

VWWP Critical Introductions and Bios Encoding Guidelines



Helpful Tools

Templates (always start with a template; lots of important boilerplate markup that is required and not described below)

- [Critical Introduction XML/TEI template](#)
- [Biographical Introductions XML/TEI template](#)

Logs

- [VWWP Intro and Bio Log](#)

- [Overview](#)
- [Encoding Workflow](#)
- [Encoding Guidelines](#)
 - [TEI Header](#)
 - [TEI Body](#)
- [Next Steps](#)

Overview

Critical and biographical introductions authored by students will accompany the source literary texts selected as part of a course, practicum or internship. The introductions are submitted as two separate files:

- critical introduction that is associated with a particular text
- biographical introduction that can be associated with multiple texts by the same author

Those essays are typically not encoded by the authors. They will be encoded by Digital Library Program / Libraries staff throughout the year.

Encoding guidelines for the introductions are documented herein with relevant bits linking out to the official [VWWP TEI P5 Encoding Guidelines](#) so not to duplicate information.

[return to top](#)

Encoding Workflow

Below are detailed steps for setting up your encoding environment and tracking problems or questions encountered along the way.

Setting Up the Files for Encoding

1. Create a working folder on your Desktop called "vwwp"
2. Download the essays (Word or PDF files) from specified location
 - a. Save the essay you are encoding in your "vwwp" folder
3. Download the [Biographical Introduction Template](#) and the [Critical Introduction Template](#) and save to your "vwwp" folder
 - a. Use this file as your "shell" for encoding, keeping in mind you will need to make changes to the [TEI Header](#) (see below)
4. Open up the Oxygen XML Editor, select File =>Open and find your XML file to begin encoding OR click on the file, which should open Oxygen automatically
 - a. **For critical introductions, Save File As:** VA# number of source text underscore intro dot xml (e.g., VAB1865_intro.xml).
 - b. **For bios, Save File As either:**
 - i. author last name.xml (or if more than one of the same last name exists, then)
 - ii. author last name_first three letters of author first name.xml
 - iii. E.g., booth.xml or keary_mau.xml

Uploading Files When Done

At the end of your encoding session, you should upload your XML/TEI file so that we can maintain a backed-up version. **Your file needs to validate against the schema before you can upload.** To do this:

1. Log into Xubmit: <http://algeron.dlib.indiana.edu:8080/xubmit/>
2. Select the "Victorian Women Writers Project" repository and click "Open Collection"
3. Click on the "Submit File" link

4. Choose the file by browsing your file system
 - a. **NOTE:** please do NOT rename your XML file, this will cause problems in the database
 - b. It is not necessary to validate the file within Xubmit since it needs to be valid before you can upload
5. Create a brief log message summarizing the work you have completed thus far or pending issues (e.g., "Encoding complete, not sure how if footnotes were encoded properly," etc.)
 - a. Set the file status to "In progress" as you work through the encoding
 - b. Once the encoding is complete, set the file status to "Pending review"
6. Click "Submit" after you have browsed for your file and drafted a log message
7. Verify that you received confirmation of your uploaded file
 - a. Check the list of file to make sure your file was indeed uploaded

Validation

You will want to validate often while encoding. This will help you uncover encoding errors early on in the process as opposed to letting errors compound, making troubleshooting more challenging.

Schema validation should be frequent and on-going.

Schema Validation

This project uses a schema customized from TEI P5, which makes sure the XML document is well-structured and valid. Your Oxygen editor automatically knows this information. The Oxygen editor will show you a green box towards the top-right of the editor if the file is valid or a red box if the file is invalid.

To validate an XML/TEI file already open in the Oxygen editor:

- Select the "validate document" icon (red checkmark)
- Or in the menus, select "Document" => "Validate" => "Validate Document"

If the file is valid, you will see:

- Green box towards the top-right of the editor
- Green box on the bottom of the editor that reads: "Document is valid"

If the file is invalid, you will see:

- Red box towards the top-right of the editor
- Red box on the bottom of the editor that reads: "Validation-failed. Errors: #"
- Error messages in a bottom pane
 - Click on each error message to position the cursor **near** the error (often the error is somewhere before the cursor, often a line or more of code above)
 - If the error message is cut off, right-click on the error message and select "Show message"

Once you fix the errors, re-validate the document. All documents must be valid to the schema at the end of an encoding session.

Encoding Problems

If you encounter a problem or a question during encoding, please document the problem/question at: [VWWP Encoding Problems](#). Angela Courtney and Michelle Dalmau will monitor this page for feedback.

[return to top](#)

Encoding Guidelines

Encoding the critical introductions should be relatively straight forward, but since not all final versions have been submitted at the time of drafting these encoding guidelines you will surely run into peculiarities that need to be documented. Please don't ignore these. Instead post the issues you encounter on the [VWWP Encoding Problems](#) page.

The introductions are essays that may contain inline bibliographic references, references to the source literary text, footnotes or endnotes and lots of biographical information. For now we won't be encoding the biographical information. When ever possible, these guidelines will link to existing encoding information so not to duplicate information. Only encoding practices that are unique to these introductions will be captured here.

Global Encoding Approaches

Adhere to the following:

- [Character Encoding](#)
- [Text Formatting and Rendition](#)
- [Lists](#)
- [Notes](#)

Page breaks WILL NOT be encoded since these documents are born-digital.

TEI Header

Every XML file contains a basic TEI Header with boilerplate information already completed. **You will complete/update the information represented by a dollar-sign variable** unless the value is in an attribute (e.g., \$Encoder's First and Last Name). **Make sure dollar signs are removed!**

Some of the information you need to capture about the source text will require consultation with the [VWWP Texts Log for Contextual Information](#) or the [VW WP Website](#) for the legacy texts.

The Log and/or the VWWP Web site will tell you the VAB# that you will need to complete in various parts of the Header and the title of the source text. For bios, it is just as important to check the encoding log as it is the web site to make sure all texts are identified and referenced accordingly in the TEI Headers for the Bios.

The following TEI Header information will need to be completed (see example header below):

- <fileDesc> <titleStmt> <title>: \$Title of Essay
- <fileDesc> <titleStmt> <author>: \$Last Name, First Name of Author
- <fileDesc> <titleStmt> <author>@xml_id: \$IU username of Author
- <fileDesc> <titleStmt> <title type="filing">: \$Title without leading articles (optional)
- <fileDesc> <respStmt> <name>: \$Encoder's First and Last Name
- <fileDesc> <respStmt> <name>: \$Editor's First and Last Name (see description below)
- <fileDesc> <publicationStmt> <idno>: \$VA#_intro (reference [VWWP Texts Log for Contextual Information](#) for VA#)
- <fileDesc> <publicationStmt> <date>: \$Year of Encoding (4 digit year)
- <fileDesc> <notesStmt> <note> <listBibl> <bibl> <title>: \$Title of source text (reference [VWWP Texts Log for Contextual Information](#))
 - repeat <bibl> as needed for the bios
- <fileDesc> <notesStmt> <note> <bibl> <title>/@ref: PURL of source text
- <fileDesc> <sourceDesc> <bibl> <title>/@type: either **biography** or **introduction**
- <fileDesc> <sourceDesc> <bibl> <title>: \$Title of Intros
- <fileDesc> <sourceDesc> <bibl> <author>: \$Last Name, First Name of Author
- <fileDesc> <sourceDesc> <bibl> <affiliation>: \$Institutional affiliation (see description below)
- <fileDesc> <sourceDesc> <bibl> <date>: \$Date in 4 digit year essay was written and submitted in final form

Contributions By IU and Non-IU Affiliates

Indiana University Affiliates

For authors, editors and encoders affiliated with Indiana University, we will use their IU network ID as the @xml:id in the <name> element.

For example: mdalmau

For Non-IU Affiliates

For authors, editors and encoders NOT affiliated with Indiana University, we use their first and last names as the @xml:id in the <name> element.

For example: chris_hokanson

Judson College

- All 9 Judson College essays were authored as part of Professor Chris Hokanson's class. Chris needs to be listed as an editor.

```

<!--Example of responsibility statement in fileDesc-->

<respStmt>
  <resp>Edited By</resp>
  <!--Name of professor or instructor who provided editorial feedback
  and editors of the project after QC-->
  <name xml:id="chris_hokanson">Hokanson, Chris</name>
</respStmt>

<!--Example of attribution fileDesc and sourceDesc-->

<sourceDesc>
  <bibl>
    <title type="biography">Lady Jane</title>
    <author>Jane Doe</author>
    <affiliation>Judson College</affiliation>
    <!--4 digit year the intro was authored and submitted in final form-->
    <date>2012</date>
  </bibl>
</sourceDesc>

```

Distinguishing Between Intros and Bios

In the sourceDesc, the title of the essay has an @type attribute with one of the following values:

- introduction
- biography

The appropriate value is **mandatory**.

Linking to Monographs

Introductions should only point to one source text, however, biographies could point to multiple texts written by the same author. Links to the source texts are handled similarly. In the case of biographies, when multiple texts are available online, you would repeat the <bibl> element.

You will need to reference the [VWWP Texts Log for Contextual Information](#) to track the monographs/authors. You should also check the web site, in the case of biographies, to make sure you link all the books that are part of the online corpus for the author.

```

<notesStmt>
  <note type="relatedItem" subtype="source_text">
    <listBibl>
      <bibl>
        <!--Reference the persistent URL for the source text. The PURL pattern
        is always: http://purl.dlib.indiana.edu/iudl/vwwp/ + VAB# -->
        <title level="m" ref="http://purl.dlib.indiana.edu/iudl/vwwp/VAB1828"> Going West :
or, Homes
          for the Homeless : A Novel</title>
      </bibl>
      <!--Repeat bibl if multiple books by same author is in the collection-->
      <!--<bibl></bibl>-->
    </listBibl>
  </note>
</notesStmt>

```

[return to top](#)

TEI Body

In addition to consulting the global encoding approaches such as rendition values, notes, etc., you will more than likely follow the [Prose](#) guidelines for these critical introductions.

Documented below are aspects of the text that will need special markup. You may encounter additional aspects that are not covered in the guidelines. Please document these in the [VWWP Encoding Problems](#) page.

General Structure of the Text

The essays will generally be structured as follows:

- teiHeader
- text
 - body
 - essay (<div type="essay">
 - section headings (<div type="section">
 - bibliography (<div type="bibliography">
 - notes (<div type="notes"><head type="supplied">Notes</head></div>

Please **note that NOT every essay contains this structure**. You may encounter other chunks of content. The schema provides a set of div type values. If none of those match any new or different chunk of content you encounter, please document the nature of the "chunk" on the [VWWP Encoding Problems](#) page.

Every "sectioned" chunk of text (headings are usually an indicator of this) should reference an appropriate division type. The entire contents will be contained within a <div type="essay">.

Every division should have a <head>. If the author did not supply a title, create a supplied title and assign the section a logical heading. For example:

- Footnotes/end notes: <div type="notes"><head type="supplied">Notes</head></div>

References to the VWWP Text

You will see references to the source VWWP text, in either full or short title form. These references should be encoded within a <bibl> <title> tags with a "ref" attribute that contains the PURL to the source text:

```
<p>
    The author of
    <bibl>
        <title level="m" ref="http://purl.dlib.indiana.edu/iudl/vwwp/VAB1828">Going West</title><
/bibl>,
    the formidably energetic journalist and
    writer Matilda Greathouse Alexander, was born in Posey County,
    Indiana, on June 14, 1842 to George W. Greathouse and Martha
    Greathouse (née Hausman). She was the youngest of four children
    and the third daughter born to the Alexanders. Less than a year after
    Matilda's birth, George Greathouse died. Martha Greathouse remarried
    an Absalom Rodenberger, bearing him two more daughters, Matilda's
    half-sisters.
</p>
```

The VWWP corpus often contains more than one text by the same author so it is **important** that when you come across another text written by the author in questions, you check the following to determine whether you can a PURL reference to the title:

- [VWWP Web site, Browse Author Index](#)
- [VWWP Encoding Log](#)

Please note that this paragraph contains lots of rich biographical information that we **won't be** encoded at this time.

In Line Citations

You will see references to other published works (by the author or otherwise) in a less structured format (narrative-style in paragraph). We will want to capture these citations using a <bibl> tag when at all possible. For example:

```
<p>Today, only two of her novels apparently survive in print or microfilm:
    <bibl>
        <title level="m" ref="http://purl.dlib.indiana.edu/iudl/vwwp/VAB1828">Going West,
        or, Homes for the Homeless</title>, published in <date>1881</date></bibl>,
    and <bibl><title>Worth Wins</title>, published in <date>1882</date></bibl>.
    The title-pages of these novels list her
    other works: <bibl><title level="m">Here and Hereafter</title></bibl>,
    <bibl><title>The Widow's Life</title></bibl> ...</p>
```

If the bibliographic information is too interspersed with explanatory text, then just capture the main bibliographic element like title. For example, a title may be mentioned followed by lots of information before a publication date is mentioned. The title and date are too far removed to place a <bibl> around the whole chunk. Instead, just use <bibl> <title> and ignore the date.

When citing an article, you may encounter any number of bibliographic information, but author name and page numbers are most common:

```
<bibl>( <author>Avery</author> <biblScope type="pp">18</biblScope> )</bibl>
```

<biblScope> is used with the "type" attribute for the following elements of a citation:

- pp = pagination
- chap = chapter title
- issue
- ll = line numbers
- part
- vol

Please note that these inline citations can occur within the main part of the essay or as part of foot or end notes. Regardless of where they appear, they should be encoded using <bibl>.

<Title>

Titles can, and should when possible, be differentiated with the use of the "level" attribute:

```
<title level="m" ref="http://purl.dlib.indiana.edu/iudl/vwwp/VAB1828">Going West,  
or, Homes for the Homeless</title>
```

The TEI provides the following values for "level:"

- m = (monographic) monographic title (book, collection, or other item published as a distinct item, including single volumes of multi-volume works)
- a = (analytic) analytic title (article, poem, or other item published as part of a larger item)
- j = (journal) journal title
- s = (series) series title
- u= (unpublished) title of unpublished material (including theses and dissertations unless published by a commercial press)

If you aren't sure of the type of title, refrain from using @level.

Please note that titles do not require a special rend attribute unless a level is unknown and the source text italicizes a title. Based on the @level, we know whether to display the title in quotes or italicized.

Bibliographies

Most if not all of these essays will have a section for Works Cited or Bibliographies. These should be contained within a division of type="bibliography" and be encoded as documented in the [Back Matter#bibliography](#) section of the VWWP guidelines.

Encode titles with "level" attributes when possible (see above). Also encode all elements of a bibliographic citation including:

- publisher
- place of publication
- dates
- volume/issue (biblScope)
- etc.

Note that some bibliographies may be annotated. If so, use the <note> element to capture this information within a <bibl>.

[return to top](#)

Next Steps

Once the essays have been encoded, **the editors of the VWWP project should:**

1. Preview the essay, proofread and review the markup
2. Mine the content to update the [VWWP Timeline](#) (encoding guidelines for the Timeline are still pending).
3. Update the source texts to point to the bio and critical introductions (currently part of our [QC process](#)).

The encoder should email [Michelle Dalmau](#) and [Angela Courtney](#) once they have uploaded the **final** version of the encoding to Xubmit.

[return to top](#)