If anyone understands the real issues relating to representation of knowledge, it is the textual editor. Born in traditions associated with transcription, comparison, collation, and classification; working in varied media and across traditions and formats within and among those media; and holding high ideals, or at least striving for accuracy often difficult to achieve in pragmatic terms, the editor practices the science of knowledge conveyance, re-presenting and representing, modelling and then transmitting the contents of textual human artefacts from one form to another, pouring old wine (as has been said very many times) into new bottles . . . and, of late, doing so into containers quite unrecognisable to all but the most contemporary of authors. If we are to believe scholars such as Northrop Frye, humanists have always “lived in a far more efficient technological world than most of their contemporaries” (7–8); this technological advantage, Frye asserts, has historically been connected to the prestige of humanists. So, too, can this be said of the editor . . . perhaps especially so in a contemporary climate that pits page against electronic-hyperlinked-scroll, and print-codex against ephemeral internet and other forms of electronic data transmission.

1 Originally presented at the RETS Josephine Roberts Panel, MLA December 2004 (Philadelphia) / RSA April 2005 (Cambridge, UK). The work presented here reflects also the involvement of Caroline Leitch, who has investigated collation tools for use on the Devonshire MS.
The rise of the electronic medium in the past two decades—especially the increase in efficiency and affordability of electronic methods to manage and disseminate textual and extra-textual materials—has allowed us to realise the ideals of one of the most important schools of contemporary thought surrounding archival representation, what is often referred to as social theories of editing, theories that expanded the notion of edited text to include what were often considered to be extra-textual elements. Such thinking, combined with computing technology, has brought about significant change in our field, perhaps most noticeably in the role of the textual scholar working in the electronic medium over the past decade, who increasingly began to accumulate information rather than synthesise it—an accumulation, made possible only by computing, also associated with the critical process of unediting, paying an increased attention to the full materiality of the texts that are, ultimately, the objects of our consideration.

Like a number of us, across the past decade, I’ve had the pleasure of editing both for print—which people have said, for several years, is moving on the way out and, yet, it shows no real sign whatsoever of doing so—and the electronic medium, which, for a decade or longer, has been heralded almost as a saviour, of sorts, but has yet to manifest itself to us in any form nearly as “intuitive” or familiar as what can be captured and navigated in print form. I’m overstating, to be sure, but do so to the good end of drawing attention to the fact that, contrary to speculation some time ago, the book is not nearly dead, nor do we yet understand the electronic medium well enough for it to live up to its promise. The reason for this, I would urge, is largely because we have yet to articulate the salient features of the new electronic book—and, more specifically, the new scholarly edition in electronic form—in a way that suggests the place of the new electronic book in our personal and professional lives, in our ways of doing things when we work alone, and in our ways of doing things when we work and meet with others.

An example not drawn from the world of the book might bear this out more clearly. Anyone participating in what has often been termed an “electronic conference” will know that it is not like a conference at all; rather, if the e-mail listserv is the computing facilitator of interaction, it is like passing notes around in a darkened room, where you can read what people have to say but not hear them or see them. Aural and visual technologies are making this better, of course, but

---

2 Such as visual presentation, incorporation of visual and aural material, &c.; the best foundational arguments for this are found throughout in McKenzie (Bibliography), and McGann (Critique, Textual Condition).
3 See Schreibman (“Computer-Mediated Texts,” 283 ff.).
4 See McLeod (“Information,” 240–245 ff., “UNEditing,” 26 ff.), and Sutherland (“Revised Relations,” 17 ff.).
there is no technologically-facilitated substitute, yet, for live contact and collegial interaction as one finds at a conference. Those who develop such technologies have not yet been able to model, accurately, and to represent, adeptly and conveniently, the processes and practices of the conference environment such that the virtual has something on the real.

So, too, perhaps, it is with the edition and the electronic edition—thus far at least. But I would suggest that we are closer in modelling the scholarly edition electronically than others have come in modelling the conference via electronic means—in large part because of groups such as that which drafted the MLA Guidelines for Electronic Scholarly Editions, and I would urge in larger part this is because the scholarly editing community has always lived with an awareness of the demands essential to the proper transmission of textual content from one object to another, from one form to another, and from one medium to another . . . ideas central both to modelling and to the re-presentation of knowledge. 5

**Miscellany Editing**

Indeed, it is such a community awareness that I’ve enjoyed alongside the experience of editing by relatively traditional means the Henry VIII MS (BL Add. MS 31, 922), then carrying out revisions on that edition while in the early stages of preparing an electronic edition of the Devonshire MS (BL Add MS 17, 492), which I’m editing as part of a joint pilot project between MRTS and ITER.

These works are both miscellanies, both originating in English royal circles within fifteen years of one another. The Henry VIII MS is one of three early Tudor songbooks, and can be dated ca. 1520–1523, reflecting events and activities from before 1509 until mid-1522. It is most notable in that it is the primary evidence of Henry VIII’s poetic and musical prowess; it allows us a glimpse into the way he chose to use these forms of expression to fashion himself, as a young

---

5 For primers into these concepts in a humanities context, please see Unsworth and McCarty. As Unsworth notes, “[k]nowledge representation draws on the field of artificial intelligence and seeks to produce models of human understanding that are tractable to computation”; further, to paraphrase, he states: While fundamentally based on digital algorithms, knowledge representation privileges traditionally-held values associated with the liberal arts and humanities: general intelligence about human pursuits and the human social/societal environment; adaptable, creative, analytical thinking; critical reasoning, argument and logic; and the employment and conveyance of these, in and through human communicative processes and other processes native to humanities. McCarty’s argument suggests that, in activities of the humanist using the computer, knowledge representation manifests itself directly in issues related to archival representation and textual editing, high-level interpretive theory and criticism, and protocols of knowledge transfer—all as modelled with computational techniques; we model our data, our intellectual processes, and beyond.
monarch championing the ideals of courtly love, in a court that all too often attempted to impose its fashions and those of his father upon him.

The Devonshire MS is best called, as Paul Remley has, “a courtly anthology”; it is of 1525–1559, with the period of its greatest activity being the mid-1530s. It has been a standard witness for Wyatt’s poetical works for several centuries, but is today most notable because it presents the earliest sustained instance in the English tradition of a literary writing community comprising both men and women and, thus, a primary site of women’s involvement in the poetic-political world captured by the early Tudor lyric.

Both editions aim to serve a similar audience, scholars interested in early Tudor poetry; therefore my treatment of each text has been in accordance with the school of documentary editing and conforming to well-accepted editorial standards. But there are differences: specifically in the areas of transcription and collation, which must be considered if we are to describe a model of the new scholarly edition in electronic form. I’d like especially to highlight those differences as they relate to the promise of modelling that new scholarly edition.

Transcription

As we know, simply put, transcription is the act of copying from one textual instance to another. When we transcribe an early modern work, we represent that work in our character set, with a set of accepted protocols to facilitate this, and an understood way of handling the expansion of early abbreviations, digraphs, non-modern characters, and so forth. These protocols we outline for the most part in a statement of editorial principles, especially so if our own practices represent a slight deviation from accepted practice. Figure 1 suggests to us some of those practices, as carried out on the text immediately underlaid the third musical voice, which comprises the first stanza of “Pastyme with good company.”

Pastime with good company
I loue and schall do tyl I dye
gruche who lust but none denye
so god be plesyd thus leue wyll I
for my pastance
hunt syng and dance
my hart is sett
all goodly sport
for my comfort
who schall me lett.
While working on the Henry VIII MS, I transcribed as many did at that time in the Manuscript Students Room at the British Museum, from original manuscript, to paper, with pencil. This passage of information capture was one that would be quite familiar to most: I kept a list of character forms and abbreviations for each of the several hands of the manuscript, and attempted to represent graphically the yoghs, digraphs, abbreviations, and other forms that I knew, later, would have to be represented in some way via our standard character set when I ultimately re-transcribed this work into the computer. I checked readings against microfilm and, ultimately, again against the original. Further, as some do, I held
and tho that I bebanest hym fro his spech, his sight, and company, yet will I in spite of his flattery love and keep my fancy.

do what they will and do the worst for all they do is vanity.

Figure 2: Image and Transcription, Devonshire MS (“My heart ys set not remoue,” ll. 9–16, 59r).
off converting the yoghs, digraphs, and abbreviated forms until I’d done the collation and was prepared to make a commitment to consistent re-presentation of these forms. Happily, the hands were legible (because the scribes were professional), the texts were clear, and the only thing that caused me significant concern at this point was the fact that I was transcribing, for a literary audience, a text which was so very clearly aligned with the music. But the music of the manuscript had been thoroughly treated by the musicological community as much as the lyrics of the manuscript had been overlooked; so, there was justification to provide a proper edition and study of the lyrics alone, for a literary-studies audience.

Transcribing the Devonshire MS was a very different enterprise. The majority of its contents are lyrics, but they are captured solely in textual form by the manuscript’s compilers. Hands were more difficult (because they are predominantly amateur), and texts were obscured at times, a function of this manuscript having very different characteristics and uses from the Henry VIII MS. Transcription was from microfilm and electronic facsimile (derived from microfilm), directly into a text editor on laptop computer. Here, I worked with two research assistants, and we independently duplicated each other’s work—comparing results at set intervals using a collation program, with the understanding that this process would highlight areas that would require special attention when I checked our full transcription against the original in the British Library. Transcribing directly into the computer meant that we had to make immediate decisions about character sets to represent earlier forms, such as digraphs and abbreviations, and to accommodate this we adopted and augmented a list of standard earlier forms and semi-standard ASCII-text representations which has been published by the Renaissance Electronic Text series at the University of Toronto.

The biggest difference in transcribing the two manuscripts, however, was not in workflow, nor in the use of computing technology itself as part of the transcription process. Rather, the biggest difference was found in the demands that a computing approach to transcription made on our group once we had finished our initial transcription and verified it against the original. We had deliberately held off as long as possible on introducing an essential component of the transcription process in the electronic medium: the application of an appropriate document encoding system. So, at the stage we might well have been making ready for other tasks associated with creating the edition, we began preparing the transcribed text for the application of a system conforming with Text Encoding Initiative XML—at the same time as a subset of the TEI Consortium
group was working toward establishing the protocols for describing the type of documents that were our focus. This involved modelling our textual data with a Document Type Definition that had as much parity as possible with the TEI—and looking further to projects such as the Digital Scriptorium at Berkeley and Kevin Kiernan’s Edition Creation Tools at Kentucky for their own modelling practices. Once modelled, we applied the DTD to the text in several passes, initially with a focus on structural and surface-level elements; that transcription and encoding is represented in Figure 2.

Time-consuming as this process was, it has added significant value to the work we will ultimately produce. As proponents of TEI and other encoding systems will say, modelling the data of an edition with a DTD forces editors to make explicit their understanding of the work being represented, in a constrained logical vocabulary that has application to other edited works as well. That DTD becomes the textual-critical vocabulary of the edition, a description of the edited text that can, by itself, facilitate study of the original artefact.

**Collation**

Collation, the second of the areas I’d like to highlight, makes for a different type of story. The collation associated with the Henry VIII MS was typical, initially involving a standard visual comparison of the base text against witnesses as I was able to encounter them. Since the works in the manuscript are polyphonic, in most cases I had “internal” witnesses to consider among the several lyrical voices of each piece. I’d transcribed all the witnesses and, in the end, I entered those transcriptions into the computer so that I could check my initial collation with the results provided by the best computing tools available at the time: early versions of the Donne Variorum Collation Program, PC-CASE, and UNITE. This was all to good end, as the use of these tools allowed for detailed comparison such that my initial work was augmented and, further, one of the tools assisted in the formatting of my textual apparatus, which is typical of the apparatus one would expect in a good print edition. The final results appear, roughly, as below:

Texts Collated:

\[
\begin{align*}
H1,2,3 & \quad (14v–15r, \text{ ll. 1–10 } H2,3 ), \quad LRit(1)1,2,3 & \quad (136v–137r, \text{ ll. 1–10}), \\
LRit(2)1,2,3 & \quad (141v–142r) \quad \text{Emendations of the Copy Text (H1):} \\
4 & \quad \text{leue} ] \quad \text{loue } H1, \quad \text{leue } H2,3, \quad \text{lyf } LRit(1)1,3, \quad \text{lyue } LRit(2)1, \quad \text{lyfe } LRit(2)2,3
\end{align*}
\]

---

I’ve discussed details of some of the more interesting problems we encountered in encoding the Devonshire MS in a forthcoming article, “The Devil is in the Details.”
Collation of the Henry VIII MS

In working on the Devonshire MS, computer-assisted collation was a process that we used as part of our transcription of the original document, to ensure the accuracy of our work at that stage. But the way in which we will collate the Devonshire MS against its witnesses is still a bit of a mystery to us. Let me explain what I mean by this. Following typical practice, we’ve gathered all the witnesses necessary for our edition, and we are nearly done with the process of transcribing and encoding them to a standard similar to that of our copy text. But we realise that the ultimate form of the textual apparatus, and the collation process we will employ to create that form, will be largely dependent on how we envision people making use of the edition of the Devonshire MS—and, further, the limitations of publication. We are considering, for example, creating a standard apparatus, based on the print model, integrated perhaps via hypertext with the various forms and instances of the texts the apparatus takes into account. Another possibility is for us to rely on our TEI encoding to house the various variants relating to the works in the Devonshire MS and to present the encoded manuscript and witnesses with software that would allow the readers themselves to carry out the collation, constructing their own apparatus in the process. In trying to find a computational solution to the problem of how best to collate the Devonshire MS against its witnesses, we experimented with a number of collation tools, including the Versioning Machine (version 2.1), Collate, and Juxta. While each of the tools we have worked with satisfied some of our needs, we found that no one tool was able to do it all.

In some ways, the tool with which we have had the most success is the Versioning Machine. The Versioning Machine satisfies our demand for open-source, freely distributed software and runs in the reader’s browser without requiring plug-ins, making it ideal for most users. It allows the user to select and display
Figure 3: Collation results and formats of the Devonshire MS, Versioning Machine.
a number of witnesses in the same browser window, view images, and see textual and bibliographic information either inline or as pop-up notes. Most importantly, the Versioning Machine makes excellent use of the TEI’s apparatus for comparing witnesses. By demanding well-formed TEI-conformant XML as its base, using the Versioning Machine encouraged us to pay close attention to our encoding standards, an important consideration if we are going to make the Devonshire MS and its witnesses available in their encoded form. As an added benefit, the Versioning Machine is a current project which means that the tool is not static. Rather, it is being updated and improved as we use it.

Although the Versioning Machine has satisfied our requirements in terms of displaying entire witnesses side-by-side, it was not the ideal solution to our problems of how to display variation on a word-by-word basis and how to regularize orthographic variations. For that capability, we turned to Peter Robinson’s program Collate, a collation tool which first appeared as a Macintosh program in the 1990s. Robinson himself admits that Collate, along with his XML publishing tool Anastasia, which we also experimented with, is very difficult to use. Robinson notes, “as their creator I think I am uniquely qualified to note that [Collate and Anastasia] are not easy to use: if everyone who wanted to make digital editions was required to use these two tools, very few digital editions would ever be made.” After using Collate for many months, we are inclined to agree. That is not to say that Collate does not have its advantages. Collate has a simple, and surprisingly intuitive, interface for regularizing variants. Our experimentation with Collate has provided us with a wealth of raw material in the form of collations that allow us to focus our attention on substantive differences between texts without getting bogged down by orthographic variations. However, we did encounter many of the problems Robinson warns of: Collate is unpredictable at times, both in its performance and its results; it requires an older version of the Macintosh OS, one which a decreasing number of editors use; Collate is an older program that is not currently being updated.

The third collation tool we used, with mixed results, was Juxta, created by Applied Research in Patacriticism. We were pleased with the elegant way in which Juxta displays visual information. The reader can easily see variation on the level of the whole text, line-by-line, and word-by-word. Juxta also allows users to view digital images of manuscripts and includes a histogram tool. Like the Versioning Machine, Juxta relies on encoded texts. The encoding required by Juxta, however, is much more simplified than that required by the Versioning Machine. This is both an advantage and a disadvantage. Encoding using the TEI’s critical apparatus tagset, as required by the Versioning Machine, while time-consuming, may be a better long-term decision because it lends itself to a wider variety of applications.
Another of our concerns with Juxta is that its performance has, at times, been inconsistent across platforms. Now that the source code for Juxta has been made available, we are looking forward to looking under the hood, so to speak, to see if we can tailor Juxta to our specific needs.

<table>
<thead>
<tr>
<th>LDev14o-TM1046_collated</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LDeyv14o-TM1046</td>
<td>My hope alas hath me abused</td>
</tr>
<tr>
<td>AAh-C7</td>
<td>abused</td>
</tr>
<tr>
<td>LGee21</td>
<td>Alas abused</td>
</tr>
<tr>
<td>LDeyv14o-TM1046</td>
<td>and vaire reioising hath me fed</td>
</tr>
<tr>
<td>AAh-C7</td>
<td>And vaio reioycinga hath fed</td>
</tr>
<tr>
<td>LGee21</td>
<td>vaine relaysing hath</td>
</tr>
<tr>
<td>LDeyv14o-TM1046</td>
<td>lust and loye have me refusid</td>
</tr>
<tr>
<td>AAh-C7</td>
<td>Luste loye refused</td>
</tr>
<tr>
<td>LGee21</td>
<td>refused</td>
</tr>
<tr>
<td>LDeyv14o-TM1046</td>
<td>and carefull playnt is in there sted</td>
</tr>
<tr>
<td>AAh-C7</td>
<td>carefull their sted</td>
</tr>
<tr>
<td>LGee21</td>
<td>carefull plaint their stede</td>
</tr>
<tr>
<td>LDeyv14o-TM1046</td>
<td>moche avancing siakte my spede</td>
</tr>
<tr>
<td>AAh-C7</td>
<td>advauncing slaked speed</td>
</tr>
<tr>
<td>LGee21</td>
<td>mucho slaked</td>
</tr>
<tr>
<td>LDeyv14o-TM1046</td>
<td>mirth hath causid my heavynes</td>
</tr>
<tr>
<td>AAh-C7</td>
<td>myrth hath caused heavynes</td>
</tr>
<tr>
<td>LGee21</td>
<td>myrth hath caused</td>
</tr>
<tr>
<td>LDeyv14o-TM1046</td>
<td>and I remaine all confortlesse /</td>
</tr>
<tr>
<td>AAh-C7</td>
<td>And remayne comforties []</td>
</tr>
<tr>
<td>LGee21</td>
<td>remain comforties []</td>
</tr>
<tr>
<td>LDeyv14o-TM1046</td>
<td>Where to ded I assure my thought</td>
</tr>
<tr>
<td>AAh-C7</td>
<td>Where to did</td>
</tr>
<tr>
<td>LGee21</td>
<td>did</td>
</tr>
<tr>
<td>LDeyv14o-TM1046</td>
<td>without displeasure stedfastelye</td>
</tr>
<tr>
<td>AAh-C7</td>
<td>stedfastelye</td>
</tr>
<tr>
<td>LGee21</td>
<td>without stedfastly</td>
</tr>
</tbody>
</table>

Figure 4: Collation results and formats of the Devonshire MS, Collate.

If we have learned one thing from our experience using a variety of current collation tools, it is that customization is in our future. There is no one collation tool that satisfies the needs of our particular project. In an ideal world, we would
obtain a tool that combines the TEI-based encoding of the Versioning Machine, the regularization interface of Collate, and the user interface of Juxta. Until such a tool exists, we must rely on the greatest advantage of all three of these tools: their ability to teach us more about how collation software works and does not work.

Figure 5: Collation results and formats of the Devonshire MS, Juxta.

Integration

In discussing my experience with transcription and collation with both of these projects, I don’t think I’ve said anything that is startling or new. But these changes in the way things are done, and changes in the “end products” of each essential editing process, are new to us as we encounter them; and these ways and end-products in the electronic medium are changing with the introduction of new technologies and the research results of our colleagues. Furthermore, what the changes in each of these processes, and others, represent to the editing community is potentially something more significant than the sum of their parts. In the case of transcription, we find that the encoding now associated with transcription in the electronic medium provides an explicit model of those data, ultimately (in the best of circumstances) adding to the level of bibliographic description we give our texts and, in turn, the utility of those texts to those in our community and beyond. In the case of collation, we find that it is not the data but rather the process which is being modelled in the electronic medium—ultimately adding
significant possibility to the way in which we represent the textual apparatus in our editions. Both have to do with strategies of the representation of knowledge, and both produce an electronic end-product whose utility is far greater than it would be if produced solely in print.

It is on this notion of greater utility that I should like to conclude. The promise of social theories of editing, and movements toward un-editing and what has been called material textuality, is a greater attention to the pertinent details of individual textual instances and the factors informing them. The promise of computation, in this context, is as facilitator for the processes involving the representation of those instances; here, transcription and collation have considerable importance, for the computer alters and enhances elements of each, both in process and in product. Further, there is also a larger context to consider here, one that involves the other actions pertinent to our discipline that are, ultimately, centred on the re-presented artefacts that our community produces; inherently connected to our representation of archival materials are the essential activities of critical inquiry and communication of results—and there are considerable gains to be realised here as well, gains the origins of which are found in these activities.

These are necessary steps toward a proper articulation of the salient features of the new scholarly edition in electronic form, a virtual object that, itself, will capture the best of our traditions at the same time as it explicitly integrates those traditions associated with archival representation with those relating to critical inquiry and the communication of results. Without doubt, the most interesting and profitable work in this regard is being done by members of our community. Returning to my earlier analogy of the electronic conference with reference to modelling practice, I’d like to point out, and happily so, that the way in which we are able to model the data of the transcription and the process of collation is well in advance of the passing-around-notes-in-the-dark model of the electronic conference.

University of Victoria
Bibliography


PC-CASE.


———. Collate.


