

Encoding Guide for Annotated Documents

These are draft guidelines for the encoding of static multimedia documents, with particular focus on documents with manuscript annotations, and using the test ground of documents annotated by Walt Whitman. The guidelines were developed under NEH Grant HD-50235-07, “Interface Development for Static Multimedia Documents,” 2007-08, by Brett Barney, Matt Cohen, and Erica Fretwell and by Brett Barney, Matt Cohen, and Nicole Gray under an NEH Humanities Collections and Reference Resources grant, “Walt Whitman’s Annotations,” 2011-2014. Except for the instances listed here, markup of manuscripts and print contents should be done in conformity with the extant *Walt Whitman Archive* Encoding Guidelines: <http://www.whitmanarchive.org/about/guidelines/index.html>.

I. Header

The header is done in conformity with the standard WWA TEI header, keeping in mind that the following elements may need special attention:

Identification Numbers

In `<publicationStm>`, each document transcribed must have an identification number. In the header template, the identification number `<idno>` is indicated as `duk.#####`. To enter the id number of your file, refer to the Whitman Archive Manuscript Marginalia spreadsheet. Enter the number in the "WWA_ID" column. Hence, if the document you are encoding is "Reading Shakespeare," you would enter the id number as `<idno>duk.00378</idno>`. Transcriptions of documents held at other repositories will have other codes, beginning with the first free ID number after those in the poetry manuscript database.

Location Finding

In `<sourceDesc>`, you will need to find the document's bibliographic finding number. This is the code used by the repository holding the document. Refer to the WWA tracking list spreadsheet to locate the "Finding #" column. Thus, the document "Reading Shakespeare" would be identified as `<idno type="callno">II-5 102</idno>`. If no number is available in the tracking spreadsheet, leave this space empty.

Source Information

In `<sourceDesc>`, you may need to include additional `<bibl>`s for base text or pasteons that are clippings. Add as much information as you can find, assigning the additional `<bibl>`s an `xml:id` and type to indicate either “pasteon,” for clippings that have been pasted on, or “base,” for base documents. A listing for a pasteon might read,

```
<bibl type="pasteon" xml:id="r2">
  <author><!-- Add paste-on author, if known. --></author>
  <title level="a"><!-- Article title, if known. --></title>
  <title level="j"><!-- Journal title, if known. --></title>
  <date></date>
  <biblScope type="vol"></biblScope>
  <biblScope type="pp"></biblScope>
  <note type="project"></note>
```

</bibl>

HandNotes

<profileDesc> is the place where we can indicate the main authorial hands responsible for the document. Under <handNotes>, indicate a <handNote scribeRef="#ww">, with scribeRef pointing to the author xml:id (assigned in the titleStmt) because it identifies the person responsible for the hand (Whitman, in this case). The medium is described by noting whether Whitman wrote in pencil or in ink, and the color of ink; hence, <handNote scribeRef="#ww" medium="black-ink">. It is necessary to provide an xml:id for Whitman's hand (which we will call "h1" standing for hand 1) so that his hand can be referenced as needed throughout the document. If other hands are to be indicated (to include unknown hands that have written codes on the document as well as hands more or less contemporary with Whitman's such as Bucke's or Traubel's), they will be added to the list of handNote scribeRefs. Different media in Whitman's hand and typescript documents should also be added to the list as separate hands. A common <handNote> tag group might read:

```
<handNotes>
<handNote xml:id="h1" scribeRef="#ww" medium="black-ink"/>
<handNote xml:id="h2" scribeRef="#ww" medium="red-ink"/>
<handNote xml:id="h3" scribeRef="#unk" medium="type"/>
</handNotes>
```

II. Describing the Physical Features of the Document

Following the header is the description of the document proper. We follow the example of the extant Walt Whitman Archive Encoding Guidelines in establishing the structural elements of the XML document, with variations as follows:

The <text> type of each document we encode should be listed as "marginalia."

Nested inside the <body> of the <text> element is <pb/>, which assigns each page an xml:id, notes the image file, listed as facs, and indicates whether the page is a recto or verso.

For example:

```
<text type="marginalia">
  <body>
    <pb xml:id="leaf001r" facs="duk.xxxxx.001" type="recto"/>
    <div type="section">
      <head type="main-derived"><handShift new="#h1"/>DOCUMENT
      TITLE</head>
      <p part="A"> TEXT</p>
    </div>
```

Title Page

For longer works with notations on the title page, the title page material should be included in a segment that appears before <body>, titled <front>. This section should not include any <div>s. The following is an example of a title page encoding structure:

```

<text type="marginalia">
<front>
  <pb xml:id="leaf001r" facs="loc.03449.001.jpg" type="recto"/>
  <titlePage>
    <note type="authorial" place="top right" resp="#h1">Walt Whitman</note>
    <docTitle><titlePart type="main"><handShift new="#h2"/>OCCASIONAL
    PIECES OF POETRY.</titlePart></docTitle>
    <byline>BY <docAuthor>JOHN G. C. BRAINARD</docAuthor>.</byline>
    <milestone unit="undeclared" rend="horbar-short-center"/>
    <epigraph><cit><quote who="Bunyan"><l>Some said, "John, print it;" others said, "Not
    so;"&#8212;</l>
    <l>Some said, "It might do good;" others said, "No."</l></quote>
    <bibl><hi rend="italic">Bunyan's Apology.</hi></bibl></cit></epigraph>
    <milestone unit="undeclared" rend="horbar-short-center"/>
    <docImprint><pubPlace>NEW-YORK:</pubPlace>
    PRINTED FOR <name>E. BLISS AND E. WHITE.</name>
    <hi rend="italic">Clayton &#38; Van Norden, Printers.</hi>
    <milestone unit="undeclared" rend="horbar-short-center"/>
    <docDate when="1825">1825</docDate>.</docImprint></titlePage></front>

```

Divs

See the Walt Whitman Archive Encoding Guidelines for an explanation of when to use `<div>`, which requires a type attribute. `<div>` marks intellectual units of a text. If the document you are encoding has multiple discernable intellectual units, separate them with `<div>`s; otherwise (if the document seems to be a single intellectual unit), omit `<div>` completely. Each `<div>` needs an `@type`: this can be “pasteon,” “notes,” “section,” for miscellaneous structural divisions; “verso,” for unrelated verso material; or “letter,” for distinct intellectual content in the form of a letter.

Structure of Printed Text

Printed text with Whitman’s annotations covers a wide array of generic conventions. Some of these are consistent with modules used by the Whitman Archive; some are not. Should you run across a structure that does not seem to fit the available elements, ask an editor whether there are additional TEI tags available that would be appropriate to that document.

HandShifts

In a document with multiple hands, note changes from one hand to the next with the `<handShift>` element. Each `handShift` should specify the new hand with the “new” attribute. So, for instance, you would encode a shift from hand #h1 to hand #h2 as `<handShift new="#h2"/>`.

1. To designate a starting hand, begin with a `<handShift>` inside the first structural element (see example above).

2. `<handShift>` will only apply when the parent element has substance, or text. So for ``, if the deletion is an overstrike and the overstrike is the only thing in a different hand, you would use the hand attribute within the ``, rather than a separate `handShift`. For example, if the word “contract,” written in hand #h1, was deleted by hand #h2, you would encode it as `<del`

`rend="overstrike" hand="#h2">contract`, whereas if “contract” was an insertion in hand #h2, while the rest of the text was in #h1, the encoding would look like: “The social `<add rend="insertion" place="supralinear"><handShift new="#h2"/>contract</add><handShift new="#h1"/>` is his natural liberty...”

3. Notes include a `resp` attribute that specifies the hand they are written in (`<note type="authorial" resp="#h2" place="top">`), so you do not need to include `handShifts` for notes.

Adding Pasteons

Pasteons should be included in the encoding approximately in the sequence that they are encountered in the document, when possible. However, if the pasteon is flippy, or if it covers existing text on the base leaf, encode the entire leaf (noting the segments missing or covered because of the paste-on as `<gap reason="covered">`), and add the pasteon(s) with the structure described below after the entire verso of the base layer has been encoded.

1. `<Add>` takes the "rend" attribute instead of the "type" attribute. So, for example, where you once encoded a paste-on as `<add type="pasteon">`, you would now encode it as `<add rend="pasteon">`.

2. Within the `<add>` element for clippings, note the source of the clipping, which should reference the appropriate `<bibl>` in the header.

Thus, an added clip should start off looking something like this:

```
<add rend="pasteon" source="#r2">
```

3. Every pasteon that is `<add>`ed will require a `q`, which will either be a "written" or "spoken" type (or, if you're not sure, "ambiguous."), as well as a `floatingText` element. `@rend` on `floatingText` should designate whether the pasteon is “flat” or “flippy.” Ex:

```
<add rend="pasteon" source="#r1">
  <q who="unknown" type="written">
    <floatingText rend="flippy">
      <body>
        <pb xml:id="leaf002r" facs="duk.xxxxx.003" type="recto"/>
        <p>TEXT</p>
```

4. You can indicate material on the front and back sides of flippy documents with `<pb/>`s, as you would for a standard leaf.

Thus, an added clip that IS a flippy should start off looking something like this:

```
<add rend="pasteon" source="#r1">
  <q who="unknown" type="written">
    <floatingText rend="flippy">
      <body>
        <pb xml:id="leaf002r" facs="duk.xxxxx.003" type="recto"/>
        <p>TEXT</p>
        <pb xml:id="leaf002v" facs="duk.xxxxx.004" type="verso"/>
```

<p>TEXT</p>

5. If the material that is on the back side of a flippy document is unrelated, you may omit the content (consult editors with questions about this). Maintain the “flippy” designation, and the <pb>, but for the verso simply include an empty <div> tag of type=“verso”, as follows:

```
<add rend="pasteon" source="#r1">
  <q who="unknown" type="written">
    <floatingText rend="flippy">
      <body>
        <pb xml:id="leaf002r" facs="duk.xxxxx.003" type="recto"/>
        <p>TEXT</p>
        <pb xml:id="leaf002v" facs="duk.xxxxx.004" type="verso"/>
      <div2 type="verso"></div2>
    </q>
  </add>
  <p>TEXT</p>
```

6. If a paragraph or other structural unit is split across two paste-ons, note the continuation using an indication of <prev> and <next>:

```
<p xml:id="p1a" next="p1b">
</p></body></floatingText></q></add>
<add rend="pasteon" source="#r1">
  <q who="unknown" type="written">
    <floatingText rend="flippy">
      <body>
        <pb xml:id="leaf002r" facs="duk.xxxxx.003" type="recto"/>
        <p xml:id="p1b" prev="p1a">TEXT</p>
```

Page Numbers

In general, if they are consistent over a series of pages, page numbers can be included in the <pb> tag as follows:

```
<pb xml:id="leaf002r" facs="duk.00182.002" type="recto" n="2"/>
```

If there is only a single number on a page, however, or if you are not sure whether it is a page number, or if the number appears in an unusual format (e.g. 2nd, 2d), you can put it in a note instead:

```
<note type="authorial" resp="#h1" place="top right">2d</note>
```

Notes

Notes require the “type” attribute (either authorial or editorial), an indication of the hand or the author in “resp,” and an indication of “place” if possible (left, right, top, bottom, inline, supralinear, &c.). **Only notes added by the Whitman Archive should be listed as editorial.** Notes should be encoded differently depending on whether or not they have a clear target. If the note does have a clear target in the text (if the text is bracketed, for instance), you can point to the bracketed or underlined text in the note. An example of a note that does have a clear target, where the targeted text has been identified as a with an xml:id, is:

```
<note type="authorial" resp="#h2" place="right" target="#s1">Egotism</note>
```

If the note has a numbered inline reference, you can assign the note an xml:id that you point to from the reference with an indication that the note is anchored. If the note is a footnote, there is no need to indicate resp. You might encode the two parts as follows:

```
<ref target="#n1" xml:id="ref1">(1)</ref>
```

...

```
<note xml:id="n1" target="#ref1" type="authorial" place="bottom" anchored="true">
    &#8212;(1.) The ablest researches into public rights often simply
    consist of the history of past abuses, and we bewilder ourselves
    to no purpose when we take the pains to study
    them deeply.&#8212;<hi rend="underline">Traite des Interests,
    France</hi>; Marquis Argenson.&#8212;This
    is precisely what Grotius has done.&#8212;</note>
```

Include notes as closely as possible to where they appear in the text. You can nest notes in structural elements, like <p> or <ab>, as necessary. You should specify a “place” if possible. For example:

```
<note type="authorial" resp="#h1" place="top">Assyria &#38; Egypt</note>
```

If the note has been circled, underlined, or otherwise marked in the same hand that the note itself is in, you can note that within <note> using rend:

```
<note type="authorial" resp="#h1" place="right" rend="circled">Is this so?</note>
```

If the circle or underline is in a different hand, you can note that by nesting a <hi> within the <note>:

```
<note type="authorial" resp="#h1" place="top"><hi rend="circled" hand="#h2">Assyria &#38;
Egypt</hi></note>
```

Brackets and Underlines

Brackets around text (print or manuscript) should be encoded as . For the element, you will need to indicate what type of bracket it is (“bracket-left” or “curlyBrace-right”), designate “hand” as necessary, assign an xml:id, and provide an indication of the material that is bracketed, as follows:

```
<span rend="bracket-right" hand="#h1" xml:id="s1" from="#s1" to="#s2"/>
```

```
    Once having become separate
    from hieroglyphics, alphabetic writing
    itself underwent numerous differentiations
    &#8212;multiplied alphabets were produced;<anchor xml:id="#s2"/>
```

For a segment that includes a structural unit with an xml:id, you can omit the xml:id in the and point. The attribute “from” specifies the beginning of the passage being annotated; if not accompanied by a “to” attribute, it specifies the entire passage, as follows:

```
<span rend="bracket-left" hand="#h1" from="#lg1"/>
```

```
<lg xml:id="lg1"><l>Envy wears the mask of love, and, laughing
sober fact to scorn,</l>
```

```
<l>Cries to weakest as to strongest, "Ye are equals,
```

equal-born."</l></lg>

If text has brackets on both sides, you can nest the s to reflect that, as follows:

```
<span rend="bracket-left" hand="#h1" from="#lg1"/><span rend="bracket-right" resp="#h2"
from="#lg1"/>
```

```
<lg xml:id="lg1"><l>Envy wears the mask of love, and, laughing
sober fact to scorn,</l>
```

```
<l>Cries to weakest as to strongest, "Ye are equals,
equal-born."</l>
```

```
<anchor xml:id="s2"/><anchor xml:id="s4"/>
```

Annotation documents contain a wide array of brackets. See Appendix II (Bracket Taxonomy) for a series of examples of how to encode different brackets. If you run into something that doesn't seem to match any of these examples, ask an editor.

Underlines should be marked using the <hi> element. Text marked as underlined would look like <hi rend="underline" hand="#h1">. If you need to point to underlined text, in the case of a corresponding note, you should add an xml:id within the <hi> (<hi rend="underline" hand="#h1" xml:id="u1">).

Formeworks and Running Heads

For printed texts, you can indicate the running heads or formework material just within the <pb> as follows:

```
<pb xml:id="leaf001r" facs="duk.00400.001" type="recto"/>
```

```
<fw type="date" place="top left">1847.]</fw>
```

```
<fw type="pageNum" place="top right">533</fw>
```

```
<fw type="header" place="top"><hi rend="italic">Washington and his Generals.</hi></fw>
```

Unusual Marks

For unusual marks, refer to the Guidelines for the Unicode abbreviation. If you cannot locate the Unicode abbreviation in the Guidelines, check <http://www.biega.com/special-char.html>. If the unusual mark is an ink spot or other unclassifiable marking, describe it in a comment.