DCS 1100 - Fall 2016
XML Assignment
Due at 5:00pm Sept. 26 via Blackboard

Assignment Overview
You will transcribe and markup a letter using well-formed and valid XML encoding.

The Big Picture
This assignment is part of our ongoing investigation of the ways that data is created and the very human decisions involved in that process. It is also related to our overall project question about how technology may be able to help us better understand the historical question about Joshua Lawrence Chamberlain's actions at the Battle of Gettysburg. We will be creating a crowd-sourced data set for later analysis.

1. Work with your partner from the JLC Letters homework.
2. Set up (Mac login).
3. Transcribe your portion of the letter to plain text (details below).
4. We will interrupt class at 12:15 on Sept. 19 to discuss your progress.
5. Encode your letter according to the TEI schema (details below). You will be tagging elements of the letter that should be identified to the class to help us better understand Chamberlain's accounts of Little Round Top. Your final letter (or prototype of the overall encoding of the letter) should have green "well-formed, valid XML" signals when opened by Profs. Hall and Irfan. NOTE: The software is only available on the computers in 304 VAC and Kanbar 101 (open 24hrs).
6. Between 9/19 and 9/26 visit Bowdoin's Special Collections & Archives to view an original letter from Chamberlain. Answer the question: how do these originals challenge datafication? (Make specific reference to at least one letter.)
7. Deliverables. Submit via Blackboard your XML file and a Word document or PDF with the reaction to the Special Collections visit and critical reflection responses.
    - Critical reflection (answers to these questions in paragraph format):
        - Your reaction to the original letter (described above).
        - How did encoding in XML change the way that you thought about your letter?
        - Is encoding an act of interpretation?
    - Your XML file: Make sure that the XML file in your Project folder has green "well-formed, validXML" signals when opened.

Evaluation
Your lab will be graded on the following components (provided the deadline is met):

40pts. - the .XML file opens with green indicators, follows the schema, replaces the 13 "INSERT" content pieces, and adds at least 4 elements related to our JLC@LRT question.

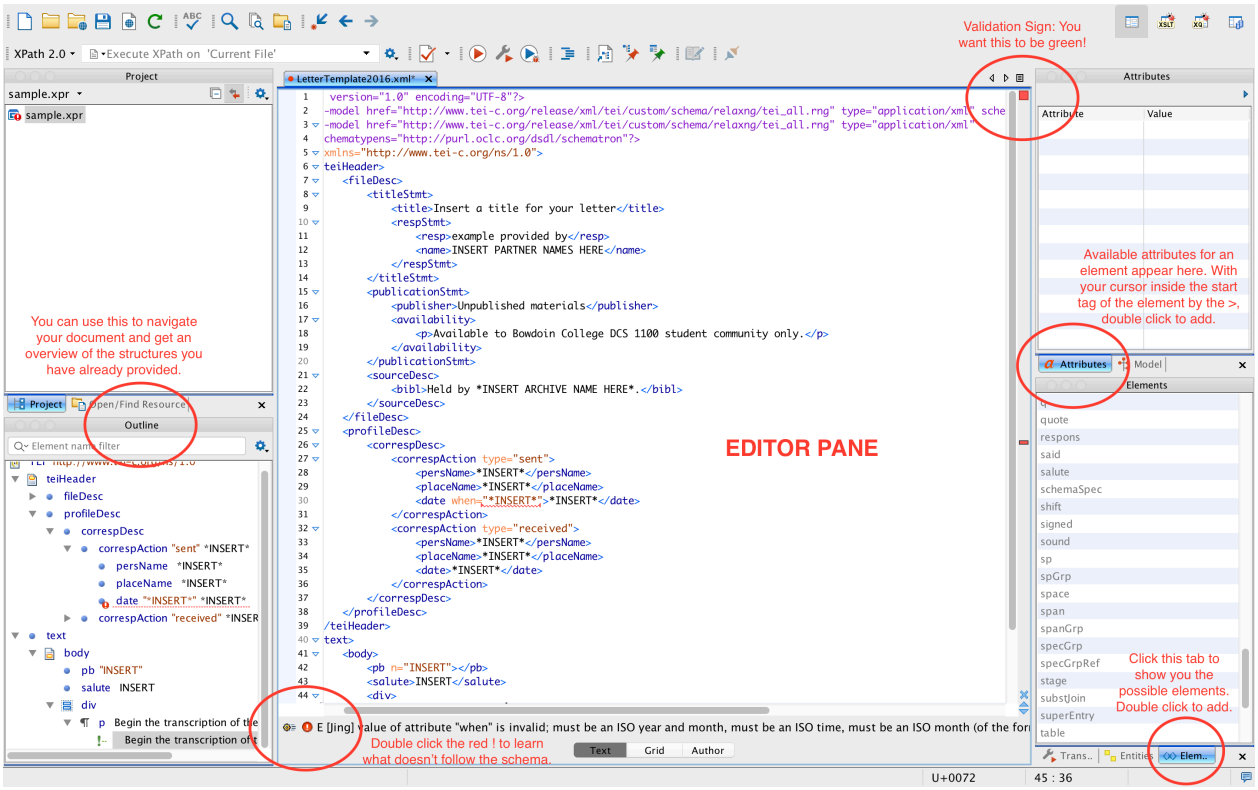
60pts. – reaction and critical reflection (2 pages max) that responds to the questions in part 7 (do not answer in essay format; response paragraphs should be on topic, concise, and organized) using evidence from readings and discussions

Transcription of your JLC Letter

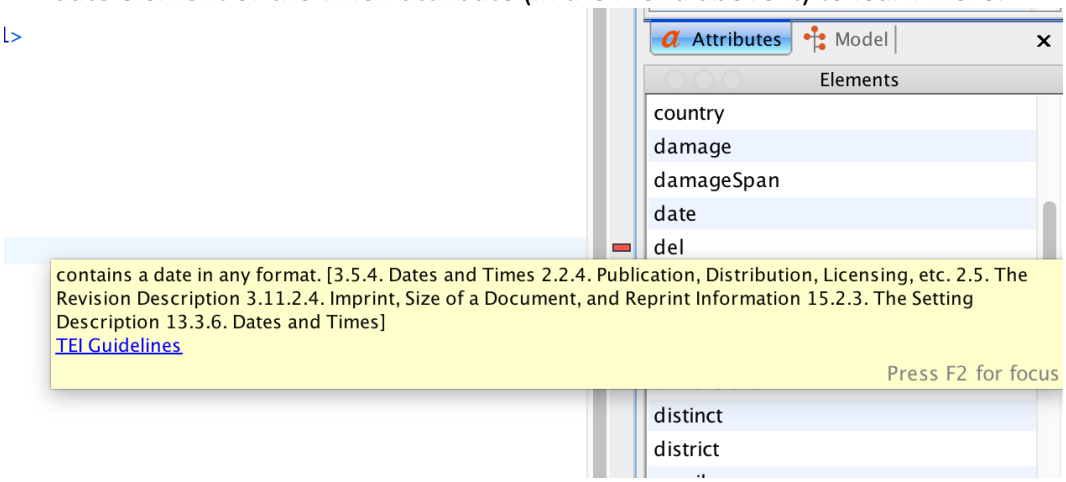
1. Access our course folder through Finder -> Go -> Connect to Server -> smb://microwave/courses/dcs1100
2. Navigate to the Projects folder. Each pair will create a new folder that has the title "XMLPartnerAPartnerB" (Bowdoin IDs are sufficient to indicate partners).
3. Templates and this assignment description are located in the folder: Materials/XML/Assignment. Copy the "Assignment" folder to your Project folder. You will store all of the materials for submission here. These files are viewable by the rest of the class.
4. Create a digital transcription of your portion of the letter that you will save as a plain text (.txt) file in your new "Projects" folder. As a reminder you will find digital images of the letters in our course folder: //microwave/courses/dcs1100/Materials/XML/JLC_Letters/
5. Document any challenges that you face, decisions you make, and standards that you set in a separate Word document. (Use the "Alphabet" handout to help with questions of deciphering the handwriting.) Be prepared for the class discussion at 12:15.

XML Markup with <oxygen/> XML Editor

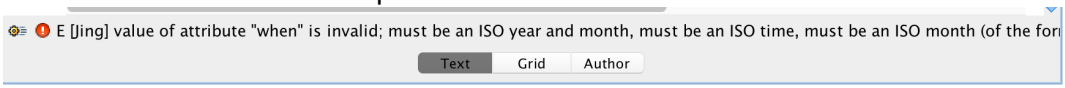
1. Navigate to your Assignment folder and double click to open the "LetterTemplate2016.xml" file. It should open using the Oxygen XML Editor:



2. Immediately save your file with a different filename, and make sure that it saves to your new folder in Projects. Choose an informative filename that will help other students (and your professors) understand what is contained in the file.
3. Out of the many panes of information that appear, we are going to be most focused on the large, central editor.
4. The name of the game with XML is the embeddedness or nesting of tags. Your template is already pre-loaded with a few tagged elements like <name>, <title>, etc. and indications about where to insert information about your letter. There are 13 locations where you will see "INSERT" to guide you. Use the "Elements" menu on the bottom right to get more information about any element. For some you may only be able to say "unknown" (and that's alright). Two are particular:
    a. **<date when= INSERT>** Right now your template is invalid because the date element is missing the "when" attribute in the proper format. Hover over the date element or the when attribute (in the menu above it) to learn more:



Hovering for more information works for any element or attribute, and is great for deciding which one to use, but not so great for troubleshooting. Instead, double-click on the red exclamation point at the bottom of the editor:



Use the information there to enter the date of your letter according to the schema validation.

b. <p><!-- INSERT… --><p/> This is the paragraph element. Most of your transcribed content will be here. What is here right now is a comment:

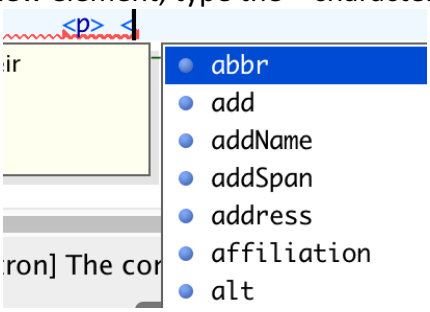
```
<!-- the format for comments (unreadable by the computer) uses these characters -->
```

If you want to signal something to future human readers of your XML document, use the comment element.

Your work in this section will include providing the structural markup for your letter and identifying four new elements related to interpretations of the content. (See below for more details.)

5. In addition to following a standard structural markup schema like TEI, our question is about how a tool like XML might help us to better apply the tools of historians (corroboration, verification, validation, reliability, tone) in order to research our question about JLC at Little Round Top. So you will be adding new elements.

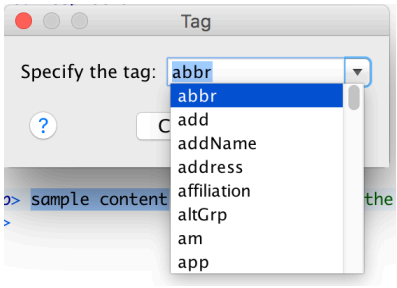
6. The schema is extremely hierarchical and <oxygen/> does its best to look out for you. This means that you should think about big structures first, then move into the details. The transcription of your letter will go inside the <body> of the template.
    a. Before marking up anything, identify paragraphs, salutations, etc. Then create those structures in your template first. Think about physical and material characteristics of the letter (margins, paper quality, spacing, etc.).
    b. You can work word-by-word to take advantage of the auto-complete feature for tagging. To add a new element, type the < character and a menu will appear:



Scroll through the options in that menu to get a sense of the standard variety of tags available at that level of the hierarchy. (**Putting your cursor within different elements gives you different options.**) By clicking on any of them, a description will appear in the yellow box (on the left above, cut off). You can also click the link in blue for further information and more examples.

Double-clicking on the tag in the Elements menu (or pressing return) will create the open and close tags of the element where your cursor is place in the editor pane. You need only transcribe the content between them in order to maintain well-formed XML.

    c. Alternatively, you can copy and paste the entire letter into the template and build the structure around it. Highlight the text that you would like to place in an element and press COMMAND+E and a menu of element tags will appear.



Just keep an eye on your end tags and make sure to maintain the hierarchy.

7. Transcribe and encode your document. In light of our research question, think in particular about the features of the text that would be difficult to find using search methods currently available (finding aids in Archives, keywords, etc.).

    Tip: You might be particularly attentive to tags such as <distinct>, <witDetail>, <terrain>, <expan>, <unclear>, <persName>, <placeName> and many more.

    Remember: the elements above are available at different levels of the hierarchy, so click within different places of the text body to see what can be indicated as an element where. This will impact your structural decisions.

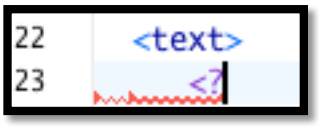
8. Your final .XML file should include the following elements:
    a. INSERT replaced by content
    b. 4 new well-formed, valid elements (either from the suggested list above or from others that are available and you choose)

TIPS

Things go wrong. You can type whatever you want, wherever you want, but you will start to see lots of warning signs on your screen if you start to step outside the hierarchical embeddedness of the TEI schema or don't complete an element tag:

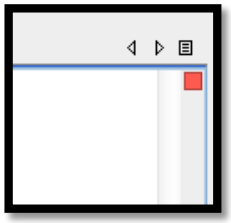
For example, if I try to put a processing instruction element (indicated by <? /?> in purple, below) somewhere that it shouldn't go, three things are going to happen.

a. Lots of red appears in the line with (or in the following line for) the not well-formed XML (because it's missing its close tag) and the invalid element (because it's not in the right spot according to the schema):



b. An error message appears at the bottom of the editor pane:

c. A red indicator of not well-formed or invalid XML appears in the top right of the editor pane:



To resolve these errors:

- Stop and look at the menus.
- Use the menus to help you to see if an element is permissible in the place where you want to put it.

- Use the red exclamation point to give you details on valid XML for the schema.

Resources

- There are many resources to help you identify appropriate tags quickly and easily. One has been developed by a colleague at Northeastern University: http://www.wwp.northeastern.edu/outreach/seminars/_current/handouts/elementList.xhtml
- More detail is provided at http://www.tei-c.org/index.xml and training modules can be found at http://teibyexample.org/