Creating a Manuscript Description

1   Learning Outcomes

During this exercise you will learn :

- how to create a basic manuscript description

- how the various components of a TEI manuscript description are organized

- how to create a detailed manuscript description

2   A minimal `<msDesc>`

A manuscript description, as the name suggests, describes a manuscript. The description can be rich and complex, or very simple. The same applies to its markup. Let's begin with the absolute minimum.

A manuscript description should always tell you

- where the manuscript is to be found : `<msIdentifier>`

- what the manuscript contains : `<msContents>` or `<p>`

- and probably something about its physical composition, size etc. (`<physDesc>` or `<p>`)

Of course it may contain much, much more, as we will see...

3   Getting started

In this exercise we will create a TEI document which contains a manuscript description, marked up with the TEI `<msDesc>` element.

- Open oXygen, and create a new XML document using the template provided for a P5 Manuscript Description

   - Select New from the File menu (CTRL-N)

   - Scroll down to Framework Templates in the New dialog

   - Open TEI P5, and select Manuscript description

- oXygen creates a template for you to fill which looks more or less like this

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
      <fileDesc>
          <titleStmt>
              <title>Title</title>
          </titleStmt>
          <publicationStmt>
              <p>Publication Information</p>
          </publicationStmt>
          <sourceDesc>
              <p>Information about the source</p>
          </sourceDesc>
      </fileDesc>
  </teiHeader>
  <text>
      <body>
          <listBibl>
              <msDesc>
                  <msIdentifier>
                   <settlement>where is it</settlement>
                      <repository>repository name</repository>
                          <idno>identifier</idno>
                   </msIdentifier>
              </msDesc>
          </listBibl>
      </body>
  </text>
</TEI>
```

- First, fill in the TEI header as follows
  - The `<title>` is something like 'TEI Conformant description of a manuscript'
  - The `<publicationStmt>` says something like 'Unpublished exercise'
  - The `<sourceDesc>` for the moment says simply 'No pre-existing source'
  - If you like, you could add a `<respStmt>` containing a `<resp>` such as TEI Encoding and a `<name>` containing your own name. Do you know where this element can go? How would you find out?

- Next, fill in the `<msIdentifier>` as follows
  - In the `<settlement>` element, give the name of the city (town, village, etc.) where your manuscript is currently stored ('Sofia')
  - In the `<repository>` element, give the name of the repository (institution, library, etc.) which holds it ('Национална Библиотека "Св. Св. Кирил и Методий"')
  - In the `<idno>` element give the shelf mark, call number or similar for your manuscript ('1144')

- This is all that is needed for a valid manuscript description. But to make it useful, you may wish to add a little more information. You could supply this as a `<head>` element, for example containing the text 'Short Apostle Lectionary in Russian, written in Cyrillic script by a single hand'. Or you could add a series of `<p>` or `<ab>` elements providing whatever information you choose. Or you could use some of the more specialised elements provided by the TEI, as follows.

- After the `<msIdentifier>`, add a `<msContents>` element to provide information about the 'intellectual content' of the manuscript. In the simplest case, you might just want to summarise this, using a `<summary>` element.

- After the `<msContents>` element, add a `<physDesc>` element to describe the physical aspects of the manuscript. In the simplest case, you could provide this as one or more paragraphs (`<p>`) elements.

Your minimal `<msDesc>` should look something like this:

```
<msDesc>
 <msIdentifier>
  <settlement>Sofia</settlement>
  <repository>National Library</repository>
  <idno>1144</idno>
 </msIdentifier>
 <head>Apostle Lectionary, in Russian, written in Cyrillic script by a single
   hand</head>
 <msContents>
  <summary>Apostle Lectionary in Russian</summary>
 </msContents>
 <physDesc>
  <p>Parchment, 39 folios, written in Cyrillic script by a single hand. Some damage
to
     outer leaves. </p>
 </physDesc>
</msDesc>
```

Since the <head> element now adds nothing to the rest of the description, you might like to remove it.

4   A more ambitious description

In this part of the exercise, we will first convert a pre-existing manuscript description from word processor format (Word or Libre Office) into TEI XML. We will then consider how each part of the description could be marked up in a TEI msDesc.

- Browse to your Work folder and click on either the file LUL_Ms40.odt (if you have Libre Office) or LUL_Ms40.docx (if you have Microsoft Word) to open it. Take a quick look at its contents and close the word processor application again.

- Now, in oXygen, choose File - Open, navigate to this file, and open it again

- A new window labelled Archive Browser opens, showing the structure of the word pro-cessor file, which is actually a collection of several files. Find the file called content.xml (for an .odt file), or Word/document.xml (for a .docx file) and open it by double clicking on it

- As you can see, this is actually an XML file, although it uses a different non-TEI set of tags. We will use oXygen to convert it automatically into a basic TEI document.

  - Select Transformation -> Configure Transformation Scenario(s) from the Document menu. Or type CTRL-SHIFT-C. Or click the little spanner icon
  - Check the little box next to DOCX TEI P5 (or ODT TEI P5) and press the Apply Associated button.
  - After a brief pause, a new window opens in which a basic TEI P5 version of the document appears. Click the Format and Indent button (CTRL-SHIFT-I) to see its structure, which should look familiar to you!

You could transform this document into a properly structured TEI MS Description in (at least) two ways. You could make a new document, using the template provided for a P5 Manuscript Description in exactly the same way as before, and then gradually cut and paste material from one document into the other. Or you could change the tagging in the document you

just created, deleting unnecessary additional material as you go. We don't mind which you choose, but we will describe the second.

- If you adopt the second strategy, you will need to change the outermost `<div>` in the body of the document into an `<msDesc>` element.
  - Highlight all the text inside the opening tag (`div type="div1"`) taking care not to include the pointy brackets.
  - Type the new name you want (`msDesc`).
  - Move the cursor outside the tag.
  - If you scroll down to the end of the file, you will see that oXygen has helpfully inserted the closing tag.
  - You will also see that the document is now invalid. Don't panic! You will fix this during the rest of the exercise.

- Begin by creating a valid `<msIdentifier>` which shows that the manuscript with identifier `Medeltidshandskrift 40` is in the holdings of a repository called `Lund University Library`. If you like, you can additionally provide the country (Sweden), the town (Lund) and the institution (Lunds Universitet), each tagged with the appropriate TEI element. Don't forget that just typing a < character anywhere in the document you are editing will pop up a list of the TEI tags available at this point: this will help you add the right tags within your `<msIdentifier>` element.

- This description includes two former identifiers (following the Latin word 'olim'), which you should mark up using the `<altIdentifier>` element. We suggest that you regard 'Bibliotheca recentior' as the name of a collection. And remember that you need to repeat the name of the host repository (`Lund University Library`) for each of these additional identifiers.

- This description also includes two summaries or abstracts, one very short, and one slightly longer. You can decide whether or not to keep them: if you do keep them, they can both by tagged using the <head> element. You could use the @type attribute to distinguish them, if you like.

The start of your document should now look something like this:

```
<msDesc>
    <msIdentifier>
        <country>Sweden</country>
        <settlement>Lund</settlement>
        <institution>Lunds Universitet</institution>
        <repository>Lund University Library</repository>
        <idno>Medeltidshandskrift 40</idno>
        <altIdentifier>
            <collection>Bibliotheca recentior</collection>
            <idno> N:o 13</idno>
        </altIdentifier>
        <altIdentifier>
            <repository>Lund University Library</repository>
            <idno>Bibl.Ms.Th. 4o 3 </idno>
        </altIdentifier>
    </msIdentifier>
    <head type="short">Russian Chronograph, redaction of 1671. Russia,
    <head type="desc">The Russian Chronograph is a cosmographic encyclc
        compilation of fragments from different works. It covers the pe
```

The remainder of the description consists of a series of `<div>` elements, each containing a `<head>` to specify its function. Some (but not all) of these `<div>` elements can be translated immediately into a more specialised element, as follows:

- replace the `<div>` tag with its corresponding MsDesc element (see table below)

- delete the <head> element

heading	element
Contents	<msContents>
Physical description	<physDesc>
Decoration	<decoDesc>
Binding	<bindingDesc>
Foliation	<foliation>
Additions	<additions>
Condition	<condition>
History	<history>
Origin	<origin>
Provenance	<provenance>
Acquisition	<acquisition>

Your document should now be valid at the start and the end, but not in the middle, because the sub-parts of the <physDesc> are not yet properly organised. We'll look at that in a moment, but first we'll tidy up the <msContents>.

5 Structuring the <msContents>

You can describe the intellectual content of a manuscript in many different ways. In the previous simple example we just supplied a <summary> element, which may be adequate in many cases. In a manuscript containing several different titles or items, it's more usual to supply a list of <msItem> elements for each one. Inside the <msItem> a number of more specialised elements are available to describe the item more exactly, such as a title, a rubric, an icipit etc. And we can also specify the part of a manuscript in which the item appears by using the <locus> element.

- First, turn the first <p> element in the <msContents> element into an <msItem> element, in the same way as you transformed the outermost <div> tag into an <msDesc>. Highlight everything inside the first <p> tag (but not the pointy brackets), and type msItem to replace it.

- The figure 1 which follows is the number of this item within the manuscript. Add an @n attribute to your <msItem> and give it this value. Then delete the whole of the <hi> element which contains the figure.

- Now look at the passage enclosed in parentheses (and). This contains a <ref> element whose @target indicates an online version of this manuscript and which contains:
  - an empty <anchor> element whose @xml:id supplies an identifier for this item
  - a <hi> element containing an indication of the folios in the manuscript concerned (ff 6r-474v)

  Proceed as follows:
  - change the <ref> element into a <locus> element
  - change its @target attribute into a @facs attribute
  - move the @xml:id attribute from the <anchor> to the <msItem> you created earlier
  - Delete the remaining parts of the <hi> elements and (if you like) the parentheses

The start of your msItem should now look like this:

```
     romanov tsar, nikhait redorovich in 1013.. </p>
 <msContents>
     <msItem xml:id="Mh_40-item1" n="1">
         <locus facs="http://laurentius.ub.lu.se/cgi-bin/image-range-db.cgi
             ff. 6r-474v</locus> Russkij Chronograf
```

We now need to tag the titles etc used to describe the work.

- Select the two words 'Russkij Chronograf' and use CTRL-E to tag them as a `<title>`

- Select the first passage within quotation marks labelled 'rubric', i.e. from 'книга глаголема□' to 'епископа гевалъскаго' and use CTRL-E to tag this as a `<rubric>`. Delete the `<hi>` element and the quotation marks.

- Do the same for the next quoted passage (from 'зачало' to 'землю') but this time tag it as an `<incipit>`

- Do the same for the next quoted passage (from 'и многолѣтно' to 'вѣки амин') but this time tag it as an `<explicit>`

- And finally tag the stretch of text labelled as a colophon as a `<colophon>`

- Don't forget to remove the redundant `<hi>` and `<anchor>` elements left behind! Your `<msItem>` element should now be valid (i.e. no red underlining anywhere in it)

To check your work, open the Outline view (Window -> Show View -> Outline) to display the structure of your msDesc. It should look something like this:

```
♀  ●   msDesc
  ○─ ●   msIdentifier  Sweden
      📄  head  Russian Chronograph, redaction of 1671.
      📄  head  The Russian Chronograph is a cosmograp
  ♀  ●   msContents
      ♀  ●   msItem "Mh_40-item1" ff. 6r-474v
          ●   locus "http://laurentius.ub.lu.se/cgi-bin/ima
          A   title  Russkij Chronograf
          ●   rubric  книга глаголемаю граног раф юж
          ●   incipit  зачало 1:го дни в начале сотвор
          ●   explicit  и многолѣтно здравствовати и
          ●   colophon  человеческих 72 четверонога
```

6   Structuring the `<physDesc>`

The TEI schema distinguishes (and therefore tags) many distinct components within a physical description that is not just a sequence of paragraphs. Specifically, it distinguishes between information about the text bearing object itself, which is marked up using the `<objectDesc>` element, and information concerning aspects of the writing and decoration it carries. In some

cases, the TEI schema also requires elements to be given in a specific order. To make our document valid, proceed as follows:

- Select the first two paragraphs (from "Support:" to "less details") and use CTRL-E to enclose these paragraphs with a `<support>` element. Delete the element `<hi>` with which it begins (both the tags and their content).

- Change the third `<p>` element (starting "Extent:") into an `<extent>` element. Delete the element `<hi>` with which it begins.

- Change the fourth `<p>` element (starting "Collation:") into a `<collation>` element. Delete the element `<hi>` with which it begins.

- Select the `<foliation>` element you created earlier and move it to follow the `<extent>` element

- Select the `<condition>` element you created earlier and move it to follow the `<foliation>` element

- Select all five elements, `<support>`, `<extent>`, `<collation><foliation>` and `<condition>`; use CTRL-E to wrap them in a `<supportDesc>` element

- Select the paragraph labelled 'Layout'; use CTRL-E to enclose it first by a `<layoutDesc>` element, and then by a `<layout>` element. Delete the element `<hi>` with which it begins.

- Select the paragraph labelled 'Script'; use CTRL-E to enclose it by a `<scriptDesc>` element. Delete the element `<hi>` with which it begins.

- Select the `<additions>` element you created earlier and move it to follow the `<decoDesc>` element

Phew. If you have done all that correctly, your document is now valid and you have a jolly green square. Check the structure of your `<physDesc>` element in the Outline view: it should look like this:

If you're still feeling strong, you might like to improve the markup of the `<support>` element as follows:

- Change the `<p>` element containing the word 'Paper' into a `<material>` element

- Change the second `<p>` element into a `<watermark>` element

- Delete the text `Three different watermarks: 1)`

- Split the `<watermark>` element into three by typing ALT-SHIFT-D before `2)` and `3)`

- Delete the numbers 2) and 3)

Our final version of this manuscript description is in the file Work/msDesc-corrected.xml.