TEI, the Walt Whitman Archive, and the Test of Time

Brett Barney

Electronic version
URL: https://journals.openedition.org/jtei/3249
DOI: 10.4000/jtei.3249
ISSN: 2162-5603

Publisher
TEI Consortium

Electronic reference
Brett Barney, “TEI, the Walt Whitman Archive, and the Test of Time”, Journal of the Text Encoding Initiative [Online], Issue 13 | May 2020 - November 2022, Online since 25 February 2021, connection on 24 November 2022. URL: http://journals.openedition.org/jtei/3249 ; DOI: https://doi.org/10.4000/jtei.3249

For this publication a Creative Commons Attribution 4.0 International license has been granted by the author(s) who retain full copyright.
ABSTRACT

“Diachronic Markup and Presentation Practices for Text Editions in Digital Environments,” a 2014–15 transatlantic collaboration funded jointly by DFG and NEH, provided an occasion to test TEI’s relatively new provisions for encoding temporality. My experiences in attempting to apply these provisions to Walt Whitman manuscripts led me to conclude that chapter 11 of the TEI Guidelines should be significantly revised and that the basic dichotomy it introduces between <sourceDoc> and <text> deserves renewed scrutiny. Some of the problems in the chapter can be traced to the way the Guidelines have evolved over time through a series of choices motivated by expediency. The Whitman Archive has, since its first foray into TEI markup in 2000, been anxious to find a way to leverage its potential to encode the temporal characteristics and relationships among various manuscript and print instances. The customization we created, however, has shown itself inadequate to describe and make processable important genetic information. Unfortunately, TEI also continues to provide inadequate guidance in the area of encoding genetic relationships, both
between documents and within them. My experiments during the Diachronic Markup project included using `<listChange>` to encode the sequence in which the text of a single document was inscribed and show that, although the Guidelines need to be developed with specific advice and examples, existing elements and attributes can be used to posit claims about the “way the page was filled,” thereby enabling useful processing and display.

INDEX

**Keywords:** scholarly editing, manuscripts, genetic editions, temporality, diachronic markup,

ACKNOWLEDGEMENTS

This essay has been substantially revised since I presented it at the 2018 TEI Conference in Tokyo. I am very grateful to Gerrit Brüning, who has offered wise and generous insights over the course of several years as we have discussed many of the specific issues I raise here. I have also attempted to benefit from the input of two anonymous reviewers, whose comments were at once kind, thoughtful, and incisive.

1 The release of version 2.0 of the P5 Guidelines in late 2011 represented a culmination of sorts for the efforts, over the course of several years, of the Workgroup on Genetic Editions. The group’s stated goal was “to develop an Application Profile for the encoding of genetic editions and, in general, genetic phenomena” (Burnard et al. n.d.). In the words of the Council members who wrote the release notes, P5 v. 2 included “several new elements ... to facilitate a more ‘document-focused’ (as opposed to ‘text-focused’) way of working.” The most dramatic innovation was the introduction of the `<sourceDoc>` element, which as an analog to the `<text>` element can be used to group both “images of a source document” and “detailed ... transcriptions of the source.” To explain the ways in which one might undertake such a representation, chapter 11, “Representation of Primary Sources,” was “extensively revised” to include the new provisions, which were intended to “improve ... support ... for documentary and genetic editing” (TEI Consortium 2011).

2 Having used TEI in my work for the Walt Whitman Archive (whitmanarchive.org) since 2000, I had been observing the progress of the Workgroup with great interest and had, in fact, been involved with them to an extent, first as one of many digital editors convened in 2009 to
provide input and later as a member of the TEI Council when the Workgroup’s proposals were being debated and eventually implemented. I was therefore enthusiastic when the editors of the Whitman Archive were approached to contribute material to a two-year project with the no-nonsense title “Diachronic Markup and Presentation Practices for Text Editions in Digital Environments,” which was to be proposed to the National Endowment for the Humanities and the Deutsche Forschungsgemeinschaft. Submitted by Anne Bohnenkamp-Renken and Malte Rehbein, the proposal states that the main purpose of the transatlantic project was “to test, improve, and make usable for a broad audience the TEI Guidelines for diachronic markup” (2012). Through a series of strange events I found myself and the Whitman Archive closer to the center of this project than I had anticipated, as I became the United States–based co-principal investigator when Rehbein relocated to Germany shortly before the project started.

The work and outcomes of the Diachronic Markup project have been the subject of several previous presentations. Gerrit Brüning, Katrin Henzel, and Dietmar Pravida, who were engaged with Bohnenkamp-Renken in creating a digital edition of Johann Wolfgang von Goethe’s Faust, have offered especially insightful reflections on their experiences in testing TEI’s then-new markup (2013). Joshua Schäuble and Hans Walter Gabler have also discussed the project and some of the interfaces that grew out of it (2018). One sobering, though instructive, outcome of our collaboration was a clearer realization of the profound differences among our individual editorial projects, all of which deal with corpora of manuscripts by single authors of the modern era. Although each of the participants was committed to the collective goal of using TEI to encode the temporal dimensions of their author’s writings, the diversity of practices among those various authors meant that each corpus presented unique kinds and amounts of evidence for textual change across time. Furthermore, the editors’ diverse theoretical and methodological commitments inclined them toward heterogeneous emphases in presenting the materials to readers and, therefore, often toward different encoding approaches.

It was my belief that the main value of our project lay not in producing an edition of Whitman’s or anyone else’s work but in its potential to advance the capacity of TEI markup for encoding genetic (and not simply “diplomatic”) information. And over the course of the grant period I did indeed formulate definite opinions about how P5 v. 2 could be revised so as to better serve the needs of those wishing to foreground the time-inflected aspects of the texts they are editing. I had occasion
to carefully reread chapter 11 of the TEI Guidelines several times and to discuss it with colleagues on the Faust project. Based on those readings and discussions, Brüning and I submitted a series of revision suggestions to the TEI Technical Council. Many of these revisions were what might be called, in TEI parlance, “phrase-level” tweaks and corrections that were implemented quickly and without need for much discussion. Others, however, involved at least chunk-level units. The most radical suggestion and the one that has generated the most discussion and debate is titled, in the TEI GitHub repository, “Restructure chapter 11” (issue no. 1427). The process of reading the Guidelines, testing what they recommend, discussing results, and then proposing changes (which have sometimes been further discussed and negotiated) has resulted or will likely result in incremental improvements that, taken together, should make the Guidelines significantly more usable. That same process, however, has also convinced me that for the TEI Guidelines to become truly adequate to the task of creating digital genetic editions of various kinds, the content models for <sourceDoc> and <text> should be thoughtfully rationalized, so as to disentangle (or at least acknowledge the entanglement of) markup most appropriate to the document (<sourceDoc>) and that most appropriate to the text (<text>). Doing so would require that the TEI Council be willing at least to consider the full range of options, not excluding those that would break backward compatibility. I say this knowing only too well that radical change is a sort of third rail for TEI, where, at least in recent years, incremental evolution has become the modus operandi. In most cases I believe this institutional conservatism to be a positive aspect of TEI and, in fact, important to its continued use and adoption, but in this case a fresh, large-scale reconsideration rather than a localized tinkering is in order, as it could help to solidify TEI’s theoretical grounding, or at least better articulate it, for the benefit of both present and future practitioners.

With regard to temporal encoding specifically, however, there is ample scope to “build out” the Guidelines’ facilities without attempting anything so fundamental and admittedly controversial as the redefinition of <text> and <sourceDoc>. In what follows, therefore, I will focus on chapter 11’s provisions for encoding a text’s diachronic dimensions, by which I mean its temporal characteristics both within a single document (that is, the posited sequence in which the textual content of a particular document was inscribed) and those between and among multiple documents (the time-sequenced relationships of text across the various components that an editor
sees as constituting its genetic dossier). During the Diachronic Markup project we sometimes talked about these two related but distinct types or levels of diachronic relationship as the “micro” and the “macro,” respectively.

Much of the weakness of TEI’s provisions for conducting genetic editing generally, and for its inadequate address of the diachronic dimensions of texts specifically, can be traced to the peculiar historical circumstances in which chapter 11 has developed. I would suggest, then, that a consideration of the genesis of the Guidelines themselves is in order, so that ongoing discussions of their remediation can proceed with a shared understanding of their historical context.

A second, closer look at the P5 v. 2 release notes reveals several significant details. First, it is worth noting that although the Workgroup carried out its remit under the banner of genetic editing, the innovations that resulted from that work were largely aimed at, and conceived in terms of, better equipping TEI for the purposes of producing narrowly documentary transcriptions, with only a brief nod to an overarching genetic purpose: the lead story in the release notes is that “[s]everal new elements have been introduced to facilitate a more ‘document-focussed’ (as opposed to ‘text-focussed’) way of working” (TEI Consortium 2011). Of course, this characterization relies on the dichotomy between the semantic, intellectual signal (viz., the text) and the material, spatial properties of the carrier (viz., the document). In the French tradition of genetic criticism, diplomatic transcriptions have been a central feature. It is probably for this reason that the transcriptions enabled by the revisions introduced in P5 v. 2 are frequently termed “genetic,” even though they are for the most part concerned with the spatial layout of text without any necessary regard to genetic relationships. Also in this regard, I would point out that the release notes mistakenly (though tellingly) refer to the new container element as “<document>,” making quite clear the connection between documentary editing and the sort of transcription provided for. The release notes’ explanation of the primary function of this new element is similarly telling: “to organize digital images” in the way previously allowed by <facsimile>, but now possible in tandem with a transcription.

Another instructive oddity of the release notes is their surprising reference to chapter 11 as the “chapter on Physical Transcription,” a slip that unwittingly evokes an earlier state of the Guidelines: The current chapter 11, “Representation of Primary Sources,” began, in the P3 version of the TEI Guidelines, as chapter 18, “Transcription of Primary Sources” (Sperberg-McQueen and Burnard
The change from “transcription” to “representation” was necessitated by the inclusion, in the 2007 release of P5 v. 1, of the new elements `<facsimile>` (with its descendants `<surface>` and `<zone>`) and `<msDesc>` (with its descendants `<msPart>`, `<physDesc>`, etc.) (TEI Consortium 2007). Members of the TEI community had long lamented the Guidelines’ deficiencies for doing codicological work, a shortcoming that the P5 v. 1 release began to address by adding both an entirely new chapter, “10 Manuscript Description,” and a relatively modest section at the beginning of what had been chapter 18, now retitled “11 Representation of Primary Sources” (TEI Consortium 2007). In its earlier incarnation this chapter was explicitly restricted in focus to “the transcription of primary sources, particularly manuscripts” (TEI Consortium 2002), but it was now expanded to demonstrate the use of `<facsimile>` as a way to provide a nontranscriptional representation of the object of interest either alone or as complement to a transcription.

Having broadened the purview of the chapter to accommodate such a treatment, the Council no doubt viewed the subsequent request for a method to allow specific components of a transcription to be topographically located—a need of those wishing to create more faithful diplomatic transcriptions—as most easily accomplished by a further retooling of the existing `<facsimile>` and associated elements, along with the addition of several new elements, such as `<metamark>`, `<transpose>`, and `<retrace>`. In this way, the chapter took on an emphasis on layout not so much through careful planning and building out of what was embryonically present to begin with, but by a sequence of choices characterized by expediency.

I recount these aspects of the historical development of chapter 11 not simply to bemoan problems that this evolution has wrought. Rather, I suggest that not only is there a logical explanation for the continuing (though unfortunate) lack of solid guidance for those wishing to encode the temporal aspects of a text’s coming-into-being, but that the chapter’s developmental history has helped to obscure the lack. In short, although chapter 11 was chosen as the location of the revisions that grew out of the Workgroup on Genetic Editions, and although its contents are largely discussed in reference to genetic editing, the chapter brings with it a continuing slant toward physical layout, with provisions for encoding temporality still relatively poorly developed.
Like most other TEI users, my interest in the development of TEI is profoundly conditioned by an investment in its use for a specific project. In the past two decades, much of my work has been devoted to using TEI, in its various iterations, to edit manuscripts for the Whitman Archive. Whitman was—and is—many things to many people: formal innovator of free verse, one of the first “truly great” American writers, pioneer of literary gay identity, poet of the working person, and master of self-promotion, among others. Certainly part of Whitman’s fame, in both his time and our own, can be attributed to the creation of an unusual public persona. Ed Folsom has claimed that “no author’s life in the nineteenth century was more continuously photographed than Whitman’s” (1986, 2–3), and it was partly through the widespread distribution of the photographs he particularly liked that Whitman worked to construct what he believed to be his best public self. In a similar vein, I have come to believe that every presentation regarding the editing of Whitman’s works should make an opportunity to present the following image, which provides visual evidence that the poet was accurate in describing his living situation: “The floor ... is half cover’d by a deep litter of books, papers, magazines, thrown-down letters and circulars, rejected manuscripts, memoranda, bits of light or strong twine, a bundle to be ‘express’d,’ and two or three venerable scrap books” (Whitman 1892, 517).
At least some of what I like about this photograph is germane to the present topic. Most obviously, it presents an arresting visual emblem of both the volume and the disarray of the record that editors of Whitman have to confront. Figure 2 shows another favorite image, a graphic created for the Whitman Archive in its early days, before I began my association with it (and before the editors adopted TEI).
One of the ideas that this image vividly conveys is the multiplicity of Whitman’s poetry collection, *Leaves of Grass*. In 1855 *Leaves* was an oddly tall volume of about one hundred pages, containing twelve poems and a long preface in prose. The last incarnation, which Whitman prepared shortly before his death in 1892, was a book more conventional in appearance, containing nearly four hundred poems on over four hundred pages and a (different) long epilogue in prose. Throughout the three and a half decades between those first and last versions, Whitman’s masterwork was
always a protean entity, moving through a series of transformations. Poems were revised; poems were added; poems were deleted; poems were retitled. The initial prose preface was turned into poetry. Poems were combined or shuffled into new groupings. Groupings were added, discarded, or retitled.

One of the first conversations that I can recall from around the time that I joined the Whitman Archive in 2000 concerned the hope of leveraging the power of the computer to analyze these various and multiple ways in which *Leaves of Grass* and its constituent works transmuted over time, across various manuscript and printed instantiations. At least part of the appeal of using the Text Encoding Initiative’s scheme (embodied in P3, which I came to call “the green books”: see Sperberg-McQueen and Burnard 1994) was its potential utility for such analysis. We found, however, that the markup they offered was inadequate to the task of indicating what seemed to us the most basic facts about the relationships among the various iterations of Whitman’s masterpiece: namely, that a certain poem in one edition developed from a certain poem in a previous edition or manuscript. Our desires to represent such relationships prompted a series of project-wide discussions that eventually led to the decision to add to our TEI DTD the custom elements `<relations>` and `<work>`, with the goal of specifying such genetic relationships. These elements we included in our transcriptions of textual instances. Here, for example, is a relevant snippet, slightly simplified, from one of our TEI-encoded transcriptions of a manuscript draft:

```xml
<TEI>
  <relations>
    <work ref="xxx.00565" cert="high"/>
  </relations>
</TEI>
```

The key to this system, of course, is the `@ref` attribute, which connects this instance to one of Whitman’s works—in this case, the poem “After the Argument.”

One might get the impression, from the photograph of the old man sitting amid chaotic mounds of papers, that Whitman saved absolutely everything—that the manuscript record of his literary production is complete, or nearly so. From another perspective, of course, the same picture provides clues as to why this is far from true, why the author’s “system” for saving things has
resulted in a less-than-ideal record from which any would-be Whitman editor must work. Still, the genetic dossiers of some of Whitman’s individual poems or essays—especially those that date from his last years—is relatively fully extant. Far fewer draft materials have survived from the 1850s than from the 1880s, and manuscripts that predate the first edition of *Leaves of Grass* in 1855 are fewer still. It is, alas, these earliest manuscripts that are of greatest interest to most Whitman scholars, who would welcome evidence that might shed light on what has been one of the great “mysteries” of Whitman scholarship: what Whitman Archive editors Ed Folsom and Kenneth M. Price have described as “his transformation from an unoriginal and conventional poet into one who abruptly abandoned conventional rhyme and meter and ... exploited the odd loveliness of homely imagery, finding beauty in the commonplace but expressing it in an uncommon way” (n.d.). As Ralph Waldo Emerson famously wrote, the remarkable poems of the 1855 edition “must have had a long foreground somewhere, for such a start,” but we have rather scant concrete evidence about the specific ways in which the text of that edition developed.

Although not many of Whitman’s early poems are attested extensively in the manuscript record, the manuscripts that do exist are often fascinating (sometimes in part because of the riddles that arise from their being partial). For the purposes of the Diachronic Markup project, I was fortunate that at least a passage from one of those first-edition poems exists in multiple manuscript forms. That passage, from the poem eventually titled “The Sleepers,” has the added distinction of being of particular interest to Whitman scholars generally as evidence of Whitman’s treatment of racial issues. At least five different manuscripts related to this passage exist. Four of these are individual leaves in the collection of the University of Virginia; the other is contained in a notebook at the Library of Congress that includes, besides the three consecutive pages of writing relevant to the passage from “The Sleepers,” much writing that is unrelated.

After a brief examination of images of these documents, nearly anyone would be able to surmise that they are related to one another. However, the exact “nature” of those relationships is far less apparent. The finding guide for the collection that houses the four manuscripts at University of Virginia (probably originally written by the famous descriptive bibliographer Fredson Bowers) describes their relationships thus: the first one shown below (figure 3) “seems to have come first or second”; the next (figure 4) “was probably second,” the next (figure 5) “appears to have come third”; and the one in figure 6 “seems to have been the last” (Special Collections, UVA Library 2011).
Figure 3. Manuscript identified as "Page 12. The sores on my shoulders are from his" in the University of Virginia Library finding aid (Special Collections, UVA Library 2011) and as "uva.00260" in the Whitman Archive's "Integrated Catalog of Walt Whitman's Literary Manuscripts" (EAD Project Team, UNL n.d.).

Figure 4. Manuscript identified as "Page 9. I am a curse" in the University of Virginia Library finding aid (Special Collections, UVA Library 2011) and as "uva.00256" in the Whitman Archive's "Integrated Catalog of Walt Whitman's Literary Manuscripts" (EAD Project Team, UNL n.d.).

Figure 5. Manuscript identified as "Page 11. Topple down upon him" in the University of Virginia Library finding aid (Special Collections, UVA Library 2011) and as "uva.00258" in the Whitman Archive's "Integrated Catalog of Walt Whitman's Literary Manuscripts" (EAD Project Team, UNL n.d.).
Figure 6. Manuscript identified as “Page 10. Black Lucifer was not dead” in the University of Virginia Library finding aid (Special Collections, UVA Library 2011) and as “uva.00257” in the Whitman Archive’s “Integrated Catalog of Walt Whitman’s Literary Manuscripts” (EAD Project Team, UNL n.d.).

Only sketchy justifications are offered for these tentative conclusions, the key one being that since the published version in 1855 (see figure 7) contains neither a version of the line “I look off the river with my bloodshot eyes ...” nor any mention of a curse, the deletion of these words on leaf 11 suggests that it represents “a reworking of the lines beginning I am a Curse in the . . . notebook” (see figure 8) combined with “verses added from Leaf 9” (Special Collections, University of Virginia Library 2011).

Figure 7. Detail of page 74 in the first edition of Leaves of Grass (Whitman 1855).
These details, plus the relatively close resemblances between leaf 10 and the published passage, make the scenario outlined in the finding guide seem quite plausible. Nevertheless, other scenarios are also imaginable. In fact, the University of Virginia finding guide itself offers one of them: On leaf 10 “the word Sleepchaser’s appears in the upper right corner,” and this word’s similarity to the title “Sleep-Chasings” (first used in the 1860 edition) might indicate that, instead of inscribing leaf 10 just after leaf 11 and just before the printed version appeared, Whitman wrote leaf 10 after the first printing—that it “is in fact a reworking of the section for the 1860 edition” (Special Collections, UVA Library 2011). Ordering of the other four manuscripts similarly turns out to be not entirely straightforward (as one might infer from the speculation that leaf 9 came either “first or second”). My point here, fortunately, is not to argue for a particular genetic sequence but to illustrate the
fact that the “dossier” with which the genetic editor must work often offers ambiguous evidence regarding the order in which documents were inscribed and, hence, might oblige one to offer multiple genetic descriptions as possibilities.

Our use of the home-grown elements `<relations>` and `<work>` to connect a particular documentary instance (manuscript or print) to a work has proven useful, most notably as the enabling presence behind our “Integrated Catalog of Walt Whitman’s Literary Manuscripts” (EAD Project Team, UNL n.d.), a union finding guide generated from individual EAD-encoded finding guides for several dozen separate collections. While it has been useful, however, our system is in no way adequate to the task of representing the sort of genetic details that I’ve outlined for the “Sleepers” manuscripts. Our encoding simply declares that the document currently under consideration (or some division of that document) is an instance of a specified work. That’s all it does. Of the many things it does not do, perhaps the most fundamental is that it offers no machine-processable way of ordering the materials—of saying that a particular manuscript was the first draft, another was the second, etc.—let alone of indicating alternative possible orderings. I sometimes shake my head at our lack of foresight in neglecting to design such a provision, but then I remember that during the nineteen years we have been using this markup the TEI community has not developed any consensus facility for associating, in a documentary transcription, all of the various instances with a given work. If the Whitman Archive were coming to TEI for the first time today, in late 2020, we would search the Guidelines in vain for advice about encoding the genetic relationships among the various manuscript instances of “The Sleepers.”

For these macro, interdocument diachronies, two basic frameworks possibly worth consideration are RDF and directed acyclic graphs. The second of these was, in fact, suggested as a possibility in the draft recommendations of the TEI’s Workgroup on Genetic Editions (Burnard et al. n.d., sec. 4: The Dossier Level). Either approach, however, would require further development to specify such things as types of relationships, degrees of certainty, and alternative theories. No already existing RDF vocabulary appropriate to describing genetic textual relationships seems to exist; likewise, to my knowledge, nothing like a universally applicable set of values for the components of a directed graph has been devised.
What the TEI Guidelines do have to offer for the encoding of diachronic information pertains to the “micro” level, that is, those traces of the passage of time that exist within the bounds of a single document. In addition to the use of the @seq attribute to specify an order among the components of <choice>, the P5 v. 2 Guidelines introduced a means of characterizing the temporal relationships among a broad range of textual alterations via <listChange>, whose <change> children can be cross-referenced via the @change attributes of the various elements used to mark up blocks of text in the transcription proper. This method (though, importantly, not the <change> element itself) was developed for the encoding of “revision campaigns” (i.e., the inferred rounds/sessions during which specific portions of a single document were inscribed), and the Guidelines discuss the markup of revision campaigns as the primary use of this set of elements and attributes. Of the four authors whose manuscripts formed the corpus under consideration in the Diachronic Markup project (Whitman, Goethe, Virginia Woolf, and James Joyce), Goethe’s writing processes most easily lend themselves to such an approach. Whitman’s individual manuscript documents rarely exhibit traces of having been inscribed in more than one writing session. That is, there are usually no changes of hand or medium to suggest that he made an initial inscription, paused, and then came back at a significantly later time or in a distinctly different “fit” to make changes. Unlike some other authors, he had no consistent habit of using, say, ink during an initial drafting phase and pencil during a proofreading one, nor did he employ scribes, as Goethe, for example, sometimes did. The lack of such typical kinds of evidence for “revision campaigns” per se, however, does not mean that the diachronic dimension is utterly imperceptible within Whitman’s manuscripts, only that the time frame is collapsed. The manuscripts are very frequently a tangle of writing and revision, and therefore pose significant challenges to understanding the processes by which they came into being. On occasion, I have spent hours examining a single manuscript image, slowly developing a theory of how the page was incrementally inscribed. Editors who want to record conclusions and/or posit conjectures about such processes might well wish to encode a sequential ordering of the parts of an inscription within a single “revision campaign.” Although it is possible to use <listChange> and other elements to encode such time-dependent relationships, the Guidelines, as currently written, offer no clear direction for how this should be done.
To illustrate some of the gaps that could be filled in future versions of the Guidelines, as well as some of the potential hermeneutical value of creating such encodings, I hereby offer my own modest experiment in employing `<listChange>` to encode hypotheses regarding “how the page was filled” for a single example manuscript, the one shown in figure 4, above. The following snippet shows part of the `<listChange>` section of the `<teiHeader>`, which asserts the order in which I believe Whitman carried out the actions that produced what is now visible as the first line of the manuscript:

```xml
<listChange>
  <change xml:id="ch01">
    <desc>writes "I am blac"</desc>
  </change>
  <change xml:id="ch02">
    <desc>overstrikes "blac"</desc>
  </change>
  <change xml:id="ch03">
    <desc>writes "a curse:"</desc>
  </change>
  <!-- ... -->
  <change xml:id="ch07">
    <desc>writes "a black slave spoke me"</desc>
  </change>
  <!-- ... -->
  <change xml:id="ch09">
    <desc>overstrikes "black slave" and writes "negro"</desc>
  </change>
  <change xml:id="ch10">
    <desc>overstrikes "spoke" and writes "felt"</desc>
  </change>
  <change xml:id="ch11">
    <desc>overstrikes "felt" and writes "thought"</desc>
  </change>
  <change xml:id="ch12">
    <desc>overstrikes "thought" and writes "thinks"</desc>
  </change>
</listChange>
```
To create this section, I populated `<listChange>` with the series of steps (one per `<change>`) that I believe Whitman took in inscribing the manuscript. For each `<change>` element I then added a `<desc>` containing human-readable prose, so that, taken together, these constitute a version of what John Bryant has called a “revision narrative” (2002, esp. 167–71). Although I was not able to look to the TEI Guidelines for a model of how to create this narrative, I found it relatively easy, by extrapolation, to arrive at a workable and perhaps even adequate approach for this part of the encoding task. The following snippet from the body of the transcription shows my attempt to make this sequence machine-readable, via references in `<change>` to the `<xml:id>` of the relevant `<change>` in `<listChange>`:

```xml
  <seg change="#ch01">I am
    <del rend="overstrike" change="#ch02">black</del>
  </seg>
  <seg change="#ch03">a curse:</seg>
  <seg change="#ch07">a
    <del rend="overstrike" change="#ch09">black slave</del>
    <add rend="unmarked" place="above" change="#ch09">negro</add>
  </seg>
  <del rend="overstrike" change="#ch10"> spoke</del>
  <del rend="overstrike" change="#ch11"> felt</del>
  <add rend="unmarked" place="above" change="#ch10">thought</add>
  <add rend="unmarked" place="above" change="#ch12">thinks</add>
  me</seg>
```

I felt (and feel) less satisfied with this part of the encoding. While it wasn’t difficult to describe a step-by-step order of inscription, the nesting of `<seg>` and `<add>` and `<del>` elements in the transcription caused me anxiety. The change associated with `<xml:id="7">`, which I describe as “writes ‘a black slave spoke me,’” offers an example of my difficulty in translating the bits of
my narrative description into unambiguous encoding. In the transcription excerpted above, the phrase is indicated as a @seg, interrupted by elements embedded at various levels and conceived as the contents of the @seg minus the child <add> elements. Child <del> elements, however, are to be considered part of this change, except when their contents are further marked as an addition. I am uncertain whether this logic is scalable to a larger set of Whitman documents, let alone to documents created by others.

More immediately, however, I was uncertain whether the logic actually was machine-readable, that is, whether it was both consistent and complete enough to allow some sort of algorithmic translation. Specifically, I hoped that two of my colleagues at the Center for Digital Research in the Humanities, Brian Pytlik Zillig and Jessica Dussault, could make use of the temporal information in my encoding to produce a meaningful display of the manuscript text. The grant period was very nearly at an end, so my expectations were modest, but in the few days that remained they did, in fact, create a JavaScript program that successfully uses the code to represent, albeit imperfectly, the temporal sequence of inscription via animation.

Figure 9. Brett Barney, Brian Pytlik Zillig, and Jessica Dussault. Temporal animation of Whitman’s manuscript changes (written in javascript). To view the animation select the original link below.
Astute readers will quickly recognize that this animation doesn’t get everything right, and I remain uncertain how much of the work to perfect it might depend on changes to the encoding and how much on further development of the script. However, when I consider it as purely a proof of concept, I find this to be a surprisingly powerful and satisfying rendering, one that suggests how an editor’s assertions about the time-dependent sequence of inscription within a document might be brought to life with an immediacy that no amount of apparatus criticus prose could.

I believe that work remains to more fully implement the aspirations and recommendations of the Workgroup on Genetic Editions, especially with regard to diachronic markup, and I will continue to advocate for further development of the TEI specification. Future versions of the Guidelines should provide direction for representing macro-level temporal relationships among documents and allow for alternate orderings and indications of levels of certainty, perhaps by leveraging the possibilities of the `<standOff>` element, which was introduced in the February 2020 release of P5 v. 4.0.0. With regard to markup for micro-level diachronic information, I see an even easier—perhaps even non-controversial—opportunity for the Council to elaborate the Guidelines’ prose to show how this can be achieved within the existing scheme. By articulating recommendations and providing encoding examples, TEI could better fulfill an important part of its stated mission to “facilitat[e] the loss-free interchange of data” (TEI Consortium 2020).

BIBLIOGRAPHY

Bellemin-Noël, Jean. 2004. “Psychoanalytic Reading and the Avant-texte.” In Genetic Criticism: Texts and Avant-textes, edited by Jed Deppman, Daniel Ferrer, and Michael Groden, 28–35. Philadelphia: University of Pennsylvania Press.

Bodard, Gabriel. 2020. “Diplomatic Transcriptions.” In EpiDoc Guidelines: Ancient Documents in TEI XML, edited by Tom Elliott, Gabriel Bodard, Elli Mylonas, Simona Stoyanova, Charlotte Tupman, Scott Vanderbilt, et al. Version 9.2. Accessed February 19, 2021. https://epidoc.stoa.org/gl/latest/trans-diplomatic.html.

Bohnenkamp-Renken, Anne, and Malte Rehbein. 2012. “Diachronic Markup and Presentation Practices for Text Editions in Digital Environments.” Unpublished grant proposal, September 26. PDF file.
Brüning, Gerrit, Katrin Henzel, and Dietmar Pravida. 2013. “Multiple Encoding in Genetic Editions: The Case of Faust.” Journal of the Text Encoding Initiative 4. http://journals.openedition.org/jtei/697; doi:10.4000/jtei.697.

Bryant, John. 2002. The Fluid Text: A Theory of Revision and Editing for Book and Screen. Ann Arbor: University of Michigan Press.

Burnard, Lou, Fotis Jannidis, Elena Pierazzo, and Malte Rehbein. n.d. “An Encoding Model for Genetic Editions.” Revised draft. N.p.: TEI Consortium. Workgroup on Genetic Editions (TEI MS SIG). Accessed February 19, 2021. https://www.tei-c.org/Vault/TC/tcw19.html.

Buzzoni, Marina. 2016. “A Protocol for Scholarly Digital Editions? The Italian Point of View.” In Digital Scholarly Editing: Theories and Practices, edited by Matthew James Driscoll and Elena Pierazzo, 59–82. Cambridge: Open Book Publishers. doi:10.11647/OBP.0095.

Cummings, James. 2019. “A World of Difference: Myths and Misconceptions about the TEI.” Digital Scholarship in the Humanities 34 (supplement 1): i58–i79. doi:10.1093/llc/fqy071.

Deppman, Jed, Daniel Ferrer, and Michael Groden. 2004. “Introduction: A Genesis of French Genetic Criticism.” In Genetic Criticism: Texts and Avante-Textes, 1–16. Philadelphia: University of Pennsylvania Press.

Folsom, Ed. 1986. “Introduction: ‘This Heart’s Geography’s Map’: The Photographs of Walt Whitman.” In Whitman Photographs, special double issue, Walt Whitman Quarterly Review 4 (2–3, Fall): 1–5. doi:10.13008/2153-3695.1138.

Folsom, Ed, and Kenneth M. Price. n.d. “Integrated Catalog of Walt Whitman’s Literary Manuscripts.” Walt Whitman Archive, edited by Matt Cohen, Ed Folsom, and Kenneth M. Price. Accessed February 19, 2021. https://whitmanarchive.org/manuscripts/finding_aids/integrated.html.

———. n.d. “Walt Whitman.” Walt Whitman Archive. https://whitmanarchive.org/biography/walt_whitman/index.html.

Gabler, Hans Walter. 2007. “The Primacy of the Document in Editing.” Ecdotica 4: 197–207.

———. 2018. “The Draft Manuscript as Material Foundation for Genetic Editing and Genetic Criticism.” In Text Genetics in Literary Modernism and Other Essays. Cambridge: Open Book Publishers, 209–19. doi:10.11647/OBP.0120.

Hay, Louis. 2017. “Genetic Criticism: Another Approach to Writing?” Chap. 29 in Research on Writing: Multiple Perspectives, edited by Sylvie Plane, Charles Bazerman, Fabienne Rondelli, Christiane Donahue, Arthur N. Applebee, Catherine Boré, Paula Carlino, Martine Marquilló Larruy, Paul Rogers, and David Russell, 531–47. Fort Collins, CO: The WAC Clearinghouse. doi:10.37514/INT-B.2017.0919.2.29.

Kline, Mary-Jo, and Susan Holbrook Perdue. n.d. “Introduction: What Is Documentary Editing? Where Did It Come From?” In A Guide to Documentary Editing. 3rd ed. Online edition. Charlottesville: University of Virginia Press. Accessed 26 February 2021. https://gde.upress.virginia.edu/01A-gde.html.
McElderry, Michael. 2013. *Thomas Biggs Harned Collection of Walt Whitman Papers, 1842–1937. A Finding Aid to the Collection in the Library of Congress*. Last updated February 2017. Washington, D.C.: Manuscript Division, Library of Congress. https://hdl.loc.gov/loc.mss/eadmss.ms002007.

Pierazzo, Elena. 2009. “Digital Genetic Editions: The Encoding of Time in Manuscript Transcription.” Chap. 11 in *Text Editing, Print and the Digital World*, edited by Marilyn Deegan and Kathryn Sutherland, 169–86. Farnham, England: Ashgate.

———. 2014a. “Digital Documentary Editions and the Others.” *Scholarly Editing: The Annual of the Association for Documentary Editing* 35: 1–23. http://www.scholarlyediting.org/2014/essays/essay.pierazzo.html.

———. 2014b. “Of Time and Space: Unpacking the Draft Page: A New Framework for Digital Editions of Draft Manuscripts.” https://hal.univ-grenoble-alpes.fr/hal-01182133.

———. 2015. *Digital Scholarly Editing: Theories, Models and Methods*. London and New York: Routledge. Preprint available at https://hal.univ-grenoble-alpes.fr/hal-01182162.

———. 2016. “Modelling Digital Scholarly Editing: From Plato to Heraclitus.” Chap. 2 in *Digital Scholarly Editing: Theories and Practices*, edited by Matthew James Driscoll and Elena Pierazzo. Cambridge: Open Book Publishers, 41–58. doi:10.11647/OBP.0095.

Schäuble, Joshua, with Hans Walter Gabler. 2018. “Encodings and Visualisations of Text Processes across Document Borders.” In *Digital Scholarly Digital Editions as Interfaces*, edited by Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber, and Gerlinde Schneider, 165–91. Norderstedt: BoD. https://kups.ub.uni-koeln.de/9116/.

Shillingsburg, Peter L. 1996. *Scholarly Editing in the Computer Age: Theory and Practice*. 3rd ed. Ann Arbor: University of Michigan Press.

Sperberg-McQueen, C. M., and Lou Burnard. 1994. *TEI P3: Guidelines for Electronic Text Encoding and Interchange*. Chicago; Oxford: Text Encoding Initiative. Available at https://tei-c.org/Vault/GL/teip3doc.tar.gz.

TEI Consortium. 2002. *TEI P4: Guidelines for Electronic Text Encoding and Interchange: XML-compatible Edition*, edited by C. M. Sperberg-McQueen and Lou Burnard. N.p.: TEI Consortium. https://tei-c.org/Vault/P4/doc/html/.

———. 2007. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 1.0.0. Last updated October 28. https://tei-c.org/Vault/P5/1.0.0/doc/tei-p5-doc/en/html/.

———. 2011. “P5 Version 2.0 Release Notes.” N.p.: TEI Consortium. https://tei-c.org/Vault/P5/2.0.0/doc/tei-p5-doc/release/tei-p5-doc/share/doc/tei-p5-doc/readme-2.0.html.

TEI Consortium. 2020. “Representation of Primary Sources.” Chap. 11 in *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 4.1.0. Last updated August 19. N.p.: TEI Consortium. https://tei-c.org/Vault/P5/4.1.0/doc/tei-p5-doc/en/html/PH.html.
NOTES

1 The term “diachronic” came into the field of editing from linguistics and signifies the historical dimension of texts (i.e., their change across time). In this way it contrasts with the “synchronic” dimension, which is static and ahistorical, like a snapshot. Marina Buzzoni has discussed the relevance of the diachronic and synchronic in digital editing in terms of Gianfranco Cantini’s requirement that an edition “aim at injecting history into the critically reconstructed text by taking into account the different synchronic states that make up its diachronic dimension” (2016, 62).

2 Their edition is now available at https://faustedition.net.

3 Although “genetic,” “documentary,” and “diplomatic” are sometimes used, in the context of digital editing, as loosely synonymous, even almost interchangeable, I wish to exploit the ways that they are distinct. These distinctions are largely a matter of emphasis (theoretical and/or methodological), which I would sketch as follows: Documentary editing foregrounds the singularity of the original upon which it is based (see Pierazzo 2014a, 2, and Kline and Perdue n.d.); diplomatic editing highlights the combination of a high degree of fidelity to the physical disposition of writing with a low degree of editorial interpretation (see Bodard 2020); and genetic editing places emphasis on the process of writing as opposed to the text produced by the process (see Hay 2017, 531).

4 See https://github.com/TEIC/TEI/issues/1427.
As I was revising this essay for publication in August 2020, in fact, I was notified that the TEI Council was planning to address issue no. 1427, our request that the chapter be restructured, by issuing a new release of the Guidelines with some significant changes to the prose of chapter 11. Among the revisions, the several new introductory paragraphs are especially welcome, as they go some way toward making explicit the basic difference in approach between transcriptions within `<sourceDoc>`, which emphasize “spatial features,” and those within `<text>`, which emphasize “logical textual structure” (TEI Consortium 2020).

The distinction that I am referring to (and advocating) here is the one perhaps most succinctly articulated by Peter Shillingsburg, who says that “document” refers to “the physical vessel ... that contains (or incarnates) the text,” whereas “text” refers to “the series of words and pauses recorded in a document” (1996, 174).

I use the term genetic dossier here as a synonym of avant-texte to refer to all of the documents that contributed to a published work. In doing so, I follow the usage of, for example, Dirk Van Hulle (2016, 46) and Elena Pierazzo (2015, 14; 2016, 49). Hans Walter Gabler notes, however, that the two terms are sometimes contrasted, the former used to mean that which comprises “transcription and image in conjunction,” thus offering “the document perspective on what, from the text perspective, is named ‘avant-texte’” (2018, 210). In fact, although in the hands of later practitioners of the French school of genetic criticism “avant-texte” has often been associated specifically with texts rather than documents, as originally coined by Jean Bellemin-Noël it seems to have been conceived as document-centric, referring to “the totality of the material written for any project that was first made public in a specific form” (2004, 31).

Note that I am using these terms to indicate temporal and notional relationships different from those indicated by the microgenetic and macrogenetic levels Van Hulle describes in his five-part scheme (the others being exogenetic, endogenetic, and epigenetic), where the contrast is essentially a matter of size or wholeness: the macrogenetic level involves “the genesis of the work in its entirety across multiple versions,” and the microgenetic “one specific textual instance across ... versions” (2016, 50).
9 Pierazzo calls attention to some of the problems of printed genetic editions, which have not only been “criticized as unreadable, unusable, time-consuming, and, in general deceptive” (2009, 171) but also as “incapable of represent[ing] the time flow” (2009, 171–73), even though, as she notes, the editors of Genetic Criticism: Texts and Avant-textes state that genetic criticism “aims to restore a temporal dimension to the study of literature” (Deppman, Ferrer, and Groden 2004, 2).

10 Equally meaningful is the interpolation in the mis-title of the word physical, which was not part of any actual version. The term physical transcription does, however, pretty accurately reflect the foregrounding of diplomatic transcription in the 2011 version of the chapter.

11 IDs are recorded, along with regularized titles, in separate files, one per work (e.g., poem or essay). A third custom element, <workParent>, is used in these files to indicate constituency relationships (for example, to declare that a poem is one of several component parts that constitute a series).

12 The quotation comes from a private letter to Whitman, which he published widely (without Emerson’s permission), even including it in some of the copies of the first edition of Leaves of Grass.

13 The Whitman Archive has, in fact, recently completed work, funded by an NEH grant, to publish a “variorum” edition of the 1855 Leaves of Grass that connects each poetic line in the first printed edition to known manuscripts (“avante-texte”) as well as to significant textual and spatial variations in different copies. In the course of that work, we identified two additional related manuscripts. These do not affect the substance of my point here, so I have not included them in the discussion. The Leaves of Grass (1855) Variorum is available at https://whitmanarchive.org/published/LG/1855/variorum/index.html.

14 See https://whitmanarchive.org/manuscripts/findings_aids/integrated.html. It is worth noting, however, that the finding guide draws on the connections between instance and work as recorded in EAD records rather than in TEI-encoded transcriptions.

15 Hans Walter Gabler has provocatively analyzed this process of “ascertaining the order in which the manuscript pages were filled” as essentially iterative, moving “loop by loop, from document to text in ever finer steps of granularity” (2007, 201).

16 It is important to acknowledge that, although populating <listChange> might be “simple,” decisions about what constitutes a discrete “step” are never simple. For example, in the case of a substitution, should each deletion and each addition be a separate step or should the two be
grouped as a single <change>? For my purposes, I chose the latter approach. It should also be noted that, in the encoding of the transcription, I chose to forego <sub> altogether, since substitutions are already implied by the corresponding @change values on the <del> and <add> elements.

17 Dussault and I have only recently resumed conversations about this and have begun exploring other, non-JavaScript alternatives for processing the TEI. As we are both doing this work in our “free time,” however, progress is slow. We would, of course, welcome suggestions or other constructive input.

18 I was, in part, inspired by work done by Elena Pierazzo and a group of collaborators to produce a prototype interactive rendering of pages from one of Marcel Proust’s notebooks. By encoding <zone>s and assigning to them a relative order, the project team was able to “create an interactive, accessible interface” (2014b, 18). Unfortunately, their prototype appears to be no longer available. Attempts to visualize the development of a text across documents have been more common, though they face their own significant challenges, of course. Such projects as Faustedition (http://www.faustedition.net/), the Samuel Beckett Digital Manuscript Project (https://www.beckettarchive.org/), and Ulysses: A Digital Critical and Synoptic Edition (http://ulysses.online/index.html) have produced various forms of visualization (including, in the case of the Beckett Project, animation) that might serve as inspiration for future work.

19 In making my case for further development I am mindful of James Cummings’s sage discussion of common “myths, misconceptions, and misunderstandings” about TEI, wherein he addresses “the criticism often levelled at the TEI” of being “too simple or too general.” He rightly points out that such complaints are often born of a “superficial understanding of the scope of the TEI” or a lack of understanding of “the way in which any individual element can be further refined by using attributes or nested levels of encoding.” Diachronic markup, however, is not such a case but, rather, an example of a “real instance[]” in which the TEI “provides only a general level of markup for some phenomena” because “the TEI community has not pushed the standard forward yet to greater detail in this area” (2019, i58, i61–i62). What I am calling for, then, is a harder collective push.
AUTHOR

BRETT BARNEY

Brett Barney is a research associate professor at University of Nebraska–Lincoln’s Center for Digital Research in the Humanities (CDRH) and has been involved in the development of the Walt Whitman Archive since 2000, first as a research assistant and more recently as senior associate editor. He has also worked on a variety of other CDRH projects, including the Willa Cather Archive and the Journals of the Lewis and Clark Expedition Online. He served two terms (2010–2013) on the TEI Consortium’s Technical Council and is especially interested in expanding TEI’s provisions for documentary and genetic editing.