TEI Encoding Standards

Seward Family Papers Project

Last updated: November 15, 2017

0.1 INTRODUCTION

TEI encoding is an essential step in the Seward Family Papers Project, which enables the transcribed and annotated letters to become part of a collection of digital texts. This encoding process allows files to be tagged with attributes such as place names, people, and works of literature. TEI files will be uploaded to the project's website where visitors can search and read items in the collection.

0.2 HOW TO READ THIS DOCUMENT

These standards follow an existing set of guidelines called the Text Encoding Initiative (TEI). They describe a set of procedural tasks used in the Seward Project TEI workflow to reveal some of the unique qualities of the manuscripts. These instructions have been adapted to fit the unique goals of the Seward Family Papers Project. Each step will help encoders engage the elements encountered when encoding a Microsoft Word document into TEI. Please refer to these guidelines as you translate your TEI document into a finished, marked-up document.

0.3 RESOURCES AND FURTHER READING

In addition to these guidelines, further information is available through the community's TEI Guidelines, here:

http://www.tei-c.org/Guidelines/

You can also reference the TEI dictionary utilized within our project here:

http://seward.lib.rochester.edu/tei/tei-dictionary.html

If you have any questions, please contact the Project's TEI Manager (Camden Burd), or the Digital Humanities Librarian (Joe Easterly)

1.1 GETTING STARTED

1. Open oXygen found in Applications.
2. Open an assigned file from the local folder on your computer or through the "Data Source Explorer" window in the oXygen application.

2.1 DOCUMENT TITLE SWAP

On Line 11, replace "Seward Family Digital Archive" with a written on description of the letter. For example if the letter is filename l_18220707LMW_FMS1v3 then write:

```
<title>Letter from Lazette Miller Worden to Frances Miller Seward, July 7,
1822</title>
```

2.2 STUDENT EDITOR AND TRANSCRIBER

First, in the title statement, add the transcriber and student editor.

Wherever you see green `<!--comment-->` tags inside TEI files you should find directions and tasks for follow-up. In the example below, you are being asked to type "spp:saz" as the value in the `@ref` attribute of the element `<persName/>`. Following the `"spp`, insert a colon and include the three-letter code of the individual who originally transcribed the document. This information is found at the bottom of the document in brackets.

```
<title type="main">Seward Family Papers Project</title>
    <respStmt>
        <resp>Transcriber</resp>
        <persName ref="spp:saz"/>
        <!--Enter three-letter transcriber ID here "ref="spp:zzz" -->
        <!--Three-letter transcriber IDs stored in the authorities folder
        inside the repository, in the file staff.xml"-->
    </respStmt>
```

Be sure to delete the green comments after completing this step.

Follow the same steps to indicate who was the student editor of the original transcription. For instance, if Bobby was Samantha's student editor, we include his initials, "spp:bdg" in this value space.

```
<respStmt>
    <resp>Student Editor</resp>
    <persName ref="spp:bdg"/>
    <!--Enter three-letter student editor ID here 'ref="spp:zzz"'-->
</respStmt>
```

Be sure to delete the green comments after completing this step.

2.3 MANUSCRIPT DESCRIPTION

Within the manuscript description we have space to add the date in which we believe the original manuscript was written. During transcription we identified this date and placed it within the header. We will enter the date using the format: yyyy-mm-dd. They will input the date according to ISO-8601 standards *(see http://en.wikipedia.org/wiki/ISO_8601).*

```
<origin>
    <date when="1861-01-01"/>
    <!-- Include date in yyyy-mm-dd-->
</origin>
```

Be sure to delete the green comments after completing this step.

2.4 PROFILE DESCRIPTION

Here we identify the sender and intended recipient(s) of a letter. We use unique identifiers (UID)s in order to "link" records for people and places to the text in the letters. In the example below, Frances sent a letter from Florida, NY, to Henry in Auburn, NY. We will also enter the date the letter was sent (this date will match the date used in step 2.3)

In the first `<correspAction/>` space, within the `<persName/>` element, add the UID for Frances Seward in the value of the attribute "ref=". Following that, you will include the place UID for Florida, NY.

```xml
<correspDesc>
    <correspAction type="sent">
        <persName ref="psn:SEWf_827"/>
        <placeName ref="pla:FLO_1351"/>
        <date type="creation" when="1830-01-01"/>
    </correspAction>
    <correspAction type="received">
        <persName ref="psn:SEWw_846"/>
        <placeName ref="pla:AUB_1319"/>
    </correspAction>
</correspDesc>
```

If the person or place are unknown then "unknown" will become the value of the ref= attribute. As in, `<placeName ref="pla:unknown"/>`.

Be sure to add the corresponding prefix to the attribute, e.g., psn: or pla:.

Be sure to delete the green comments after completing this step.

XML: IDs of people and places can be found at the following web pages:

http://seward.lib.rochester.edu/tei/persons.xml

http://seward.lib.rochester.edu/tei/places.xml

2.5 REVISION DESCRIPTION

The revision description is where members of the Seward Project log changes that have been made to the digital versions of the letter. Dates for `<change type="transcription">` should reflect the turn-in date for your v3 letters. Dates for `<change type="revision">` the date that you completed and submitted your TEI-encoded letter. In the example below, Samantha submitted a letter at the end of the spring 2014 semester and Mike converted it to TEI in September.

```xml
<revisionDesc>
    <listChange>
        <change type="transcription" who="spp:saz"/>
        <change type="revision" when="2014-09-01" who="spp:mhr"/>
    </listChange>
</revisionDesc>
```

Be sure to delete the green comments after completing this step.

3.1 PAGE NUMBERS

During the transformation from transcription into an XML documents, page breaks are marked by the `<pb/>` tag. We then add a new element to specify the page our code correlates to on the original manuscript. For example to indicate that we are the first page of the letter or document:

```xml
<pb n="1"/>
```

3.2 EDITORIAL NOTES

There are times when we want viewers to be aware of editorial changes we have made in regards to the document. For instance, if we change the date of letter despite what is written on the original manuscript we want to share our editorial decision with readers. Editorial notes will be visible on the website and should contain the note from the original transciber. Editorial notes should not be confused with Comments (See 4.3)

For example, we tag date changes in our editorial notes:

```
Friday October 8th 1831 <note type="editorial">Frances was incorrect, Friday was the 9th</note>
```

Editorial notes should also be used to designate when there are attached printed materials or pictures included within letters. Because we create a separate page for these items, we use editorial notes to draw readers' attention to the fact this material is not transcibed, but simply attached. You can use the editorial note to add context or understanding to the attached materials.

Other editorial notes may apply based on the context of the note and should be used at the discretion of the TEI manager.

4.1 INDIVIDUAL PERSON

After you submitted your v3 letter in Word it was transformed into an XML document. Your footnotes were inserted into the text as `<note>` tags. The primary task in tagging these letters is to wrap the information from those original notes around the people and places to which these notes apply.

For example, Mr. Cummings is tagged in the sentence below. Thefore, we will tag the name by surrounding Mr. Cummings with <persName>, and a @ref should be added from the TEI Persons link (http://seward.lib.rochester.edu/tei/persons.xml)

```
I have just returned from a walk to <persName ref="psn:CUMc_4029">Mr Cummings</persName> the good man himself came home with me.
```

4.2 INDIVIDUAL PERSON (LESS THAN CERTAIN)

For an individual whose identity is either unknown or whose identity is less than 100% certain, we have a few certainty values at our disposal.

Unknown

```
We have just returned from a visit to <persName ref="psn:unknown">Deacon Smith's </persName>
```

Possibly / Maybe

```
<persName ref="psn:HYDp_4953" cert="low">Mrs Hydes</persName>
```

Probably / Likely

```
<persName ref="psn:EVAm_4614" cert="medium">Cousin Mary</persName>
```

Note: though `@cert` allows for a `"high"` value, we don't use it. If we are highly certain it means that the `@xml:id` stands without an attribute value.

4.3 COMMENTS

Because this project has grown through the years it is possible earlier versions of letters that we encode have not gone through the same rigorous editorial process as more recent years. Some individuals found in letters are annotated incorrectly, vaguely, or the individuals have not been added to the Drupal database. If the note has information that allows the TEI encoder to add the individual to Drupal, please do so. However, if the note provides little information the TEI encoder should tag the individual as "unknown" but leave a note alongside the tag for future clean-up.

As a TEI encoder, if you see individuals tagged with any ambiguity, we would rather tag individuals as "unknown" and leave an editorial comment to the side with the original note to ensure that we have the most accurate tags we can provide with the information given. This way, future editors can take the time to parse out some of the ambiguities in identification. At the same time we can best identify, to our ability, the person or persons mentioned in a letter.

For example, if you come across an annotated individual, William, and the note contains vague historical evidence such as:

```
I express my joy, the bliss of embracing of my brother<note>Could be Charles
Covington Tuthill or Edward Ely Tuthill </note>.
```

It is clear the transcriber struggled to find definitive evidence to claim either person as the brother in the letter. If there is little evidence to support either individual then we would like to mark the person as "unknown" but keep the note as an editorial comment for any editors looking to clean-up the project in the future. Such a note would look like:

```
I express my joy, the bliss of embracing of my <persName
ref="psn:unknown">brother</persName><note type="editorial"> Could be <persName
ref="psn:TUTc_4661">Charles Covington Tuthill</persName> or <persName
ref="psn:TUTe_4663">Edward Ely Tuthill</persName></note>.
```

4.4 TAGGING GROUPS

In cases where a word or phrase in the text refers to multiple persons or places, tag these references with `<name>`, add `@type` `"group"`, and add a `@ref` attribute for the identifiers, separating each with a space.

```
The <name type="group" ref="psn:HOFj_4200 psn:HOFp_4201">Hoffmans</name> dined with
both <name type="group" ref="psn:unknown psn:unknown">Hockenglockenses</name>.
```

As with the footnotes used in transcriptions, we only tag the first instance of a name in a letter. All subsequent instances of a person's name are not tagged.

Note that each UID receives its own psn: or pla: prefix.

4.5 PLACES

Places follows a practice similar to persons.

Known

```
<placeName ref="pla:LUD_4971">Ludlowville</placeName>
```

Unknown

```
I will not undertake to spell the name of the<lb/>
<placeName ref="pla:unknown">place where he resides</placeName>
```

4.6 BIBLIOGRAPHIC WORKS AND OTHER PRINTED MATERIALS

For printed material mentioned use `<title>` around the title of the book, journal, or paper. If we have a Drupal ID for the book or printed material we also include XML ID similarly to People and Places.

```
My eyes are exhausted from <persName ref="psn:GIBe_410">Gibbon's</persName> <title
ref="bib:DECL_14407">Decline and Fall</title>, the print is…
```

4.7 HANDSHIFT

In TEI, changes in authorial hand are marked by the element `<handShift>`. We ask that when known, you include the `@xml:id` of the scribe. If the scribe's identity is unknown or less than certain, you can use `scribe="psn:unknown"` or add a `@cert` attribute.

It is important to remember to **switch back** using the `<handShift>` tag if the author's hand changes more than once.

```
<handShift scribe="psn:SEWf_826"/>Dearest Father, I've asked mother to allow me to
```

```
<handShift scribe="psn:SEWf_829" cert="medium"/>Dearest Cousin, I've asked Mary to
allow me...
```

```
<handShift scribe="psn:unknown"/>two potatoes and a firkin of butter
```

4.8 STAMP

We note the postmark of a letter if one exists. These are often found on the last page of a letter.

```
<stamp type="postmark">12 April 1838</stamp>
```

5.1 SUPPLIED

If a transcriber has placed square brackets around text or punctuation it means that he/she has "supplied" those characters. We place these supplied pieces in the appropriate XML tags. For example, in the following case the "so" isn't readable to one who views the manuscript, as in the case of the wax stamps removing paper or if the transcriber is using context to supply a word.

```
intimated<lb/> that you was in love with her once is it <supplied>so</supplied> you
must have been<lb/> very young as she has been married sixteen years.
```

Wax-Seal Related Tags

Sometimes the wax-seal has ripped part of a letter from the manuscript and only part of the word is visible. Encode such instances as:

Today I saw a [moose] [1] walking through the woods.

or

Today I saw a [m][2]oose walking through the woods.

The portion that is visible on wax-seal will be tagged as:

```
<supplied reason="wax-seal" n="orphan">m</supplied>oose
```

This tag indicates that a hole (orphan element) is the reasoning behind the supplied text. This describes the physical structure through the encoding process.

Supplied Based on Contextual Clues

Sometimes words or phrases can be assumed based on the context of the letter. If we can reasonably "guess" a word, we demonstrate those supplied words or phrases in coding. For example:

Today, I saw a [moose] [3] walking through the woods. The moose was large…

We demonstrate the supplied text as:

```
Today, I saw a <supplied reason="hole">moose</supplied> walking through the
woods. The moose was large…
```

5.2 ALTERNATE/MISSPELLING

Misspelling

Transcribers also mark misspelled words with a unique tag. Follow the format in the example below:

Today, I saw a muse [moose][4] walking through the woods. The moose was large…

We would tag this as:

```
Today, I saw a <choice><sic>muse</sic><corr>moose</corr></choice> walking
through the woods. The moose was large...
```

5.3 HIGHLIGHTED TEXT

Superscript

Superscript will be tagged using the `<hi>.` For example:

May 19th

This function should occur automatically, however if not, the TEI would appear as:

```
May 19<hi rend="superscript">th</hi>
```

Underlined Text

Some passages may occasionally be underlined—these generally should be encoded using `<hi>`, tag. For example:

---

[1] located on wax seal

[2] located on wax seal

[3] hole, supplied

[4] misspelling

when you first saw him he might have been <u>drunk</u> and forgotten his name.

This would appear in the TEI document as:

```
when you first saw him he might have been <hi rend="underline">drunk</hi> and
forgotten his name.
```

If there are quotations underlined, we tag that differently as seen below:

He in all probability frisked when he was introduced to two such "<u>interesting and intelligent</u>" young ladies.

```
<hi rend="underline quote_marks">interesting and intelligent</hi>
```

Inserted and Cancelled Text

`<add>` indicates text that was added by an author into a passage after the passage was written. It may be indicative of an afterthought or a revision by the author. It does not necessarily refer to text that was interlined simply because the author ran out of room.

For example:

"We went ^to^ Canandaigua yesterday."

In TEI this would appear as:

```
We went <add>to</add> Canandaigua yesterday.
```

It is important to note that all cancelled text appeara as `<hi rend="strikethrough">` when it is converted into XML documents. We replace `<hi rend="strikethrough">` with `<del>` tag. For example a sentence:

"We went to Canandaigua ~~yesterday~~ two days ago."

After conversion will appear as:

```
We went to Canandaigua <hi rend="strikethrough">yesterday</hi> two days ago.
```

After will correct the conversion process so that the sentence appears as:

```
We went to Canandaigua <del>yesterday</del> two days ago.
```

At times there will be words crossed out and replaced with other words or phrases. We use a different tag to indicate this occurrence. For example:

"any politeness ~~that he was~~ ^and so^ much fatigued."

Appears as:

```
any politeness <del>that he was</del><add>and so</add> much fatigued
```

6.1 GAPS

In instances where a transcriber has placed "[hole]" or "[illegible]" you will add the tag "`<gap>`". TEI-C defines "gap" as "a point where material has been omitted in a transcription." In our project, such cases occur where the transcribers cannot provide a transcription due to difficult handwriting, or invisible text due to holes in the page, wax seals, or other obstructions.

6.2 ILLEGIBLE

Sometimes during the transcription process a word or phrase can be difficult to read. In this case, transcribers have identified it as "illegible." This will receive a <gap> tag as seen below.

Today, I saw a [illegible][5] walking through the woods. The moose was large…

    This will tagged as:

```
Today, I saw a <gap/> walking through the woods. The moose was large…
```

6.3 GAPS CAUSED BY DAMAGE TO THE DOCUMENT (HOLE)

There are instances when damage (such as tears) to the physical document has made it difficult or impossible to transcribe passages in a document. These gaps often cover more than one line of text. In TEI we ascribe a value to the space in the document. This would be an instance when the note indicates that there is a "hole." For example:

    The same time I had the account of our great [hole][6]and affliction I received a letter from John Thompson...

Since the hole in the transcription was caused by physical damage to the document we tag that space as seen below.

```
The same time I had the account of our great <gap reason= "hole"/> and
affliction I received a letter from John Thompson...
```

6.4 MARGINALIA

Some letters have text written in the margins of letters and documents. Sometimes when the file is converted from Word into an XML document, those appear as standalone `<pb/>` It is our job to move that text into the correct page in the script. For example text written in the margins of page two will appear with the section of the code on page two of the script right before the `<pb/>` tag begins for page 3. We then place the text within a tag as seen below:

```
<seg type="marginalia" subtype="top" n="1">A free form note.</seg>
```

The `subtype="top"` tag indicates where the marginalia begins on the document. We can choose either "`top`," "`bottom`," "`left`," or "`right`." The `n="1"` distinguishes the order in which the marginalia appears in that specific subtype. For marginalia written at the top of page one:

```
<seg type="marginalia" subtype="top" n="1">A free form note.</seg>
```

For second marginalia written on the top of page one. While this is rare, it sometimes occurs with a hand shift or a particularly full letter:

```
<seg type="marginalia" subtype="top" n="2">Another free form note on page
2.</seg>
```

6.5 TABLES

Some letters have numbers or table-like structures written into the letters and documents. In these instances, we simply wrap the contained letters and words in a `<seg type="table">` tag to draw attention to the uniqueness of these figures. Always check with the image of the original document

---

[5] Illegible
[6] Hole

to confirm the beginning and ending of the table in order to accurately place the tag. We then wrap the text within a tag as seen below:

```
<seg type="table">2        14     6
                      13     46     Hats </seg>
```

7.1 TELEGRAMS

We treat telegrams slightly differently than letters. While the basic mark-up will be the same, there are a few key differences noted below.

1)   In the title for the document use "telegram" instead of "letter." (see 2.1) For example:

```
<title>Telegram from Lazette Miller Worden to Frances Miller Seward, July 7,
1822</title>
```

2)   You will also need to change the object description from "letter" to "telegram" in the header of the document. This change will occur at line 27 in the XML document. Example:

```
<objectDesc form="telegram"/>
```

3)   Italics will be used to designate the printed text visible on the telegram document. However, italics will not be used in the spaces where handwritten text is used. See example below:

```
<hi rend="italic">Office U.S Military Telegraph. </hi><lb/>
<hi rend="italic">WAR DEPARTMENT.</hi><lb/>
<hi rend="italic">The following Telegram received</hi> 10 30 A M March 7 1864
<hi rend="italic">  at Washington </hi>
```

Notice how the italic tag is placed around the formal text whereas the written text does not have italics.

7.1 MISCELLANEOUS (JOURNALS, FRAGMENTS)

Because not all manuscript items in the digital archive are correspondence, we vary the TEI Standards slightly to accommodate the unique nature of journals and fragments. The process for tagging persons, places, and books is the same as noted above. The changes for documenting journals and fragments take place in the header of the XML document.

7.2 JOURNALS

1)   In the title for the document use "journal" instead of "letter." (see 2.1) For example:

```
<title>Journal of Frances Adeline Seward, December 25, 1860 to April 4, 1860
    <title>
```

2)   You will need to also change the object description from "letter" to "journal" in the header of the document. The change will occur at line 27 in the XML document. Example:

```
<objectDesc form="journal"/>
```

3)   The origin date of the document (which can be found on line 31 of the XML document) will represented the first entry date of the journal. Think of the origin date as the start date for the journal. Using the example from above:

```
<origin>
    <date when="1860-12-25"/>
</origin>
```

4)   Because a journal is not correspondence there is no use for the corresponding action component of the XML document. We will use the `<!--comments-->` tag to comment-out the corresponding action in   the header of the XML document. (appears around line 71 of XML document) For an example:

```
<!--<correspAction type="received">
        <persName ref="psn"/>
        <placeName ref="pla"/>
</correspAction>-->
```

7.2 FRAGMENTS

1)   In the title for the document use the title of what the manuscript item has been identified by the transcriber. (recipe, letter fragment, etc.) (see 2.1) For example:

```
<title>Recipe by Clarinda McClallen Miller<title>
```

2)   You will need to also change the object description from "letter" to "miscellaneous" in the header of the document. The change will occur at line 27 in the XML document. Example:

```
<objectDesc form="miscellaneous"/>
```

3)   Because many of the fragments have unknown dates, you will use the `<!--comments-->` tag to comment-out the fields that ask for dates of the associated with the manuscript item. The origin date of the document (which can be found on line 31 of the XML document) will appears in the example below:

```
<origin>
    <!--<date when="1801-01-01"/>-->
</origin>
```

       You will also use the same `<!--comments-->` tag to comment-out date associated with the sending correspondent tag near line 72 in the XML document. See example below:

```
<correspAction type="sent">
    <persName ref="psn:MCCc_592"/>
    <placeName ref="pla:unknown"/>
    <!--<date type="creation" when="1801-01-01"/>-->
</correspAction>
```

4)   Because a miscellaneous item is not necessarily correspondence some manuscript items will have no corresponding action. As a result we do not need the corresponding action component of the XML document. We will use the `<!--comments-->` tag to comment-out the corresponding action in the header of the XML document. (appears around line 71 of XML document) For an example:

```
<!--<correspAction type="received">
        <persName ref="psn"/>
        <placeName ref="pla"/>
</correspAction>-->
```

8.1 CLOSING OUT THE XML FILE

Before you finish your letter, you must remove a few transitional pieces of scaffolding and check to be sure your document is "valid."

1)   Remove the transcriber's initials from the bottom of the letter.

2)   Remove the letter's header (that was created in the original Word Document) from the top of the first page.

3)   Run a check for any square brackets and <notes> that you did not remove as you processed the letter (left over from supplied text) or persons, places mark-up process.

4)   Double-check to be sure that your document's header and file name are accurate and labeled correctly.

5)   Remove the `<!--comments-->` which served as prompts for <u>procedural</u> tasks that have already been completed. The `<!--comments-->` that are added for future annotation will stay in the document.

8.2 COMPLETING THE EDITING PROCESS

When you have finished the letter or are done editing for the day, complete the following steps to ensure your work is retained.

1)   Save Document.