

# Book, Body and Text: The Women Writers Project and Problems of Text Encoding

*Julia Flanders*

---

*Women Writers Project, Box 1841, Brown University, Providence, RI 02912*

KEYWORDS: Women Writers Project, SGML/TEI, editing

AFFILIATION: Brown University Women Writers Project

E-MAIL: Julia\_Flanders@brown.edu  
FAX NUMBER: 401-862-7329  
PHONE NUMBER: 401-862-3835

## Introduction

SGML and text encoding arose from the need to make paper disappear. Documents whose size, complexity, or frequency of revision rendered them unmanageable as physical texts were the first and strongest impetus behind the development of SGML, and perhaps provided some of the crucial emotional component as well, so to speak. The conception of a text as an "ordered hierarchy of content objects," which completely separates the text from its physical vehicle, does so almost with a sense of relief: the repudiation of procedural markup and the cumbersome mechanics of pages seems like an entry into a purer, simpler world where the real structures of the text become tractable.

With the advent of the Text Encoding Initiative, SGML began to look additionally like a way to give paper a new life. Humanities text encoding projects often focus on rare or inaccessible documents, which are encoded in order to give them wider accessibility and greater flexibility of use. Projects of this nature include the Canterbury Tales Project and Project Electra at Oxford; the SEENET edition of Piers Plowman; the Women Writers Project at Brown University; the Rossetti Archive; and numerous other similar undertakings, all of which aim at reproducing existing physical texts in electronic form. Approaches to this reproduction vary; some provide digital images of the original text along with the encoded version; others attempt to include aspects of the physical structure and appearance of the text in the encoding; still others encode only the aspects of the physical book which are seen to have linguistic meaning. All of them, however, are aimed primarily at audiences whose scholarly work requires reference to the physical text; the electronic text is a facilitator, but ultimately it only stands in front

of, in place of, the real object which is seen as grounding its existence.

As Jerome McGann and others have pointed out, the aims of projects like these are often difficult to accommodate using the structures provided by SGML, since the physical and textual structures of the book often overlap and cannot both be represented. A number of methods have been developed which offer workarounds and practical solutions, but the essential conception of the SGML document as having a single hierarchy remains unchanged. Text encoding projects, then, must choose a method which best suits their particular commitments and needs.

This paper places the Women Writers Project's own approach within the context of these alternatives, and also within a larger theoretical context; editorial theory is rooted in notions of the status of the physical book and its relation to the linguistic text, and this theory is reflected in the assumptions made by text encoding projects about the relation between the electronic transcription and the original.

## 1. Editorial Theory and the Physical Text

The physical document plays a crucial practical and metaphysical role in editorial theory. Literary editing proposes to itself the task of consulting physical exemplars of a text in order to determine, by their variations, how much of the authorial message they carry and what authority they have within the textual tradition. In an almost ironic way, the real goal of this activity is to transcend the physical document and get at the textual message that it carries. The "best text" which the resulting new edition attempts to present, and the "copytext" which in some theories is the projected/hypothesized authorial ur-text from which all others vary, are both in a sense ideal and disembodied texts which none of the existing documents fully produces. The physical document, then, is metaphysically a failure, a corrupted falling-away from the ideal of the authorial text. For the editor it is like a dirty window through which one struggles to see the real original which never was. Practically, it becomes a fetish, since it replaces the unattainable object of one's desire; an edition becomes authoritative by attention to every physical exemplar in their minutest details.

Emotionally, our culture has a deep attachment to physical books, particularly to old ones, which are made to represent an earlier age which respected learning and textuality rather than shopping malls and video. Accompanying the new wave of fascination with electronic texts is a vein of deep distrust at their perceived replacement of the physical book. Among the criticisms of the electronic text is that it has nothing grounding it; that it is changeable and hence unreliable; and that it is potentially anarchic in allowing multiple readings

and user modification. It is important to bear in mind that these reservations do not regard the electronic text's disembodiment as moving it closer to the ideal, but rather think of it as a failed version of the physical, one which has all the potential for corruption that books suffer from, but none of their ability to serve as a solid point of grounding. Thus editors who address the task of producing an electronic scholarly edition or an electronic version of a primary source document are faced with a crisis of confidence; they must endow the electronic text with an authority which it has not as yet been accorded by the general imagination. Furthermore, in preparing an electronic version of a primary source, they are under the special burden of making this text function – practically and metaphysically – like its physical source. At the same time, they are under pressure to build in the kinds of functionality that come from foregrounding the linguistic text – enabling searches, navigation, and retrieval based on the conceptual structures of the text – since this is part of the point of encoding the document in the first place. These two imperatives, unfortunately, do not always pull in harness.

## **2. Encoding the Book, Encoding the Text**

The WWP's own commitment within this landscape is somewhat troubled. To begin with, our primary goal is to create the electronic equivalent of an archive – a repository of primary source documents which are available to researchers for a wide range of uses. It was the very inaccessibility of archival materials that fueled the WWP's beginnings; early modern women's texts exist in abundance but are too often only available in rare book libraries, and our intention has been to create another, more accessible and functional archive which would broaden access and research. This commitment points us naturally towards fidelity to the physical book; our audience needs both completeness and accuracy of description of features like pagination, line breaks, typographical errors, front and back matter, damage and illegibility, and so on. At the same time, part of the seduction of encoding quantities of textual material is that one has the opportunity to add function to it, enabling users to fulfill their wildest research dreams of searches and comparisons and easy navigation – indeed, one could not get grant funding without conceiving of the project in these terms. At once, then, we are thrown back towards the other end of the spectrum, committed to encoding the conceptual structure of the text and addressing the nuances of its analytical hierarchies. Beside our desire to create an archive – texts which meet our expectations of their physicality – is the equally strong desire to create texts which respond to our expectations of their textuality.

Our accommodation of these two motives ends by privileging the textual structure, and encoding the physical structure using milestone tags. This is partly because the TEI's provision for tagging this structure is so much more richly conceived and nuanced, and partly because – in the absence of a workable method of encoding both simultaneously – the textual structure facilitates more basic activities. Users rely on notions of act, scene and verse, of chapter and paragraph, for their basic navigation in the document, while their use of physical details amounts to something more like consulting: an event rather than an ongoing process. Furthermore – and this is indicative of something quite basic about our conceptions of documents – the user is more apt to use the textual structure to compare two versions of the same work. One thinks of Chapter 2 of *Clarissa* as being the same item from edition to edition, whereas page 37 may be altogether different. The “work,” we think, transcends the book precisely because it is abstractable from it. This sense, however, may say more about our emotional and intellectual commitment to this notion of difference than about the nature of things; why should this transcendence appeal to us so overwhelmingly? The answer, I would argue, has to do with the way that books and authors have been written into our culture; about this I will say more in the finished paper, concluding that cultural institutions rely on the authority vested in notions of authorship and transhistorical textuality.

By and large, the WWP has found that milestone tags fill our basic needs for encoding the physical structure of the book. However, in some cases they introduce considerable complexity, and require considerable ingenuity in execution. One case of particular interest is the problem of continued footnotes, which require, in some sense, two separate milestones indicating the end of the page. I will outline this in more detail in the finished paper, showing how it is necessary to think of the footnote as orthogonal to both the textual and physical structures in order to avoid difficulty. Furthermore, other projects who need to record even more information about the physical text – projects dealing with manuscripts, for instance – might find that using milestones is unworkably cumbersome. Clearly, additional research on these kinds of encoding needs will help advance our thinking about overlapping structures – particularly those of the physical and the textual book – and how to handle them.

## **Conclusions**

Although it is possible to address the issue of overlapping structures in SGML as a purely practical problem, I believe that this is a short-sighted solution. I have tried to indicate the ways in which

overlapping structures can indicate important methodological issues, and also the degree to which these methodological issues are implicated in the larger cultural and political arena. Text encoding projects ignore such matters at their peril; in creating data, we both create and rely on implicit intellectual structures about which we can choose to be naive or canny.

### **Bibliography**

- DeRose, Steven, J., David Durand, Elli Mylonas and Allen H. Renear. "What is Text, Really?," *Journal of Computing in Higher Education*. 1:2 (1990).
- Jed, Stephanie. *Chaste Thinking: the Rape of Lucretia and the Birth of Humanism*. Bloomington: Indiana University Press, 1989.
- McGann, Jerome. "What is Critical Editing?," *The Textual Condition*. Princeton: Princeton University Press, 1991.
- O'Keefe, Katherine O'Brien. "Text and Work: Some Historical Questions on the Editing of Old English Verse." *New Historical Literary Study: Essays on Reproducing Texts, Representing History* Jeffrey N. Cox and Larry Reynolds (eds.). Princeton: Princeton University Press, 1993.
- Shillingsburg, Peter. *Scholarly Editing in the Computer Age : Theory and Practice*. Athens: University of Georgia Press, 1984.
- Tanselle, G. Thomas. *Textual Criticism and Scholarly Editing*. Charlottesville: University Press of Virginia, 1990.