

Digital Scholarly Editing

Susan Schreibman

Over the past twenty years there has been an evolving body of scholarship exploring the standards, theories, and methodologies of digital scholarly editing. Scholarship from the early-to-mid-1990s maintained a bifurcated focus. On the one hand, many textual scholars found themselves in the slightly unusual position of writing primers, guidelines, and documentation laying the groundwork for basic digital tasks, such as Peter Robinson's *The Digitization of Primary Textual Sources*, published in 1993, and *The Transcription of Primary Textual Sources Using SGML*, which followed a year later, or the largely uncredited scholarship created by the many hands that went into the guidelines of the Text Encoding Initiative (TEI), first officially published in 1994. Editors and staff members from many scholarly editing projects spent a great deal of time in documenting encoding practices. This documentation proved to be an essential resource to ensure consistent encoding and to help those entering the field understand how a standard like the TEI was put into practice.¹

On the other hand, as the decade progressed there was an ever-growing body of research exploring what this new medium could bring to the study of the transmission of texts, and thus editorial practice normative to this environment began to develop. Norman Blake and Peter Robinson's *The Canterbury Tales Project: Occasional Papers I*, published in 1993, was one of the earliest, but by the time *The Canterbury Tales Project: Occasional Papers II* was published in 1997 other collections were appearing, such as Richard J. Finneran's *The Literary Text in the Digital Age*. Published in 1996, it was one of the first collections of essays devoted to electronic scholarly editing. Many of its chapters are still required reading, such as C. M. Sperberg-McQueen's "Textual Criticism and the Text Encoding Initiative," John Lavagnino's "Completeness and Adequacy in Text Encoding," and Robinson's "Is There a Text in These Variants?"

The same year, a conference was held in Ann Arbor exploring the semantics of the page in print, in manuscript, and on screen. The results of that meeting appeared in the 1998 publication edited by George Bornstein and Theresa Lynn Tinkle entitled *The Iconic Page in Manuscript, Print, and Digital Culture*. In that volume, Kevin Kiernan explores the use of new technologies to restore what is invisible to the eye in medieval manuscripts (18), and Martha Nell Smith argues that new developments in technology have the potential to create an entirely new reading environment for a poet like Emily Dickinson, many aspects of whose work are still contested. An electronic edition, Smith argued, could be designed so that readers could explore multiple orders of the poems and choose between contrasting representations,

as well as provide a library of secondary sources within the same reading space (“Corporealizations” 215).

In 1997 Kathryn Sutherland’s *Electronic Text* appeared. Like Finneran’s collection, it explored the methods and theories of digital textuality. In Sutherland, Jerome McGann’s “The Rationale of Hypertext” first appeared, as well as Allen Renear’s “Out of Praxis: Three (Meta)Theories of Textuality,” in which the OHCO (ordered hierarchy of content objects) theory was explored alongside other theories of textual representation.

All the authors cited above were also practitioners who created scholarly electronic editions. These first-generation digital editions did not appear, at least on the surface, as scholarly or critical in the ways that the textual editing community had come to understand the terms for print publication. Frequently these early editions did not include the apparatus, critical notes, and contextual material that, by the late 1990s, had come to signal the apotheosis of scholarly editing. Moreover, claims about technology’s potential, such as those made by Smith, while common in theory, were difficult and thus rarely implemented in practice (Schreibman 83).

Until the end of the 1990s, there was only one software program, *DynaWeb*, that allowed texts encoded in standard generalized markup language (SGML) to be displayed on the World Wide Web. SGML is the language from which the now more ubiquitous extensible markup language (XML) and the language of the Internet, hypertext markup language (HTML), were derived. *DynaWeb* was a commercial tool developed for early adopters of SGML (including the pharmaceutical and defense industries) to search, display, and organize large textual corpora. When the software was owned by INSO Corporation and Electronic Book Technologies, it was possible to apply for a free version of the software, and several—including Oxford University; University of California, Berkeley; Duke University; and University of Virginia—took advantage of this opportunity. It thus became possible for academics to present their digital scholarship within a database environment that allowed for dynamic searches and transformations to HTML for delivery.

With the development of XML and associated standards, it became easier to realize some of the earlier theoretical goals claimed for electronic editions. By 2006, when the Modern Language Association published *Electronic Textual Editing*, edited by Lou Bernard, Katherine O’Brien O’Keeffe, and John Unsworth, a confidence reflecting over a decade of practice was evident. John Bryant’s somewhat earlier *The Fluid Text* (2002) and Peter Shillingsburg’s *From Gutenberg to Google* (2006), for example, explore the tensions between traditional editorial practice and this new medium. Moreover, the Blackwell *Companion to Digital Humanities* (2004; Schreibman, Siemens, and Unsworth) and *Companion to Digital Literary Studies* (2008; Siemens and Schreibman) included essays (Smith, “Electronic Scholarly Editing”; Price) that were dedicated to electronic scholarly editing and examined aspects of scholarly editing practice, reflecting its centrality to digital scholarship. *Theories of Digital Textual Scholarship: New Norms, New Paradigms, New Modes of Analysis*

In his monograph on the history, nature, and practice of textual studies, *Bibliography and the Sociology of Texts*, D. F. McKenzie grappled with the relevance of traditional bibliographic practice in the light of the wider field of literary studies. If bibliography, he argued, was simply the practice of describing or enumerating texts, it would have little relevance to contemporary theoretical concerns:

The problem is, I think, that the moment we are required to explain signs in a book, as distinct from describing or copying them, they assume a symbolic status. If a medium in any sense effects a message, then bibliography cannot exclude from its own proper concerns the relation between form, function, and symbolic meaning. If textual bibliography were merely iconic, it could produce only facsimiles of different versions. (10)

McKenzie defines text as “verbal, visual, oral, and numeric data, in the form of maps, prints, and music, of archives of recorded sound, of films, videos, and any computer-stored information, everything in fact from epigraphy to the latest forms of discography” (13). He goes on to suggest a new definition of bibliography as a practice that not only recognizes but also seeks to describe how the form of the text affects meaning. Moreover, it is not simply the material embodiment of the text that informs meaning but also the social process of transmission, “[its] physical forms, textual versions, technical transmission, institutional control, . . . perceived meanings, and social effects” (13).

First given as a series of lectures at Oxford University in 1985, some seven years before the advent of the World Wide Web, McKenzie’s writing seems strikingly prophetic. Here, McKenzie meticulously develops a vocabulary to describe a not-yet-invented medium:

In terms of the range of demands now made of it and of the diverse interests of those who think of themselves as bibliographers, it seems to me that it would now be more useful to describe bibliography as the study of the sociology of texts. If the principle which makes it distinct is its concern with texts in some physical form and their transmission, then I can think of no other phrase which so aptly describes its range. (13)

In many ways, the great enterprise of the last decade and a half to digitize our cultural heritage has been an exploration of the sociology of texts. It has fostered new theories of editing and new modes of editorial practice. Some of these theories share much with their print counterparts; others exist only in the digital realm. Textual scholars were among the first to explore the visual, computational, and navigational possibilities offered by this new medium. By creating a docuverse that flattened all cultural artifacts to binary code, the materiality of the original object was brought into sharp relief. Deciding what was essential to re-present provided the distance that textual scholars needed to reconceive the field’s theories, methodologies, and practices. As McGann, one of the field’s most eloquent critics, has written, it was only when textual scholars had the opportunity of editing in a medium other than the book that they were able to realize the constraints that the medium imposed on them:

This symmetry between the tool and its subject forces the scholar to invest analytic mechanisms that must be displayed and engaged at the primary reading level—for example, apparatus structures, descriptive bibliographies, calculi of variants, shorthand reference forms, and so forth. The critical edition's apparatus, for example, exists only because no single book or manageable set of books can incorporate for analysis all of the relevant documents. (*Radiant Textuality* 56)

There are, of course, many shared goals between print editions and first-generation digital editions: above all, the creation of new works by means of a re-presentation of the works of the past. Issues of authority, textuality, and representation are of concern in both mediums: editors must decide to what level of fidelity the linguistic codes (the linguistic elements of the text and paratext) are maintained and whether—and if so, to what degree—the bibliographic codes (the material aspects of the text: the typography, advertisements, illustrations, decorations, etc.) are captured. What editors of first-generation Web-based editions discovered was that a plethora of new intangibles also preoccupied them, such as whether HTML was expressive enough to create a digital scholarly edition and, if it was not, what other markup scheme might be appropriate; how best to encode structural divisions of texts, such as paragraphs, titles, footnotes, and lines of verse; and how closely a digital surrogate of a print publication should, or indeed could, capture and make clear to users essential qualities of the material object.

Many first-generation digital editions explored whether it was more apposite to realize a project's goals through a unique encoding scheme rather than to use a standard like DocBook or the Text Encoding Initiative Guidelines. Moreover, issues that had heretofore been the preserve of publishers and typesetters preoccupied many literary scholars, such as how to represent characters for nonstandard text (e.g., accented characters or special symbols) or how to deal with edition-specific typographic features (e.g., running headers, font, and pagination). Moreover, this new medium allowed an exploration of textuality beyond the printed word, creating editions of other cultural objects, such as images (still and moving) and audio. The digital environment leveled the playing field for multimedia and text, creating a holistic environment within which to seamlessly navigate between primary objects and the contextual, between the visual and the aural, engendering a reevaluation of the social and material ontologies of the text (Loizeaux and Fraistat 5).

Much of this exploration has gone on within a new genre of scholarly production, the thematic research collection (TRC). One might argue that TRCs became the framework within which editors conceived, explored, and realized these new editions. In 2000 Unsworth set out essential characteristics of TRCs. Above all, they are electronic, contain heterogeneous data types, are extensive but thematically coherent, and are structured but open-ended. They are designed to support research, are written by at least one if not many authors, are interdisciplinary, and are collections of digital primary resources. Carol Palmer further refined the definition, distinguishing collections created by libraries or other cultural heritage organizations from

those created by scholars:

In taking a thematic approach to aggregating digital research materials, they are producing circumscribed collections, customized for intensive study and analysis in a specific research area. In many cases these digital resources serve as a place, much like a virtual laboratory, where specialized source material, tools, and expertise come together to aid in the process of scholarly work and the production of new knowledge. (348–49)

TRCs take many forms. Many early digital editions explored the notion of unediting—that is, reproducing the text in documentary form, typically in the form of facsimiles. In print this was a fairly expensive undertaking reserved for the most canonical of authors. It worked particularly well where there was one copy of a manuscript, such as T. S. Eliot's *The Waste Land*, in which Eliot's typescript was edited by Ezra Pound, Valerie Eliot, as well as Eliot himself. The more recent facsimile editions of Windham Lewis's *Blast* capture the anger and arrogance of the original typography as no other representation could.

Unediting is, in many ways, fairly trivial in digital form. Digital images are relatively inexpensive to create and store. Many projects choose to take this route rather than to transcribe and encode text. But current technology can also make these editions clunky. It can be difficult to ascertain the level of engagement one needs to make with a text published as a series of PDFs, a problem not encountered with my print copy of *Blast*. Too many projects use unimaginative strategies for browsing (such as displaying twenty or even one hundred thumbnails to a page) or simply provide a page-turning strategy (à la *Google Books*).

Other facsimile editions, such as *The William Blake Archive*, are works of extreme editing. *The William Blake Archive* was one of the first projects of the Institute for Advanced Studies in the Humanities at the University of Virginia. It explored the technical, editorial, and legal issues surrounding the creation of an image-based digital scholarly edition. The goal was to make publicly available Blake's nineteen illuminated manuscripts. Begun in 1992, it is among the earliest of the TRCs. The archive's adherence to strict digitization standards, to capturing the fidelity of the original artifacts, and to creating a vocabulary that would allow unprecedented access to the complexity of the rich visual vocabulary that Blake employed has set the gold standard for image-based electronic editions.

The more recently published *In Transition: Selected Poems by the Baroness Elsa von Freytag-Loringhoven*, edited by Tanya Clement and available through the University of Maryland Libraries, seamlessly melds the facsimile tradition with rigorously edited, TEI-encoded text that is surrounded by scholarly apparatus. Clement argues that the electronic re-presentation of the twelve poems in different versions charts the text's composition history through a textual performance in an electronic environment. She calls the presentation of networked text (networked through material space, reception, and theme) in a networked environment "textual performance theory."

Other TRCs are based on alternative theoretical approaches. George Landow's *The Victorian Web* was an early electronic edition that used hypertext theory as its philosophical underpinning. Hypertext theory describes "an ideal textuality," in which content items (text, images, audio, etc.) are linked through "multiple paths, chains, or trails in an open-ended, perpetually unfinished textuality" (Landow 3). This notion of textuality was viewed as an embodiment of poststructuralist theory, in which "readers could navigate between inter- or intra-textual lexias and engage in multisequential reading" (Schreibman 78). *The Victorian Web*, unlike the other SGML and XML projects discussed here, is not database-driven but maintained as a series of static HTML pages interlinked through the ubiquitous HTML `<a href>` tag. The "Credits" page of *The Victorian Web* demonstrates just how different the challenges are for digital scholarly editors than for those who edit for print publication. This page, with its litany of standards, protocols, and software, documents the changes in technology, audience expectations, and skills needed to create and maintain the site.

In the production of this new knowledge space, many literary scholars have found themselves not only assembling and editing the content but also building the tools and software that enable this scholarship. For example, *The Versioning Machine*, which I developed with colleagues over many years at several institutions, is a framework for creating electronic scholarly editions of texts that exist in various versions. It was available to Clement and allowed her to develop her theories of the text. While *The Versioning Machine* provides for features typically found in critical editions, such as annotation and introductory material, it also takes advantage of the opportunities afforded by electronic publication to allow for the comparison of diplomatic versions of witnesses and the ability to easily compare an image of the manuscript with a diplomatic version.² Like much software created by the digital humanities community, it is freely available and open-source, so that Clement was able to download it and make changes to it to represent her theories of performativity.

Technologies and Standards

Before the visual capabilities of the World Wide Web, most of the work in the literary studies that used computation was in the service of text analysis. Linguists and literary scholars used software for the creation of concordances and for text retrieval. Some of the issues facing these early adopters have relevance today—methods of alphabetization, the size and range of context units, and the treatment of ambiguous symbols (Hockey 49). Computers traditionally read text as a sequence of characters, and thus when plain (unencoded) ASCII text is processed, many ambiguities can arise. For example, software may not be able to disambiguate the Roman numeral *I* from the personal pronoun *I*. In the 1980s the artificial intelligence community's expert systems and more recently the Semantic Web community's linked data have sought to develop powerful algorithms to allow software to construct meaning through context. But much of this context still relies on structured text. Also known as markup or encoding, this extra intelligence allows computers to more effectively locate and process semantic

textual units. For example, the now fairly ubiquitous <p> sign represents a paragraph. By making explicit that a block of text functions as a paragraph, the computer can be instructed to style it using generic style sheets or to search for a specific term only within that unit as opposed to another such as <title>.

Individuals as well as communities of practice developed encoding schemes and the systems to support them to mark, search, and display features of text that their disciplinary area deemed important. By the mid-1980s, however, the academic community realized this cacophony of signs, symbols, and standards did not serve the individual scholar (who frequently had to invent a system) or the broader scholarly community (since these schemes were typically incompatible). All too frequently work was undocumented, making it impossible for other scholars to build on. In other cases systems were so specific to platforms and standards that when those systems were no longer usable, neither was the scholarship embedded in them. Other scholars chose to work in proprietary formats, thinking their work had better chances of longevity, but too frequently their scholarship was locked in systems that proved to be nonmigratable when the company went out of business or abandoned the software.

Thus in November 1987, a group of scholars from the humanities, information studies, libraries, and computer science came together at Vassar College in Poughkeepsie, New York, to discuss the idea of creating an open, nonproprietary standard that would be created by and sustained by the academic community for which it was developed. At the close of the conference, nine design principles, known as the Poughkeepsie Principles, were articulated. These became the intellectual foundation for TEI Guidelines.

These design principles have served the scholarly text-encoding community well, although in places they show signs of their age. For example, principle 6 established subcommittees to draft guidelines for text documentation, text representation, text interpretation and analysis, and metalanguage definition and description. The implicit understanding was that text was the only media practicable to work with in 1987. However, other principles, such as the first four, are still relevant to the scholarly editing community:

- .
- . The guidelines are intended to provide a standard format for data interchange in humanities research.
- . The guidelines are also intended to suggest principles for the encoding of texts in the same format.
- . The guidelines should
 - define a recommended syntax for the format,
 - define a metalanguage for the description of text-encoding schemes,
 - describe the new format and representative existing schemes

both in that metalanguage and in prose.

- . The guidelines should propose sets of coding conventions suited for various applications.

An early theory emanating from the textual editing community was the concept of texts as an ordered hierarchy of content objects (OHCO). This theory of textuality was informed by and informed the development of the TEI Guidelines and was developed by Allen Renear, Elli Mylonas, and David Durand in the late 1980s and went through several revisions through the early 1990s. According to it, text is composed of nesting objects, such as chapters, sections, paragraphs, lists, and so forth. Like a set of Chinese boxes, these content objects fit neatly into one another, from the smallest (a letter or a word) to the largest (a book or monograph), with a myriad of other nested units in between (sentences, paragraphs, chapters, sections, etc).

While this theory was instrumental in the development of the TEI, it never fully accounted for the problem of overlapping hierarchies. Overlapping hierarchies breaks the neatly nesting pattern described above. For example, a metaphor in a poem may cut across two or more lines (marked by the tag `<l>`). It might seem like a purely technical issue that a language like XML requires one element to close before another opens, as in the following:

```
<l>text text text</l>
```

```
<l>text text <metaphor> text text</metaphor></l>
```

as opposed to the following:

```
<l>text text <metaphor> text text</l>
```

```
<l>text text </metaphor></l>.
```

The creators of the OHCO theory concede that this may be more than a technical issue and that it may point to some of the thorniest issues surrounding text encoding as an intellectual endeavor. Text encoding, like any other area of textual scholarship, is not theory-free. It is subjective, theoretical, and interpretative. Texts, particularly literary texts, have competing hierarchies, all of which may have equal claim to being represented as they express different views of the text. For example, the hierarchy that SGML and hence TEI most eloquently expresses is what one might term the editorial or bibliographic; that is, representing the text in terms of sentences, paragraphs, chapters, front and back matter, and so on. This is not surprising given SGML's roots as a language written to publish documentary texts in electronic form. From this point of view, one might deduce that the documentary view of text can be read as its *only* structure.

Yet there are many textual features that do not conform to this hierarchy. As

mentioned previously, metaphors may span many lines or stanzas of verse. Narrative events may span many paragraphs and indeed may overlap. Verse drama contains dialogue lines (speeches), metrical lines, and sentences. But these sentences and metrical lines overlap in the case of enjambment or when a character begins talking and another interrupts (Renear 119–21). All these hierarchies have equal claim to representation.

The TEI has risen to the challenge of accommodating alternative views of the text. Since its establishment as a consortium in 2000, it has created opportunities (particularly through its chartering of special-interest groups) for new communities of practice to propose additional tags, as well as to inform and influence its intellectual growth. For example, the Manuscript Special Interest Group has proposed a methodology to create a genetic edition of a text. Within this view of textuality, the editor not only identifies what is on the page but also attempts to reconstruct the process by which those linguistic and bibliographic codes came into being.

Even if the cultural objects of an edition are primarily multimedia (rather than full-text), searching, browsing, and to some extent display are performed on metadata attached to the objects as opposed to the objects themselves. Some media-based projects use the TEI Header to encode bibliographic information for nontextual objects. Others use standards, such as VRA Core (developed by the Visual Resource Association for the cultural heritage community), a rich encoding scheme for image-based material. Like the TEI, VRA Core can be used as an interchange format, can be integrated with other XML standards through the use of XML name spaces, and can also be mapped onto a less-expressive standard, such as Dublin Core, which has become the de facto interchange format for basic metadata.

But the encoding standard used is only one piece of the framework for electronic scholarly editions. The encoded text or other digital object typically resides in a database that enables the sophisticated searching and browsing we have come to expect from online editions. Some editions simply use scripts—a combination of PHP or JSP with XSLT to transform texts to HTML for Web delivery. Open-source, XML-aware databases, such as *eXist*, are frequently used for single-author or themed editions. An enterprise-level solution can be found in *FedoraCommons* (*Fedora* stands for “Flexible Extensible Digital Object Repository Architecture”), an architecture for storing, managing, and accessing digital objects. *FedoraCommons* is media-agnostic: its architecture defines a set of abstractions for expressing these objects, for asserting relationships among digital objects, and for linking behaviors (i.e., services). *FedoraCommons* provides a framework for multiple projects to be housed within one repository, allowing for greater reusability of code, but even more important is that editions within a common framework do not exist as digital silos.

Possibly the most important lesson from the past twenty years of digital scholarly editions is that it is necessary to separate content from display and to present the objects of our contemplation—the full-text files, the images, the audio, and moving images—according to well-established standards. There is no doubt that the editions we create today will be migrated into new platforms

and formats in the future. There are also more opportunities for derivative works to be created by harvesting objects from a variety of sites into new compilations in which the object's original context is lost. Knowing that this type of reuse is possible, it is even more important for editors to ensure that essential information, such as copyright, reuse statements, and provenance, is attached to every object as opposed to having that information reside in an "About the Project" Web page.

Digital editions are typically never complete: there is frequently no point that marks the end of a project, as with book publication. Rather, digital editions tend to be open-ended. The relative ease in adding new material as well as in correcting old content allows for an expansionary editorial model, much like hypertext itself, with shifting centers and nodes. Editors of digital editions need to be cognizant of systems and applications that no longer function in newer Web environments and keep abreast of standards and protocols to decide when and how to migrate resources. Moreover, in recent years we have been more aware of the fragility of the works we create: how easy it is, for example, for a server to be turned off, resulting in years of work disappearing into the ether or in the migration of a custom-made encoding scheme and Web application into a newer environment being too great an effort, no matter how valuable the resource.

Digital scholarly editions are not simply the encoded text that can be glimpsed through a browser's reveal codes or, indeed, that can be hidden from view completely as the text is transformed from the database to HTML for Web delivery. Digital scholarly editions are a myriad of standards and transformations, of scripts and software that are brought into existence when a reader comes to a site and makes a request. On a server, possibly halfway around the world, a series of commands are called into play that seemingly instantaneously deliver the results of a query. But the decisions that enable this Internet alchemy are practical and rooted in the theoretical and philosophic concerns of three disciplinary areas: computer science, information studies, and the humanities. This triad of expertise contains within it the underpinnings of the most successful digital scholarly editions.

The Futures of Digital Scholarly Editing

In the last decade and a half, digital scholarly editing has matured as a field. No longer does it seem herculean to create an electronic edition, although the barriers to entry are still high. But as we enter a period in which the born-digital artifact is more frequently *the* literary or cultural artifact, and as more derivative works are created without reference to an analog original, scholarly editing may not solely or typically be about migrating the analog into the digital or about re-presenting print norms in digital format. In closing, I will discuss just a few of these new modalities and the issues they raise for new genres of digital scholarly editions.

Electronic Literature

The Electronic Literature Organization (ELO) defines electronic literature as “works with important literary aspects that take advantage of the capabilities and contexts provided by a stand-alone or networked computer.” The ELO also identifies a number of forms of practice, including hypertext fiction and poetry, kinetic poetry presented in *Flash* or other software, computer-art installations with prominent literary aspects, interactive fiction, and collaborative writing projects that allow readers to contribute to the text of a work.

The ELO has created an online anthology of electronic literature. Like more traditional anthologies, contextual information—such as biographical information and a short précis—is included for each work. Also included is information such as the software and plug-ins needed to run the work. What would an electronic scholarly edition of one of these works look like? How would variants be presented? the work’s genetic history? What metadata and code are important to capture about these works when the software and hardware they were created for no longer function? Will textual editors become forensic scientists examining the palimpsests of hard drives as assiduously as they now examine watermarks? The ELO has begun to tackle these questions in their *Acid-Free Bits: Recommendations for Long-Lasting Electronic Literature* (Montfort and Wardrip-Fruin) and *Born-Again Bits: A Framework for Migrating Electronic Literature* (Liu et al.).

Crowdsourcing and the Social Edition

Arguably, other disciplines have been more creative than literary studies in engaging the public in large community-based projects. Projects such as *Galaxy Zoo*, *Old Weather*, and *Foldit* allow anybody with an Internet connection to help solve questions of contemporary science. The Australian and Finnish national libraries developed tools to facilitate correction of OCR (optical character recognition) errors in newspaper conversion projects, while, to date, over one million dishes have been transcribed for the New York Public Library’s *What’s on the Menu?* project. *Transcribe Bentham* is one of the first projects to engage the public in contributing to a scholarly edition. By customizing *MediaWiki*, people with no experience in editing, editorial theory, or scholarly transcription are provided with a platform to easily transcribe the letters of Jeremy Bentham adding (unbeknownst to them) light TEI encoding. Over 2,975 letters were transcribed by volunteers during the initial seven-month period of the project (Causser, Tonra, and Wallace 126).

We have already seen how successful distributed editorial projects, such as the *Perseus Digital Library* or *Romantic Circles*, can be. Robinson goes further in his 2010 article “Editing without Walls” to suggest a new model in which the traditional idea of the scholarly editor overseeing the production process is replaced by a distributed workflow in which anybody can participate according to their skills, interests, and abilities. This type of edition, much like *Wikipedia*, relies on the intelligence of the crowd to correct errors—as opposed to the erudition of the editors. As holding institutions make images available for reuse, as the tools to make editing easier are developed, and as

we engage the imagination of the public, the number of (scholarly) editions will increase: new models will proliferate, and exigencies, beyond the economics of the canon, will guide their creation.

Virtual Worlds

Increasingly, the scholarly community is investigating virtual worlds as research and teaching spaces. Virtual worlds are immersive environments that can model, annotate, and stage critical encounters with temporal and spatial realities. These modeled worlds create their own ecosystems that provoke and encourage evolving thought about the material, aesthetic, and cultures of the real-world events they simulate.

Spatial reconstructions are common in virtual worlds such as *Second Life*. One can visit a model of the Globe Theatre or of Great War battlefields. But how could the immersive power of a virtual world be harnessed to provide new insights into works of literature? Just as architectural historians have used the power of virtual-worlds modeling to test theories of building construction and use, literary scholars could manipulate the spatial and temporal aspects of the material evidence surrounding a work's genesis and reception to test their assumptions. TRCs often bring together a wealth of documentary evidence from disparate sources, but these environments, like the codex, flatten time and space onto a one-dimensional reading surface. Immersive environments provide a venue to raise issues of how the phenomenology of place and space can be used to design a new language of scholarly editions, one that has the ability to model sensorial experience lost because of technological and evidentiary constraints.

Mass Digitization

The large-scale mass digitization projects currently under way may be shifting some of the locus of scholarly activity in literary studies from the creation of relatively modest, finely crafted TRCs to developing methods and models to answer the question, What do we do with a million books? To be sure, the creation of TRCs, as well as the tools and software to create and present these collections, are as robust as ever. But mass-digitization projects, beginning with *Google Books* in 2004 (then *Google Book Search*), have created opportunities for literary scholars to engage with more books than any one person could read in a lifetime. What does this mean for our field? What services and software are needed to engage with such massive data sets?

Mass digitization seems, in many ways, the antithesis of scholarly editing. But can we envision a new type of interactive edition that harnesses the distant reading theories of Franco Moretti and others to create dynamic variorum or genetic editions? We will need new editorial structures, ones that do not rely on centralized control but instead are algorithmically generated to exploit, study, and analyze these corpora. Editors working in this environment will not only work with but also design new analytic tools and displays to allow users to engage with an ever-shifting, unbounded textual field. They may include

secondary materials impossible to convey in print form, such as interactive maps and geographic data sets, moving images, and audio, as well as other born-digital derivatives (e.g., folksonomies, *Twitter* feeds, crowdsourcing). This superset of interlocking multitexts and services will allow deeper and wider corpora to be generated and studied than previously available or even imaginable, fostering new forms and theories of textual and bibliographic scholarship.

Digital scholarly editors work in an environment of abundance; they are no longer bound by the constraints of the codex and the economics of print publication. New dialectics, such as the mutual pressure of the alphabetical, figural, and aural within a single representational space (Flanders), have taken the place of material and economic limitations. Digital editions also frequently embed in them a dialectic between what we might consider the more traditional and intuitive bases of literary interpretation and the disambiguating premise of stylometrics, attribution studies, and other statistical methodologies common to computational and algorithmic processing (Drucker 687). But it is the very provocations that these encounters provide that enable new forms of analysis, meaning, and insights. Next-generation digital scholarly editing may well provide some of the most exciting theories for our discipline by melding disciplinary concerns and practices of fields such as computer science, information studies, and human-computer interaction with traditional theories in literary scholarship.

Notes

1. Early examples of this are my own documentation from *The Thomas MacGreevy Archive*, the *Victorian Women's Writers Project*, and *DALF: Digital Archive of Letters in Flanders*, an extension to the TEI Guidelines, to encode modern correspondence.
2. Diplomatic editions are those in which, as far as possible, the marks on the page (any of type of textual witness beyond a typeset page with no emendations) are represented typographically. Emendations, such as additions and deletions, are represented to mirror the edited page. Thus the goal in diplomatic editing is not to present the reader with a finished text but to give the reader insight into the revision process. A witness is one of n number of documents in a text's composition history. For example, a poem might exist in five states: three manuscript drafts, an editor's proof, and a final published text. Each of these five versions would be considered a witness, a chain of physical documents that demonstrates the author's evolution of a single text.

Works Cited

- Bernard, Lou, Katherine O'Brien O'Keeffe, and John Unsworth, eds. *Electronic Textual Editing*. New York: MLA, 2006. Print.
- Blake, Norman, and Peter Robinson, eds. *The Canterbury Tales Project*:

- Occasional Papers I*. Oxford: Office for Humanities Communication, Oxford U Computing Services, 1993. Print.
- . *The Canterbury Tales Project: Occasional Papers II*. Oxford: Office for Humanities Communication, Oxford U Computing Services, 1997. Print.
- Bornstein, George, and Theresa Lynn Tinkle, eds. *The Iconic Page in Manuscript, Print, and Digital Culture*. Ann Arbor: U of Michigan P, 1998. Print.
- Bryant, John. *The Fluid Text: A Theory of Revision and Editing for Book and Screen*. 2002. Ann Arbor: U of Michigan P, 2005. Print.
- Causser, Tim, Justin Tonra, and Valerie Wallace. "Transcription Maximized; Expense Minimized? Crowdsourcing and Editing *The Collected Works of Jeremy Bentham*." *Literary and Linguistic Computing* 27.2 (2012): 119–37. Print.
- Clement, Tanya, ed. *In Transition: Selected Poems by the Baroness Elsa von Freytag-Loringhoven*. U of Maryland Libs., n.d. Web. 8 Jan. 2010.
- Design Principles for Text Encoding Guidelines*. 9 Jan. 1990. Web. 17 Sept. 2012. <<http://www.tei-c.org/Vault/ED/edp01.htm>>.
- Drucker, Johanna. "Theory as Praxis: The Poetics of Electronic Textuality." *Modernism/Modernity* 9.4 (2002): 683–91. Print.
- Eliot, T. S. *The Waste Land: A Facsimile and Transcript of the Original Drafts, including the Annotations of Ezra Pound*. New York: Harcourt, 1971. Print.
- Finneran, Richard J., ed. *The Literary Text in the Digital Age*. Ann Arbor: U of Michigan P, 1996. Print.
- Flanders, Julia. "The Productive Unease of Twenty-First-Century Digital Scholarship." *Digital Humanities Quarterly* 3.3 (2009). Web. 5 Jan. 2010.
- Hockey, Susan. *Electronic Texts in the Humanities: Principles and Practice*. Oxford: Oxford UP, 2000. Print.
- Kiernan, Kevin. "Alfred the Great's Burnt *Boethius*." Bornstein and Tinkle 7–32.
- Landow, George P. *Hypertext 2.0: The Convergence of Contemporary Critical Theory and Technology*. Baltimore: Johns Hopkins UP, 1997. Print.
- Lavagnino, John. "Completeness and Adequacy in Text Encoding." Finneran 63–76.
- Lewis, Wyndham, ed. *Blast*. 1914. Foreword by Paul Edwards. Berkeley: Gingko, 1981. Print.
- Liu, Alan, et al. *Born-Again Bits: A Framework for Migrating Electronic Literature*. Vers. 1.1. Electronic Lit. Assn., 5 Aug. 2005. Web. 5 Jan. 2010.
- Loizeaux, Elizabeth Bergmann, and Neil Fraistat. *Reimagining Textuality: Textual Studies in the Late Age of Print*. Madison: U of Wisconsin P, 2002. Print.
- McGann, Jerome. *Radiant Textuality: Literature after the World Wide Web*. New York: Palgrave, 2001. Print.
- . "The Rationale of Hypertext." Sutherland 19–46.
- McKenzie, D. F. *Bibliography and the Sociology of Texts*. Cambridge: Cambridge UP, 1999. Print.
- Montfort, Nick, and Noah Wardrip-Fruin. *Acid-Free Bits: Recommendations*

- for *Long-Lasting Electronic Literature*. Vers. 1. Electronic Lit. Assn., 14 June 2004. Web. 5 Jan. 2010.
- Moretti, Franco. *Graphs, Maps, Trees: Abstract Models for a Literary History*. London: Verso, 2007. Print.
- Palmer, Carol L. "Thematic Research Collections." Schreibman, Siemens, and Unsworth 348–65.
- Price, Kenneth. "Electronic Scholarly Editions." Siemens and Schreibman 434–50.
- Renear, Allen. "Out of Praxis: Three (Meta)Theories of Textuality." Sutherland 107–26.
- Renear, Allen, Elli Mylonas, and David Durand. "Refining Our Notion of What Text Really Is: The Problem of Overlapping Hierarchies." N.p., 6 Jan. 1993. Web. 17 Sept. 2012.
<<http://www.stg.brown.edu/resources/stg/monographs/ohco.html>>.
- Robinson, Peter. *The Digitization of Primary Textual Sources*. Oxford: Office for Humanities Communication, Oxford U Computing Services, 1993. Print.
- . "Editing without Walls." *Literature Compass* 7.2 (2010): 57–61. Web. Aug. 2012.
- . "Is There a Text in These Variants?" Finneran 99–116.
- . *The Transcription of Primary Textual Sources Using SGML*. Oxford: Office for Humanities Communication, Oxford U Computing Services, 1994. Print.
- Schreibman, Susan. "The Text Ported." *Literary and Linguistic Computing* 17.1 (2002): 17–87. Print.
- Schreibman, Susan, Ray Siemens, and John Unsworth, eds. *A Companion to Digital Humanities*. Oxford: Blackwell, 2004. Print.
- Shillingsburg, Peter L. *From Gutenberg to Google*. Cambridge: Cambridge UP, 2006. Print.
- Siemens, Ray, and Susan Schreibman, eds. *A Companion to Digital Literary Studies*. Oxford: Blackwell, 2008. Print.
- Smith, Martha Nell. "Corporealizations of Dickinson and Interpretive Machines." Bornstein and Tinkle 195–221.
- . "Electronic Scholarly Editing." Schreibman, Siemens, and Unsworth 306–22.
- Sperberg-McQueen, C. M. "Textual Criticism and the Text Encoding Initiative." Finneran 37–62.
- Sutherland, Kathryn, ed. *Electronic Text: Investigations in Method and Theory*. Oxford: Clarendon, 1997. Print.
- "TEI: History." *Text Encoding Initiative*. N.p., n.d. Web. 5 Jan. 2010.
<<http://www.tei-c.org/About/history.xml>>.
- Unsworth, John. *Thematic Research Collections*. N.p., 28 Dec. 2000. Web. 10 Jan. 2010. <<http://www3.isrl.illinois.edu/~unsworth/MLA.00/>>.