

H. Lewis Ulman
Department of English
The Ohio State University
164 W. 17th Avenue
Columbus, OH 43210-1370
(614) 292-2275
ulman.1@osu.edu



Will the Real Edition Please Stand Out?: Negotiating Multi-Linear Narratives Encoded in Electronic Textual Editions

". . . we stand on the edge of a period that will see the complete editorial transformation of our inherited cultural archive. That event is neither a possibility nor a likelihood; it is a certainty. As it emerges around us, it exposes our need for critical tools of the same material and formal order that can execute our other permanent scholarly function: to imagine what we don't know in a disciplined and deliberated fashion. How can digital tools be made into prosthetic extensions of that demand for critical reflection?"

—Jerome McGann, *Radiant Textuality* 18.

My title, of course, evokes the long-running game show "To Tell the Truth," in which three challengers, all claiming to be the same person, tried to fool a panel of celebrity contestants by providing plausible, if made-up, answers to questions aimed at revealing the "real" person. Scholarly editions are rarely so playful, but electronic editions of printed and manuscript texts, given the flexibility with which they can offer multiple representations of a work and its material, textual, historical, and archival contexts, foreground the problem, already present in printed textual editions, of negotiating multiple states of the "work" being edited (e.g., manuscripts written by authors, copies of manuscripts by other hands, various printed editions, and so on). Moreover, as Jerome McGann's reference in my epigraph to digital prostheses for critical reflection suggests, an electronic edition can transform the text, the act(s) of reading it, and the editor and reader. For example, by providing readers with access to high-resolution images of original texts as well as multiple editorial interpretations of those texts, the roles of the editor and reader begin to merge: readers who can manipulate—zoom, rotate, adjust color and contrast—high resolution, color images of original texts have an unprecedented opportunity to revisit the evidence for editorial decisions. Learning to encode and decode electronic editions, then, requires

editors and readers to negotiate—through what McGann calls critical reflection—multi-linear narratives of creating, interpreting, preserving, and disseminating texts.

My Dear Elizabeth

Letters from Sophia Peabody Hawthorne to Elizabeth Palmer Peabody, ca. 1837-1868

I will illustrate my argument with an online electronic edition of five previously unpublished letters written by Sophia Peabody Hawthorne to her sister Elizabeth Palmer Peabody between 1837 and 1868. The occasion for the project was an undergraduate course entitled “Digital Media and English Studies,” in which the students and I collaborated on an edition of the Hawthorne letters, the originals of which are held in the Rare Books and Manuscripts Department of The Ohio State University Libraries. I focus on how the encoding of our online electronic edition enables us to foreground variously the *personal and cultural milieu* evoked by the text (in this case, samples from a decades-long correspondence), the *material text* (i.e., the manuscript artifacts in the library’s collection), and our *editorial interpretations* of both.

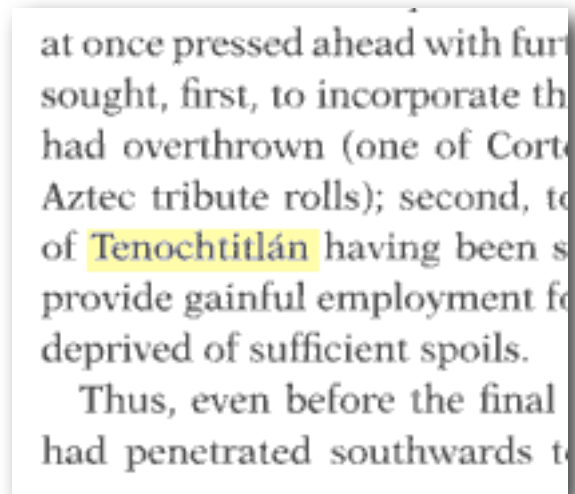
```
<!-- Page 1 -->
<pb id="Letter1_P1a" n="1"/>
<!-- The opener tag contains any dateline and salutation apper
  <opener>
    <dateline>
      <date value="1837-8-8">8th August</date>
    </dateline>
    <salute>My dear <name type="recipient">
      <name key="EPP" reg="Elizabeth Palmer Pe
    <seg>
      <note target="EPP">1</note>
    </seg>
  </salute>
</opener>
<lb/>
```

Before delving into the Hawthorne edition, however, it is instructive to consider a widely publicized contemporary online text project—Google Book Search (GBS). Embroiled in legal battles over its policy of scanning and providing limited online access to books from academic and public libraries that are still in copyright, the GBS project has elicited optimistic visions of wider public access to information (Carlson and Young) and wary forecasts of the erosion of copyright, damage to irreplaceable books, and an increase in plagiarism, among other worries (Herring). For the purpose of this essay, however, I am interested in the degree to which the encoding of texts in GBS addresses the need McGann identifies for "critical tools of the same material and formal order" as electronic texts and supports critical reflection. First, let's think about the sorts of encoding that GBS involves, then about the roles and actions that encoding supports.

Of course the texts included in GBS contain the semantic codes of natural language, but that fact does not distinguish them from printed text. Nor am I interested here in the transliteration from the Roman alphabet to ASCII or Unicode, though that particular encoding is necessary to extend the reach of our critical tools into the realm of the digital. Rather, I am concerned in this essay with information embedded in, or associated with, digital files that describe the contents of those files in formats that computer-based systems can process in response to human queries. In short, metadata.

On its opening screen, GBS presents the user with a text entry box and an invitation to "Search the full text of books (and discover new ones)." In other words, the user is invited to search the texts of books for characters strings that may occur in the books. No metadata is required for such a search. For example, curious to learn more about Tenochtitlán, the Mexica city destroyed by Cortés, which is discussed in a book I am currently reading, Charles C. Mann's 1491: New Revelations of the Americas before Columbus (Knopf, 2005), I enter the city's name in GBS. 1491 doesn't turn up in the first 20 screens of returns, presumably because the book is too new or because the publisher has not given GBS permission to include 1491 in its database.

However, with 20 screens of results to choose from, each result containing an image of a book's cover along with its title, author's name, year of publication, length in pages, and a sample passage containing my search term and identified by page number, I learn something about the metadata associated with the text of each book. If I click on one of the returns, I am presented with one of three views, depending on the copyright status of the text and the permissions granted to GBS by the book's publishers: [Snippet View](#), [Sample Page View](#), or [Full Book View](#). Again, I learn something about the metadata associated with each text—in this instance, rights management information. In all three views, the search results are presented as a page image—not text—with the search term highlighted in yellow, which tells me that the database consists of images and texts, referenced to one another.



An "Advanced Book Search" link reveals a bit more about the structure of GBS's database. In addition to some Boolean options for modifying how the search engine handles the full text search strings, this page allows me to limit my search by title, author, publisher, publication date, and ISBN fields. Other links direct me to specific pages of the book (front cover, copyright, table of contents, index, back cover), and an "About this Book" link provides me with a final glimpse of the database structure as revealed in the following list of information: synopsis, title, author(s), publisher, publication date, subject, format, pages, dimensions, ISBN.

All in all, GBS associates a fair amount of *bibliographic information* with each book and *pagination* with the textual contents of each page. But what does that information allow me to *do*? I can locate every instance of a particular text string in GBS's vast and rapidly growing collection, but I can't ask GBS to list every place name contained in a book or every *reference to* a particular place (using some phrasing other than its name) because the system apparently "understands" its

texts only as strings of characters in structural containers (books, pages)—it contains no encoded semantic information that might enable more sophisticated critical analysis. For all its scope and raw search power, GBS encodes only a single narrative of how users might interact with books: we can use the system to identify books that contain text strings in which we are interested and then learn where we might borrow or buy those books. To its credit, GBS promises no more, noting its limitations on its "Common Questions" page, though in terms of constraints on reading rather than the paucity of analytical tools:

Why can't I read an entire Library Project book online? Google Book Search is designed to help you discover books, not read them from start to finish. We want to make it easier for users to find and buy books while protecting the rights of copyright holders. So when you find a book that is still under copyright, you'll see the Snippet View which, like a card catalog, shows you information about the book plus a few snippets – a few sentences of your search term in context. However, there are some books that are out of copyright which means we can display the full book in the Full Book View.¹

Clearly, GBS does not, and is not intended to, meet the need for "critical tools of the same material and formal order" as the transformed cultural archive (i.e., the large-scale digitized library collections) for which it serves as such a popular icon. What might those tools look like?

One promising tool is the Text Encoding Initiative's Guidelines (Sperberg-McQueen and Burnard, eds.), an XML-compliant document markup language designed by scholars over the past decade and a half in order to "provide means of representing those features of a text which need to be identified explicitly in order to facilitate processing of the text by computer programs. In particular, they specify a set of markers (or tags) which may be inserted in the electronic representation of the text, in order to mark the text structure and other textual features of interest."² For example, the parenthetical citation at the beginning of this paragraph might be

¹ <http://books.google.com/googlebooks/common.html>

² <http://www.tei-c.org/P4X/AB.html>

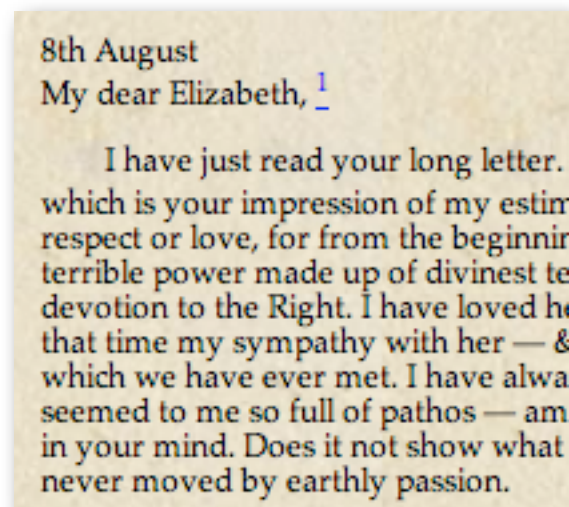
encoded as follows: (<name type="person" reg="Sperberg-McQueen, C. M.">Sperberg-McQueen</name> and <name type="person" reg="Burnard, Lou">Burnard</name>, eds.). That encoding allows a computer program to identify the occurrence of persons' names and regularized versions of those names (in this case, the forms used by the Library of Congress catalog). Once such features are identifiable by a computer program, readers can process those features in any way that a computer can be programmed to process information: e.g., compile, count, and sort a list of personal names in a document in a regularized format, no matter how they are referred to in the document; define the formatting of those names or references; create links from or to those names; graphically represent where in the document those names appear, and so on. Detailed technical descriptions of the TEI [Guidelines](#) and XML publishing lie well beyond the scope and purpose of this essay.³ The most important detail to appreciate about the approach my class took with the Hawthorne letters is that we created a *single* XML file for each letter, encoded it with *all* of the editorial information we wished to embed in the text, then created various filters that extracted and presented different information for different scenarios of use.⁴ Aside from the efficiency of that editorial strategy (you can create 15 views of the letters with 8 files—five letters x 3 filters), it encourages students to think of the encoding not as a mechanical process of marking up a text according to some rules but as a process of enabling potential narratives in which real users read and study the letters for different purposes.

Foregrounding the personal and cultural milieu evoked by the text. Many readers, we realized, will have little or no interest in the manuscripts as such. Their focus will be primarily on the narrative evoked by the text of the letters. Accordingly, our “[literary](#)” presentation makes few visual references to the manuscript page—it presents the words of the letter as they might appear

³ Readers interested in the technical details of the Hawthorne editions will find links to information about the technical standard and editorial practice under the "Technical Standards" and "Project Update" headings on the project home page.

⁴ Links to information about the manuscript features, editorial analysis, and metadata encoded in our XML markup can be found on the site's home page.

printed in a book. For instance, line breaks are determined by the width of the text area in the Web browser, not by the lineation of the original manuscript. Cancellations (text deliberately scratched out or otherwise obscured by the writer), though recorded and encoded as such in the XML transcripts of the letters, are "silently" suppressed in this view. Interlinear additions (text added between lines of text) are also recorded and

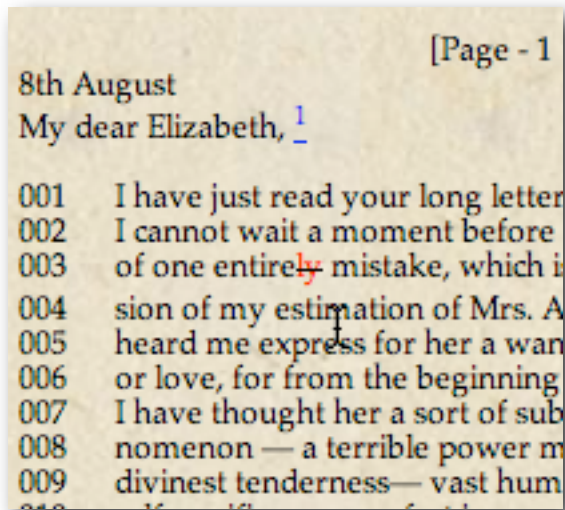


encoded in the XML transcript but included inline without any comment in this view. Footnotes reside in their conventional location at the bottom of the page, and the reader is directed to illustrations of details that cannot be transcribed (a postmaster's stamp and a wax seal) in much the same way that readers are directed to figures in print. There is, of course, nothing intrinsic about this presentation that ties it directly to the personal and cultural milieu evoked in the text. Rather, we choose this format for this audience because it is "invisible" by convention. We experience the sensation of getting "lost in a book" because long exposure to the conventions of the printed page make its materiality recede from our attention—especially our critical attention (when lost in a book, the typical reader is not wondering why a book is printed in a particular font or worrying about where the author might have reconsidered his or her wording). For that reason, we call this a "literary" view of the letters and hope that it enables readers to immerse themselves in the world of Sophia Peabody Hawthorne in 1837.

Beyond the obvious fact of its display in a Web browser, however, this view differs in various ways from a printed page. Several textual and visual clues point or hint at the presence of additional views, additional scenarios for reading the letter. First there is the menu of views at the top of the page, suggesting that the reader can choose an alternative format. Next, the background of the browser is constructed by tiling a small image of the manuscript page free of writing. And

through the inclusion of the bibliographer's abbreviation for an "Autograph Letter Signed" (ALS)—a letter written and signed in the hand of the writer (rather than, say, by a scribe to whom it was dictated)—the headnote to the letter points to the manuscript "authority" for this text.

Foregrounding the material, archival existence of the text. By contrast, a "[philological](#)" view of the letter re-presents various features of the manuscript page such as lineation, cancellations, and interlinear additions, all accompanied by editorial apparatus signaling the

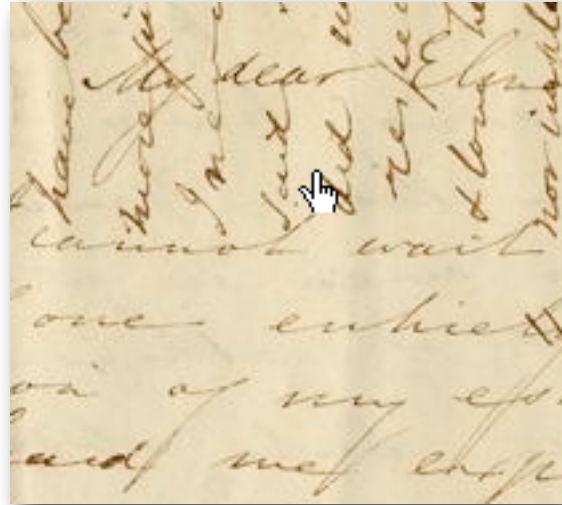


nature of each feature ("recovered" cancellations are set in a strikethrough typeface, lines are numbered, and interlinear additions are preceded and followed by arrows that indicate their displacement from a line of text). Further, each page break is marked by a link that opens an image of the manuscript page from which the following text is transcribed. The image can be magnified and repositioned on the screen, allowing the user to read the handwritten

script or examine physical features (e.g., approximate ink and paper colors, creases, stains, tears) in great visual detail. A "[Text/Image](#)" view juxtaposes the philological view with images of the manuscript pages in the right margin, linked to full page views of the images. These two views arise from attempts by the designers to construct a narrative of use that foregrounds scholars—or curious readers—who wish to examine the material, archival existence of these textual artifacts.

Many other approaches to this (or any other) narrative of use are possible. A tool currently under development at Ohio State will allow users to invoke what Jerome McGann has termed *deformations* of the manuscript page images that might reveal certain characteristics of the page better than a more faithful representation. For instance, during our work with the Hawthorne manuscripts, my students and I found that an image containing only the blue channel of the scanned digital image made it possible to read some text that we couldn't decipher in a full-

color image. Electronic editions that provide readers with this many powerful tools for critical analysis left my students unsure about how to negotiate their status as editors. "Why would anyone read our edition?" they asked, "Why wouldn't they simply look at the images?"



Interestingly, some seasoned textual editors asked me the same question when I presented a version of this essay at a conference—though they were more confident that readers who bypassed the editor's work would be shortchanging themselves. My answer to both students and colleagues points to another narrative of use: engaging readers with the process of editorial interpretation, much as a textual introduction or table of variants does in a print edition but with far richer access to original materials.

Foregrounding interpretation of the text. A third possible representation of the edition foregrounds our TEI-compliant encoded transcript of the letters and the interpretive work involved in creating that transcript. Because any particular view of the letters derived from the encoded transcript may not reveal all of the encoded information, the code itself constitutes the

```
<opener>
  <dateline>
    <date value="1837-8-8">
  </dateline>
  <salute>My dear <name type="text">
    <name key="EPP" reg="1">
  <seg>
    <note target="EPP">1
  </seg>
  </salute>
</opener>
```

fullest record of our interpretation of the letter, including declarations of encoding standards. For instance, at one point we might encode "correct" or "modernized" punctuation for a sentence but not report that interpretation in any views of the letter. Further, the technology that determines how we present any particular view of an encoded text is a style sheet written in Extensible Stylesheet Language (XSL) that selectively suppresses, reveals, orders, and

formats the encoded transcript. That style sheet is another instantiation of the editor's interpretation of the text, but few editions make their XML transcripts and XSL style sheets available to readers. At present, our edition of the Hawthorne letters provides access only to excerpts from the transcripts and style sheets, and relatively few readers would be prepared to read those files critically.⁵

Conclusion. As we design electronic editions, I ask my students the following questions: How do we communicate to readers of electronic editions the nature and flexibility of the relationships that exists between the visual/aural representations they are experiencing and the multiple layers of encoding and meaning-making associated with it? How do we provide users with tools to manipulate and evaluate those relationships themselves? One of our strategies for answering those questions is to imagine narratives of use and accommodate our designs to them.

⁵ Links to excerpts from the XML transcripts and XSL style sheets are available under the "Project Update" heading on the edition home page.

Works Cited

- Carlson, Scott, and Jeffrey R. Young. "Google Will Digitize and Search Millions of Books From 5 Top Research Libraries." Chronicle of Higher Education 7 Jan. 2005: A37.
- Google Book Search. 2006. Google, Inc. 27 Feb. 2006 <<http://books.google.com/>>.
- Hawthorne, Sophia Peabody. My Dear Elizabeth: Letters from Sophia Peabody Hawthorne to Elizabeth Palmer Peabody, ca. 1837-1868. 30 Apr. 2005. The Ohio State University Libraries. Ed. H. Lewis Ulman, et al. 26 Feb. 2006 <http://people.cohums.ohio-state.edu/ulman1/Sophia_Hawthorne_Letters/>.
- Herring, Mark, Y. "Don't Get Goggle-Eyed Over Google's Plan to Digitize." Chronicle of Higher Education 11 Mar. 2005: B20.
- Mann, Charles C. 1491: New Revelations of the Americas before Columbus. New York: Knopf, 2005.
- McGann, Jerome J. Radiant Textuality: Literature after the World Wide Web. New York: Palgrave, 2001.
- Sperberg-McQueen, C. M., and Lou Burnard, eds. TEI P4: Guidelines for Electronic Text Encoding and Interchange, XML-compatible edition. 2004. Text Encoding Initiative. 28 Feb. 2006 <<http://www.tei-c.org/P4X/>>.
- TEI: Yesterday's information tomorrow. N. d. The Text Encoding Initiative. 28 Feb. 2006 <<http://www.tei-c.org/>>.