**Notes on Initial Transcription of *Missionary Travels* for XML Template**

**Justin Livingstone 8th Jan 2014**

Transcribed Sections

Sect 01, pp.0004-0008

Sect 05, pp.0092-0093

**Note on XML File Names**

The manuscript of the *Missionary Travels* is divided into sections of varying lengths, often between 20-30 folios. In the manuscript, the section divisions are indicated by a Roman numeral (see for instance, the numeral on the top left of p.0004). We are creating one xml document for each section and the xml file will be named according to the section number (we’d thought of using Arabic numerals ie Sect01, Sect 02 etc).

JC: Seems entirely reasonable to me.

As well as myself, there are two assistants working on the project and the xml file names need to make the workflow clear. In each case there will be an initial transcription by one assistant, which will be proofread by the other, and then finally passed on to me. This is how we intend to name the files, using our initials:

Initial transcription by 1st assistant (Kate Simpson):

Sect01\_ks

Proofing by 2nd assistant (Johanna Green):

Sect01\_ks\_jg

Final proofreading and additional encoding by me (Justin Livingstone):

Sect01\_ks\_jg\_jl

Do you have any comments on this? We might face problems if either of the assistants do further work after I’ve looked at the document. If the simple linear workflow is disrupted, I’m not sure how I would name the files.

JC: This is a fine way to do it, and once final name it ‘Sect01.xml’ or similar. Working in phases in this way where multiple people see the same file is good practice and I’ve seen used to good effect before. Also note that they can add a <change> to the <revisionDesc> at the bottom of the header to note the major stages of work that has been completed.

**Major Features**

* Do we need to encode the manuscript section breaks within the xml document? At the moment Kate (the primary research assistant) has just transcribed the Roman numeral as part of a line of text, but should we do this in another way? The section divisions will of course be signalled in the xml file names.

JC: That is fine, though I would suggest that either the enclosing <div> immediately inside the <body> have an @xml:id indicating the section. (This way if the files are combined then they will have easy identifiers.)

* The manuscript page numbers in the top right hand corner do not correspond to the jpg file numbers which we’ve decided to encode as the document’s page number (ie. 0004). Do these numbers in the top right hand corner need to be encoded in a particular way? There are sometimes two such numbers, in different hands (see p.0009).

JC:I’ve given a suggestion in the file putting running heads of any sort, including the page number in a <fw> with an <add> as you’ve done is probably good enough. For the actual page break <pb n=”3” facs=”0004.jpg”/>

* There are multiple hands used in the text. At the moment, Kate has marked up unknown hands in this way: <add place="inline" hand="U">. There are at least two of these in the manuscript; one that marks in red (John Murray’s copyeditor, I think – see 0093) and another that marks in pencil (possibly a later curator or archivist – see 0092). Kate suggests using a mark up tag such as element <edit> attribute <pencil>. What do you think? How shall we differentiate between these hands, particularly if the individuals remain unknown? Is there perhaps a way of numbering them as unknown hand 1 and 2 for instance? See also comments/questions in the xml documents.

JC: When marking up unknown hands you have the option of not specifying a hand <add place=”inline”> but that is usually assumed to be the same as the main hand (e.g. author). I would mark these as known hands and use @rend to give red/pencil description. <add place=”above” hand=”#murray” rend=”red”>greatly</add> If you want to record the revision campain use @change to point to a <change> element in the header (in ‘creation’ element in the header). If unknown hands then describe the hand in a <handNote> in the header with an @xml:id=”hand1” and refer to hand=”#hand1” on add/del/etc. In that hand.

* Much of the material added by Livingstone to the body of the text is in the margins. However, he often draws a line from the margin to the place in the text that the material should be inserted. Shall we encode such lines and arrows? Another example of a mark we might want to encode is the double line Livingstone draws above the first paragraph on 0092.

JC: As noted in the Sect05 Document I would I personally wouldn't record the arrow itself except perhaps by classifying with @type or describing with @rend on <add> because my bias is to record the intended reading of the document whilst noting the physical features which instantiate that. (e.g. deletions, etc.) So I wouldn't have the <add place="below">^</add> but would have <add place="margin" rend="arrowed">this did not...</add>. Partly because there are so many types of arrow etc. If you \*do\* want to record it, this is precisely the type of thing that the new <metamark> element is for. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-metamark.html>

**Other Issues**

* How shall we mark up other nationalities? “Tribe” doesn’t seem appropriate for this.

JC: I would have said: <name type="nationality">Boer</name> and in other cases if they are a band of people acting like tribe then <orgName type="tribe">Bakwena</orgName> You could use <term> for this as you have, but <name> or <orgName> gives you an additional annotation - this is a name / showing nationality / 'Boer'  
- this is a name / acting as a single organisation / of type 'tribe' / Bakwena

* We intend to correct Livingstone’s incorrect spelling (using the <choice><sic><corr> tags). But would you recommend correcting punctuation? In other words, if Livingstone has left out a full stop should we add it as a correction or simply ignore it and stay true to the manuscript?

JC: If you want to generate a reading text, then yes but using<supplied reason="omitted">.</supplied>

* How shall we encode indented paragraphs? See comment in xml document Sect01.

There are two things to mark: One is that the paragraph as a whole is on the right-hand side, the other is that the first line is indented. I would record the location on the page with the @rend and provide <space> to indicate the indent. <p rend="right"><space dim="horizontal" extent="5" unit="char"/>... But that is if you want to make sure to record the exact length of the indent. You \*could\* just provide two @rend values: <p rend="right indent"> to say the paragraph is on the right and also the first line is indented.

* Kate wonders if we should retain the American style of date to ensure consistency with Livingstone Online, or use the international date standard?

JC: Ick. I hate American dates, internationally we never know if 2/4/14 is 2nd of April or 4th February. But if you \*must\* use American dating, provide the 'real' machine-readable format in the @when attribute: <date when="2013-11-18">18.11.13</date>

* We have marked up the Cape as a “region” and we have done the same for the Hebrides (in contrast, Ulva is marked up as an “island”). In Livingstone Online, East Africa is also marked up as a “region”. It seems odd to mark up these very different areas in the same way. Is this alright, or would you suggest something different?

JC: You can have <region> both inside and outside a <placeName> What you might do is classify these further by doing <placeName type=”islands”><region>Hebrides</region></placeName>  
But otherwise using <region> for these disparate areas of regions is fine. Though you could argue you should use <bloc> for something that covers multiple countries.

* Is capturing Adrian’s role as an “editor” in the xml document the right way to proceed? He is co-director of Livingstone Online (which is the publishing my edition) and a consultant on transcription issues, so this should be captured in the document somehow?

JC: Yes, this is ok, but if you want more fine-grained control then using <respStmt> as I've demonstrated below gives an unlimited number of possibilities (and means you can give everyone and anyone the credit they deserve):  
<respStmt>  
 <persName xml:id="AW">Adrian S. Wisnicki</persName>  
 <resp>Transcription Consultant</resp>  
 <resp>co-director of Livingstone Online</resp>  
 <resp>director of staff morale</resp>  
</respStmt>  
  
The benefit of the @xml:id is any time AW has done something you can point to it (@resp=”AW”)