t work: in XML such things must be explicitly marked-up. You could of course just type the tags `</p><p>` but there is another and a better way.

4 Element splitting

Suppose you now have a few hundred words in your single paragraph, which you’d like to split up into several smaller paragraphs.


- Put the cursor at the place where you want to close the first paragraph.
- Type ALT + SHIFT + D or choose XML Refactoring – Split Element from the Document menu
- Repeat this for the next paragraph

You can also do the same thing to divide up your single chapter (`<div>`). Take care that the cursor is in the right place when splitting the division: it must be *between* two `<p>` elements (after a `</p>`), not within a `<p>` element.

Don’t forget to add a list (tagged with a `<list>`) to one of your chapters. You will see that this can contain an optional `<head>`, followed by a sequence of `<item>` elements.

5 Displaying tagged text

Is the amount of tagging in your document beginning to worry you?

You can tidy up the way your tagged text is displayed by clicking on the Indent button (this one: ) on the toolbar, or typing CTRL-SHIFT-P, or selecting Source->Format and Indent from the Document menu. But maybe you just want to hide the tags?

- Click on the Author button at the bottom of the display. The display mode changes and a new TEI P5 menu appears.
- Choose Tags display mode from this menu, and inspect the effect of the various options this provides.
- Your document is also now styled: for example, headings are distinguished visually from text, and paragraphs are left justified.
- Several useful short cuts are available in Author mode. For example, you can divide an element simply by typing RETURN twice. You can also insert a new paragraph or section element simply by choosing an option from the TEI P5 menu.

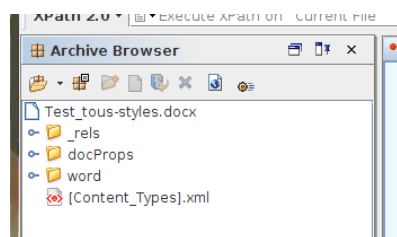
The way your document is displayed is determined by a *style sheet*. Later on we will see the effect of using different stylesheets.

6 Importing a wordprocessor file

Most digital documents in the real world seem to be entirely innocent of XML tagging. For example, a document prepared with tools like Microsoft Word or Open Office seems to contain nothing but nicely formatted text. Can we convert such documents to TEI XML in oXygen? Of course we can!

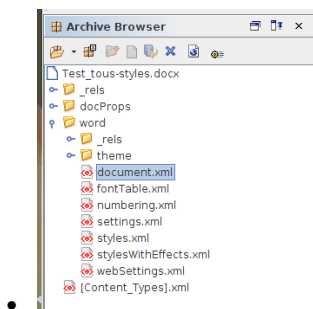
A Word docx file is actually a compressed archive containing several other files, including XML files which can be manipulated and converted in the same way as any other XML file. oXygen knows this, and comes with predefined XSLT stylesheets for converting from docx to TEI P5. The conversion can also take into account any Word styles found, which makes it possible to customize the way your Word file, is converted, in order to produce more intelligent markup than would otherwise be possible.


- In oXygen, use the usual Open dialog to select the docx file `convertMe.docx` which is available in your Work folder.
- A window opens to the left of the main screen headed "Archive browser". It shows the file structure of the docx archive.

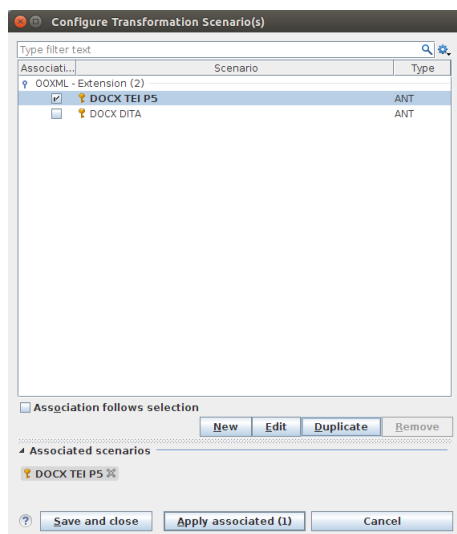



- Click the blue key next to the folder called word to see the contents of this folder

6 IMPORTING A WORDPROCESSOR FILE



- Select the file called `document.xml` and double click to open it (may take a few moments if the file is large)
- As you see, this is an XML document, with lots of XML tags, using many different namespaces, most of them defined by Microsoft. Let's convert it!
- With the `document.xml` file open in your main editing window, select **Transformation -> Configure Transformation Scenarios** from the Document menu. Or type **CTRL-SHIFT-C**. Or click the little spanner icon (this one : )
- oXygen shows you the transformation options available for this kind of document. This should include one called **DOCX TEI P5**, which is the default TEI conversion.



- To see what it does, check the little box next to it, and press the **Apply associated** button at the bottom. A new window opens at the bottom of the screen in which oXygen displays messages relating to the conversion process ; you may safely ignore these.
- In the main editing window, you should now see a TEI-conformant document. (If it is displayed as a single line on the screen, choose **Source -> Format and Indent** from the Document menu, or type **CTRL-SHIFT-P** or click the **Indent** button (this one: )). Scroll down to see how the Word styles have been converted.
- Note that many of the paragraphs have a `rend` attribute. The values of this attribute correspond with the Word style used in the original docx file. An XSLT stylesheet will use these values to make a more intelligent conversion, as we will see.