

## XML TEI P5 Tagging Guidelines for Livingstone Online

*A Cheat Sheet*, version 12-11-2013

Created by Adrian Wisnicki  
Updated by Heather F. Ball

Coding of the following elements is covered in this document (click on an entry's page number to skip down to it):

0. Marking Up Text – Basic Rules .....	2
1. File Naming.....	2
2. the Header (part 1): Adding Bibliographical Data .....	3
3. the Header (part 2): Adding Your Name .....	3
4. a New Page .....	3
5. the Beginning of a Manuscript Line .....	4
6. the Beginning of a New Paragraph.....	4
7. a Date .....	4
8. Text that is perpendicular to the rest of the text.....	5
9. Textual Formatting.....	5
10. A Word Broken Up Over Two Lines.....	6
11. More Than One Space between Words .....	6
12. Text that is Not Clear.....	6
13. a Missing Portion of the Manuscript Page or Text that is Illegible .....	7
14. a Portion of the Text that You Have Supplied.....	7
15. Text Added to or Deleted from the Original Manuscript.....	7
16. Grammatical, Typographical, and Other Mistakes Made by the Author .....	7
17. Abbreviations .....	8
18. Fractions .....	8
19. Quotation Marks, the Ampersand, and Other Special Characters.....	8
20. a Person's Name .....	9
21. Geographical Elements .....	9
22. Nationalities .....	9
23. an African Tribe .....	9
24. Words in a foreign language.....	10
25. a Medical Reference.....	10
26. Figures, Drawings, Calculations, and anything else unusual .....	10
27. Notes on Tagging Order .....	10

tags are in blue  
 attributes are in red  
 values are in green

(Note: the colors are provided only to assist you in distinguishing between tags, attributes, and values. You do not need to change font colors in your transcriptions as you are transcribing).

## 0. Marking Up Text – Basic Rules

1. When text is marked up, a tag is placed at the beginning of the text as well as at the end, per the following format:

*Original text:* the man turned out to be Livingstone

*Marked up:* the man turned out to be `<persName>Livingstone</persName>`

Note that the first tag denotes the beginning of the marked up text, while the second tag begins with "/" and denotes the end of the marked up text.

2. If the opening tag includes an attribute, the attribute is NOT repeated in the closing tag:

*Original text:* we were greeted by the Nyanwezi

*Marked up:* we were greeted by the `<term type="tribe">Nyanwezi</term>`

3. If there are two or more tags on one bit of text, the closing tags are placed in reverse order:

*Original text:* we saw the Lualaba

*Marked up:* we saw the `<placeName><geogName>Lualaba</geogName></placeName>`

4. An attribute is always accompanied by a value, for instance `<term type="tribe">`

5. If an opening tag ends with a "/", then it is an open tag and need not be used again at the end of whatever is being marked up.

*Original text:* we saw the Lualaba

*Marked up:* `<lb/>`we saw the Lualaba

6. Any problems related to the transcription should be addressed directly in the transcription at the points where they occur using the "comment out" feature.

*Original text:* He called me over.

*Marked up:* He called me over. `<!-- Adrian, the text is very illegible here and I've done the best I could. However, this might also be transcribed as "He willed in on me." What should I do? Debbie 19/2/11-->`

In other words the commented out bit begins with "`<!--`" followed by a space and ends with "`-->`" not preceded by a space. Always add your name and the date of your comment, with dd-mm-yy format.

## 1. File Naming

When you begin transcribing, use the date of the letter to find the corresponding entry in the Livingstone Online Catalogue from this specific page:

<http://www.livingstoneonline.ucl.ac.uk/catalogue/browse.php?cat=date>

The first column in the tables that you access through this page will give you the C&C number. This is the number from which you will create the file name. So, for instance, if the letter number is 864, the file name will be LETT864

## 2. the Header (part 1): Adding Bibliographical Data

Once you have found the corresponding entry for the letter you're transcribing in the Livingstone Online tables, you are also read to format the first part of the header as the same entry has all the information you'll need.

Take a look at the header text below. The data in red is what you'll need to input each time you start transcribing a new letter in XML. This red data is taken from Livingstone Online. As you'll see, you'll need to add the recipient twice and the letter number twice, everything else just once.

```
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:id="LETT1422">

  <teiHeader><fileDesc>
    <titleStmt>
      <title>Letter to George H. Richards</title>
      <author xml:id="DL">David Livingstone</author>
    </titleStmt>
    <publicationStmt>
      <authority>Livingstone Online</authority>
      <pubPlace>London, UK</pubPlace>
      <publisher>University College London</publisher>
      <date>2012</date>
    </publicationStmt>
    <sourceDesc>
      <bibl xml:id="shelfmark">Bodleian Library of Commonwealth and African Studies at Rhodes House, MSS.Afr.s.16/1, fol. 1-6</bibl><!-- The foregoing is the name of the repository, followed by the shelf mark -->
      <bibl xml:id="ccnumber">1422</bibl>
      <bibl xml:id="recipient">George H. Richards</bibl>
      <bibl xml:id="date">10 February 1864</bibl>
      <bibl xml:id="where">Shupanga</bibl><!-- Place where the letter was written. -->
    </sourceDesc>
  </fileDesc>
```

---

## 3. the Header (part 2): Adding Your Name

The second part of the header contains entries that look like the following:

```
<change>
  <date>2011-11-29</date>
  <name xml:id="AW">Adrian Wisnicki</name> initial transcription and XML encoding.
</change>
```

You'll need to complete the corresponding entry (usually the first one if you're making the initial transcription). You need to change the date to the correct YYYY-MM-DD entry. You need to change `xml:id="AW"` to match your initials and, of course, you need to put your name in the right place.

---

## 4. a New Page

After the initial `<text><body><div>` tags (already in the diary template document), place the following tag at the beginning of each new page/image that you transcribe

```
<pb n="13r"/>
```

Note that this tag ends with a "/" and that `n="13r"` records that this is folio 13 and the "recto". If you were transcribing the "verso" you would use "v".

## 5. the Beginning of a Manuscript Line

Whenever you come to the beginning of a manuscript line, use `<lb n="1"/>`. The number value is the number of the line and should increase with each successive line. No corresponding tag is needed at the end of the line.

e.g. Original text: As we traveled  
I came  
cross two men.

Marked-up: `<lb n="1"/>`As we traveled  
`<lb n="2"/>`I came  
`<lb n="3"/>`cross two men.

## 6. the Beginning of a New Paragraph

When you start a new paragraph, use the following tag: `<p>`

The following attributes are also available (though not usually necessary) for this tag, though their use is a bit of a judgment call regarding the placement of the paragraph on the page:

`rend="center"` or `rend="right"`  
(note: default is `rend="left"` so it is not necessary to code this)

e.g. `<p rend="right">`

Likewise, at the end of the paragraph remember to use the following tag: `</p>`

The `</p>` tag will also always be used when you come to the end of a paragraph.

Note that the text that opens and ends a letter will usually be tagged as in the following examples:

### letter opening:

```
<p rend="right">Hadley Green</p>
<p rend="right"><date when="1857-08-22">22 August</p>
<p rend="right">1857</date></p>
```

### letter closing:

```
<p rend="center">I am & </p>
<p rend="center"><persName>David Livingstone</persName></p>
```

It is somewhat subjective whether text is "center" or "right," so use your best judgment.

For the `<date>` and `<persName>` tags used in the examples above, see below.

## 7. a Date

For marking up dates use the following tag and attribute

```
<date when="1871-02-05">
```

The value of the attribute should always be given in this format, namely four-digit year, two-digit month, two-digit day. If day is not given, then omit this in the value.

e.g. we visited last year on `<date when="1994-06-16">June 16, 1994</date>`

e.g. we visited in `<date when="1994-06">June 1994</date>`

However, if the day (and month) are given, but not the year, yet you know the year provide it. For instance, if Livingstone begins a diary entry with "16th" but you know from previous entries that this actually means 16 June 1871, then code it as follows:

e.g. `<date when="1871-06-16">16th</date>`

e.g. `<date when="1871-06-16">16th June</date>`

## 8. Text that is perpendicular to the rest of the text

There are three possible situations here, so here's how to handle them:

1. The page text is horizontal, then it clearly \*continues\* to vertical text in the margin. Use `<cb/>` in either one of the two following ways (line numbering continues in the normal way):

a. If the vertical text in the margin is a new paragraph put `</p>` at the end of the horizontal text, then start the vertical text with `<cb/><p>` and when you finish the vertical text close it with `</p>`

b. If the vertical text in the margin continues the horizontal text without a paragraph break, do not end the horizontal text with `</p>`. Rather start the vertical text with `<cb/>` and when you finish the vertical text close it with `</p>`.

2. The page text is horizontal, but ends at the bottom of the page. You believe that the vertical text continues from the horizontal text, but it seems to be a new "section" (how you define section is a bit subjective). Then close the horizontal text with `</p></div>`. Begin the vertical text with `<div><p>`, then close the vertical text with `</p></div>`. In other words, you're breaking this up into two sections that you believe are continuous. Line numbering is continuous.

3. The page text is horizontal and may or may not end at the bottom of the page. The vertical text has no clear relationship to the horizontal text. End the page text with `</p>` (whether or not it continues to the next page), then begin the vertical text with `<ab>` and close the vertical text with `</ab>`. Then on the next page, whether or not it's a new paragraph, begin with `<p>`, etc. Restart the line numbering in this new section.

**\*\*Please note:** When dealing with `<cb>` or `<ab>`, it is possible to use the "rend" element within the tag:

`<ab rend="vertical-line"> </ab>`

## 9. Textual Formatting

If the original text is formatted in some manner such as underlining, superscript, etc., use the `<hi>` tag plus the appropriate attribute from the list that follows:

`rend="italic"`

`rend="bold"`

`rend="cap"` for marking capitalized script

`rend="sc"` for marking small caps script

`rend="sup"` for marking superscript

`rend="sub"` for marking subscript

`rend="ul"` for marking underline

`rend="overbar"` for marking overline

e.g. Original text: I saw the man  
 Marked up: I saw `<hi rend="ul">the man</hi>`

If there is more than one type of formatting, for instance both underline and superscript, use the `rend=` attribute only once, but provide two values separated by a semi-colon:

e.g. Original text: He was the N<sup>umber</sup> 1 man  
 Marked up: He was the N`<hi rend="ul;sup">umber</hi>` 1 man

## 10. A Word Broken Up Over Two Lines

Use the `<w>` element to mark words that Livingstone breaks up over two lines. You 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="27"/>by which the irresponsible <w>con-
<lb n="28" break="no"/>-clave</w> brought the Indian command
```

## 11. More Than One Space between Words

If you are coding text where the author has placed more than one space between words, do not insert additional space in your transcription. Rather insert only one space plus the following tag:

```
<space dim="horizontal" extent="4" unit="chars"/>
```

The value for the `extent=` attribute will of course be a bit of a judgment call. The value `"chars"` stands for "characters"; you can also use `"words"`.

## 12. Text that is Not Clear

Use `<unclear>` to mark up any text that is not clear; furthermore, use `cert="low"` or `cert="medium"` or `cert="high"` to denote the degree of confidence in the transcription you have made:

e.g. he said he would `<unclear cert="low">say</unclear>` me tomorrow

---

### 13. a Missing Portion of the Manuscript Page or Text that is Illegible

We use `<gap>` to tag text that is wholly illegible due to deletion or some other factor, or missing all together from the manuscript page due to physical damage. The `<gap>` element never contains text:

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

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:

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

The value of the `@extent` attribute is always an approximation.

---

### 14. a Portion of the Text that You Have Supplied

If a portion of the original text is missing for whatever reason and you supply it, use `<supplied>` to mark up this addition; furthermore, use `cert="low"` or `cert="medium"` or `cert="high"` to denote the degree of confidence in the addition you have made:

e.g. Original text: he saw man  
Marked up: he saw `<supplied cert="high">the</supplied>` man

Note: The `<supplied>` tag is never nested in the `<gap>` tag.

e.g. he saw `<gap reason="damage" extent="3" unit="letters"></gap><supplied cert="high">the</supplied>` man

---

### 15. Text Added to or Deleted from the Original Manuscript

If a portion of the text has been added by the author or someone else other than you (see above), use `<add>` to mark up that text. Possible attributes and values include: `place="inline"` or `place="above"` or `place="below"` or `place="marginleft"` or `place="marginright"` or `place="margintop"` or `place="marginbottom"`

e.g. Original text: he saw us [in this case it's clear in the ms. that "saw" was added later]  
Marked up: he`<add type="inline">saw</add>` us

If a portion of the text has been deleted by the author or someone else, use `<del>` to mark up that text. Possible attributes and values include: `type="striketrough"` (deleted with a single line) or `type="cancelled"` (deleted with multiple lines)

e.g. Original text: he saw us  
Marked up: `<del type="striketrough">he saw us</del>`

In both the cases of added or deleted text, if you believe that the work has been done by someone other than Livingstone but cannot identify who that person is, use the `hand="U"` attribute:

e.g. he`<add type="inline" hand="U">saw</add>` us

---

### 16. Grammatical, Typographical, and Other Mistakes Made by the Author

When you encounter some sort of textual mistake made by the author, mark it with the `<choice>` tag PLUS both the `<sic>` tag to mark the erroneous version PLUS the `<corr>` tag to indicate the correct version that you provide, as follows:

e.g. Original text: I recieved your letter  
 Marked up: I `<choice><sic>recieved</sic><corr>received</corr></choice>` your letter

If there is an uncertainty as to the correction that you provide, use `cert="low"` or `cert="medium"` or `cert="high"` to denote the degree of confidence in the addition you have made:

e.g. `<choice><sic>recieved</sic><corr cert="high">received</corr></choice>`

**\*\*Please note:** Livingstone tends to leave off his apostrophes when it comes to his contractions. The proper way to code this is:

`<choice><sic>dont</sic><corr>don't</corr></choice>`

---

## 17. Abbreviations

When you encounter an abbreviation made by the author, mark it with the `<choice>` tag PLUS both the `<abbr>` tag to mark the abbreviation PLUS the `<expan>` tag to indicate the expansion that you provide, as follows:

e.g. Original text: my no. 1 friend  
 Marked up: my `<choice><abbr>no.</abbr><expan>number</expan></choice>` 1 friend

If there is an uncertainty as to the expansion that you provide, use `cert="low"` or `cert="medium"` or `cert="high"`.

e.g. `<choice><abbr>no.</abbr><expan cert="high">number</expan></choice>`

**\*\*Please note:** any prefix pertaining to a personal name, i.e. Dr, Mr, Miss, Rev, need not be expanded.

---

## 18. Fractions

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

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

---

## 19. Quotation Marks, the Ampersand, and Other Special Characters

When you transcribe quotation marks, use `&quot;`; instead of actual quotation marks:

e.g. Original text: the man said, "I have spoken."  
 Marked up: the man said, `&quot;`I have spoken.`&quot;`;

When you transcribe an ampersand, use `&amp;`; instead of an actual ampersand:

e.g. Original text: I saw John & Jim  
 Marked up: I saw John `&amp;`; Jim

If you come across any other unusual characters that are not found within the "Edit—Insert from Character Map" option from the editing toolbar, use the comment out feature and I'll put them in.

`<!-- Adrian, there's a weird character you need to insert here. -->`



---

## 20. a Person's Name

When marking up a person's name, use the following tag: `<persName>`

e.g. I saw `<persName>`David Livingstone`</persName>`

However, note the following example:

e.g. I saw `<persName>`David Livingstone`</persName>`'s

**\*\*Please note:** any prefix pertaining to a personal name, i.e. Dr, Mr, Miss, Rev, should be included within the `<persName>` tag, and does NOT need to be expanded.

---

## 21. Geographical Elements

When you mark up a geographical element, use the `<placeName>` tag PLUS one of the following tags:

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

`<region>` to denote a region like East Africa

`<country>` to denote a country like England

`<settlement>` to denote cities, town, villages, and other settlements. Possible attributes include `type="city"` or `type="town"` or `type="village"`

e.g.: `<placeName><settlement type="village">`Nyangwe`</settlement></placeName>`

`<geogName>` to denote a geographical entity not covered by the above. Possible attributes include `type="feature"` or `type="lake"` or `type="river"`. An attribute may not always be necessary.

e.g.: `<placeName><geogName type="river">`Nile`</geogName></placeName>`

**\*\*Please note:** When coding "Cape" and "Interior", both should be coded with the `<region>` tag.

---

## 22. Nationalities

When dealing with nationalities, use the `<term type="nationality">`

e.g. the `<term type="nationality">`Portuguese`</term>` settlements

**\*\*Please note:** Boers is tagged as a nationality (not a tribe).

---

## 23. an African Tribe

When marking up an African tribal name, use `<term type="tribe">`

e.g. I saw the `<term type="tribe">`Nyamwezi`</term>`

**\*\*Please note:** Boers is not tagged as a tribe, but as a nationality (see previous section).

---

## 24. Words in a foreign language

When dealing with foreign words, there are three steps to complete:

1. Find the abbreviation for the foreign language at the following webpage:  
<https://www.iana.org/assignments/language-subtag-registry/language-subtag-registry>
2. In the header, add in a tag for the foreign language under the English tag as follows:

```
<langUsage>
  <language ident="en">English</language>
</langUsage>
<langUsage>
  <language ident="la">Latin</language>
</langUsage>
```

3. Within the text, when the language comes up, you code the word/phrase as follows:

```
<foreign xml:lang="la">et id genus omne</foreign>
```

**\*\*Please note:** If you are unsure of the language/dialect, but know that it is not English, simply use the "Undetermined" tag, which is abbreviated as "und".

## 25. a Medical Reference

When marking up a medical reference, use `<term>` plus the appropriate attribute and value, such as `type="ailment"` or `type="prescription"` or `type="medicament"` or `type="bodypart"` or `type="bodyprocess"`

e.g. I had `<term type="ailment">malaria</term>` for a month.

## 26. Figures, Drawings, Calculations, and anything else unusual

If you come across any figures, drawings, calculations, or anything else unusual, alert me using the comment out feature.

```
<!-- Adrian, there's a drawing here. -->
```

```
<!-- Adrian, take a look at the manuscript. There's something odd here. -->
```

## 27. Notes on Tagging Order

A) Tags signifying textual formatting always surround tags signifying some kind of informational tagging.

e.g. Original text: the Lualaba  
Marked up: the `<hi rend="ul"><placeName><geogName type="river">Lualaba</geogName></placeName></hi>`

However, take note of the following exceptions:

e.g. Original text: Syd bin Majid  
Marked up: `<persName><choice><sic>Syd</sic><corr>Syde</corr></choice> bin Majid</persName>`

e.g. Original text: 19<sup>th</sup> May 1871  
Marked up: `<date when="1871-05-19">19<hi rend="sup;ul">th</hi> May 1871</date>`

In other words, if the textual formatting applies to only part of the text coded by the informational tagging, then the textual formatting can be nested within the informational formatting.