Exercise 4. Named entities

1. Handy oXygen tricks

Enclose the selection with a tag:

- highlight the characters which you want to tag
- type CTRL+E to display the menu of available tags
- type pe (for <persName>) or pl (for <placeName>) and then press RETURN

Split the long chunk of text into a sequence of elements of the same kind:

- highlight the long chunk and wrap it with your desired element (say <p>)
- move the cursor to a place within that chunk that is the start of next element of the same kind
- type ALT+SHIFT+D to split elements (it inserts closing and starting tags at the cursor position)
- repeat as many times as needed

If you forget the key combination to perform the trick, try the right-click and see what's in the Refactoring section

It can help a lot to format and indent your work automatically via CTRL+SHIFT+P or clicking the Format and indent icon.

2. Find your files

In this exercise, we will use the file created in the previous exercise. The aim is to encode:

- persons
- places
- the metadata about these entities
- dates

3. Goal: Marking Names

Reading the letter, you will note that people and places are mentioned in it. Now go through the document and any time you come across the name of a person, place, or organisation mark them up using <persName>, <placeName>, and <orgName> respectively.

Remember the oXygen trick of enclosing the selection with a tag with CTRL+E combination.

4. Adding Metadata for Persons

In our <teiHeader> we are able to store metadata about the people, places, and organisations mentioned in the text. Where exactly? In the <profileDesc> (after the
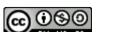
<fileDesc>). For persons and organizations we add a <particDesc> and for places we add a <settingDesc>. Let's do it step by step.

We will now encode metadata for two of the persons that you have just encoded.

- add a <particDesc> element in the <profileDesc>, and inside it a <listPerson>.

```
<profileDesc>
     <particDesc>
          <listPerson> … </listPerson>
     </particDesc>
</profileDesc>
```

- Let's start with a person who appears in all letters: the writer **Stratís Tsírkas**. Add a<person> element with the **xml:id** 'ST'. Add a <persName> inside it; you may use the 'sub-elements' <forename> and <surname>.

```
<listPerson>
     <person xml:id="ST">
          <persName>
               <forename>Stratís</forename>
               <surname>Tsírkas</surname>
          </persName>
     </person>
</listPerson>
```

- Don't forget the **xml:id** attribute! This identifier is necessary in order to point at the person metadata; the **xml:id** can contain any value, but it is more convenient to give it a recognizable one, like the initials or the full name of the person that you are encoding ("ST" or "StratisTsirkas"). Note that a **xml:id** cannot start with a number.
- You can now add other data about Stratís Tsírkas. Start adding the date of birth, the date of death (if any) and the sex, using the elements <birth>, <death> and <sex>.
- If any place, organization or other person appear in the data that you are encoding (for example, a place of birth inside the element <birth>), do use the elements <persName>, <placeName> and <orgName>. Don't forget to use the attribute **when**, if needed.
- Provide an identifier from an external Authority List is a good practice. Use the element <idno>, specifying the attribute @type and the identifier. Start looking for the person in VIAF at <http://viaf.org/>.
- Your <person> element should look like this:

```
<person xml:id="ST">
          <persName>
               <surname>Τσίρκας</surname>
               <forename>Στρατής</forename>
          </persName>
          <birth when="1911-06-23">
                    <placeName>Cairo</placeName>, 23 July 1911</birth>
```

```
              <death when="1980-01-27">
                      <placeName>Athens</placeName>, 27 January
              1980</death>
              <sex value="M">male</sex>
              <idno type="VIAF">64012589</idno>
      </person>
```

- Very good! Now choose another person in the text and add a <person> to the <listPerson>.
- Info about the persons mentioned in the letters can be found … on the web :)

5. Adding metadata for places

In this step we will encode the metadata for two of the places that you have found in the letter.

- Inside <profileDesc> and after <particDesc>, add a <settingDesc> element and inside that a <listPlace>. (Technically the <listPlace> is unnecessary, but it is a good habit as it allows you to group related places.)

```
<profileDesc>
      <settingDesc>
              <listPlace> … </listPlace>
      </settingDesc>
</profileDesc>
```

- Let's start with the first place that appears in your letter (for example, 'Αλεξάνδρεια'). Inside the <listPlace> add a <place> element with an **xml:id**; the value of the **xml:id** attribute can be the name of the place.
- Inside this first <place> add a <placeName> (in this example, 'Αλεξάνδρεια'). You can enter the correspondent <country> (and also going into more details, with <region>)
- In order to point to a place without ambiguity, you can use the Geo-coordinates. A lot of website give the geo-coordinates, as http://www.latlong.net/. The coordinates need to be entered 'lat, long' (with a comma in between). We may use the Geo-coordinates of these places for the exercises in the following days, so do provide them!
- Now add another place after the first one in a similar manner.

```
<place xml:id="athens">
        <placeName>Athens</placeName>
        <country>Greece</country>
        <location>
            <geo>37.983810, 23.727539</geo>
        </location>
```

`</place>`

6. Linking Names and Metadata

- Having marked all these names, and created metadata about them, it seems a shame not to link the names to this metadata. So let's do that!
- Go to the first persName you marked: probably Στρατής. Move the cursor into the tag name just before the closing > and press space. *oXygen* should prompt you with a list of attributes allowed at this point. Add the **ref** attribute and, after doing it, you should get a drop down list of all the xml:id values: '#ST' should appear in the list. Select it.
- Your text should be now something like:

Αγαπητέ μου `<persName ref="#ST">`Στρατή`</persName>`

- Bravo! Continue on and for each `<persName>` and `<placeName>` (for which there is a `<person>` or `<place>` element in the Header) go through and add a **ref** attribute pointing to the correct **xml:id**.
- 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 that the place it is pointing to is in the same document. You could also have stored the listPerson in a separate document; in this case you can:
  - point at it. For example, **ref**='listPerson.xml#ST'
  - include it, as we have seen, using the element `<include>`. In this case (which is our case), you can directly point to the **xml:id**, as if it was in the same document. For example, **ref**='#ST'
  - point at it if it is online. For example, **ref**='http://www.example.com/listPlace.xml#ST'
- 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, linking, or displaying this information).

7. Self-Assessment

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

- Which elements are used to mark personal, place, and organizational names?
- How do you store metadata in the header about the entities these names refer to?
- What values does the @ref attribute allow? How can this be used to point to external files or URLs?

8. Results

You can find an example of encoded letter here:

- view online
  <https://github.com/DiXiT-eu/Digital-Scholarly-EditionsGR-workshop/blob/master/results/correspondence_tsirkas/19570325t.xml>

9. Save often!

Don't forget to save the file you have created! We will continue to work on it in the next couple of exercises.

If you have time, add more persons and places to the <listPerson> and the <listPlace>.

- If you want to know more, here are some handful links to related portions of the TEI Guidelines:
- How to encode personal characteristics, at
  http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html#NDPERSEpc
- How to encode personal events, at
  http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html#NDPERSEpe
- How to encode geo-political and geographic names, at
  http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html#NDPLGU
- How to encode the location of a place, at
  http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html#NDGEOGva
- We haven't seen much of the Organizational Names. You can find info about how to encode these names in the text, at
  http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html#NDORG; how to encode the metadata about them in the Header, at
  http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html#ND-org