

Encoding Guidelines 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-50236-07, "Interface Development for Static Multimedia Documents," 2007-08, by Brett Barney, Matt Cohen, and Erica Fretwell. 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 TEI header, keeping in mind the following elements:

Identification Numbers

In <publicationStmt>, each document transcribed must have an identification number. In the header template, the identification number <idno> is indicated as duke####. 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>duke.00378</idno>

Location Finding

In <sourceDesc>, you will need to finding the document's bibliographic finding number. Refer to the WWA tracking list spreadsheet to locate the "Finding #" column. Thus, the document "Reading Shakespeare" would be identified as <idno type="location">II-5 102</idno>

HandNotes

<profileDesc> is the place where one indicates the main authorial hand responsible for the document. Under <handNotes>, one indicates a <handNote scribe="Walt Whitman"> because it identifies the person responsible for the hand. The medium is detailed by noting whether Whitman wrote in pencil or in ink; hence, <handNote scribe="Walt Whitman" medium="ink">. Moreover, it is necessary to provide an xml id for Whitman's hand (hand 1 or "h1") so that his hand can be referenced as needed throughout the document. If other hands are to be indicated, they will be added to the list of handNote scribes. Thus far, the entirety of handNote reads:

```
<handNotes>
  <handNote scribe="Walt Whitman" medium="ink" xml:id="h1">
  </handNote>
</handNotes>
```

II. Describing the Physical Features of the Document

Following the header is the description of the document proper. The TEI P5 guidelines recommend use of the <facs> tag. However, we found this structure to be inadequate for capturing the complex relationships of moveable surfaces in Whitman's annotations. Instead we will mark up the physical structures as structural elements of the xml document. The tags associated with these structures are: <page>, <graphic>, and "x" (the tag for the coordinates).

The text type of each document is given the attribute "manuscript." Nested inside the <body> of the <text type> element is <page>. The <page> approach complicates the markup of intellectual structures. To allow the use of <divs> for capturing major intellectual sections we have implemented a special attribute structure. <Page> tags correspond to any leaf, whether referring to the support for a document or the pasteon attached to a support <page> tag that then can be nested. The <graphic> tag points to the image file associated with a <page> entity. Most <page> tags will have a <graphic> element associated (for instance, if the document is flippy); however, it is not required. Attributes for the <page> element include: "whole" for a side of a leaf that serves as a support; "pasteon" for a paste-down; and "flippy" for a moveable leaf attached to a support. Coordinate attributes may be attached to structures of the types "pasteon" or "flippy." <Page> tags of the type "flippy," as well as graphics associated with it, require coordinate attributes. Each <div> is associated with an attribute "section" that is either numbered or lettered. The intellectual framework of this structure allows us to use <divs> in a non-hierarchical fashion, yet still links the <divs> together across the <page>.

For example:

```
<text type="manuscript">
  <body>
    <page type="whole">
      <graphic url="duk.44444.001.jpg">
        <div type="section" part="A">
          <head type="main-derived">DOCUMENT TITLE</head>
          <p> TEXT</p>
        </div>
```

Adding Clippings

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

2. Every clipping that is <add>ed will require a quotation, which will either be a "written" or "spoken" type (or, if you're not sure, "ambiguous." Ex: <q who="unknown" type="written">

3. Clippings are to be encoded as floatingText types, rather than text types. Hence <floatingText type="clipping">.

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

```
<add rend="pasteon" place="inline">  
  <q who="unknown" type="written">  
    <floatingText type="clipping">  
      <body>
```