

LEAP Transcription Pre-Kick Off – Documentation

Adrian S. Wisnicki (28 April 2014)

Overview

This document sets out the procedure for transcription for LEAP. It reviews the information presented in the LEAP transcriptions pre-kickoff (25 April 2014).

To carry out transcription for the project, four elements are needed: 1. The *Livingstone Online* Encoding Manual (LEM), 2. The appropriate LEAP TEI P5 template, 3. The LEAP RelaxNG Schema, and 4. Images of the item to be transcribed.

1) The *Livingstone Online* Encoding Manual (LEM)

The *Livingstone Online* Encoding Manual (LEM) is currently found at the following URL:

<http://research.it.ox.ac.uk/LEAP/> The LEM can be used online or downloaded and used locally. To download either select the “ZIP of all the above” option from the URL above or right click on the LEM, select “Save As” and save it to your hard drive. If you use the LEM locally, you should also download the leap-doc.css file from the URL above in the same way as the LEM and place it in the same directory as the LEM. The css file provides styling for the LEM.

The LEM was developed in collaboration with James Cummings and represents a revised, enhanced, and expanded version of the original *Livingstone Online* cheat sheet. The LEM consists of two principal parts: 1) The numbered opening section and 2) the appendices. The numbered opening section (about 30 pages when printed) is self-explanatory, is what you will normally use for transcriptions, and will cover the majority of encoding issues encountered. The appendices supplement and expand upon the numbered section and can be ignored unless needed. The LEM will eventually be incorporated into the new *Livingstone Online* site.

2) The LEAP TEI P5 Transcription Templates

There are currently two LEAP TEI P5 transcription templates, one for letters, one for diaries. I have distributed them by email, but they will eventually be made available at the URL given in point #1 above. The LEM outlines the procedure for using the templates with additional guidance provided via comments within the template themselves.

The templates are based on the previous *Livingstone Online* transcription template. Each template requires you to complete two distinct sections: the <teiHeader> and the <text> sections. The <teiHeader> section is partly completed, but requires you to revise and/or complete various parts. For some of the bibliographical parts in the <teiHeader> you will need the DLC Digitisation Masterlist, v3.xlsx file (see point #4, below). The <text> section will consist of your transcription of the given *Livingstone* document based on the LEM.

3) The LEAP RelaxNG Schema

The LEAP RelaxNG Schema is also available for download from the URL given in point #1 above. This schema will allow us to validate our transcriptions against our new *Livingstone Online*/LEAP TEI specifications. The schema needs to be “associated” with each document you transcribe. James Cummings has provided the following simple instructions for creating this association via oXygen:

- Go to the Document -> Schema -> Associate Schema menu item and note the shortcut icon that looks like drawing pin/red square/blue triangle. Select this menu item, or use the icon on the toolbar.

- Use the folder icon on the dialog box to navigate to leap.rng and choose it. If you have an option to "Add additional association for embedded schematron rules" then include that as well. (This turns on additional rules that make the schema even tighter.)

4) Images of the Item to be Transcribed

For LEAP the principal objects that we are transcribing are Livingstone's Field Diaries and the Unyanyembe Journal. The University of Glasgow is creating images of these items (as well as many, many other items from the David Livingstone Centre) for our project.

a) Locating the DLC Digitisation Masterlist,v3.xlsx

To access these images, first navigate to the following directory:

[removed]

In this directory, find the DLC Digitisation Masterlist,v3.xlsx file, download it to your hard drive, then open it. At the bottom of the Masterlist spreadsheet you will find five tabs which correspond to the five categories of items found in the Glasgow-Final-Images directory.

b) Additional Bibliographical Information on the Masterlist

In any of the tabs on the Masterlist, Column C contains the title of the item, so here you need to look for the item you've been assigned to transcribe (for instance, "Field Diary II"). Other cells of a given row in the spreadsheet will provide additional bibliographical data for that item. This is the bibliographical data you will need to complete the <teiHeader> for the given item when transcribing it (see point #2 above).

c) File Names on the Masterlist

In any of the tabs on the Masterlist, Column B (in purple) contains the LEAP file name (without the suffix) for the first image of the given item. For instance, for "Field Diary II" it is:

liv_000002_0001

The first segment (always "liv") is short for "Livingstone." The second segment ("000002" in this case) is the item number, is arbitrary, and will change from item to item. The third segment ("0001" in this case) is the page number and will, of course, be incremented by one number for each successive page of the given item (_0001, _0002, _0003, _0004, etc.)

As a result, the full name for the JPEG image for this particular page would be liv_000002_0001.jpg. The full name for the TIFF image would be liv_000002_0001.tif. A transcription of this page would be liv_000002_0001.xml. Metadata for the page would be liv_000002_0001.txt.

The second segment of the file name, therefore, allows us to group all elements for a given item together, while the third segment allows us to group all elements for a given page of the item. Items with similar names are differentiated by the file suffix (.jpg, .tif, .xml, and .tif). To better understand this, consider the following example, which contains a mixture of items, pages, and page types, but which can be readily sorted out via the file names:

```
liv_000002_0001.jpg
liv_000002_0001.tif
liv_000002_0001.xml
liv_000002_0002.jpg
liv_000002_0002.tif
liv_000003_0001.jpg
liv_000003_0002.jpg
```

d) Locating Images of the Given Item

The first two segments of the LEAP file name will also allow you to find images of the given item. For instance, the directory containing the images of the example item above will be called “liv_000002”

As a result, once you are in the “Glasgow-Final-Images” directory, open the “01-Journals” directory and find the directory that corresponds to the item for which you are looking (“liv_000002” in our example). Open that directory, and you will see a final four subdirectories:

```
JPEGs
JPEGs-txt
TIFFs
TIFFs-txt
```

The first directory contains JPEGs of the item (these are the images you will need for your transcription; you can ignore the other directories here). The second directory contains metadata files for the JPEGs. The third directory contains TIFFs of the item. The fourth directory contains metadata files for the TIFFs.

Once you have found the images for the item you are transcribing, copy them to your hard drive.

e) Naming the File of Your Transcription

Section 2.3 of the LEM (<http://research.it.ox.ac.uk/LEAP/leap.html#file-naming>) contains our general file naming rules for *Livingstone Online*. For LEAP, each of us will always be transcribing only a segment of a given item. As a result, please review the rules in the LEM and follow them, but please take note that when you name your transcription for LEAP, you will be using the page number segment in the file name, and this page number segment should correspond to the first image of the file that you are transcribing.

So for instance, if I were transcribing pages 10 to 20 of Field Diary II, the name of the XML file containing transcriptions of all these pages after my initial transcription would be “liv_000002_0010_aw.xml”. After it had been proofread by Megan Ward it would be “liv_000002_0010_aw_mw.xml”. Once all the separate transcribed parts of this diary were assembled into the final integrated file, the file would be called “liv_000002.xml”.