1 Theory: Text

The impact of digital media on early modern studies in general, and editorial scholarship in particular, has been long and profound. From scanned facsimiles to editions of single texts or authors to large-scale, transcribed corpora like the Early English Books Online – Text Creation Partnership (EEBO-TCP), the ease of access and flexibility offered by digital texts has transformed everyday research and pedagogical practices. Especially in the field of scholarly editing, the anticipation and acceptance of digital texts has a long prehistory.[1] Textual studies embraced and theorized the possibilities of digital texts quite early and shaped significant aspects of technological development that paved the way for the creation of massive archives of digital media. The development of TEI (Text Encoding Initiative), first as an SGML (Standard Generalized Markup Language) and then as an XML (Extensible Markup Language) protocol, was at the forefront of this innovation and influenced technologies that would be at the core of the World Wide Web and the explosion of the internet.[2]

Thanks to multidecade, multi-institution digitization projects such as the EEBO-TCP, early modern scholarship finds itself in the unique position that the vast majority of the texts it covers as a field are now available in

[1] For an early example of corpus scale scholarship, often described as the first digital humanities project, see R. Busa, "The Annals of Humanities Computing: The Index Thomisticus," *Computers and the Humanities* 14, no. 2 (1980): 83–90. For early articulations of the possibilities offered by digital texts, see Peter L. Shillingsburg, *Scholarly Editing in the Computer Age: Theory and Practice* (Athens: University of Georgia Press, 1986), and Jerome J. McGann, *The Textual Condition* (Princeton: Princeton University Press, 1991). Sukanta Chaudhuri, *The Metaphysics of Text* (Cambridge: Cambridge University Press, 2010), offers a nuanced overview how digital texts have broadened the theoretical horizons of scholarly editing.

[2] Nancy Ide and C. M. Sperberg-Mcqueen, "The Text Encoding Initiative: Its History, Goals, and Future Development," *Computers and the Humanities* 29, no. 1 (1995): 5–15.

a searchable digital format.[3] Moreover, Shakespeare's texts have held a privileged position even within the broader field of early modern studies and have been among the earliest ones available in well-curated digital versions.[4] The availability of texts, combined with the broader accessibility of Shakespeare, has spurred computational work and served as testing grounds for many digital humanities methodologies. Perhaps the most publicly visible intervention of this steady stream of digital and quantitative scholarship has been the use of stylometric analysis in the *New Oxford Shakespeare* edition to make radical claims about the authorship of several plays.[5] The initial hum of excitement about digital texts, therefore, has grown into a veritable roar over the last two decades. No aspect of early modern scholarly work – from editing and research to pedagogy and performance, from text encoding and bibliographic research to stylometrics and cultural analytics – remains untouched by digital technology.

But a lingering anxiety continues to mark this encounter with technology. Neither the pervasive presence of digital texts in everyday scholarly practice nor the depth and vitality of scholarship spurred by digital technology can alleviate the note of uncertainty, perhaps even apprehension, that remains in many assessments of its impact. On the one hand, many digital techniques and strategies are deeply familiar, and their use in scholarly practice – especially in the field of scholarly editing – is so ingrained that they present well-traversed territory. And yet, the flexibility and scalability of digital texts that make them

[3] "Early English Books Online – Text Creation Partnership," http://quod.lib.umich.edu/e/eebogroup/ (accessed April 21, 2023). I will return to the question of EEBO-TCP's coverage later, but it is worth noting that while it is certainly comprehensive enough to accommodate most research and teaching as well as statistical analysis, it is not quite a randomly sampled dataset and, in addition to certain editorial preferences, echoes the exigencies of book history and survival rates.

[4] For a brief overview of early digital editions, see Toby Malone and Brett Greatley-Hirsch, "Digital Shakespeare," in Paula Rabinowitz, ed., *Oxford Research Encyclopedia of Literature*, 2021, https://doi.org/10.1093/acrefore/9780190201098.013.1192.

[5] Gary Taylor and Gabriel Egan, eds., *The New Oxford Shakespeare: Authorship Companion*, 1st ed. (Oxford: Oxford University Press, 2017).

such a valuable resource also raise some concerns. Suzanne Gossett, in her discussion of "Textual Studies After the Digital Turn" in *Shakespeare and Textual Theory*, gestures at this sense of unease in terms of the fast-changing and disorienting nature of the technological terrain. She notes that even though her coda on the "immaterial text" and its role in Shakespeare scholarship was written last, it will likely be the first to be outdated. The move to digital media, she suggests, "has caused considerable disruption and has radically altered the communication circuit," even as she acknowledges the great impact of technology – "both theoretical and practical" – on scholarly editing. She points to "concerns about the loss of intellectual accuracy," and notes that new tools and techniques pose significant challenges for "traditional Shakespeare scholars."[6] Interestingly, Gossett admits that textual scholarship has always required scholars to master somewhat esoteric skillsets, such as "creating collation formulas, tracing the reuse of skeleton formes, or operating a Hinman collator."[7] What, then, one might ask, is so unique about the challenges presented by digital technologies? Why do digital texts, which have been around, and well theorized, for decades now, continue to evoke such a wide range of reservations?

In this Element, I intend to interrogate the sense of the uncanny – a deeply unsettling strangeness within familiar terrain – that haunts this anxious relationship with the digital. By rethinking our underlying assumptions about digital texts and computation, I suggest, we can open up new ways of exploring individual texts and their place within the broader corpus – ways that transcend mere automation and can accommodate fundamentally humanistic modes of thinking. In this section I argue that while contemporary editorial practice has wholeheartedly embraced the flexibility offered by digital texts, it has not come to terms with such texts as truly computational objects – objects that are not mere electronic proxies of material texts but uniquely flexible computational artifacts in their own right. Editorial theory has paid attention to the ways in which the digital

[6] Suzanne Gossett, *Shakespeare and Textual Theory*, The Arden Shakespeare (London: Bloomsbury Publishing, 2022), 215.
[7] Gossett, *Shakespeare and Textual Theory*, 217.

format makes it possible to capture the complexities and inherent instability of the processes of textual transmission. However, by treating texts as computational objects, I want to emphasize their affordances in a fundamentally new medium. We often treat such affordances as happy byproducts of the electronic format, something the digital medium innately lends itself to: search, retrieval, enumeration. In other words, these problems are transferred to the domain of technical implementation rather than humanistic conceptualization. The next section, expanding our scope from "text" to "corpus," foregrounds computation as not only a set of strictly procedural goal-oriented processes, but as a heuristic that enables transmutability, intertextuality, and scalability, and that explores the ways in which conceptualizations of scale overlap with humanistic modes of enquiry.

The final two sections – "Search" and "Discovery" – explore related but, I shall argue, fundamentally distinct modes of information retrieval that we associate with computation. While "search" seems to be a deeply familiar paradigm in a world where we are inundated with information, my purpose is to render it somewhat strange. By dissecting examples of certain kinds of search that scholars encounter regularly – either for catalogs or for, say, the ProQuest EEBO website – I highlight the complex and often messy historical trajectories and intellectual assumptions that mediate what might at first glance appear to be a thoroughly dry, technical process. Having problematized the concept of search, or at least rendered it less stable than a mere technological black box that simply retrieves bits of information, I extend my study of its possibilities under the rubric of "discovery" to accommodate a set of approaches that are more open ended, flexible, and, often, serendipitous. These approaches, I argue, have the potential not only to align with but also to extend humanistic modes of inquiry and transform the kinds of questions we can ask of the early modern corpus in the first place. To be sure, such approaches can often be highly technical, involving statistical modeling, data mining, and machine learning. Nevertheless, I shall argue that breaking away from static notions of text and corpus and seeing computation not only as a mere technological handmaiden but also as a distinctive mode of knowledge reveals affinities with the kinds of subtlety, ambiguity, and intertextuality that humanists value. It is a phenomenon that requires distinctive kinds of scholarly

attention, and it has the potential not only to align with but also to extend humanistic modes of inquiry and transform the kinds of questions we can ask of the early modern corpus in the first place. Throughout this Element, I will distinguish and disturb terms that are usually collapsed together or implicitly treated as slightly different ends of the same spectrum: digital/computational; text/corpus; editing/curating. My intention in troubling these binaries is not to suggest some kind of qualitative/quantitative divide or fundamental incommensurability between the humanistic scrutiny that individual texts invoke and the technological apparatus required to make large numbers of texts tractable to computation. In fact, I hope to show that these terms denote not so many different objects or activities but distinct and complementary perspectives, each with its unique scholarly purchase. Only by reconciling these perspectives – by seeing what is distinctive about scale and computation as modes of humanistic (rather than technical) knowledge – can we begin to undo the strange unfamiliarity at the heart of our encounter with digital textuality.

Digital Text

From its very outset, the appeal of digital texts has been their highly procedural and hierarchical nature. Computers are good at implementing well-defined repeatable procedures and were deemed ideal for taking over what Peter Shillingsburg called the "idiot work" of scholarly editing: "tedious jobs . . . most liable to careless error," such as "collation, typesetting, and proofreading."[8] But the excitement of handing over such tasks was tempered by concerns about computers ultimately overstepping these mechanical bounds and somehow infringing on the more critical aspects of scholarly editing. Poststructuralist re-evaluations of textual theory and the critiques of the New Bibliography it has produced have helped to renegotiate this hierarchy between the procedural and the critical. The editor is no longer – or as explicitly – tasked with "critical analysis" or with teasing out some unique insight about authorial intention. Freed from this burden, scholarly editing has become more collaborative, enlisting the reader as a participant in the critical process of negotiating the problems of textual transmission rather than aspiring toward an ideal, fixed text. This paradigmatic shift has been

[8] Shillingsburg, *Scholarly Editing in the Computer Age*, 135.

facilitated in part by the malleability and fluidity of the digital medium, which makes it possible for the text to be radically multiple.[9] As Kathryn Sutherland puts it, the "digital vantage point" allows the shifting of interpretive agency that was silently assumed by the New Bibliographers.[10] Gossett has suggested that the flexibility of digital editions facilitates the ideals of the "postmodern edition" and makes it possible to represent "the challenge of poststructuralism to any concept of textual stability."[11]

These are lofty aspirations for what an electronic edition should be, though often tempered by skepticism about the underlying technology's ability to accommodate critical nuance. This tension between liberatory embrace and apprehension has been the driving dialectic of modern textual scholarship's encounter with digital technology. Digital editions require relatively complex technical, financial, and institutional infrastructures to create and maintain. But what makes the digital medium so appealing to scholarly editors, and also what provides a conceptual vantage point from which to contemplate textuality itself, is a core set of technologies that are in themselves elegantly simple: XML and hypertext. Both are, in fact, information organization protocols that make use of more generalizable underlying text or data-processing technologies. An XML or hypertext file or data-stream is no different from any other stream of text information that computer processing, storage, and transmission hardware can handle. Of course, the "simplicity" I attribute to this innovation is deceptive. We need only remind ourselves of the explosion that the addition of hyperlinks to previously text-based networks caused in the form of the World Wide Web to realize that immensely complex systems can be built out of strikingly simple core innovations.

[9] For examples of digital projects that foreground the polyvalent nature of texts, see "The James Merrill Digital Archive: Materials for The Book of Ephraim," accessed May 13, 2024, http://omeka.wustl.edu/omeka/exhibits/show/jamesmerrillarchive/; "Bichitra: Online Tagore Variorum," accessed May 13, 2024, https://bichitra.jdvu.ac.in/index.php.

[10] Kathryn Sutherland, "Being Critical: Paper Based Editing and the Digital Environment," in *Text Editing, Print and the Digital World*, ed. Marilyn Deegan and Kathryn Sutherland (Farnham: Ashgate, 2009), 16.

[11] Gossett, *Shakespeare and Textual Theory*, 218.

The appeal of both these protocols lies in the ways they can break both the linearity and the transparency of the textual encounter: XML introduces organizing structures to texts as markup, paratextual information in the form of metadata, and the ability to encode multiple versions or states of text. Hypertext, on the other hand, punctuates the linearity of text flow without quite dismantling it. It puts at the reader's disposal the text's potential axes of connection and cross-pollination to its various outsides, and signals that every text exists within a larger matrix of material and cultural conditions. It is not difficult to see, even from such a schematic outline, why such an information architecture would seem liberating to editorial scholars who have always worked within the limitations of the printed codex. Even though print as a technology has, over its long history, developed a formidable array of apparatuses that reconfigure linear reading – notes, marginalia, indices, tables, concordances – the representation of multiple states and nonlinear organization still seem like convoluted accommodations rather than primary affordances of print.

The information infrastructure available in the digital space provides editors with the building blocks for moving from a notion of the text as some abstracted version of an originary stable object to an account of textuality as process. Electronic editions, in other words, are models of textual phenomena rather than representations of particular instantiations. One might suspect that, now that the editor is no longer burdened with divining authorial intention and can recruit the reader to navigate the labyrinth of textual states and variations, they might assume the more limited role of collator of evidence. It would be fair to say that technological innovation – the emergence of the editor as model-builder – has inspired (and, in turn, been spurred by) increasingly sophisticated theorizations of textuality, each challenging us to broaden our frameworks of textual analysis and the technical apparatus' capacity to accommodate ever more detail, ambiguity, and play.[12] If the editor's task is to capture what scholars

[12] See, for example, Elena Pierazzo, *Digital Scholarly Editing: Theories, Models and Methods*, 1st ed. (London: Routledge, 2015); and Katherine Bode, "The Difference an Editor Makes," *Modern Language Quarterly* 82, no. 3 (2021): 401–4.

like McKenzie and McGann have popularized as the processes of a text's "socialization," then they must attend not only to the text as a linguistic construct but also as an artifact mediated by multiple actors, processes, and materialities.[13] Shillingsburg's insistence that we further widen our conception of textual transmission to what he calls "script-acts" – which would include not only the customary sites of a text's socialization (editor, printer, publisher, bookseller, and so forth), but every interaction, including and especially readerly ones, that relate to texts – broadens this scope even further. A physical rendition of a text within such a framework is not only partial and necessarily provisional; it does not stand in any particularly privileged position. The emphasis rather lies on processes of textual encounter – creation, transmission, reception – which form an endless and ever-incomplete chain.

The road from textual authority to open-ended play, therefore, has been a fraught one that, in many ways, takes the ambit of play and polyvocality to its postmodern limits. Far from being relegated to the easily mechanizable "idiot work" of scholarly editing, an increasingly sophisticated machinery of textual representation seems to be evolving to meet the demands of even more astonishingly ambitious and nuanced conceptualizations of textuality. The expansiveness of the text-as-model paradigm puts ever-increasing pressure on the digital editions' capacity for representing ambiguity in ways that have made many editors anxious about losing sight of the original, somewhat pragmatic goals of scholarly editing. On the one hand, many have embraced the erosion of authority and argued that, once such absolute privileged insight is disavowed, the main utility of the digital text becomes its ability to demonstrate the unfixity of text and thus the contradictions inherent in its processes of production. Katherine Rowe termed such editions "good enough" texts, wherein textual instability offers an opportunity to renew the reader's engagement and intervention in the editorial process: "Yet new digital editions also invite us to return to editorial first principles, replacing single textual authorities with ambiguous

[13] D. F. McKenzie, *Bibliography and the Sociology of Texts* (Cambridge: Cambridge University Press, 1999); and Jerome J. McGann, "The Socialization of Texts," in *The Textual Condition* (Princeton: Princeton University Press, 2020), 69–87.

alternatives, and including the reader in the editorial process."[14] The moment of editorial intervention to fix a text in an apparently stable form no longer precedes, nor is separated from, the readerly encounter. Instead, reading subsumes within its hermeneutic instabilities the processes by which a text comes to be.

Despite being theoretically challenging and pedagogically productive, textual instability – and the radical celebration of it – also raises concerns. Thus, it is unsurprising to find that while Gossett shares some of Rowe's excitement about the opportunities provided by good-enough texts, she reminds us that textual editing remains a deeply pragmatic discipline, and that any theorization of text, however profoundly influenced by poststructuralist notions of epistemic fluidity and linguistic instability, must, in the last instance, be grounded in praxis. "Textual theory is different," she argues: "it focuses most often on developing an informed inference about the nature and history of a surviving text." It must be articulable in terms of formalized procedures of selection, elimination, and organization of various textual states. Gossett notes that while editors are aware that "philosophical premises are often implicit in textual work . . . much textual theory is primarily concerned with methodology and procedure."[15] Digital editions give textual scholars the opportunity to "open up" the text, to invite the reader in as participant and co-creator in assessing the complex material and intellectual histories of transmission. But if the task of the editor is to be distinguished from that of the free-ranging literary theorist, the procedural foundations of textual editing need to be emphasized. One might say that the polarities of insightfully critical and merely procedural have been dissolved to an extent as the evolution of text technologies has proven that the digital text can rise to the challenges of literary reading.

[14] Katherine Rowe, "Living with Digital Incunables, or a 'Good-Enough' Shakespeare Text," in *Shakespeare and the Digital World: Redefining Scholarship and Practice*, ed. Christie Carson and Peter Kirwan (Cambridge: Cambridge University Press, 2014), 148.

[15] Gossett, *Shakespeare and Textual Theory*, 2–3.

Computational Text

The text as model: an engine of meaning-making that carries within it the archaeology of its own material and social evolution. The metaphor is not quite a novel one. We are accustomed to thinking of the book as an evolving set of technologies that both facilitate and set limits on the production and circulation of language. Jerome McGann describes the book as "a machine of knowledge" and compares its capabilities with those of digital editions that he suggests can transcend "the formal limits of all hardcopy's informational and critical powers."[16] The electronic edition, McGann argues, opens up ways of interrupting hermeneutic procedures, urging us to encounter textuality as a generative process, as an unfolding performance of formal innovation: "electronic tools in literary studies don't simply provide a new point of view on the materials, they lift one's general level of attention to a higher order."[17]

The central point of thinking of the book as a machine is that, while it has metaphorical overtones (a book is *like* a machine in the way it amplifies and transforms ideas), it is also a literal description that draws attention to the materiality of the book. A book, as so much contemporary scholarship reminds us, is, first and foremost, a physical object, a technology of information. From the material intersection of paper, ink, thread, and glue to its conceptual innovations such as the random access the codex form facilitates, its easy reproducibility, and archival stability – the book doesn't merely supply a new template – a convenient technical upgrade from scroll and manuscript – it redefines our understanding of text. The technology of the book has imprinted itself invisibly on our notion of what a text can be.

It might seem that the text-as-machine comparison would be more obvious in the case of electronic texts, and we would not need to belabor it. After all, our basic access to such texts depends on our ability to negotiate considerably more complex piles of metal and plastic, not to mention the bewildering array of information transactions that Matthew G. Kirschenbaum

[16] Jerome McGann, *Radiant Textuality* (New York: Palgrave, 2004), 1153, 1201, Kindle edition.
[17] McGann, *Radiant Textuality*, 1173.