

Putting Renaissance Women Online

Paul Caton

Brown University

Providence, Rhode Island

USA

e-mail: paul@ursula.stg.brown.edu

ABSTRACT

Since its inception in 1986 the aim of Brown University's Women Writers Project (WWP) has been to bring to much wider attention the extensive body of pre-Victorian women's writing in English. In an initiative funded by a grant from the Andrew W. Mellon Foundation, the WWP is creating "Renaissance Women Online" (RWO), a collection of SGML-encoded transcriptions of works in English by women writers of the sixteenth and seventeenth centuries. This full-text database or "textbase" will be integrated into a Web site designed to help teachers and students access and appreciate these long-neglected works.

This undertaking has three main aspects: preparing the electronic texts, organizing access to them, and presenting them in a way appropriate to the target audience. Preparing the texts involves both transcribing them and encoding them in a manner conformant with Standard Generalized Markup Language. The WWP has a close association with the work of the Text Encoding Initiative and has adopted its Guidelines for Electronic Text Encoding and Interchange as the Project's encoding scheme. The advantages of the TEI scheme for the WWP lie in its breadth of application, its flexibility, and its widespread adoption among other humanities text-encoding projects. To offer wide access to the RWO materials, the WWP will publish them on a dedicated Web site. In preferring this option over CD-ROM, flexibility is again a prime consideration. A textbase like RWO gets revised and updated with a frequency which would make the timely distribution of update discs burdensome and impractical. Also, RWO has some material which is available to anyone, and some which can only be accessed with the payment of a licence fee. A Web server can easily offer different levels of access to different users of a single site. The RWO site will not be simply a passive resource. While it does have archival value, it aims to be pedagogically useful and proactive. The design of the site must help students read the works, literally and critically. For the first, RWO uses the DynaWeb interface. Its ability to exploit documents' SGML encoding while serving them to a Web browser offers readers easy navigation through, and searching of, the textbase. To help students and non-specialist readers understand and critically approach what they read, supplementary materials specially written by scholars in the field introduce the works and explain the ideas and issues of concern to the women writers represented in the textbase.

Publishing the Renaissance Women Online electronic resource is a crucial step for the Women Writers Project progress in fulfilling its aim, and not just because a Web site offers the chance of truly widespread distribution. For the WWP to continue its work it must move from grant-dependency towards financial self-sufficiency. RWO represents the Project's first large-scale effort to market its most valuable product, its textbase.

Introduction

The humanities have woken up to electronic texts. As recently as five years ago, projects whose main purpose was to make primary documents available in electronic form provoked a mixture of curiosity, suspicion, and disdain in academics, and anxiety in librarians. But with the rapid growth of the Internet the advance of new technology into humanities research has reached critical momentum, and in college and public libraries rows of computers appear like shrines to the gods of the ethernet as old-style reference and reading rooms get retasked into 'e-text centers'¹. Recognized now as the prophets of the digital library, electronic text projects have gained credibility and respectability. Old ones are treated with reverence, new ones spring up daily².

The text-encoding project I work for, the Brown University Women Writers Project (or WWP)³, had its beginnings in the mid-nineteen-eighties, which is the digital equivalent of having been with Lenin in Siberia. But only now are we taking the first steps towards making our electronic presence felt. The time seems propitious; nevertheless, we step with fingers crossed because no one has clearly blazed the trail for what we want to do. An academic project with altruistic aims, we now find ourselves in the hard-nosed world of business plans and market surveys, because we want--need-- to make money from our electronic publishing venture: its financial success is crucial to our future. Naturally, we hope to achieve our nobler goals as well as turn a profit. In this talk I describe our situation and our thinking as we find ourselves in the early stages of the initiative. After some necessary background information, I shall discuss three things: the basic product we offer; our plans for delivering it; and how we will package it.

Background

In 1986 a group of scholars, dismayed at the underrepresentation of early women's writing in popular teaching anthologies such as those published by Norton, began planning a project to redress the balance. They wanted to compile an anthology of primary works demonstrating the

¹ Two good examples of what e-text centers can offer may be found at the University of Virginia (www.lib.virginia.edu/etext/ETC.html) and Princeton (cethsun.princeton.edu/menu.html). On the kinds of work implementing an e-text center involves, see Burrows (1995).

² I exaggerate, of course, but one has only to look at the popularity of recent Summer Seminars by the Center for Electronic Texts in the Humanities, or the work of the Model Editions Project, or of Brown's Scholarly Technology Group, or at the expanding list of projects on the SGML pages, to appreciate the growth taking place.

³ <http://www.wwp.brown.edu/>

breadth and depth of writings in English by women from the earliest times up to the beginning of the Victorian era (from which point women's writing has received at least a little more attention). Transcribing the works onto computer was a feature from the start, but the original intention was to produce a traditional scholarly print publication--familiar in everything but its selection of works.

One of the Project's advisors, Allen Renear, persuaded the scholars that, instead of simply putting flat transcriptions onto the computer, encoding the works using an SGML-conformant markup language would create a resource which people could exploit in numerous ways, of which print publication was merely one option. The focus of the Project's efforts shifted towards the creation of a full-text database, or textbase. Its scope expanded to include the capture of all writings by women in English prior to 1830. Both Renear and the then Project Director, Elaine Brennan, were involved in the early work of the Text Encoding Initiative (TEI), so the Project naturally moved towards following the evolving TEI Guidelines on Electronic Text Encoding and Interchange in its encoding scheme.⁴

Over the past ten years the Project has encoded well over two hundred works. During this time it has made its versions of the works available in two forms. From early on it has supplied on request and for a nominal fee unbound printouts, mainly to academics wanting to teach the works in their classes. And for a while it participated actively in a commercial print venture with Oxford University Press, which uses WWP transcriptions for the 'Women Writers in English' series. However, neither of these offer satisfactory long-term methods of getting our texts to large numbers of people. The barebones printouts the Project supplies are not as user-friendly or informative as a traditional print edition, and as for helping to produce that kind of edition, the Project spent far more staff time and effort preparing its electronic files for book publication than the returns justified. Most importantly, in neither case does the medium exploit the added value of the encoding.

Because of these disadvantages of print, we at the WWP have for a long time been looking towards some form of electronic delivery as the best way to bring our texts before the public eye. Ideally, the means of delivery would generate revenue for the Project to reduce our reliance upon grant funding. In the autumn of 1996 an award from the Andrew W. Mellon Foundation allowed us to start work in earnest on a system for electronic publication. We plan firstly to take a coherent subset of the textbase and put it online in a format geared towards scholars and undergraduate users. Because writings by women from the early modern period constitute a particularly rich and neglected body of work, we aim within the grant period to finish encoding and then make available on a Web site one hundred of these works. This initiative has the somewhat salacious title 'Renaissance Women Online'. Built into the plan is an evaluative component to meet the Mellon Foundation's particular interest in comparing the cost of traditional print-based scholarship with that of using electronic resources.

⁴ For a full description of the TEI implementation of SGML, see Sperberg-McQueen and Burnard (1994).

The WWP 'product'

The WWP produces diplomatic editions of specific copies of primary works. Our advising scholars recommend which copy to use; we obtain a photocopy and from that we transcribe the work. The person who transcribes simultaneously adds the encoding, attempting to faithfully record all the important features of the work. In the TEI Header at the front of each encoded transcription we make clear what we have not recorded, and what choices and changes we have made in preparing the text. As a result, users know where the electronic versions come from and how they relate to the originals, so that if necessary, they can check one against the other. We assume scholars have two basic needs with primary sources: they want to know exactly what they are looking at, and they want it to be as complete as possible. When sources do not meet these needs, scholars cannot make authoritative statements about the work. We do not add any scholarly annotation or interpretive comments to the text, because the purpose of the Women Writers Project is not to produce highly finished pieces of scholarly work itself, but to provide raw materials for others to use in research and teaching.⁵

We record a lot of information in our encoding--some people would say, more than is necessary⁶--but we omit much to do with the printed work's appearance; we leave that to microfilm and facsimile editions. They come closest to recreating an unmediated reading experience and reproduce to a degree acceptable to all but those doing the most exacting bibliographical research the physical details of the book, which are either extremely time-consuming or simply impossible to reproduce through transcription. But their usefulness ends there: you cannot easily search them, do collations, reproduce them in different media, etc. With facsimiles and microfilm, what you see is all you get. And this is only a little less true of plain electronic transcriptions, such as those offered by Project Gutenberg.⁷ As electronic texts they are of course open to some forms of automated processing, but their potential for exploitation will always be limited because they possess no meta-text. A plain transcription presents a computer with a largely undifferentiated mass of words and gives it no electronic equivalent to the kind of information that human readers rely upon to make sense of an unfamiliar text. *The plain transcription offers no enabling mediation:* for that, we need text encoding.

I cannot talk about the specifics of our encoding scheme here, but I would like to stress one point before moving on. Encoding is an input of intellectual work that greatly facilitates further intellectual work. In this sense it resembles the activity of traditional scholarly editing. But it also resembles that activity in being interpretive and interventionist. An encoded text results from choices made by someone other than the author, and in some cases those choices are subjective. For example, encoding the use of italics for certain words and phrases in an otherwise Roman typeface work brings us uncomfortably close to attempting to interpret the author's intentions (never mind the fact that the work's original printer may have already done the same). Legacy documents of the kind we deal with constantly present these difficulties,

⁵ For more on the practical challenges of producing an encoded diplomatic edition, see Mah et al(1996). For discussion of some of the other, more theoretical and political considerations behind our decision to produce this kind of edition rather than another, see Flanders (forthcoming), and Lavagnino (forthcoming).

⁶ For example, this is the opinion of Mark Olsen of the ARTFL project.

⁷ On Project Gutenberg, see <http://www.promo.net/pg/>

and because we want a usefully rich layer of encoding we cannot avoid making decisions; we just try to base them on a rational, explicit set of principles applied consistently throughout the textbase. Documenting the history of these decisions is important to us, not just because it preserves a record of the Project's growth, but also because we feel strongly that users of our electronic texts need to be aware that while what they see on their computer monitor may seem a straightforward reproduction of the original work, it did not get there by a wholly mechanical, 'objective' process.

Delivery

The factor of human intervention has another importance- cost. Producing a richly-encoded text is slow, painstaking work--highly labor intensive whatever your technological support. Who pays for this labor? At the WWP, we have been fortunate in consistently securing base funding from the US National Endowment for the Humanities, and we have been solidly supported by our parent institution, Brown University. However, grant funding, particularly in the humanities, so often subject to the vagaries of the political climate, does not provide a stable foundation for a long-term project; and rarely does it provide sufficient money for a project to do its job in the manner and at the rate it would like. The WWP keeps its costs down by employing Brown students as encoders, and making do with a mostly part-time staff. But this hampers our efforts, and the constant need to apply for base funding takes resources away from the work itself. All these factors make us committed to achieving a reasonable economic return from our activities. We now have to reconcile our central mission of making our texts widely available with the need to generate income from and for our intellectual work. The two are not exclusive: publication and commerce go back a long way together.

We believe our product has value and people will be prepared to buy it if the price is right. However, reckoning the cost of electronic products like ours involves more than counting pennies. What will getting to it and using it cost in people's time and patience? Our means of electronic delivery has to offer easy access for as many people as possible to as many texts as possible. We plan to create a Web-site⁸ and sell licences for full access privileges, with educational institutions and large public libraries as our main customers. The Web is by far the most easily accessible platform for electronic delivery and already has a wide base of users who feel comfortable using it. A Web-based resource has advantages for both buyer and seller. From our point of view, commercial servers allow a finely grained control of access--we can make some of the product freely available to everybody without having to duplicate work or destroy the integrity of the design. Also, we can continuously correct, update, and improve the product without having to recall it from purchasers or send out multiple replacements. Buyers also reap benefits: purchasing a licence gives them in effect a single 'copy' which can be accessed by multiple users simultaneously. It involves no specialised or dedicated equipment, no maintenance or training on their part, no problem in keeping up with new issues, and no worry about the product physically deteriorating, being stolen, vandalised, or otherwise abused.

⁸ The under-construction version may be found at <http://www.wwp.brown.edu/rwo/home.html>

'Packaging'

Delivering sought-after works via the Web gives us a good start in selling Renaissance Women Online, but we cannot rest there. A large part of our marketing pitch will be that because we encode our texts more richly than comparable projects, users can do more with them. But realistically, at least in the beginning, we could expect only a few users to possess the specialised software and expertise to work with the texts independently. This means the site itself must help users take full advantage of the functionality of the texts. Two factors become important here: the software used and the material that is added. In the RWO site, we use DynaWeb to serve our SGML files as HTML to browsers, so that the works can be read and searched onscreen in real time. We prefer DynaWeb over Panorama for many of the reasons we prefer Web-based delivery over CD-ROM: we can deal with problems at the source; we can make changes for everyone simultaneously; users do not have to own, operate, and upgrade any more software than their Web browser; and we do not have to provide support for programs we may know nothing about.

We also have to recognize and work around some software limitations. First and worst among these is the inadequacy of HTML as the standard Web encoding language. Let me mount a soapbox for a second and urge the universal adoption of XML (Extensible Markup Language), the streamlined subset of SGML, as the new standard for Web encoding.⁹ Secondly, for the onscreen version we have to compromise our tagging somewhat to accommodate the DynaWeb search engine. While that search engine is quite adequate for most users' needs, some people will want the more powerful capabilities of an engine like PAT. They may also want to run concordancing programs. They might even want to alter or add to the actual markup. For all of these reasons, we will make available the base SGML files, together with a copy of our DTD, so that users can download the texts and use their own particular applications. Again, users may not want to read long texts using the computer (who does?), so we will also provide for downloading files that will print out into formatted texts without any markup. All of these will be freely available; only the online texts will be restricted to licence holders.

If the RWO site consisted only of the textbase of primary works, then no matter how wonderfully a piece of software might display them, the site would never be used regularly by anyone other than academic specialists. We want more than that, we want our site to persuade people to explore and intellectually engage with the primary works. We feel we have two target audiences with different needs. On the one hand we have scholars who specialise in the field. Their research provokes three needs: firstly, to be able to quickly browse a wide range of texts; secondly, to be able to study a single text intensively; and thirdly, to get information which will help them formulate and support assertions. On the other hand we have non-specialist users, mainly undergraduate students. Their needs will be, firstly, to be able to actually read the texts with relative ease (without having to get over the hurdles of unfamiliar typefaces, special character glyphs, etc.), and secondly, to have expert guidance--written at a non-specialist level--in comprehending and appreciating the works.

RWO meets those needs in the following ways. Providing a wide range of texts presents no problem, given the WWP's comprehensive scope. Ease of browsing is one of the strengths of

⁹ For more on XML, see <http://www.textuality.com/sgml-erb>

the DynaWeb interface, which has a table of contents feature that allows users to move relatively quickly between points in a book. Also, it is much easier to access a large number of texts in a short time in an electronic resource than it is in a rare book library or using a microfilm reader. DynaWeb has the common search facilities built into it, and it also enables customized pre-formatted queries. For example, in RWO we have a ready-made search for personal names: users don't need to know any search syntax, they simply type a name in a field and click a button to start the search. For information about the works, we use Tango software to link Web-page query forms to a Filemaker database. Users can find what the textbase contains according to author, title, date of publication, Short Title Catalogue number, and so on. For each work, they can access a set of basic bibliographical data. None of this pretends to be exhaustive; it exists only to help scholars make quick decisions about what works to look at, and to save them some of the time and effort that researching facts about primary works involves.

Making our online versions of the texts easy to read is not only desirable, but also unavoidable. The limited formatting capabilities of the Web make it impossible to accurately reproduce the appearance of the print original; nor does that particularly concern us, because scholars with a special interest in visual details are better served by consulting an original copy or viewing the work on microfilm. And while our encoding records things such as missing, turned, or incorrect letters in words, or the use of 'i' for 'j', 'u' for 'v', etc., we omit that encoding from the onscreen book version because having tags in the middle of words makes it hard to search for those words. A lot of the characters peculiar to old typography (long 's', macrons, thorns, etc.) we replace with their modern equivalents, because most web browsers have no way of resolving the entity references that we use to encode them. The only obviously 'Renaissance' feature of the onscreen version is the period spelling. Although modernizing the spelling would make specifying search strings easier, we feel too much would be lost in the process, not only in terms of information of potential value to scholars, but also in the reading experience.

While the free-for-all of spelling in these works may be a new thing to undergraduate readers, it will by no means be the most unfamiliar thing they have to cope with. With little experience of sixteenth- and seventeenth-century writing outside of Shakespeare, today's undergraduates could easily open an RWO text and find themselves lost in a sea of unfamiliar thoughts generated by unfamiliar contexts and expressed in a syntax so unfamiliar as to be almost a different language. And all this from authors who, I confidently assert, have not been, nor are ever likely to be the subjects of big-budget Hollywood treatments. So RWO has to do more than just bring students and works together, it has to ease the one into the other. We try to do this by providing a variety of contextual materials that help orient the student with regard both to individual works and to the principal topics that women writers of the time thought about, responded to, published on.

Each individual work comes with its own little 'suite' of materials. First, a very brief description of the work helps users decide whether they want to investigate this work further or not. If they do, a short introduction, very like the kind of introductory paragraph found in print anthologies, describes the work. Next comes another short section saying something about the textual history of the work. Was it printed abroad and smuggled into England? Did different copies have different dedications? How many editions did it run to during and after the author's lifetime? etc. Then, to help students (and scholars) begin to connect that work

with others, another section points out connections users can follow to increase their understanding of how that particular work fits into its era's overall scene of writing. For example, the author may have been part of a coterie of women writers and her work might be a response to something another member wrote. Or the work might be an example of a particular *kind* of writing (e.g. prophetic writing) and it would be useful to know what other instances of that kind of writing the textbase contained. It might be that the work treats a particular theme or issue, or is one in a series of works arguing a particular question, like the famous polemical debate about the 'woman question' which followed the publication of Joseph Swetnam's *Arraignment of ... women* in 1615. The theme, or topic, also guides users into the textbase from the other direction, because we have a collection of short essays, concise summations of the essentials of a topic, similar to encyclopaedia entries. So, for example, a student who wants to write a paper on the early seventeenth century debate about women can go first to a short essay on that topic. From there, links will lead to the particular works in the textbase which directly relate to the topic.

Like the database information, these contextual materials do not try to be comprehensive. For one thing, they come from scholars volunteering some of their scarce free time to write them for us, and there's a limit to what we can ask of people to whom we cannot offer any recompense. Also, the purpose of the RWO initiative remains the same as that of the WWP as a whole, namely, to bring people to the texts so they can read them, work on them, discover new things about them. The ends, not the means, give electronic resources their value--especially those with pretensions to educational utility. Whenever we feel ourselves getting carried away with some item of hardware or software that performs a function the old one did not, we should remember that, at least in humanities studies there is almost no operation a computer can perform that a scholar cannot also. And while computers perform mechanistic operations faster than any human ever will, they cannot do anything clever, creative, and illuminating with a text without a human intelligence guiding them. Scholars tell computers what to search for, count, collate, etc., and scholars assess the significance of what computers find.

Conclusion

I wish the title of this talk could read 'How the WWP Broke the Bank While Changing the Face of Scholarship'. Starting this venture feels exciting, but it also makes us a little nervous. For a long time now the Women Writers Project has promised that one day it will deliver the electronic goods. When we formally launch Renaissance Women Online a little over two years from now, we put ourselves as a project not just online but *on the line*. For years scholars have been asking for our texts for their research and especially for their classes. They have been willing to pay for crudely-formatted and error-ridden printed versions. We tell ourselves that naturally people will want to use our cleaner, better, more functional electronic texts, and they will clamour for their institutions to buy access to the resource as soon as possible. But then, Ford were convinced they had a winner in the Edsel. We hope not to go the way of that sophisticated failure, and we believe that in focussing our efforts on rich text encoding and the provision of contextual materials for our textbase, we have chosen the right route.

References

- Burrows, Toby. Educating for the Internet in an academic library: the Scholars' Centre at the University of Western Australia. *Education for Information* 13 (1995): 229-242.
- Burrows, Toby Using DynaWeb to Deliver Large Full-Text Databases in the Humanities. *Computers and Texts* 13 (1996): 15-17.
- Flanders, Julia. The Body Encoded: Questions of Gender and the Electronic Text. In *Electronic Textuality* . Eds. Marilyn Deegan and Kathryn Sutherland. Oxford: OUP, forthcoming.
- Lavagnino, John. Electronic Editions and the Needs of readers. *New Ways of Looking at Old Texts II: Papers of the Renaissance English Text Society, 1992-1996* . Ed. W. Speed Hill. Tempe, Arizona: MRTS for RETS, forthcoming.
- Mah, Carole, Julia Flanders, and John Lavagnino. "Some Problems of TEI Markup and Early Printed Books." *Revue Informatique et Statistique dans les Sciences humaines*, 32 (1996): 183-202; also revised version forthcoming in *Computers and the Humanities*.
- Sperberg-McQueen, C.M. and Lou Burnard, eds. *Guidelines for Electronic Text Encoding and Interchange (TEI P3)*. 2 vols. Chicago: Text Encoding Initiative, 1994.
- Paul Caton is Electronic Publications Editor at the Brown University Women Writers Project. Currently on leave from the Graduate School at Brown, he hopes to complete his dissertation on text encoding and literary studies before the next millenium.*