The Confederation Debates

XML Encoding Tutorial

Table of Contents

Introduction:

This brief tutorial will introduce you to simple XML mark-up tasks using Oxygen XML editor. Coding conventions will follow the Text Encoding Initiative (TEI). For a brief tutorial on XML please see the W3schools online course [here](). Read through "XML Home" to "XML Attributes." FAQ for TEI can be found [here]().

By the end of this tutorial you will:

- Learn how to use Dropbox, download your files, and store your changes.

- Learn how to use Oxygen.

- Learn simple XML mark-up using TEI conventions.

Setup

Before you can begin reviewing documents, you will need to perform the following steps on the computer that you will use for your encoding work.

A) Dropbox Setup

If you do not already have a Dropbox account, please first:

1. go to https://db.tt/cjjrMDVp.
2. Create a Dropbox account.

Regardless of whether you needed to do steps 1 and 2, please:

3. Send the email address associated with your account to dheidt@uwaterloo.ca, so that he can add you to the Dropbox folders (One folder will be "To Do" the other folder "Complete").

B) Oxygen Setup

1. Dan will provide you with a licensed copy the "editor" edition of Oxygen, which can be downloaded from here.

2. Once Oxygen is installed on your computer contact Dan for a license key. If Dan has not yet provided you with a key, please sign up for a trial, use the key that Oxygen emails you to get started on your work, and email Dan notifying him that he needs to send you a product key.

   *Some third-party programs are not recognized and denied download access by Mac OSX. If Oxygen cannot download refer to these instructions on how to install third-party software.

3. Please note that nearly all screenshots in this tutorial were taken using a Linux operating system. Oxygen also runs on Windows and Mac computers and uses the same interface, but colours and icons may differ slightly.

Managing your Project Files

1. Create a "TCCD" folder on your computer. After receiving your invitation to join the TCCD Encoders Team Dropbox" folder on Dropbox, go to this directory and find your zip package with your name on it in the "To Do" folder. Place the zip package file in your TCCD folder and unpack/unzip/extract the files using whatever program that is supported by your operating system. A brief explanation of zipping and unzipping can be found here.

2. Open Oxygen and go to Project/Open Project and find the package provided to you and select the .xpr file in your TCCD folder.

Note: When you first open your package you will receive a message that says: "project contains options that will change the application interface and behavior. To continue opening it, please confirm that the project comes from a trusted source." Click "Yes" as your source is the UVic team.
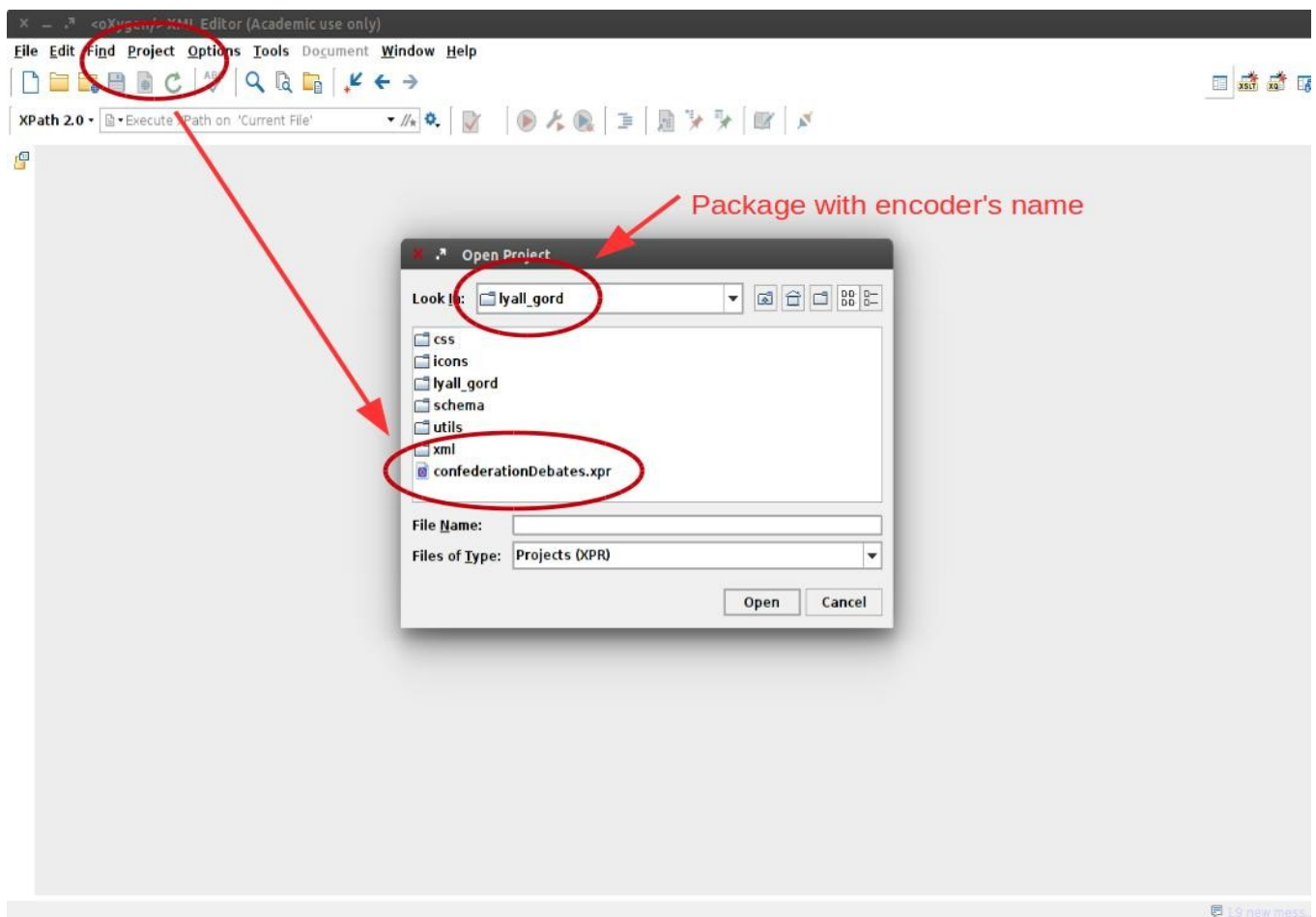
3. Once the project is open, your folder should appear in Oxygen at the sidebar with all files associated with the project. (Note your project will have only the files provided to you by the UVic encoding team).
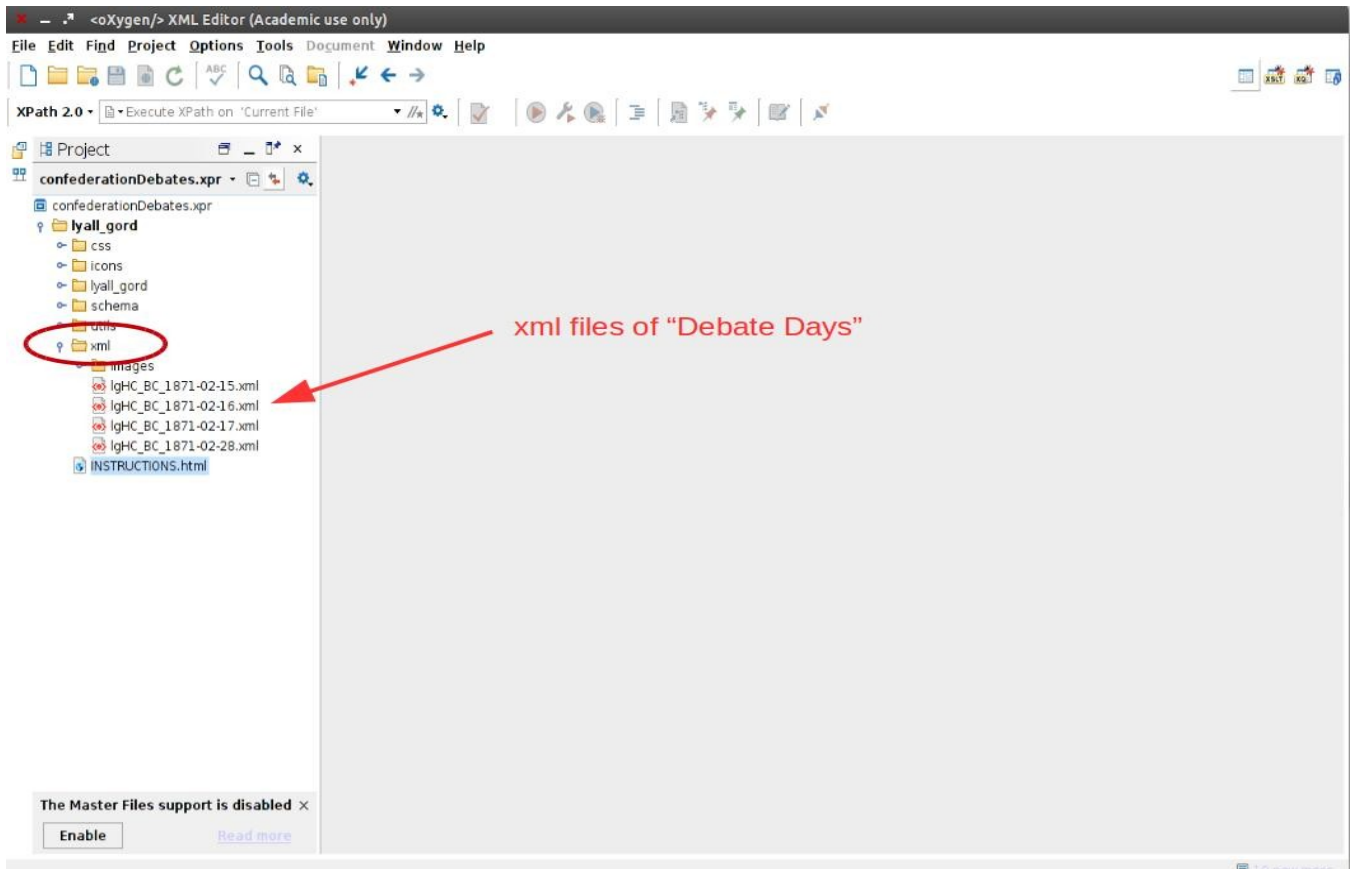
4. If it is your first time using Oxygen, you should change the settings to wrap text within the window. You can find this option in: Options/Preferences/Editor/Edit Modes/Text - toggle on "Line wrap."

Marking-Up Your Documents

Please refer to the W3schools here for explanation of xml terminology and the TEI FAQ here. Key terms to recognize are "elements" and "attributes" .

Select a Page to Review

1. To edit a Debate Day, open Oxygen and find the .xml folder on the left, and double click the xml file you want to work on. You will then see a screen that looks like this:
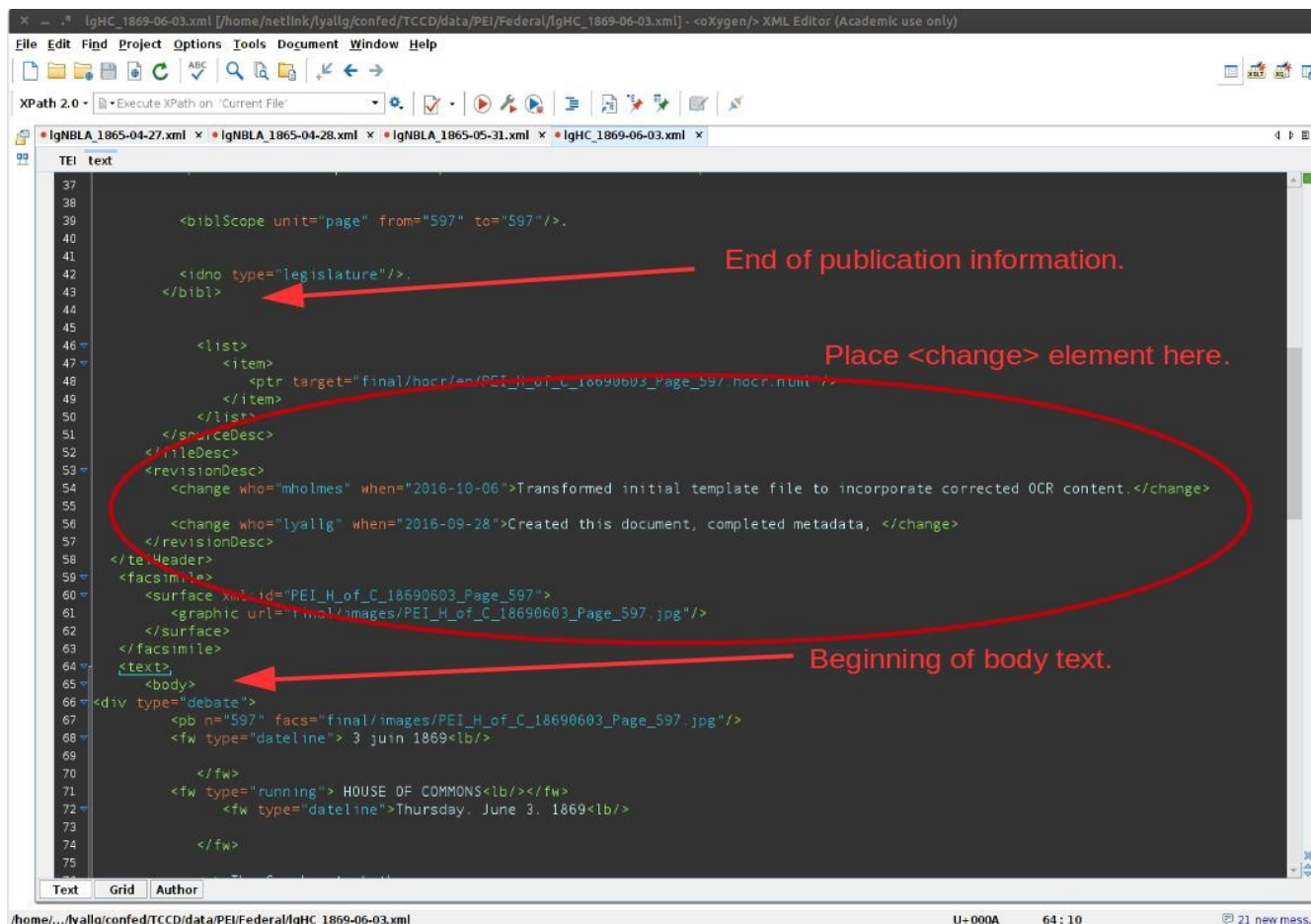
2. Scroll down to the <revisionDesc> section. This can be found in the metadata after the publication information </bibl> and before the <body> text.

3. In the <revisionDesc> element you will see two embedded <change> elements with "who=" attribute keys "mholmes" and "lyallg". Directly below "lyallg", after the closing tag </change>, input your name (last_first) and date in a <change> element with "who=" and "when=" attributes. Within this tag type the words, "Marked-up people." I, for example, wrote:

<change who="lyall_gord" when="2016-09-28">Marked-up people. </change>

General Guidelines

Please consider the following guidelines when marking-up the text:

1. You are only concerned with finding proper names. Only mark-up the names of people who speak. Additional individuals will, sometimes, be referenced (for example: "the honorable member from Argenteuil"), but you will not mark-up these additional references.

2. To identify speakers: compare the XML to the images provided in your package by clicking on the <u>image icon</u> in *Author* mode. Speakers are typically in bold, and/or all capitals, and opening a paragraph. All other instances of names can be ignored. <u>Return to *Text* mode for mark-up</u>.
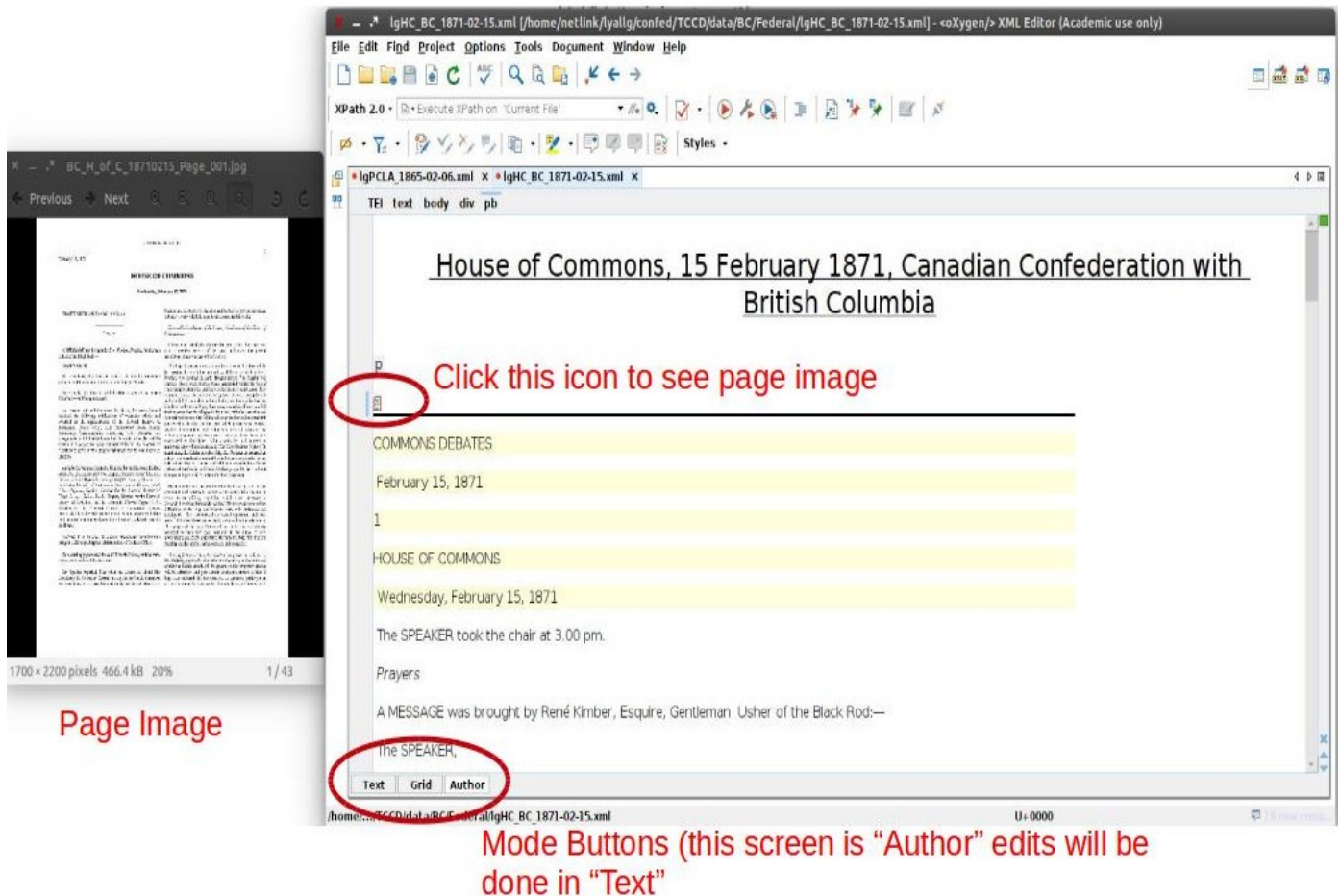
Encoders only need to know 1 basic element and attribute.

<u>persName</u>: This is the element that will be used to identify people in the document. <persName></persName>. The attribute is a ref="" tag where you will input the correct, unique, xml ID, which will appear in a dropdown list (see xml IDs below).

Note: the titles and roles preceding the name of a person will be included within the element tag.

So, when you spot names like those shown in the following example documents…

…you will encode the entire name and title, but you will *not* include the riding as shown by the examples below:

`<persName ref="pers:ABBO1">Hon. Mr. Abbott</persName>` (Argenteuil) ✓

NOT

Hon. Mr. `<persName ref="pers:ABBO1">Abbott</persName>` (Argenteuil) X

OR

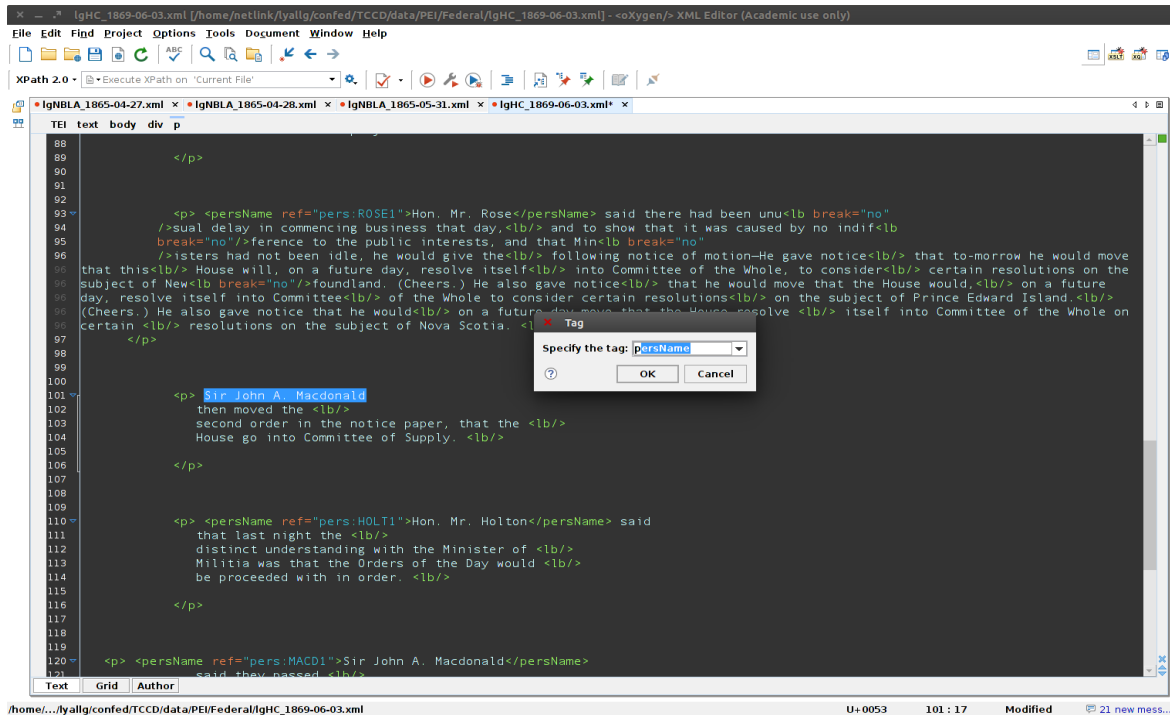`<persName ref="pers:ABBO1">Hon. Mr. Abbott (Argenteuil)</persName>` X

To Mark-up a name:

1. Highlight the entire name, including any role or title, and hit "CTRL E" ("CMD E" on Macs). This will bring up a dialogue box with element options. Begin to type "persName" and when it is selected, hit enter.

2. The name should now be surrounded with an opening tag and a corresponding closing tag.
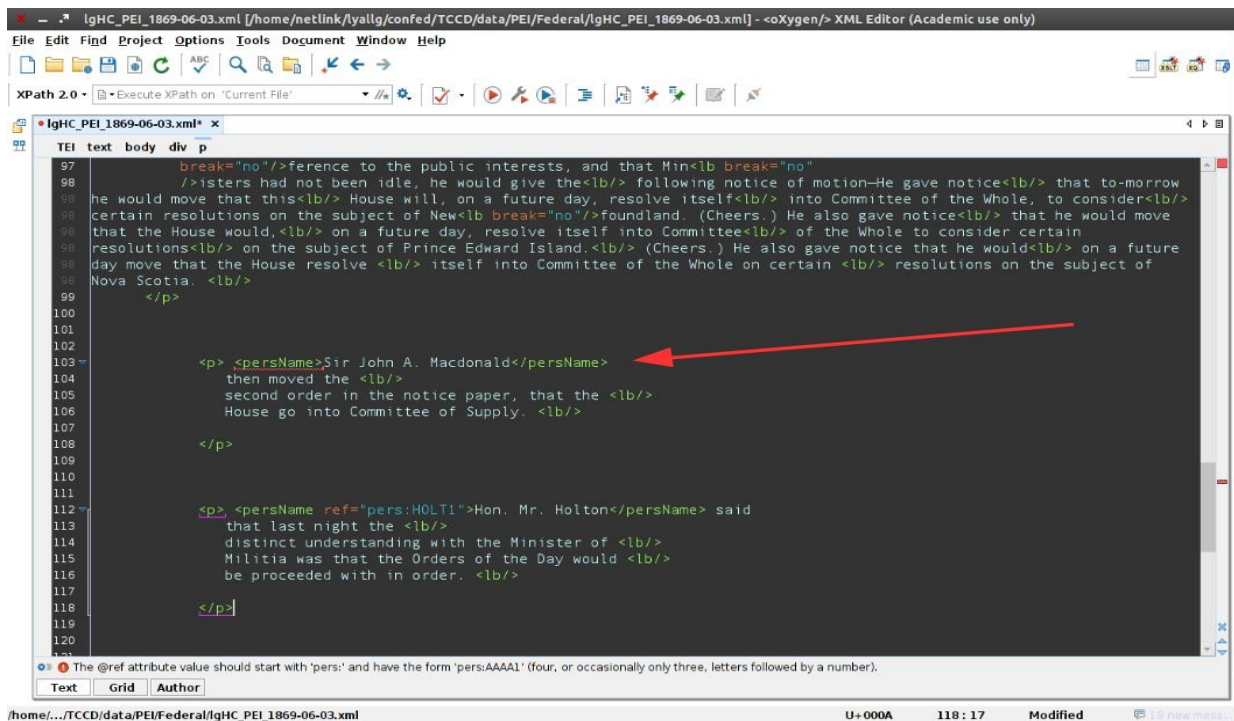
3. Next, the ref attribute has to be added.

4. In the opening tag, preceding the name, place your cursor immediately after "persName".

5. Hit space and start typing "r" for "ref". If done correctly a box should pop-up with options for attributes starting with the letter "r".
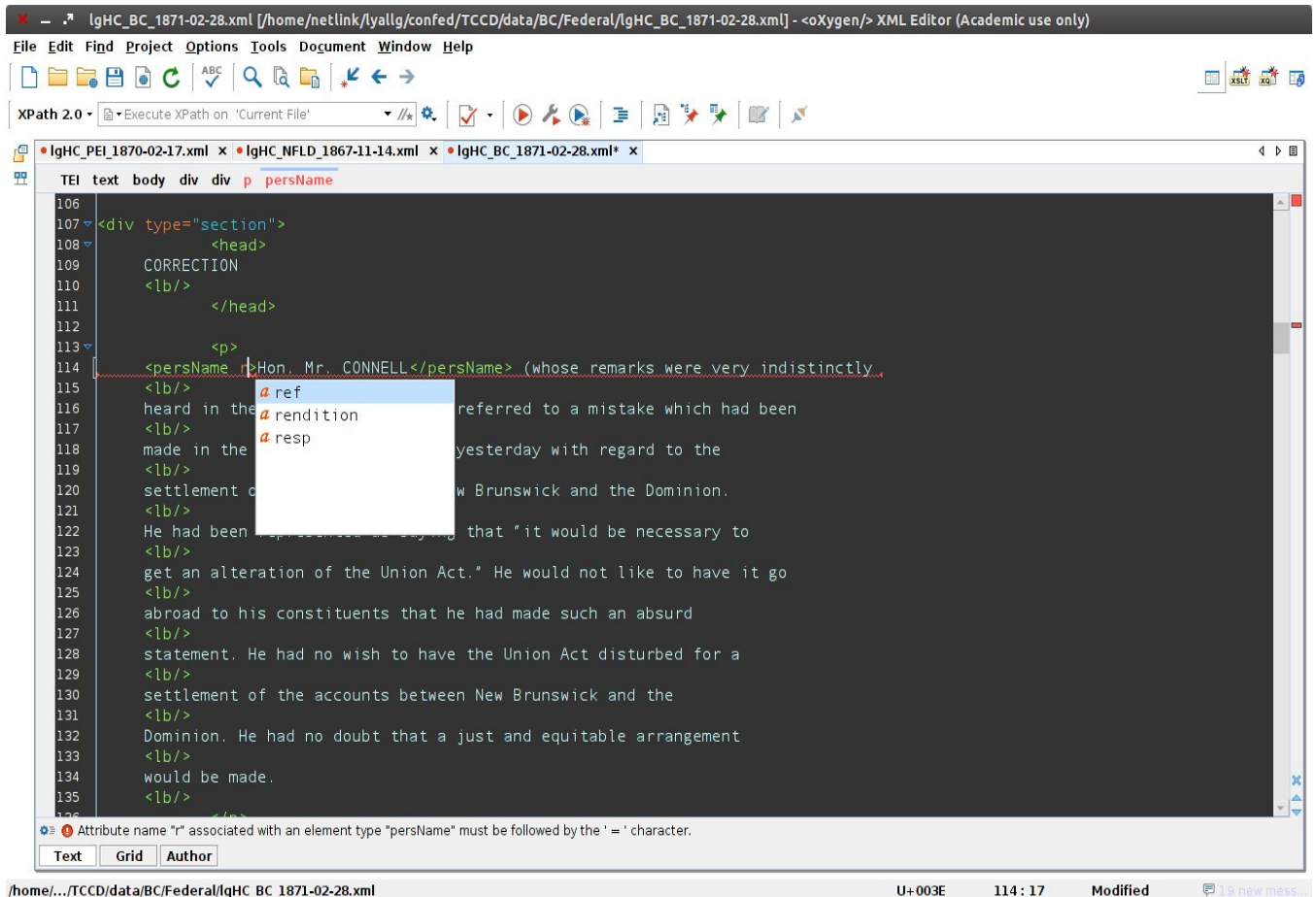
6. Select "ref" in the dialogue box and fill this attribute in with an id using the dropdown box to find your correct xml id for that person (see "xml ids" below).

   a. Alternatively you can complete the ref attribute manually by typing out ex) ref="pers:ABBO1" (ABBO1 is the id for Sir John Abbot).

Determining the right ID Via "Debate Day":

1. Sometimes more than one debater had the same last name. We have therefore created sequential IDs with the same alphabetic prefix, but determining which one matches your line of text can be tricky. For example, more than one historical figure with the last name "Fisher" participated in our debates. We have two ids for "Fisher": FISH1 and FISH2.

2. Sometimes we can find the right historical figure by using the "**debate day**" (date of debate) of the .xml you are working on. The date can be found within the filename of the .xml document (ex. lgHC_BC_1871-02-15). Note that dates are standardized in this order: year-month-day.

   a. If we click on FISH1 and FISH2 we can see the ridings and years that each of the historical figures represented. Because FISH2 did not appear in parliament until 1905 and FISH1 represented York in 1865 we know that the correct id for this MR. Fisher is FISH1.
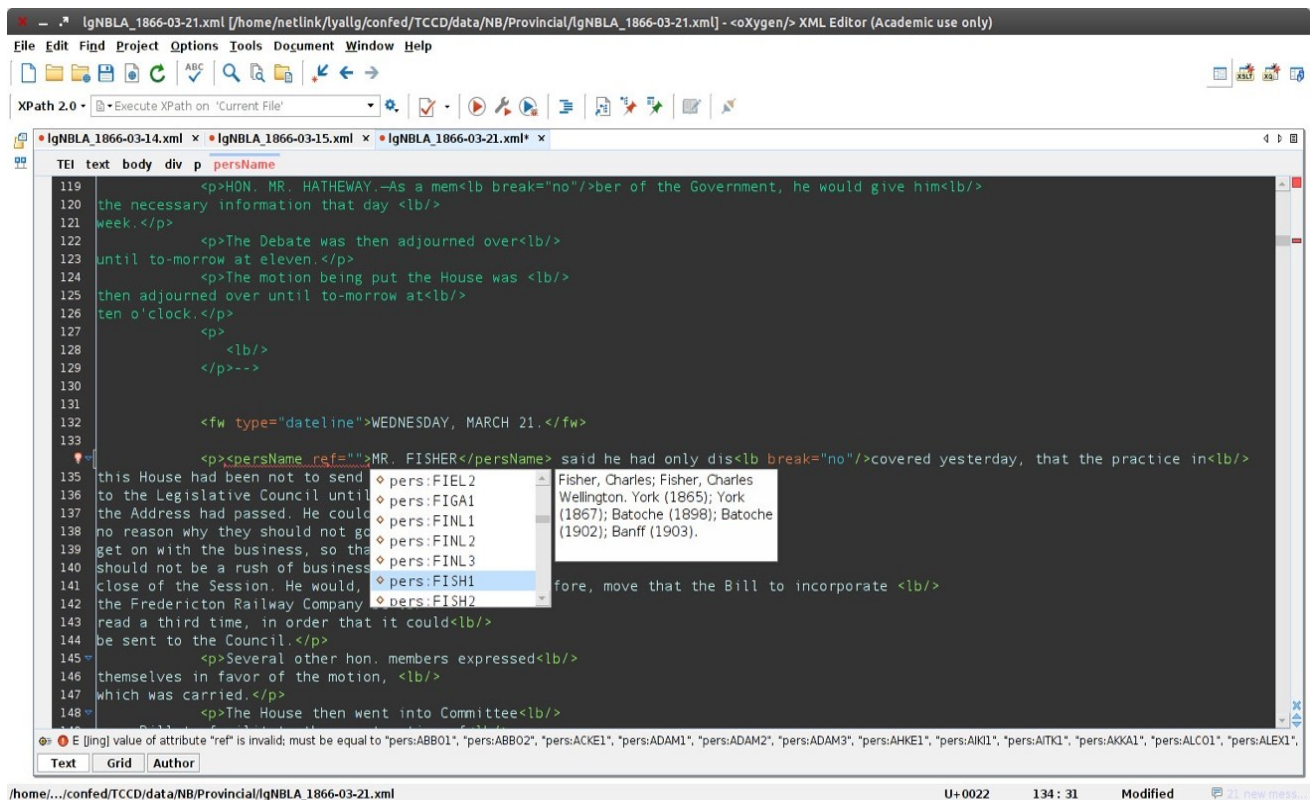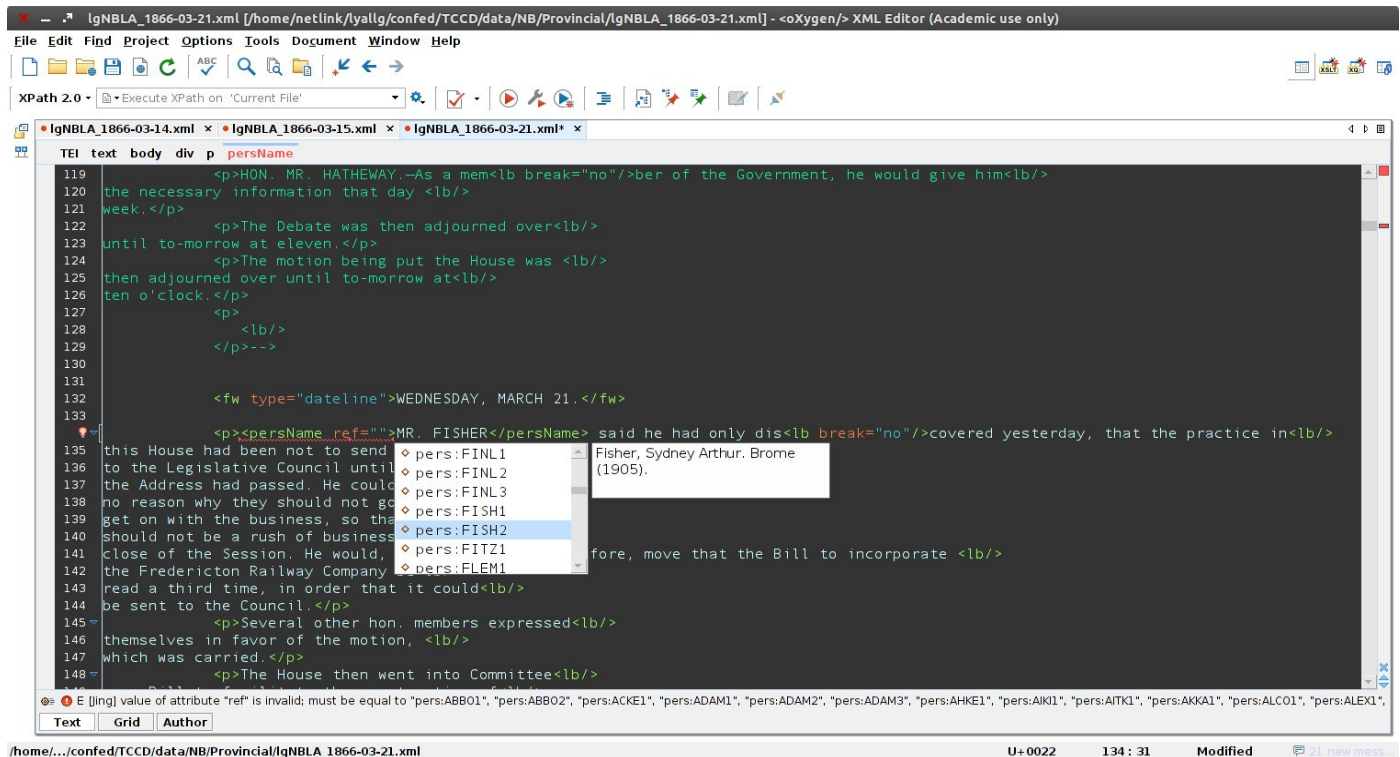
So, the correct tag for this example will look like this:

`<persName ref="pers:FISH1">Mr. Fisher</persName>`

Determining the right ID Using Other Information:

1. But what happens when two historical figures with the same last name spoke during the same debate day?

2. In this second example, we have three different Macdonald's speaking on the same day:

    a. Sir John A. Macdonald (MACD1)

    b. John Sandfield Macdonald (MACD2)

    c. Donald Alexander Macdonald (MACD6)

3. Determining which ID is correct can be tricky, but there are often clues that can help including:

    a. initials preceding the last name

    b. a title that indicates the office held by the member (ex. "<u>Sir</u> John A. Macdonald"). Your document may have been created before the historical figure received their title (ex. John. A Macdonald was not knighted until 1867, but he participated in the Ontario-Quebec debates in 1865). If you found a comment by Sir John A. Macdonald from 1865, you would nevertheless use MACD1.

    c. the member can be identified by a parenthetical reference to his riding, ex: MACD3 = McDonald (Toronto)

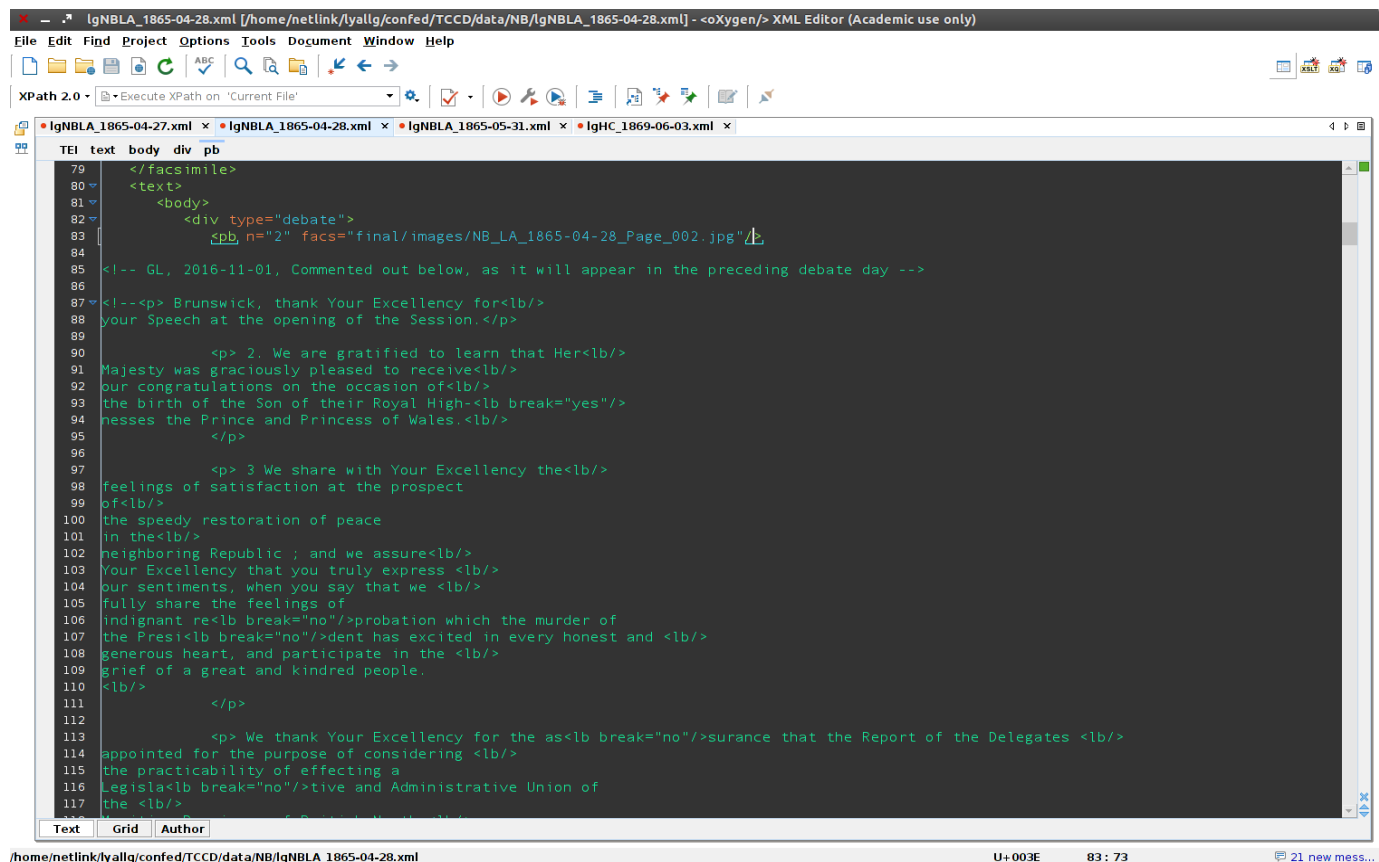4. Using these methods you can tag the "Macdonalds" correctly, as seen here:

5. Note: if a speaker is mentioned again in a subsequent paragraph that is the unbroken continuation of his speech you *do not* need to mark-up his name again.

6. If you are unable to determine which historical figure you are encoding, you can tag them ref="UNSPECIFIED", which appears at the top of the id box. When doing this, please also contact Dan for verification. When sending Dan an email, please make sure you specify the person's name, the debate day, and the page number (in the original document) that the name appears on.

7. If there is no mention of the speaker (often due to the transcription starting in the middle of a speech), or if they are identified simply as "The Speaker", please contact Dan with the debate day and he will instruct the UVic team on who is speaking so that they can add the speaker with an <add> element.

14

Dealing with Comments

Comments look like this: <!-- "COMMENT HERE" -->.

Comments are only visible at the code level, and do not appear in the rendered version of the XML page. They can be used to ask questions or give directions to other encoders. We do not anticipate that you will need to use them.

It is important, however, for you to recognize commented out sections of the text. In most cases, the UVic team has commented out sections of text that will appear on a different debate day. Because this text will not appear in the final rendering of the website it is *not* necessary to mark-up the text within. Therefore, Do not mark-up content that is commented out (i.e. green text within <!-- -->).
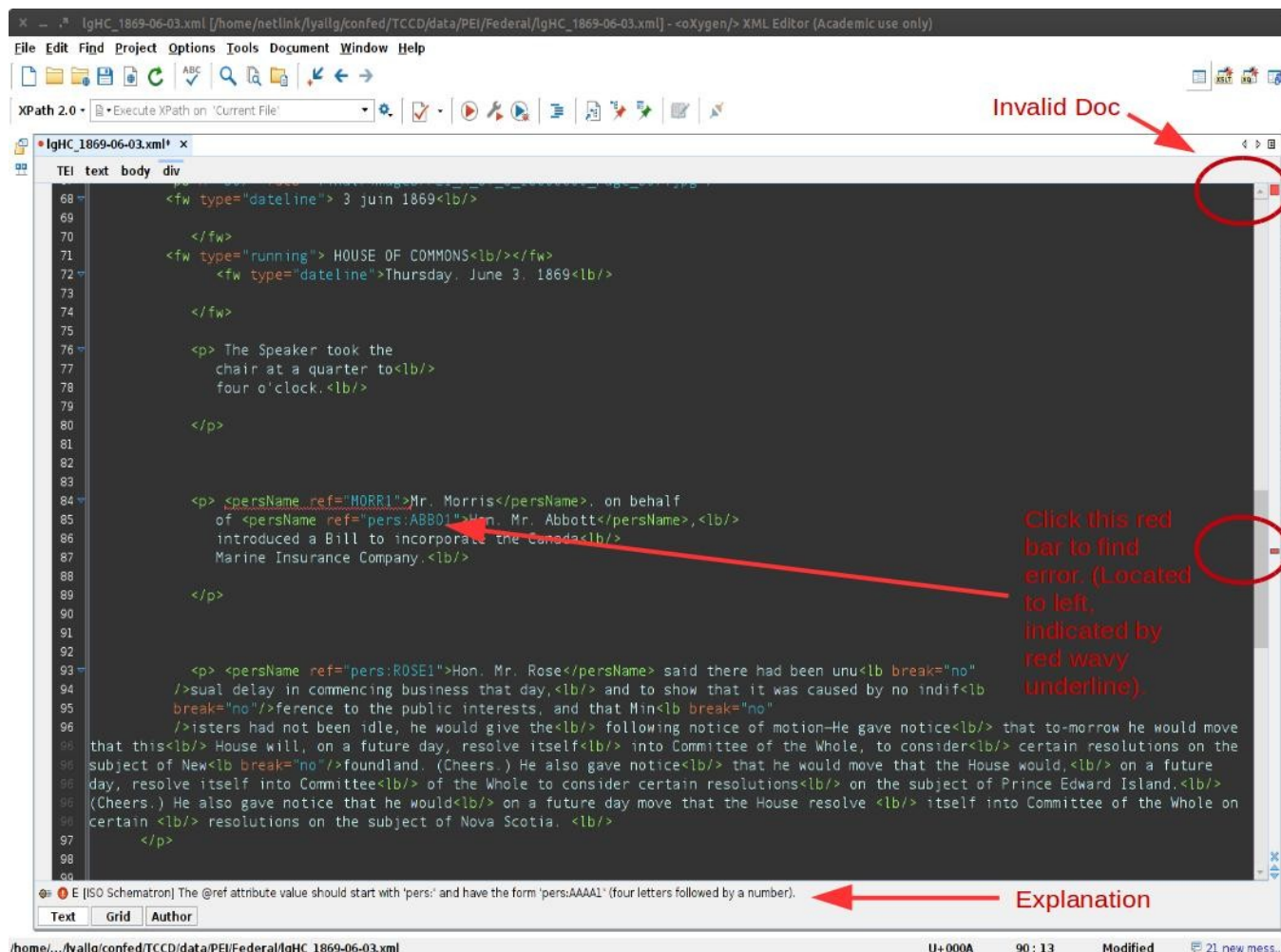
Validation

1. If there is a red box at the top right corner of Oxygen your document is invalid and there is a problem with your coding.

2. Do not be alarmed if this appears when you first enter a <persName>; the warning will disappear once you correctly enter the ID.

3. Given the minimal amount of mark-up, your invalid mark-up is likely caused by one of the following scenarios:
   a. There is no corresponding close tag to the element
   b. an element is not properly contained by angle brackets < >
   c. a spelling error in the element tag or an incorrect attribute (for example, missing the "pers:" before the xml id).

4. To view the error, find the red bar along the right side of the browsing window and click on it. This will take you to the error and a technical explanation will be displayed at the bottom of the screen.

16

5. When your document is valid, the box at the top right hand corner will be green.
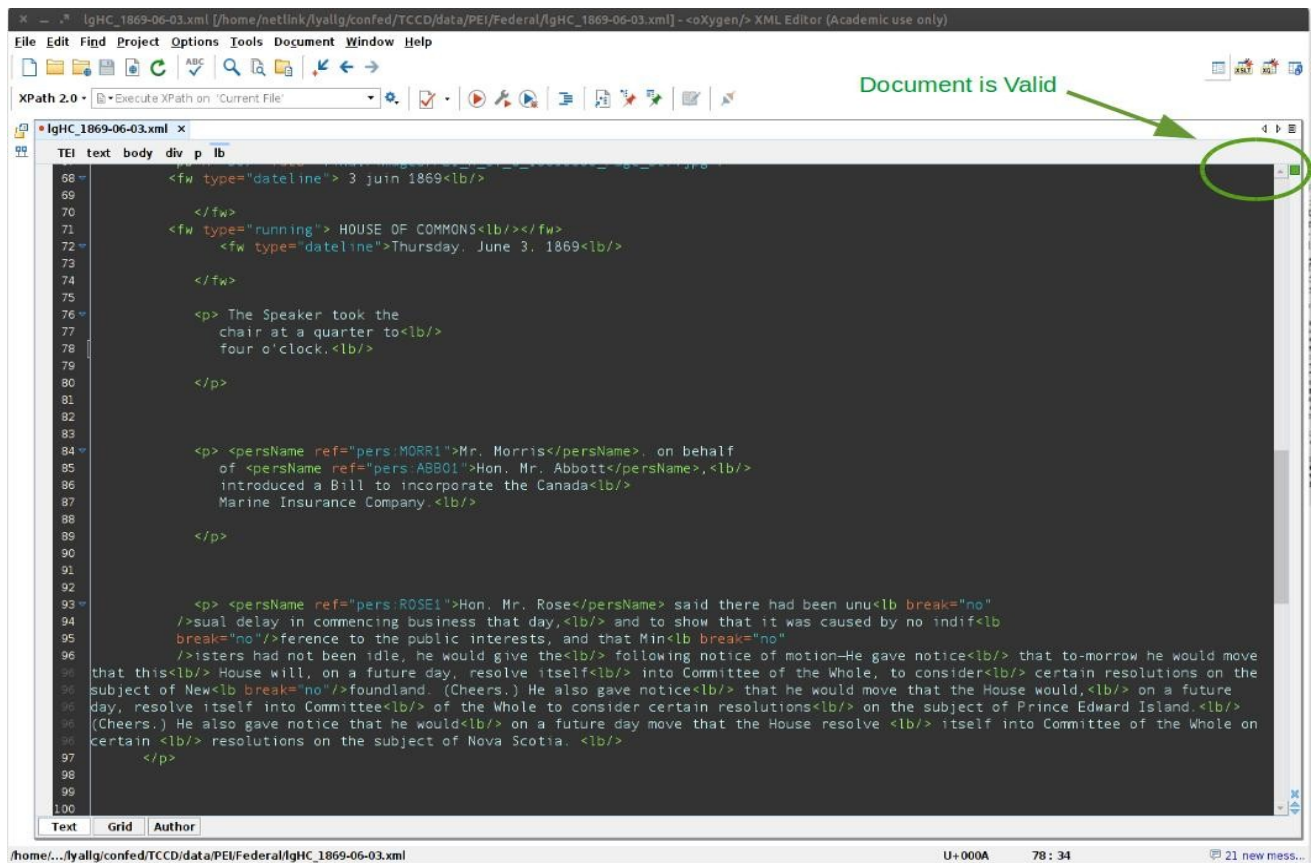
Final Steps

1. Please ensure your documents are valid before moving on to the next Debate Day.
2. If you have any questions about the validity of a document contact Dan or, XML Encoding Team leader, Gord Lyall, lyallg@uvic.ca.
3. Once you are done marking up all of the XML files that came in your .zip file, in Oxygen, open the INSTRUCTIONS.html file for re-zipping package. Execute the transformation by clicking on the red triangle button and place the file in the "Complete" folder in Dropbox and the UVic XML team will merge it into the project.