

David Livingstone's 1871 Field Diary – XML TEI P5 Encoding Practices

We have transcribed and encoded the text of David Livingstone's 1871 Field Diary into XML according to the P5 Guidelines of the Text Encoding Initiative (TEI, see <http://www.tei-c.org/Guidelines/P5/>). Our encoding practices also draw on the tagging guidelines developed by *Livingstone Online* (<http://www.livingstoneonline.ucl.ac.uk/technical/guide.html>), although we have introduced a variety of modifications and have updated these guidelines from TEI P4 to TEI P5. Finally, our work is also indebted to the advice of Lisa McAulay, the Librarian for Digital Collection Development in the UCLA Digital Library Program; James Cummings, Manager of InfoDev, University of Oxford; and Doug Emery, our data manager.

Each XML document we have produced contains two major structural elements: the header and the text. A summary of the tagging practices used for both of these elements follows. Note: in the text below "folio" refers to one side of a leaf of Livingstone's diary, whereas "page" may denote all or only a portion of that folio. In all cases, a leaf of Livingstone's diary contains two folia (one on the recto side, the other on the verso), while a folio may contain either one or two diary pages.

Table of Contents

I. File Naming

II. Document Header

A. <fileDesc>

1. <titleStmt>
2. <publicationStmt>
3. <sourceDesc>

B. <encodingDesc>

C. <profileDesc>

D. <revisionDesc>

E. <facsimile><surface><graphic><zone>

III. Body Text

A. Major Structural Elements

1. <text><body><div>
2. <ab>
3. <cb/>
4. <milestone>
5. <fw>
6. <p>
7. <lb/>
8. <w>

B. Textual Layout and Formatting

1. <space/> and Spacing between Words
2. <hi>

C. Additions and Deletions

1. <add>
2.

- D. Illegible Text
 - 1. <unclear>
 - 2. <gap>
- E. Editorial Interventions
 - 1. <supplied>
 - 2. <choice>
 - a. <abbr><expan>
 - b. <sic><corr>
 - c. <unclear>
 - d. <orig><reg>
- F. Content Tagging
 - 1. People
 - 2. Tribes and Villagers
 - 3. Places and Geographical Entities
 - 4. Dates
 - 5. Medical References
 - 6. Foreign Words and Phrases
 - 7. Drawings, Maps, Calculations, Seals, Squiggles, Hebrew/Arabic
- G. Special Characters

I. File Naming

The electronic transcription of Livingstone's 1871 Field Diary consists of 44 individual XML files. The name of a typical XML file takes the following format:

DLC297c_133-102_001r_TEI.xml

The *first segment* consists of the initials of the institution holding this portion of the diary (in this case, the David Livingstone Centre) followed by the institutional shelfmark (297c).

The *second segment* indicates Livingstone's own page number(s), if provided. Roman numerals have been changed to Arabic numerals to ease reading. For any portions of the diary where Livingstone does not provide page numbers, we have numbered the folia consecutively beginning with 001, with the same number being used for the recto and verso:

NLS10703_001_036r_TEI
 NLS10703_001_039v_TEI
 NLS10703_002_037r_TEI
 NLS10703_002_038v_TEI

The *third segment* indicates the institutional page number(s), if provided. In addition, this segment includes the letter "r" (recto) or "v" (verso). For any portions of the diary lacking institutional page numbers, we have numbered the folia consecutively beginning with 001, with the same number being used for the recto and verso:

DLC1120b_001r_001r_TEI
 DLC1120b_001v_001v_TEI

The *fourth segment* indicates that the given folio of the diary has been transcribed according to TEI P5 guidelines.

The *extension* indicates that this is an XML document.

II. Document Header

The document header is divided into four main sections, all of which are contained within the <teiHeader> element:

A. <fileDesc>

The <fileDesc> element records the description of the original document and computer file. This element contains three major sections:

1. <titleStmt>

This element gives the title, authorship, and editor of the file. Since the 1871 Field Diary is part of the Manyema Field Diary, we use the latter rather than the former title in all of our files. This element also includes the date of the original document, Livingstone's page numbers (where available) in Arabic numerals, and the first three segments of the file name (see the section on "File Naming," above).

2. <publicationStmt>

This element describes how, where, when, and by whom our transcription is being published. The element also provides the rights and conditions of use.

3. <sourceDesc>

This element describes the original source, including the collection in which it is held and the shelfmark as well as bibliographical details of Livingstone's "undertext." We include the <sealDesc> and <seal> elements only if the original document contains a seal.

B. <encodingDesc>

The <encodingDesc> element describes the relationship between the transcription and the original document. This element contains the <projectDesc> which records our encoding objectives, including a description of those features of the original documents that were and were not encoded. In general, our practice has been to encode any text written by Livingstone in the Body Text (see below), while transcribing (but not encoding) any text by other hands within the <projectDesc>. Finally, the <projectDesc> also contains information about material characteristics of Livingstone's original manuscript and how these relate to the "undertext."

C. <profileDesc>

The <profileDesc> element contains classificatory information about the text, including the languages included.

D. <revisionDesc>

The <revisionDesc> element describes the file's revision history. This element contains information describing all the actions undertaken by the editorial team, such as when and by whom a folio was transcribed and proofread. Each of these separate actions are enclosed within a <change> element containing detailed information on the action taken and the individual responsible for it. The <date> element here contains dates in the yyyy-mm-dd format.

E. <facsimile><surface><graphic><zone>

These elements are dynamically inserted into the file (directly after the <teiHeader> element) and assist in linking the transcription, line by line, to the corresponding areas of the registered spectral images.

*

The document header will typically take the following structure in the XML files we have produced:

```
<teiHeader>
```

```
<fileDesc>
```

```
<titleStmt>
```

```
<title>David Livingstone, The Manyema Field Diary, 1871, pp.149/146, DLC297b_149-146_012r</title>
```

```
<author>David Livingstone</author>
```

```
<editor>Adrian S. Wisnicki</editor>
```

```
</titleStmt>
```

```
<publicationStmt>
```

```
<authority>Livingstone Online</authority>
```

```
<pubPlace>London, UK and Los Angeles, California, USA</pubPlace>
```

```
<publisher>Digital Library Program, University of California, Los Angeles </publisher>
```

```
<date>2011</date>
```

```
<availability>
```

```
<p>All materials are licensed for use under the <ref  
target="https://creativecommons.org/licenses/by-nc/3.0/">Creative Commons Attribution-  
Noncommercial 3.0 Unported License</ref>. (c) Dr. Neil Imray Livingstone Wilson, 2011<p>
```

```
</availability>
```

```
</publicationStmt>
```

```
<sourceDesc>
```

```
<msDesc>
```

```
<msIdentifier>
```

```
<collection>The Scottish National Memorial to David Livingstone Trust</collection>
```

```
<idno>297b, fol. 12</idno>
```

```
</msIdentifier>
```

```
<physDesc>
```

```
<objectDesc>
```

```
<ab>Written over:
```

```
<biblStruct>
```

```

<monogr>
<title level="j">The Standard (London)</title>
<imprint>
<biblScope type="pp">1</biblScope>
<date>24 November 1869</date>
</imprint>
</monogr>
</biblStruct>
</ab>
</objectDesc>
<sealDesc>
<seal>
<p>[Describes any seals the original document may contain.] </p>
</seal>
</sealDesc>
</physDesc>
</msDesc>
</sourceDesc>
</fileDesc>

<encodingDesc>
<projectDesc>
<p>David Livingstone writes his diary entries over a portion of a page from The Standard
(London). Livingstone's "overtex" is what is transcribed and represented in Text Encoding
Initiative (TEI) P5 XML encoding. In addition, this folio includes the number "10" written in pencil
at the top center of the page in a second, unknown hand.</p>
</projectDesc>
</encodingDesc>

<profileDesc>
<langUsage>
<language ident="en">English</language>
<language ident="und">Undetermined African Language</language>
</langUsage>
</profileDesc>

<revisionDesc>
<change>
<date>2011-06-03</date>
<persName>Adrian S. Wisnicki</persName> initial transcription and XML encoding.
</change>
<change>
<date>2011-08-03 </date>
<persName>Kathryn Simpson</persName> proofreading.
</change>
</revisionDesc>

</teiHeader>

```

III. Body Text

A. Major Structural Elements

1. <text><body><div>

We record the transcribed text of each folio of Livingstone's diary within the <text><body><div> elements. Each XML file contains only one <div> element, even if the given folio contains two of Livingstone's numbered pages.

2. <ab>

We use the <ab> element to mark up portions of text that appear alongside but are not clearly related to the diary entries on a given page of Livingstone's manuscript (see DLC297c_133-102_001r_TEI and DLC297b_151-144_011r_TEI for examples).

We also use the <ab> element to mark a few parts of the manuscript where Livingstone writes over his own writing, in effect producing "undertext" (DLC297b_133-162_005v, DLC297c_111-124_005v) and "overtext" (DLC297c_121-114_007r). In these cases, the <ab> element takes both the @rend and the @xml:id attributes, each with a value of layered-text. Except for the first instance (see below), which covers a significant amount of text, we do not encode the second layer of text in any way.

The <ab> element is contained within the <div> element when used for the foregoing purposes.

3. <cb/>

In one instance, we use the <cb/> element to mark the beginning of a significant portion of "overtext" that Livingstone writes at a perpendicular angle to the rest of the diary entries on the page (DLC297b_133-162_005v). In this case, the <cb/> element also takes the @corresp attribute with a value of #layered-text.

4. <milestone>

We use the <milestone> element to mark:

- a. The beginning of each folio:

```
<milestone unit="folio" n="001r"/>
```

The @n attribute indicates the institutional page number (see "File Naming," above) and whether this is the recto or verso.

- b. The beginning of each diary page:

```
<milestone unit="page" n="CII"/>
```

The @n attribute records Livingstone's original page number or, if not provided, a sequential page number (1, 2, 3, etc.) assigned by us. This latter practice, however, only applies to the Additional Diary Pages, which were not assigned Roman numerals by Livingstone.

c. The beginning of each column of text on those folia that have more than one column of text but have not been provided individual page numbers by Livingstone (e.g., DLC1120b_001r_001r); or on numbered pages that contain two or more adjacent columns of text on a single page (e.g., DLC297c_105-130_002v):

```
<milestone unit="column" n="1"/>
```

The @n attribute specifies the column number.

The <milestone> element is also used to mark where a portion of text with multiple columns ends:

```
<milestone unit="endColumns"/>
```

d. The horizontal line(s) Livingstone draws across his diary pages to separate individual diary entries:

```
<milestone unit="section" rend="line"/>
```

The @rend attribute specifies the number of lines Livingstone has drawn. Values used include: half-line (to mark a line that only goes halfway across the page), line, 2 lines.

5. <fw>

We record Livingstone's headers and, in one case (DLC297c_131-104_002r), the footer within the <fw> element:

```
<fw n="heading_CXXIV">CXXIV Journal </fw>
```

The @n attribute specifies that this is a heading and denotes Livingstone's original Roman numeral. We use the <fw> element to encode Livingstone's original page number and the words "Journal" or "Note" (which Livingstone typically includes at the top of each diary page). We record any succeeding dates within a separate <p> element.

The exceptions to this practice are as follows. In two instances where Livingstone places the word "Journal" on a line other than the first line of the page (DLC297b_159-136_007r, DLC1120b_001r_001r), we do not encode the word as part of the <fw> element but rather treat it as part of the first dated diary entry on the page. In two other instances (DLC297c_113-122_006v, DLC297c_115-120_007v), we include a date that marks the opening of a new diary entry in the <fw> element because the date appears between Livingstone's Roman numeral and the word "Journal."

6. <p>

We encode each of Livingstone's paragraphs within the <p> element. Typically, the <p> start-tag is followed by the date for a new diary entry. If, at the beginning of a diary page, the date follows the words "Journal" or "Note" and represents the beginning of a new diary entry it is encoded within the <p> element. We use the <p> end-tag at the end of the diary page, even if the last paragraph continues to the next folio/page.

7. <lb/>

We use the <lb/> element to mark the beginning of each line of text:

```
<lb n="1"/>
```

The @n attribute indicates the line number. We number each line on each diary page, including the line that includes the <fw> element regardless of whether the <fw> element takes up all or just part of the line.

Example #1

```
<lb n="1"/><fw n="heading_CLIV">CLIV. Journal = </fw><p><date when="1871-07-22">22<hi  
rend="ul;sup">nd</hi> July 1871</date> off  
<lb n="2"/>at daylight about six miles to  
<lb n="3"/>village of <placeName><settlement  
type="village">Mañkwara</settlement></placeName> where I  
<lb n="4"/>spent the night in going - the
```

Example #2

```
<lb n="1"/><fw n="heading_CXLVII">CXLVII Journal - <date when="1871-07-15">15<hi  
rend="sup;ul">th</hi> July</date> continued</fw>  
<lb n="2"/><p>The canoes were jammed in a creek at  
<lb n="3"/>the bottom of the market place
```

Line numbering on any given page is continuous, except for text contained within the <ab> element, where the line numbering is restarted.

When we encode two columns of adjacent text on a numbered diary page using the <milestone> element, we continue the line numbering from column to column:

```
<lb n="19"/><milestone unit="column" n="1"/><p><date when="1871-04-01">Monday  
<lb n="20"/>1<hi rend="sup;ul">st</hi> April  
<lb n="21"/>1871</date>  
<lb n="22"/><milestone unit="column" n="2"/><space dim="horizontal" extent="1"  
unit="chars"/>Rain early every morn<supplied>-</supplied>  
<lb n="23"/>-ing <space dim="horizontal" extent="1" unit="chars"/>I fear it will be  
<lb n="24"/>difficult to buy a canoe – The
```

We do not assign a separate line number to any text added by Livingstone above or below the line of text proper:

```
<lb n="2"/>what the Egyptians priests <add place="below">^</add> <add  
place="above">learned men</add> of remote antiquity
```

Finally, we dynamically insert the @fac attribute into each <lb/> element in order to help link the given line of text to the corresponding area of the relevant spectral images.

8. <w>

We use the <w> element to mark words that Livingstone breaks up over two lines. We also add the @break with the value of no to the <lb/> that falls between the two halves of the word.

Livingstone's use of hyphens is erratic. Sometimes he hyphenates a word at the end of the first line, sometimes at the beginning of the second, sometimes in both places, sometimes in neither. We code each of these instances in the same way, but we also always supply a hyphen at the end of the line if Livingstone himself has failed to provide it.

Example #1 (no hyphen at all)

```
<lb n="7"/>I look <gap reason="deletion" extent="1" unit="chars"></gap> on the drove they brought <w>un<supplied>-</supplied>  
<lb n="8" break="no"/>chained</w> with a sort of pleasure after
```

Example #2 (hyphen only in first line)

```
<lb n="30"/>by one of <persName>Dugumbe</persName>'s people after <w>finish-  
<lb n="31" break="no"/>ing</w> a piece of work = said he was tired
```

Example #3 (hyphen only in second line)

```
<lb n="27"/>sorely needed to be employed <w>him<supplied>-</supplied>  
<lb n="28" break="no"/>-self</w> in something else than penny
```

Example #4 (hyphen in both lines)

```
<lb n="16"/>by which the irresponsible <w>con-  
<lb n="17" break="no"/>-clave</w> brought the Indian command
```

B. Textual Layout and Formatting

1. <space/> and Spacing between Words

a. We record the presence of an unusual space between words using the <space/> element:

```
<space dim="horizontal" extent="5" unit="chars"/>
```

The @extent attribute represents an approximation of character spaces between words. We have tagged all unusual spaces between words, whether the spaces appear to have semantic value or not:

Example #1

```
which the <space dim="horizontal" extent="2" unit="chars"/> ancients may not have
```

Example #2

```
the general level <space dim="horizontal" extent="3" unit="chars"/>It is covered with
```

In one instance (DLC297c_103-132_001v), we also record an unusual space between lines using the <space> element:

```
<space dim="vertical" extent="1" unit="lines"/>
```

However, in general we do not record unusual spaces at the beginning of manuscript lines, unless these spaces mark the opening of a new diary entry or paragraph. Although Livingstone attempts to write his text flush left, he is not always successful in this endeavor and, additionally, is sometimes forced to follow the ragged contours of his diary pages. We believe the effort

needed to represent such highly erratic and inconsistent spacing practices would not be worth the benefit to be gained.

b. Any space between the material contained in the <fw> element and the <p> element is recorded within the <fw> element:

```
<fw n="heading_CXXXIX">CXXXIX Journal = </fw><p><date when="1871-06-20">20<hi  
rend="sup;ul">th</hi> June 1871</date> - Two
```

Otherwise, the space at the beginning of a paragraph is always recorded with the <space/> element and within the <p> element, even if the space is only the extent of one character:

```
<p><space dim="horizontal" extent="1" unit="chars"/><date when="1871-06-29">29<hi  
rend="ul;sup">th</hi></date> <persName>Manilla</persName>'s foray burns ten
```

c. Livingstone's spacing before and after the n-dash and the equal sign (=), his two main forms of punctuation, is highly erratic. At times he places a space before and/or after these characters, sometimes not. Our practice has been to place one space before and one space after these characters in every instance, unless the presence of a larger space necessitates the use of the <space/> element:

Example #1

large goat = <foreign xml:lang="und">lokolia</foreign> colour or skin - Horns

Example #2

palm oil - fowls - <space dim="horizontal" extent="2" unit="chars"/>Each is intensely in

In encoding a few instances where Livingstone uses a particularly long dash, we have used a series of three n-dashes. We have placed a space before and after the series, but have not inserted spaces between the individual n-dashes:

30 feet of depth at flood --- which

However, no space has been placed before an n-dash if it is used to hyphenate a word at the end of a line or at the beginning of a line:

```
<lb n="16"/>by which the irresponsible con-  
<lb n="17"/>-clave brought the Indian command
```

d. On occasion Livingstone introduces or closes a citation by using quotation marks directly over an n-dash. In these cases we transcribe this as an n-dash directly succeeded or preceded by a quotation mark, without any intervening spaces:

sold it -"It is an ill wind that blows no body good" Thanks to the

2. <hi>

We record textual formatting using the <hi> element:

of `<hi rend="ul">vanilla</hi>` pods which the natives mix

The values we use for the `@rend` attribute include: `sup` (superscript), `ul` (underline), `double-underline`, `overbar` (overline), `sc` (small caps), `cap` (capital letters), `vertical-line` (for text that Livingstone writes perpendicularly to the rest of the text on the page), `circled` (for text Livingstone has circled, usually when editing). Multiple values are separated with a semicolon:

`<date when="1871-04-16">16<hi rend="sup;ul">th</hi> April</date>`

C. Additions and Deletions

1. `<add>`

We use the `<add>` element to record additions made to the manuscript by Livingstone himself:

your Lordship that `<add place="below">^</add>` `<add place="above">at last</add>` I have succeeded in

In the majority of cases, Livingstone marks the place of addition with a caret, which is recorded as above. The values we use for the `@place` attribute include: `above`, `below`, `inline`, `marginright`, `marginleft`, `overright`, `overleft`, `subbelow`, `subright`, `subleft`, `supabove`, `supright`, `supleft`, `underright`, `underleft`. (Note: when an addition appears between lines, below or above the line of text to which it, the addition, has been added, the `@place` attribute takes the value of below or above. If the addition appears in the previous or next line or any preceding or subsequent line, the `@place` attribute takes the values of `subbelow`, `subright`, `subleft`, `supabove`, `supright`, or `supleft`. If an addition is oertext or undertext, the `@place` attribute takes the values of `overright`, `overleft`, `underright`, or `underleft`.)

If Livingstone deletes one or more characters by writing over them, we note the addition directly after the deletion, as in the following example:

foray - `The<add place="inline">We</add>` met the horde

2. ``

We use the `` element to record deletions made to the manuscript by Livingstone himself:

putting between the `<del type="cancelled">object` lenses of the object

The `@type` attribute indicates the type of deletion, either `strikethrough` (a single line) or `cancelled` (multiple lines). If Livingstone deletes text by writing over it, we do not use an `@type` attribute.

D. Illegible Text

1. `<unclear>`

We use <unclear> to tag text that can be read with some degree of certainty:

which supports the <unclear cert="high">long grass</unclear> and is

The values we use for the @cert attribute include: low, medium, high.

2. <gap>

We use <gap> to tag text that is wholly illegible due to deletion or some other factor, or missing altogether from the manuscript page due to physical damage:

```
<lb n="18"/><gap reason="damage" extent="6" unit="chars" agent="blotting"></gap> sugar –  
candles
```

The <gap> element never contains text. The values we use for the @reason attribute include: deletion, illegible, and damage. If the @reason value is damage, then an @agent is also provided. The values we use for the @agent attribute include: hole, blotting, fading, stain, overwriting. The value of the @extent attribute is always an approximation.

E. Editorial Interventions

1. <supplied>

We use <supplied> to mark text that has been added to the transcription by the present editors:

went a long way to see <supplied cert="high">a</supplied> canoe but

The values we use for the @cert attribute include: low, medium, high. The type of text we have supplied falls into eight general categories:

a. Illegible text marked with the <gap> element. In some cases, the text may be partly visible, for instance when the top portion of a folio has crumbled away, but the lower half of the first line is still visible (e.g., DLC297b_135-160_006v).

b. A word illegible or otherwise damaged in part or in whole that can be reconstructed from the context with some degree of confidence:

```
only when the rains have supersat<gap reason="damage" extent="2" unit="chars"  
agent="fading"></gap><supplied cert="high">ur</supplied>ated
```

This category also includes one instance where a leaf of the original document has been torn in two and a missing letter on one half of the document can be supplied from the other half of the document (see DLC297b_118_004v and DLC297b_117_003v). In this case we have used the @source attribute to point to the place from which we have taken the missing letter:

```
<lb n="23"/><gap reason="damage" extent="1" unit="chars" agent="hole"></gap><supplied  
cert="high" source="DLC297b_117_003v_TEI.xml">s</supplied>mall insignificant rivers
```

c. A portion of a word or a whole word accidentally omitted by Livingstone that can be restored to the context with some degree of confidence:

ten and twelve <supplied cert="high">degrees</supplied> South Latitude is between

d. A portion of the text missing due to textual damage that can be reconstructed based on Livingstone's usual method of composition (e.g., DLC297b_137-158_007v).

e. Missing punctuation that can be restored with some degree of confidence.

shall return the goods to him<supplied cert="high">"</supplied> - this is

f. A portion of a proper noun that is missing due to manuscript damage or slip of the pen:

that <gap reason="damage" extent="2" unit="chars"
agent="stain"></gap><persName><supplied cert="high">Ha</supplied>ssani</persName> had
played him false

g. A hyphen at the end of a line denoting that a word has been broken up over two lines. Occasionally, Livingstone also includes this type of hyphen at the beginning of a line, but in that case we have not supplied it where it is missing.

h. Text that can be recovered by comparison with the parallel passage in Livingstone's Unyanyembe Journal (1866-72). See DLC297b_151-144_011r for an example.

2. <choice>

We use the <choice> element along with additional, nested tags to indicate when a portion of text may be encoded in more than one way.

a. <abbr><expan>

We use the <abbr> and <expan> tags to provide the expanded version of text that Livingstone has used in abbreviated form in his manuscript:

<choice><abbr>R</abbr><expan>River</expan></choice>

b. <sic><corr>

We use the <sic> and <corr> tags to offer corrections to typographical mistakes in Livingstone's manuscript. Our use of these tags falls into five general areas:

i. Words that Livingstone consistently misspells such as "receive" and "conceive":

<choice><sic>recieve</sic><corr>receive</corr></choice>

ii. Obvious slips of the pen:

<choice><sic>the</sic><corr>then</corr></choice> a mile beyond

iii. Omission of the possessive apostrophe:

The <choice><sic>headmans</sic><corr>headman's</corr></choice> house

iv. Omission of the apostrophe in contractions:

I <choice><sic>dont</sic><corr>don't</corr></choice> trust

v. Accidentally repeated words:

that we <choice><sic>we</sic><corr></corr></choice> have allowed a secret

c. <unclear>

We use multiple <unclear> tags to mark variant readings of the same word, i.e., alternate readings of a word that is unclear:

<choice><unclear>Zouady</unclear><unclear>Zonady</unclear></choice>

d. <orig><reg>

We use the <orig> and <reg> elements to encode fractions:

<choice><orig>½</orig><reg>1/2</reg></choice>

F. Content Tagging

1. People

We tag all personal names (including all titles) using the <persName> element:

Example #1

<persName>Dugumbe</persName>

Example #2

<persName>Lord Palmerston</persName>

2. Tribes and Villagers

We tag African tribes and villagers with the <term> element:

<term type="tribe">Bagenya</term>

The @type attribute values in this case include: tribe, villagers. (Note: Livingstone uses "Manyema" to refer either to the tribe or the region, so our coding of this word always depends on context.)

3. Places and Geographical Entities

We tag all places and geographical entities with the <placeName> element:

```
<placeName><geogName type="river">Nile</geogName></placeName>
```

We also tag all place names and geographical entities when used in an adjectival form:

```
<placeName><settlement type="village">Ujiji</settlement></placeName>an slaves
```

Nested tags, which we always use, include the following:

a. <bloc> to denote a large entity like Africa

b. <region> to denote a region like East Africa or Zanzibar. (Note: Livingstone uses “Manyema” to refer either to the tribe or the region, so our coding of this word always depends on context.)

c. <country> to denote a country like England

d. <settlement> to denote cities, town, villages, and other settlements:

```
<placeName><settlement type="village">Nyangwe</placeName></settlement>
```

The @type attribute values in this case include: city, town, village.

e. <geogName> to denote geographical entities not covered by the above. The @type attribute values in this case include: river, lake, feature. For entities not covered by these values or in cases of uncertainty the @type attribute is not used:

```
<placeName><geogName>Kalahari Desert</geogName></placeName>
```

4. Dates

We tag all dates using the <date> element:

```
<date when="1871-05-20">20th May 1871</date>
```

The value of the @when attribute takes the following format: yyyy-mm-dd. Although Livingstone does not always include the year and/or month when recording dates, we include these in the value when known:

```
<date when="1871-05-31">31<hi rend="sup;ul">st</hi></date>
```

Likewise, we code any other temporal information provided by Livingstone within the <date> element:

Example #1

```
<date when="1871-04-01">Monday 1<hi rend="sup;ul">st</hi> April 1871</date>
```

Example #2

```
<date when="1870-11">November last year</date>
```

We also code dates from the Arab calendar when Livingstone provides them by using the @calendar and the value of Muslim:

```
<date calendar="Muslim">Arab fifth month</date>
```

Finally, in a few instances, Livingstone begins diary entries with the time rather than the date. In these cases we also use the <date> element rather than the <time> element, but give the @when attribute a value that includes the time:

```
<date when="1871-07-16T13:00:00">1 PM</date>
```

5. Medical References

We code medical references using the <term> element:

Example #1

```
<term type="medicament">Quinine</term>
```

Example #2

```
<term type="ailment">dysentery</term>
```

The @type attribute values in this case include: ailment, medicament, bodypart, bodyprocess.

6. Foreign Words and Phrases

We code foreign words and phrases, except for those in Hebrew or Arabic, using the <foreign> element:

```
<foreign xml:lang="und">chitoka</foreign>
```

We use the @xml:lang attribute to identify foreign languages. The following values are used in our transcription of Livingstone's diary: en (English), swh (Swahili), la (Latin). If the African language quoted by Livingstone is unknown, we use: und (Undetermined African Language). Any languages used by Livingstone are also recorded in the <profileDesc> in the document header (see above). For Hebrew and Arabic, see the next section below.

7. Drawings, Maps, Calculations, Seals, Squiggles, Foreign Text (Hebrew/Arabic)

Livingstone's manuscript contains a number of drawings, maps, calculations, seals, and squiggles to correct ink flow. We never encode any part of these, but rather note their presence and describe them using the <figure> and <figDesc> elements:

```
<figure><figDesc>Livingstone records a series of calculations on the left-hand side of the page.</figDesc></figure>
```

We also dynamically insert the @facts attribute to help link the <figure> to the corresponding area of the relevant spectral images.

In three instances (DLC297c_111-124_005v, DLC297c_121-114_007r, NLS10703_002_038v), we also encode bits of Hebrew, Arabic, Nagari text, some of which may not be in Livingstone's hand, using the <figure> and <figDesc> elements.

The presence of a seal is also noted in the <sealDesc> element in the document header (see above).

G. Special Characters

We represent a few commonly used special characters with the hexadecimal character entity. These characters include:

Character	Entity Reference
"	"
&	&

All other special characters have been inserted as unicode characters directly through oXygen:

Edit > Insert from character map...

For our coding of fractions see the section on <choice>, above.