

Exercise 2.

Using the oXygen XML Editor: Creating an XML document

1. Learning Outcomes

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

- recognize an XML declaration
- mark up basic features of a poem
- create a well-formed XML document

2. Summary

This exercise will walk you through creating an XML document in the oXygen editor and introduce a variety of ways to mark this document up. You will first start a new document, then insert some unmarked up text into the editor, and then mark up the stanzas or line-groups (lg) and lines (l). You will learn to check that your document is well-formed or not.

Note: read each step before you do it

--What you'll need for this exercise:

- Download and set up the Oxygen editor using your license key.
- Find and open the file `poems_text→karyotakis_soneto.txt` in your Materials folder in Unit 2.

3. Starting A New XML File

Let's start a new XML file by following the following steps:

- 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 you may wish to close any additional popup windows, and optionally the 'views' on the left and right-hand side. This will give you more space.
- From the 'File' menu select 'New' and double click on 'New Document' select 'XML Document' This should open up a blank document with an XML Declaration added.
 - An XML Declaration looks like:

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

The XML declaration at the top of the file tells anything processing your XML file, including the editor, that this is an XML file and what version of XML you are using through the @version attribute. The XML declaration



needs no closing tag as it takes the form of a special processing-instruction that starts and ends with an angle-bracket and a question mark.

4. Creating a Division

Let's create a division of a text using the `<div>` element. This is a generic division or section element.

- On the line below the XML declaration type: `<div>`

5. Inserting some text

- Notice what happens when you type the final `'>'`. oXygen tries to help you and inserts the closing `</div>` tag. This is because it knows the rules of XML, and knows that if you type an opening `<div>` you are required to have a closing `</div>` sooner or later.
- We haven't said what type of division this is, which means it is just a generic division. Let's categorise it as 'verse' by adding a `@type` attribute. (Note: although I say `'@type'` the attribute's name is `'type'`. The at-sign, '@' is used as a convention to denote an attribute in prose.) Move the cursor back until just after the letter 'v' in the opening tag. Press space, and then type: `type="` and notice what happens when you type the quotation mark. oXygen is again trying to help you by putting the closing quotation mark, because it knows that attribute values must always be quoted.
- In between the quotation marks type `verse` to categorise our division as being in verse.
- Move back until you are directly in between the opening `<div>` and closing `</div>`. Press 'enter' a couple times to give yourself some space inside the element.
- We are going to put one or more sonnets by Κώστας Καρυωτάκης inside the `<div>`. Add a `<head>` element inside the `<div>` with the heading `'Ποιήματα'`. Your file should now look like:

```
<div type="verse">
<head>Ποιήματα</head>
<!-- We will add the sonnet here -->
</div>
```

We are going to use Καρυωτάκης Σονέτο as an example for this exercise. This uses:

- `karyotakis_soneto.txt`
- Make sure your cursor is in-between the opening `<div>` and the closing `</div>` underneath the `<head>` element. Open `karyotakis_soneto.txt` with any text editor and highlight the entire poem (click, then control-a) and copy (control-c) and then in the oxygen window paste it (control-v).
- Your division should now look like:



```
<div type="verse">
<head>Ποιήματα</head>
Σονέτο
Τα φύλλα κιτρινίσανε και, πέφτοντας στο χώμα,
ξεχύνονται σ' αφάνταστο, σ' απότρελο χορό.
Τα λουλουδάκια κλείσανε το θεϊκό τους στόμα
και γέρνουν άθελα στη γη, κοπάδι θλιβερό.

Ο ήλιος που άλλοτε, που χθες, προχθές ακόμα,
εσκόρπιζε το γέλιο του, ψηλά, το λαμπερό,
σκυθρώπασε, και έπνιξε τ' ολόχρυσό του χρώμα
στα μολυβένια σύννεφα που πέρα εκεί θωρώ.

Η θάλασσα εφούσκωσε κι εθέριεψε το κύμα,
στο πεζοδρόμι η βροχή χτυπάει ρυθμικά
και του διαβάτη βιαστικό ακούγεται το βήμα.

Απ' τη βροχή ετρόμαξαν τ' αθώα χελιδόνια
και σαν πετάνε φαίνεται πως λένε μυστικά:
«Εμπρός, εμπρός, να φύγουμε, θ' αρχίσουνε τα χιόνια.
</div>
```

- Now we are going to use the line-group element `<lg>` not only to group the sonnet as a whole but the individual quatrains. Highlight the text of the poem including the 'Σονέτο.' but no other elements. Either press control-e as a shortcut key, or right-click and under 'Refactoring' select 'Surround with Tags'. A box should pop up. Type `lg` into it and press enter. Notice how oXygen helps you again by putting the opening tag before what you had highlighted and the close tag afterwards. This should surround your entire sonnet with the `<lg>` element. Add a `@type` attribute to this `<lg>` with the value of 'sonnet'.
- Use the same surround-with-tag shortcut to highlight and surround '18.' with a `<head>` element
- It should now look like:

```
<div type="verse">
<head>Ποιήματα</head>
<lg type="sonnet">
  <head>Σονέτο</head>
  Τα φύλλα κιτρινίσανε και, πέφτοντας στο χώμα,
  ξεχύνονται σ' αφάνταστο, σ' απότρελο χορό.
  Τα λουλουδάκια κλείσανε το θεϊκό τους στόμα
  και γέρνουν άθελα στη γη, κοπάδι θλιβερό.
  <!-- rest of poem -->
</lg>
```



</div>

6. Encoding the lines (using surround-with-tag and split element)

- In addition to the very useful “surround-tag” shortcut, there is another one called “split-element” which splits the current element. It inserts an end tag and follows it immediately with a start tag of the same kind where the cursor is. This can be used to mark up lots of elements very quickly. In our case we want to wrap each of the lines in the metrical line element `<l>`.
- Highlight from the first phrase ‘Τα φύλλα’ all the way down to the last word ‘χρόνια’ then use the control-e surround-with-tag shortcut to wrap the whole sonnet in a single `<l>` element. This means that our encoding is claiming that the whole sonnet is one metrical line. This is clearly false.
- Place the cursor at the start of the second line, just before ‘ξεχύνονται’ and press alt-shift-D to split the line element that you are in. This will put `</l><l>` at your current position.

7. Format and Indent

Our poem is marked up, but some of the markup might be a bit messy because of the split-element shortcut we used.

- Make sure that your file is ‘well-formed’. You’ll be able to tell it is well-formed because oXygen will have a happy green square in the upper right-hand corner. If it is an angry red square, you better find the problem (where a red bar on the right-hand side is) and correct the mistake! (Note: if you are red/green colour blind, ask me and I’ll show you how to change this to colours you can differentiate between!)
- Now let’s format and indent our file. This tidies up some of the white-space and indents elements based on their place in the hierarchy. Either select the ‘Format and Indent’ icon from the toolbar (it looks like some indented lines), or go to the menus: ‘Document’ -> ‘Source’ -> ‘Format and Indent’.
- Formatting and indenting your markup is not necessary, it could all be on one big long line, but it makes it much easier for people to read.
- You can delete the extra space at the end of each line, but it is optional.

8. Adding the stanzas

Now we have each `<l>` marked up we want to mark up the stanzas. A shakespearean sonnet is made of three quatrains and a final couplet.

- Around each set of four lines and the last two wrap an `<lg>` element. Use the surround-with-tag shortcut again.
- Give each `<lg>` a `@type` attribute with the values of ‘quatrain’ or ‘couplet’ (for the last one).



- Your poem should now look like:

```
<div type="verse">
  <head>Ποιήματα</head>
  <lg type="sonnet">
    <head>Σονέτο</head>
    <lg type="quatrain">
      <l>Τα φύλλα κιτρινίσανε και, πέφτοντας στο
χώμα,</l>
      <l>ξεχύνονται σ' αφάνταστο, σ' απότρελο
χορό.</l>
      <l>Τα λουλουδάκια κλείσανε το θεϊκό τους
στόμα</l>
      <l>και γέρνουν άθελα στη γη, κοπάδι θλιβερό.</l>
    </lg>
    <lg type="quatrain">
      <l>Ο ήλιος που άλλοτε, που χθες, προχθές ακόμα,
</l>
      <l>εσκόρπιζε το γέλιο του, ψηλά, το λαμπερό, </l>
      <l>σκυθρώπασε, και έπνιξε τ' ολόχρυσό του
χρώμα</l>
      <l>στα μολυβένια σύννεφα που πέρα εκεί θωρώ. </l>
    </lg>
    <lg type="tercet">
      <l>Η θάλασσα εφούσκωσε κι εθέριεψε το κύμα, </l>
      <l>στο πεζοδρόμι η βροχή χτυπάει ρυθμικά</l>
      <l>και του διαβάτη βιαστικό ακούγεται το
βήμα.</l>
    </lg>
    <lg type="tercet">
      <l>Απ' τη βροχή ετρόμαξαν τ' αθάνα χελιδόνια</l>
      <l>και σαν πετάνε φαίνεται πως λένε μυστικά: </l>
      <l>«Εμπρός, εμπρός, να φύγουμε, θ' αρχίσουνε τα</l>
χιόνια. </l>
    </lg>
  </lg>
</div>
```

Don't worry about the closing tag being on the wrong line for now. After you've done the first couple lines it should look like:

```
<div type="verse">
  <head>Ποιήματα</head>
  <lg type="sonnet">
    <head>18.</head>
    <l> Τα φύλλα κιτρινίσανε και, πέφτοντας στο χώμα,
</l>
```



```
<l>ξεχύνονται σ' αφάνταστο, σ' απότρελο χορό.:  
</l>  
<l>Τα λουλουδάκια κλείσανε το θεικό τους στόμα  
και γέρνουν άθελα στη γη, κοπάδι θλιβερό.<!-- rest of  
poem -->  
</l>  
</lg>  
</div>
```

Do the rest of the poem.

9. Adding more markup

We've marked up the structure of this sonnet, but not much else. Let's add more markup.

- The `<rhyme>` element has a `@label` attribute and is used to mark up the rhyme word of each line and what letter it is in the rhyme scheme.
- Using the surround-with-tag shortcut highlight the last word in a line and wrap the `<rhyme>` element around it. Then add a `@label` attribute for the part it plays in the rhyme scheme. For example the first quatrain should look like:

```
<lg type="quatrain">  
  <l>Τα φύλλα κιτρινίσανε και, πέφτοντας στο <rhyme  
label="a">χώρα</rhyme>,</l>  
  <l>ξεχύνονται σ' αφάνταστο, σ' απότρελο <rhyme  
label="b">χορό</rhyme>.</l>  
  <l>Τα λουλουδάκια κλείσανε το θεικό τους <rhyme  
label="a">στόμα</rhyme></l>  
  <l>και γέρνουν άθελα στη γη, κοπάδι <rhyme  
label="b">θλιβερό</rhyme>.</l>  
</lg>
```

Note how we've not included the punctuation in the `<rhyme>` element. (This is just an editorial decision.)

- There are other features that we could add. For example the last line appears to contain a direct quote - You could surround it with `<q>` to indicate that. Note that the `<q>` is wholly contained within the line.

```
<l><q>«Εμπρός, εμπρός, να φύγουμε, θ' αρχίσουνε τα  
<rhyme label="c">χιόνια</rhyme>.</q></l>
```

We could choose to include the quotation mark, or remove it and indicate what to use with a `@rend` attribute.

```
<q rend="initial-double-angle-quote">Εμπρός, εμπρός,  
να φύγουμε ...</q>
```



10 The result

In the end your file should look something like:

```
<div type="sonnet">
  <head>Ποιήματα</head>
  <lg type="quatrain">
    <l>Τα φύλλα κιτρινίσανε και, πέφτοντας στο <rhyme
label="a">χώμα</rhyme>,</l>
    <l>ξεχύνονται σ' αφάνταστο, σ' απότρελο <rhyme
label="b">χορό</rhyme>.</l>
    <l>Τα λουλουδάκια κλείσανε το θεικό τους <rhyme
label="a">στόμα</rhyme></l>
    <l>και γέρνουν άθελα στη γη, κοπάδι <rhyme
label="b">θλιβερό</rhyme>.</l>
  </lg>

  <lg type="quatrain">
    <l>Ο ήλιος που άλλοτε, που χθες, προχθές <rhyme
label="a">ακόμα</rhyme>,</l>
    <l>εσκόρπιζε το γέλιο του, ψηλά, το
<rhyme>λαμπερό</rhyme>,</l>
    <l>σκυθρώπασε, και έπνιξε τ' ολόχρυσό του <rhyme
label="a">χρώμα</rhyme></l>
    <l>στα μολυβένια σύννεφα που πέρα εκεί
<rhyme>θωρώ</rhyme>.</l>
  </lg>

  <lg type="tercet">
    <l>Η θάλασσα εφούσκωσε κι εθέριεψε το <rhyme
label="c">κύμα</rhyme>,</l>
    <l>στο πεζοδρόμι η βροχή χτυπάει <rhyme
label="d">ρυθμικά</rhyme></l>
    <l>και του διαβάτη βιαστικό ακούγεται το <rhyme
label="c">βήμα</rhyme>.</l>
  </lg>

  <lg type="tercet">
    <l>Απ' τη βροχή ετρόμαξαν τ' αθώα <rhyme
label="c">χελιδόνια</rhyme></l>
    <l>και σαν πετάνε φαίνεται πως λένε <rhyme
label="d">μυστικά</rhyme>:</l>
    <l><q>«Εμπρός, εμπρός, να φύγουμε, θ' αρχίσουνε τα
<rhyme label="c">χιόνια</rhyme>.</q></l>
  </lg>
```



</div>

If you didn't manage to finish the exercise in time, don't worry, there is a pre-completed version [here](https://github.com/DiXiT-eu/Digital-Scholarly-EditionsGR-workshop/blob/master/receipts/poem/karyotakis-soneto-Unit2.xml) [<https://github.com/DiXiT-eu/Digital-Scholarly-EditionsGR-workshop/blob/master/receipts/poem/karyotakis-soneto-Unit2.xml>].

11 Checking well-formedness?

We are not using a particular schema at the moment, so we can't check validity, but we can check whether it is well-formed or not. That is, if our file follows the rules of XML. If you have a happy green square in the upper right-hand corner of your oxygen scroll-bar then you are well-formed! If not, and you have an angry red square, you must look at the error message at the bottom, at what is underlined in red, and try to solve the problem. (We can help you!) If you are well-formed then delete the first part of any closing tag and see the error produced. Once you are comfortable you know how to tell if your document is well-formed or not, ensure that it is and that you have a happy green square.

12. 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?
- 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 'exercise01.xml' or another name of your choice.

13. Self-Assessment

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

- How do you start a new XML document in oXygen?
- What is an XML declaration?
- What is a well-formed document?
- How do I 'Surround with tag' and repeat that action quickly?
- Why might using the 'Split element' approach be useful?
- What is the function of each element and attribute in your current file? Why have we encoded it this way?
- What is the advantage of formatting and indenting your markup?



14. Next?

Your XML file may be well-formed but it is not yet valid because it doesn't validate against a particular schema (such as those which are customisations of the TEI). Next we will have a short introduction to the structure of TEI documents and some of the most frequently-used elements. 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.html>. In particular you might want to look at the Elements appendix for reference pages for individual elements. Consider looking up all the elements you've used in this file to see how they are defined.

