Encoding as Editing as Reading

Alan Galey

One of the great things about digital textual scholarship is that its difficulty makes us all students again. I was reminded of this recently when I happened upon an account of the creation of one of the earliest digital databases for early modern theatre history, thanks to Ben Ross Schneider Jr.’s 1974 book Travels in Computerland; or, Incompatibilities and Interfaces: A Full and True Account of the Implementation of the London Stage Information Bank. Schneider, a professor of English at Lawrence University, undertook in 1970 to digitize – or to computerize, in the parlance of the time – the eleven-volume reference work The London Stage, 1660–1880 and from it to build a database to support queries by actor, role, theatre, performance, and other parameters. This would be no small enterprise even today, and Schneider’s account offers a glimpse of the experience of humanities computing in the days before email, the Web, social media, Skype, cell phones, and even inexpensive long-distance telephone calls. Yet Schneider’s book also speaks to the future by triangulating with the past in the form of his playful adoption of the eighteenth-century novel’s picaresque form, as in the argument for one of the opening chapters, titled ‘Of Feasibility’:

The author becomes entangled in the London Stage Information Bank & conceives of a Pilot Project. He questions Administrators & Computer Scientists, visits Widener Library & the Association for Computing Machinery, where he encounters a Living Advertisement. He considers the scarcity of Funds, and draws to an Inside Straight.  
(Schneider 1974: 7)

Those who have undertaken a large humanities computing project – and the fewer still who have completed one – may have trodden some of the same roads, and thus will have no difficulty imagining the travails of such work in the form of a novel (if not an eighteenth-century satire such as Schneider’s possible model, Tristram Shandy, then perhaps Bleak House or Ulysses). Schneider’s appropriation of literary form to narrativize a digital project says something about his experience as an early modernist venturing into an unfamiliar world. His fascinating account also reads like a message in a bottle sent from the distant shore of an earlier era of humanities computing – one that helps us see our present with new eyes.
The most striking difference between Schneider’s experience as a computing humanist in the early seventies and my own today (beginning in the late nineties), was that computing in his time required much greater amounts of time and energy for the task of data entry. Indeed, *Travels in Computerland* reads not as an early seventies reflection on the whole life-cycle of a digital project, but specifically as an adventure in humanities text encoding. This adventure includes truly heroic feats of data entry: at one point, Schneider mechanically modifies an IBM type ‘golfball’ to prevent the capital letter O from scanning as the numeral zero (149–57). Reading as someone who grew up amid the luxuries of personal computing, I was reminded that computing in 1970 was not something one did at home on a sunny back porch with a dog at one’s feet, on a relatively inexpensive machine that an individual might own, glutted with enough surplus processing power to permit inefficient programming and memory-management, and with barely seconds (not hours or days) intervening between the act of writing a bit of code and testing its success or failure. Even screens were a luxury, prompting Schneider to record his delight upon seeing a Cathode-Ray Tube monitor for the first time (104–5). Computing in 1970, like printing throughout its history, was a mechanical art. For Schneider in the episode of the IBM type golfball, computing even became something one did with a sliding saw-mount, vicegrip, and grinding wheel (pictured in Schneider’s book), all for the sake of altering a tiny bit of metal that skewed his data model.

What can all this teach us about encoding, editing, and reading Shakespeare today? All three of the terms in my title have scholarly lineages and literatures which deserve a fuller treatment than I can give in this short chapter, and which cannot be conflated merely by stringing them together with the word ‘as’. Reading and editing, for example, are terms that Eric Rasmussen and W. B. Worthen productively complicate in their chapters on apparatus and post-print performance, respectively. Encoding, likewise, takes on different connotations in the context of the highly networked scholarship that David Weinberger advocates in his chapter. The present chapter should not be mistaken as an introduction to text encoding, though I hope it will introduce new and established digital Shakespeareans to alternatives to project-centric best practices in the digital humanities, in which the imperatives of a funded project (and the acquisition of more funding) can often foreclose the broader processes of intellectual inquiry that distinguish scholarship from business.¹

Shortly before the mainstream humanities suddenly began to embrace digital humanities around 2009, Willard McCarty published his foundational book, *Humanities Computing*, in which he warned against thinking of computers as mere ‘knowledge jukeboxes’ (2005: 27). Although he was well aware of the value and potential of digitization projects – and had won that knowledge the hard way, as a coder and scholar long before humanists showed that combination much respect – McCarty nonetheless felt compelled to warn that ‘the torrent rushing out of computers into the various disciplinary heartlands pull[s] attention away from the difference between cultural artifacts and the data derived from them – away from the analytic concerns of earlier work . . . to a great stocking of the shelves’ (5). That was 2005, but today any digital

¹ The literature on text encoding is very large, but two complementary starting points I recommend are Deegan and Sutherland 2009: 59–88 and the TEI By Example Project.
Shakespeare project that seeks to investigate that very difference between artifacts and data could find itself strongly at variance with funding models. The result is that the shelf-stocking instinct has been internalized by many digital humanists, and project management often overshadows the other forms that digital scholarship can take. Might there be a middle ground between the extremes of business-model projectism, on the one hand, and ungrounded speculation about digital technologies, on the other?

My approach seeks that middle ground at the intersection between method, theory, and pedagogy – including the reflexive pedagogy of scholars willing to become students again within their own projects. Specifically, I take *Travels in Computerland*’s historical alterity as a spur to consider the idea of thinking through making in relation to some of its specific implementations in Shakespearean textual scholarship, namely editing and digital text encoding. Schneider would almost certainly have understood his project as a form of thinking through making, and his book is full of reflections on the nature of information, computing, history, the London stage, and the conditions of research in the humanities and the sciences. However, in his account, moments of critical engagement with digitized materials happen under pressure – even the literal pressure of vicegrips – and emerge in tension with the project’s size, complexity, and need to produce results. The luxuries of twenty-first-century computing afford us greater opportunities to slow down and explore the moments of critical understanding that pass by so fleetingly in Schneider’s account. This chapter considers what Schneider called data entry – or, as I prefer to call it, encoding – as a crucial moment when we can understand our materials anew through the act of making digital representations of them. Close reading is often dismissed if not maligned in big-data approaches to digital humanities, but I would like to consider how digital text encoding, like the more traditional activities of textual criticism and editing, lead back to granular engagements with texts that resist, challenge, and instruct us.

***

Many disciplined forms of reading begin with acts of writing in the form of transcription. Textual scholarship and its subfields have long depended on various kinds of transcription – paleography cannot be taught without it – but what can textual scholars learn from transcription as a practice that spans domains? The multiple meanings of transcription in music, for example, mirrors encoding in humanities computing, in the sense that transcription can refer to the practice of writing out pieces of music in standard notation for the sake of documentation, or as a process by which students learn the music itself. As in many digital Shakespeare projects, the one rationale for transcription may blur into the other. For example, one of the great exemplars of transcription in rock music is the guitarist Steve Vai, who as a student in 1979 sent Frank Zappa a transcription of a particularly difficult Zappa drum instrumental along with a sample of his playing, which prompted Zappa to hire Vai to create transcriptions of several of Zappa’s labyrinthine guitar solos, most of which had never been written out in musical notation. Vai’s now-legendary transcription feats led to *The Frank Zappa Guitar Book* in 1982 and, more significantly, to Vai developing such a deep understanding of the music that he joined Zappa’s band (Watson
Galey, “Encoding as Editing as Reading” / 4

1993: 390). The lesson in this example is not only the link between discipline and art, but also the value of transcription in coming to understand complex materials. Here transcription serves as an intermediate stage between the music and the musician, and exploits the imperfections of the medium of writing to slow down the act of listening, forcing the transcriber to notice the details and patterns that otherwise go unperceived, though not unheard. Vai listened by writing.

I learned how to read by encoding while an undergraduate research assistant with the Internet Shakespeare Editions (ISE) in the late nineteen-nineties, which involved long hours of proofing transcriptions of the Folio and quarto play texts that had been marked up in SGML and needed conversion to XML. It was not exactly the same as musical transcription, but it called forth similar forms of attention. One formative experience came in the form of a problem in the 1623 Folio text of *A Midsummer Night’s Dream*, when the four lovers lost in the forest awaken and may or may not share the stage with an awakening Bottom (around 4.1.195 in most modern editions). Figure 1, below, shows how this moment is represented in the First Folio (sig. O1v; TLN 1725–8):

![First Folio page showing stage directions](image)

Modern editions tend to move Bottom’s stage direction to the line below, fixing it in temporal sequence after the lovers leave the stage. I realized that XML, which requires that tagged entities not only be named explicitly but also ordered in non-overlapping hierarchies, would impose a logical structure in which Bottom and the lovers might share a fascinating dramatic moment of mutual awareness on stage. Any query of the sequence of stage directions in the scene (tagged as separate `<sd>` elements in the ISE’s encoding scheme) would return ‘Bottom wakes’ in order before ‘Exit Lovers’. Furthermore, any digital model or visualization of the play’s entrances and exits based on this digital transcription would incorporate this intriguing overlap, but digital models based on most modern editions, including the ubiquitous Globe text, would not.

As an undergraduate just discovering the performance criticism of W. B. Worthen and Alan Dessen along with the unediting sorties of Randall McLeod and Leah Marcus, my reading had primed me to notice a textual ambiguity that permitted the worlds of Bottom and the lovers to intersect as they awoke. The placement of the Folio’s stage directions may simply be an effect of a compositor trying to save space – perhaps with an eye to the long prose passage that follows these lines – or it may be an artifact of the simultaneity of action available to the stage, wrenched into sequentiality by modern editions and XML structure alike. My experience as an editorially minded encoder in this case was borne out many years later in Stephen Ramsay’s description of the value of thinking through
Galey, “Encoding as Editing as Reading” / 5

making in code: ‘Encoding texts in XML . . . places one in a simultaneously cooperative and antagonistic relationship with the codes that already subsist in written works. . . . Rather than hindering the process of critical engagement, this relentless exactitude produces a critical self-consciousness that is difficult to achieve otherwise’ (2011: 34). I could not have arrived at the instance of ‘critical self-consciousness’ Ramsay describes had it not come in the course of transcription as sustained labour. However, it was equally important that the ISE’s coordinating editor, Michael Best, and I afforded ourselves the opportunity to pause and learn from the problem, as opposed simply to putting it aside as an inconvenience that slowed down the project.² For textual scholars, like scientists, the discovery of illuminating problems is a form of success that outweighs project milestones; indeed, the rewards of curiosity are the reason many of us undertake digital projects, and why they are worth funding.

For digital Shakespeareans and aspiring musicians alike, there is also a lesson here in the value of the constraints that apply in translating a text from one medium to another, and from one notation system to another. Those constraints, and the transcriber’s response to them, make the difference between intellectual inquiry and the unreflective rule-following demanded by assembly lines. For example, Vai’s transcriptions, like Zappa’s original compositions, dealt creatively with the shortcomings of traditional musical notation in representing the rhythmic nuances of jazz, blues, and other musical forms that reached beyond the European tradition (Watson 1993: 415–18). Similarly, Friedrich Kittler describes a creative solution to the problem of non-Western music’s supposed unintelligibility to European ears in 1904, in the form of a musicologist’s proposal to use the phonograph to slow down recordings, thus making strange melodies and rhythms discernable through deformation (2006: 3–4). Both examples use an intermediate stage of representation to tease out complexity through constraint. Both also blend technology – the gramophone and writing itself – with technique, reminding us that both words share the same Greek root: techne. This productive tension between technology and technique has long been exploited by musicians, as the jazz saxophonist Dave Liebman describes in his introduction to a transcription exercise:

This mode of thought holds especially true in jazz because outside of the specific notes and rhythms, the intangible essence of this music cannot be noted exactly. This includes but is not limited to the subtleties of rhythmic feel and how the artist interprets the beat as well as the use of expressive nuance in one’s sound, aspects of which are usually lumped under the word ‘phrasing.’ In transcribing, a musician is forced to hear and duplicate everything – even the intangibles. (Liebman n.d.)

Liebman’s rationale in this last sentence comes remarkably close to Fredson Bowers’s rationale for the value of transcriptional thinking in the very different domain of descriptive bibliography, specifically the practice of title page

² Michael offers his own reflections on this encoding crux in Best 2009: 34–5. His article alludes to a simple Javascript animation I created at the time for this and similar variants, which I am developing through my Visualizing Variation project: www.visualizingvariation.ca.
transcription of printed books. In describing the intellectual value of title page transcription over photographic reproduction, Bowers points out that imperfections such as flyspecks may appear as punctuation in photographic facsimiles, but the act of transcription ‘forces one to determine what every letter and punctuation mark is’ whereas photographic facsimile reproduction ‘tempts one to ignore uncertainties’ (1962: 136).

Early modern print, like jazz and its musical relatives, productively challenges the rules of the intermediate notational systems we use to transcribe it. W.W. Greg offers the example of the early modern half-italic colon as a transcriber’s bête noir. As Greg puts it in his contribution to a 1926 round table article (in what could be the title of a Frank Zappa song), ‘There is a peculiar terror lurking in the colon’ due to the deliberate visual and typographic ambiguity built into this piece of type, which was a solution to the nuisance of needing separate roman and italic sorts for punctuation in typographically hybrid early modern books (Pollard, et al. 1926: 326). According to Greg, ‘it apparently struck some bright spirit to cast a colon that would do for either [roman or italic], being intermediate between the two. It is not quite upright like the roman one, but much less sloped than in italic’. This example of a printed intangible also offers a window into an encoding insight of an early modern type-founder, who was thinking digitally by exploiting contextuality and ambiguity to save resources. However, as Greg also points out, ‘the result was not quite what was intended, for in a roman passage of course it looked like italic, and in an italic passage like roman! What is the unhappy facsimile printer [or transcriber] to do?’

Greg asks a good question. As anyone who has worked with Optical Character Recognition (OCR) methods in digitization can attest, teaching a machine to transcribe a text can become an exercise in frustration and illumination alike. A primitive version of OCR existed at the time of Schneider’s project, and his Travels in Computerland relates a moment of insight that arose out of the problem of the OCR scanner’s difficulty in interpreting white spaces in the source. At certain points their text used two consecutive white spaces as a field delimiter to indicate that a title followed immediately afterward, yet the scanner registered these two-space units became semantically indistinguishable from spaces between words. As Schneider puts it, the scanner ‘was aware of space but did not measure it. So all space had to be one space’ (1974: 111). This problem, however, triggers a passing reflection upon the nature of writing and human-machine interfaces: ‘this machine’s failure to register meaningful space probably resulted from the designer’s overlooking one of our unconscious assumptions about writing – that space is a delimiting character. It was one of those cases in which computer logic forced into my consciousness an assumption about the nature of things that I wasn’t aware of making’ (111). Part of the answer to Greg’s question, then, is that an encoder can pause to recognize encoding cruxes as teachable moments.

One imagines the lessons to be learned from transcribing texts that make even more complex use of inline white spaces, such as postmodern poetry. A related Shakespearean example can be found in the New Variorum Shakespeare (NVS) editions’ policy of not including the white spaces that editors since George

---

3 See, for example, the Early Modern OCR project: emop.tamu.edu.
Steevens have added to the play texts to indicate – or impose – metrical part-lines shared among speakers. This form of legacy markup from the eighteenth century has become so embedded in Shakespeare’s editing infrastructure that it determines modern line numbering, as in this example from the 2007 Cambridge edition of *The Winter’s Tale* that numbers two typographical lines as a single pentameter line:

POLIXINES . . .  
The art itself is nature.  
PERDITA  
So it is.  
(4.4.97)

The effect of the added white space is to imply an underlying formal unity to argumentative, even hostile exchanges between characters who nonetheless complete each other’s pentameters. In this example, the dialectic between art and nature enacted by Polixenes and Perdita is played out within an overriding metrical order shared by all – or at least by characters entitled to speak in verse.

The NVS edition, by contrast, presents the play text of *The Winter’s Tale* as a lightly corrected transcription of the 1623 First Folio text, and does not add the structuring space that would normally precede Perdita’s ‘So it is’ in a modern edition (Turner, Haas, Jones, et al. 2005). Maurice Hunt connects this editorial policy to postmodernism’s ‘purposeful creation of irregular aesthetics’ out of ‘white spaces ... and related concepts of absent presences / present absences’ (1999: 64). In that light, the NVS policy presents alert readers with a Shakespeare text that raises questions about its own nature. Is any given short line actually part of an adjacent line or just a loose fragment? Is aesthetic cohesion the product of innate linguistic patterns or of imposed editorial desires? Do metrical forms, like numbers and harmonies, have an ideal existence that precedes writing, or does inscription call them into being?

These kinds of questions are pedagogical opportunities, which is why many textual scholars have constructed transcription and editing exercises that ask students to engage with editorial challenges. Anne Hawkins’s valuable collection *Teaching Bibliography, Textual Criticism and Book History* offers several examples, some specific to Shakespeare. Tatjana Chorney’s chapter, for example, describes an exercise based on the multiple-text plays *Romeo and Juliet* and *King Lear*, in which students collate authoritative versions, and then, like good editors, consider the interpretive possibilities that each version opens up and forecloses. As Chorney describes, the exercise prompts students to consider the consequences of granular details such as word choices, and to notice gaps and omissions of longer passages that might not otherwise be apparent (Hawkins 2006: 168–9). Another contributor, Erick Kelemen, has his students work even more closely within the editorial tradition in an assignment he calls a ‘collaborative edition’, in which students work as a group on the apocryphal play *The London Prodigal*, carrying out stages in the editorial process: ‘transcribing, proofing, preparing the text, preparing the textual notes, preparing the content

---

4 On this editorial policy, see Knowles 2003: 12–13. On the handling of metrical part-lines in Shakespeare editing, see Werstine 1985 and Bertram 1981.
notes, and preparing the editor’s preface’ (163–4). Although Kelemen’s students don’t face a transcriptional challenge on a par with Vai’s notation of Zappa, the transcription stage forces them to begin their process by encountering the strangeness of early modern typography and orthography, including the dual forms of the letters s, ifj, and u/v. A key element in this stage of Kelemen’s assignment is that the students then proofread each other’s transcriptions, which elevates transcriptional errors into teachable moments. That debate continues in later stages, as students must decide which conventions of capitalization and italicization to preserve in the edited text (164). The point in these exercises, as Matthew Kirschenbaum describes with regard to his own digitization exercises elsewhere in the volume, is that ‘the students constantly [make] decisions about what counts as information’ (158). The act of making something – a transcription, an edition, a digital archive – shapes the forms of attention that we pay to our materials.

***

My own counterparts to the editing exercises described above attempt to close the ouroboros implied in my title by linking reading back to encoding. The two exercises I describe below owe a debt to Jerome McGann’s game-like deformation experiments with poetry in Radiant Textuality (2001a: 106–35, 139–46, 150–1), and to Rob Pope’s playful exercises in literary production as described in his book Textual Intervention (1995) – with the difference that mine is not primarily a literature classroom, but an information studies classroom in the University of Toronto’s iSchool. As such, we begin with the history of digital markup and its roots in electronic typesetting, and then consider the emergence of descriptive markup as implemented in the metalanguages SGML and then in XML. One of the key concepts here is the distinction between procedural markup and descriptive markup (TEI Consortium n.d.: part v.1). Procedural markup consists of embedded imperatives for the interface or information system; for example, I might mark up the end of my preceding sentence procedurally as

```
<italic>the distinction between <italic>procedural markup</italic> and <italic>descriptive markup</italic>.
```

The <italic> tag tells an information system not what an element is, only what to do with it (i.e., to render it in italics). However, procedural markup cannot distinguish between the different contextual meanings of italics, which may indicate emphasis, book titles, foreign words, quotations, subheadings, references to a word or letter as such, or (as in my example) the highlighting of an introduced term. Descriptive markup, by contrast, seeks to describe what the marked-up elements mean, not how they appear:

```
<term>the distinction between <term>procedural markup</term> and <term>descriptive markup</term>.^5
```

This distinction is fundamental to markup theory and practice, as is the idea that descriptive tagging constitutes interpretation.

Descriptive markup thus raises even more opportunities for interpretive ambiguity and debate than musical transcription. Supposedly a C♯ is a C♯ for everyone, though the aural nuances of a note or chord may change when played in different locations on a stringed instrument, the same chord may be voiced in different ways, and rhythmic patterns may be expressed differently in notation. Descriptively tagging a Shakespeare text can be just as subjective and intellectually rigorous as writing a critical essay, yet even prior to tagging there is the question of distinguishing markup from text. Sperberg-McQueen defines *markup* as ‘all the information in a document other than the “contents” of the document itself, viewed as a stream of characters’ (1991: 35). However, he then goes on to show the inadequacy of the scare-quoted notion of ‘contents’ by demonstrating that the presence of markup in all texts makes the notion of pure content chimerical: ‘word boundaries are interpretive if the source is in *scriptio continua*. Vowels are interpretive if the source is unpointed Hebrew or Arabic. Verse boundaries are interpretive in all editions of *Beowulf*, since the manuscript does not mark them’ (35). What would be the equivalents in Shakespeare’s texts? We have considered some already and will return to others before the end, but what matters here is the process of finding them.

My first exercise tests the notion of pure content by attempting to replicate the process by which James Coombs, Allen Renear, and Steven DeRose arrive at an example of supposedly unmarked text, one of a set of examples of the same Milton quotation represented in four different states of markup: descriptive, procedural, presentational (i.e., using indentation rather than explicit tagging to indicate a block quotation), and, finally, ‘no markup’ in the form of Sperberg-McQueen’s ‘stream of characters’:

\[
\text{miltonexpressethisideamostclearlylaterinthetacticannotpraiseafugitiveandcloisteredvirtue (936)}
\]

Coombs and his co-authors present this ‘no markup’ example at the beginning of their continuum, with descriptive markup at the other pole as the most overtly interpretive and therefore the most powerful for humanities computing. My first exercise with the students, however, interrogates the continuum itself by presenting the classroom with a copy of Gerard Manley Hopkins’s sonnet ‘As Kingfishers Catch Fire’ and asking students to subtract everything that could be considered markup, metadata, bibliographic codes, or elements otherwise extraneous to the pure scare-quotable ‘content’ of the poem. Students usually identify the stress marks on the vowels first, then move on to the italics in the final lines of the octet, which gives us an opportunity to consider the procedural/descriptive markup distinction described above. The dismantling of the poem carries on with the removal of the assumed title and author’s name (as editorial metadata), indentations to indicate line groupings, the space between octet and sestet, capitalization, punctuation, and finally spaces between words.

---

6 Traditional notation as a musicological tool has been criticized for its reductiveness and formalist bias; for critiques relevant to Shakespearean page/stage/code debates, see Randel 1992: 11–13 and Citron 1993: 37–9.
(often suggested by a student mindful of Sperberg-McQueen’s point about *scriptio continua*). The resulting text looks as disconcerting as the unspaced Milton example above, but the process of getting there helps us all to notice the textual equivalents of Liebman’s musical intangibles.

One of the virtues of using literary texts to teach digital encoding is that the interpretive richness of certain texts, coupled with the material strangeness of their early forms, can combine to subvert the notion that all texts are simply content to be copied and pasted from one form into another. One could perform the same exercise with just about any literary text, including Shakespeare’s, but this Hopkins sonnet offers two particular advantages. One is that it also appears in an excellent companion reading for this exercise, Jerome McGann’s ‘Visible and Invisible Books in N-Dimensional Space’, in which he uses this sonnet to illustrate the idea that ‘because works of imagination are built as complex nets of repetition and variation, they are rich in what informational models of textuality label “noise”,’ which challenges all text encoders to question ‘the informational content and expository structure’ of texts that digital tools tend to assume (2001a, 175). Assigning this companion piece also gives us the opportunity to perform some textual criticism on one of our own readings, thanks to the anomaly of McGann’s article having been published in at least five places in print and online, but with no two versions reprinting the Hopkins sonnet exactly the same in terms of layout and accidentals (2001a: 176; 2001b: 290; 2002: 68; 2004: 150–1). A related advantage of using Hopkins is that students can view Hopkins’s manuscript thanks to Norman MacKenzie’s facsimile edition, which shows the elaborate diacritical marks that Hopkins (anticipating Zappa) used to indicate his sprung rhythms (1991: 106–7).

The second exercise – likely a familiar one to many Shakespeareans – asks students to restore the punctuation and capitalization to a sonnet, using their own judgment to make decisions that best suit the text. The exercise is similar to those of Shakespearean actors and dramaturges, for whom de-punctuating verse can help actors to find their breathing patterns (Donnellan 2002: 255–8). In my classrooms, we use this exercise in a similar manner to explore the TEI Guidelines’ premise that ‘encoding a text for computer processing is, in principle, like transcribing a manuscript from *scriptio continua*; it is a process of making explicit what is conjectural or implicit, a process of directing the user as to how the content of the text should be (or has been) interpreted’ (TEI, part v). The perilous gap between ‘should be’ and ‘has been’ is where encoders, editors, and readers alike do their best work. To illustrate this idea I present students with paper copies of a stripped-down original-spelling transcription of Shakespeare’s Sonnet 129:

```
  th expence of spirit in a waste of shame
  2 is lust in action and till action lust
     is periurd murdrous blouddy full of blame
  4 sauage extreame rude cruell not to trust
      injoyd no sooner but dispised straight
  6 past rea  son hunted and no sooner had
```

---

7 See also the undated HTML version on McGann’s personal website: [www2.iath.virginia.edu/jjm2f/old/nlh2000web.html](http://www2.iath.virginia.edu/jjm2f/old/nlh2000web.html).
Students work in groups for ten to fifteen minutes to punctuate and capitalize the text, during which time they weigh the interpretive consequences of each other’s choices. Stephen Booth’s helpful extended commentary in his edition/facsimile of Shakespeare’s Sonnets (1977), along with Randall McLeod’s provocative unediting of the poem (247–53) (1991: 247–53), both elucidate some interesting challenges for encoders, including the unmetrical trainwreck of adjectives that pile up in lines 3–4, the orthographical and typographical dilemma of ‘mad’ and ‘made’ in lines 8–9 (the same word spelled differently? different words yet homonyms to early modern ears?), and the McLeodian visual homonym (typonym?) in line 11 between the quarto’s ‘proud’ and the word we would pronounce as proved, which modern editors often emend to prov’d. Showing the students the poem as it appeared in the 1609 quarto helps to settle some questions, raise others, and establish that our real objective isn’t the right answer but the right frame of mind.

All encoding involves choice, but the proud/prov’d case in particular challenges encoders with a provocative dilemma when they consider McLeod’s point that the particular sequence of typographic marks in the 1609 quarto represents two words at once in early modern typography, which Shakespeare, his compositors, and his readers would have known. Digital humanists could learn a thing or two from early modern readers’ willingness to accommodate ambiguity and hybridity.

***

Presented with this chapter’s examples of typographic homonyms, overlapping stage directions, half-italic colons, Zappa-esque rhythms, and other encoding cruxes, a sober, professional, results-oriented encoder might call such detailism a distraction from the needs of a digitization project – an indulgence in special cases at the neglect of the world’s overwhelming normality. After all, Schneider’s businesslike mantra for the London Stage Information Bank, impressed upon him by a well-meaning colleague, was ‘get results’ (1974: 17). However, moments of thinking through making afford encoders an experiential lesson in how to avoid the ossifying mistake, as McLeod describes it, of providing ‘an answer to the question, while hiding the fact that there is a question: at which point an answer becomes the answer becomes the Truth’ (1991: 253; emphasis in original). For textual scholars – including students, who are scholars in the most fundamental sense – the understandings that arise in these moments are results.

For another worthwhile companion reading for this exercise, see Sherman 2008.
The mistake McLeod points out happens all too often in digital projects today, especially those that simply ingest literary texts as normalized data. The problem is worsened by many digital humanists’ narrow conception of their field as fundamentally large project-based, which, wittingly or not, plays into the neoliberalization of the academy on a business model. I suggest that a well-conceived digital project should afford, even demand, reflection upon materials and tools alike, such that one emerges with new insights into both, and which would not be possible without the hands-on activity of making. The size, shape, and computing resources of the London Stage Information Bank forced Schneider and his team into a relentlessly progress-driven program of data entry, with many false starts and Herculean efforts required to render the text of The London Stage machine-readable and computationally tractable. Those pressures are amplified today, but it is always worth asserting the choice to approach encoding as an intellectual exercise that rewards the interdisciplinarity that happens not just between individuals, but within them.

Those two kinds of interdisciplinarity have been named, respectively, ‘normal’ and ‘deviant’ by the sociologist Steve Fuller. As he puts it, ‘Normal interdisciplinarity is designed for teamwork, as each disciplinary expertise is presumed to make a well-defined contribution to the final project, whereas deviant interdisciplinarity assumes that the differences in disciplinary expertise themselves pose an obstacle to the completion of the project’ (2013: 1901). Deviant interdisciplinarity, which I am championing along with Fuller, therefore runs counter to capitalist divisions of labour and specialization, and hews closer to what I would identify as traditional humanistic values – even specifically Shakespearean values – of eclecticism, curiosity, and lifelong education. We are all encoding, editing, and reading subjects, whatever role we might play in a project, and to be a humanist is to appreciate the potency of literary works in shaping our subjectivity, for better and for worse. Deviant interdisciplinarity calls us to listen to our texts, even if it takes our attention away from the project manager at the head of the conference table. The challenge facing the digital humanities, and digital Shakespeare studies in particular, is to evolve new forms of scholarship – and models of education – that do not force this unnecessary dilemma upon us.

McLeod’s point about the disambiguation of proud/prov’d also shows the value of slowing down and of welcoming high-friction texts into digital scholarship: as he says, ‘the older forms of texts remain questionable for us: at moments like this they posit knowledge in a ratio with uncertainty’ (1991: 253). Preserving that ratio allows encoding to progress reciprocally with reading, leading to the realization, as McGann puts it, that ‘there is no such thing as an unmarked text, and [that] the markup system laid upon documents to facilitate computerized analyses are marking orders laid upon already marked up material. (Thus all texts implicitly record a cultural history of their artifactuality)’ (2001a: 138). In other words, every act of transcription requires us to read the prior acts of encoding (typographical, scribal, digital, graphical, or otherwise) that persist in documents, and to play the role of editor in a controlled collision between information systems that may be centuries apart in origin. To these

---

9 I critique the misapplication of the term data to Shakespeare texts in the final chapter of The Shakespearean Archive (see Galey 2014: 248–57).
ends, I usually conclude my sequence of exercises by taking a page from Peter Stallybrass and unpacking the myriad forms of markup in the opening page of Genesis from the 1611 King James Bible (2011: 91–4; cf. Norton 2005: 47–51). Its typographic encodings of interpolations, glosses, and cross-references – as well as the political agendas of the book’s makers – remind us that for early modern readers, decoding complex markup was not only a matter of reading competence, but also part of the textual machinery of salvation. New media and digital tools have apparently become the latest machinery of salvation for the humanities in the early twenty-first century, and their associated truisms deserve critical scrutiny all the more. To see the artifactuality of texts is to apprehend the same quality in our tools: we use artifacts to understand other artifacts. There is no standing outside that circle, but it makes for an expansive horizon.

I am grateful to Alan Stanbridge for his advice on musical transcription, to Ashley Bodiguel for her advice on acting exercises, and to Rebecca Niles for the many conversations about Shakespeare encoding that inspired this chapter. I am also grateful to the participants, organizers, and funders of the NEH-supported ‘Early Modern Digital Agendas’ summer institute at the Folger Shakespeare Library in July 2013, and especially to Julia Flanders and Heather Wolfe for the opportunity to teach text encoding and transcription with them. Any errors are my own.

Works cited


