Book Publishing Final Research Paper   English

5-2015

Critical Editions and the Promise of the Digital: The Development and Limitations of Markup

Alexandra Haehnert
Portland State University

Critical Editions and the Promise of the Digital: The Development and Limitations of Markup

Alexandra Haehnert

Reading Committee

Dr. Per Henningsgaard
Abbey Gaterud
Adam O'Connor Rodriguez

Critical Editions and the Promise of the Digital: The Development and Limitations of Markup[1]

Contents

---

[1] The research question addressed in this paper that was approved by the reading committee on April 28, 2015, is "How has semantic markup evolved to facilitate the creation of digital critical editions, and how close has this evolution in semantic markup brought us to realizing Charles L. Ross's vision of a layered hypertext with which reader-users can productively interact?"

Introduction

My earlier paper "Digital Critical Editions: Early Prospects and Current Realities" (Haehnert) provides an overview of traditional editorial theory in order to highlight the inadequacy of printed critical editions to represent transparently the diachronic character of the works they present to the reader, and it argues that the digital environment can, in theory, facilitate the creation of user-friendly critical editions that flexibly display the development of literary works.

This paper seeks to explore the technology underlying current digital editions—which I maintain are wanting in their exploitation of the digital medium—in more detail; its focus is squarely on editorial praxis, not on editorial theory (leaving aside the fact that all praxis is informed by specific theoretical assumptions). Neither am I concerned with questioning whether digital techniques render traditional critical editing obsolete; in fact, I am convinced that they do not, and my paper proceeds from this standpoint (cf. Robinson$_{2004}$).[2] Similarly, I am only marginally interested in exploring the augmentation of digitized texts with multimedia elements as described by Boot and van Zundert (142). My main concern is with the representation of purely textual material and the technologies underpinning that representation—semantic markup languages—in the digital realm. I argue that the limitations of semantic markup contribute to current digital editions falling short of expectations.

The Promise of the Digital

Virtually all texts that scholarly editors deem worthy of critical editorial treatment have a varied, oftentimes messy, history of genesis and transmission that any scholarly editor must come to terms with. Texts may exist in manuscript form, as typesetter proofs, in various editions created before and after the author's passing; they may be fragmentary documents; and many different agents besides the original author may have had a hand in their development. "Textual fluidity," according to John Bryant, is the "inherent condition of any written document" (1). I already introduced two

---

[2] For better readability, I will deliberately diverge from MLA style throughout this paper and distinguish works by the same author by appending to in-line references their year of publication in subscript rather than including a short form of their title.

significant traditional approaches to critical editing—the Anglo-American school, preferring eclectic text editions, and the German-inflected school, striving to produce genetic text editions—to demonstrate the profound inadequacies of both approaches to represent the text's inherent fluidity in print. The eclectic edition chooses to represent the text as a static and choate entity as a clear reading text, with explicatory material relegated to an often unwieldy and incomplete apparatus. The genetic edition, on the other hand, sacrifices readability for a complicated representation of the text's fluidity, thereby making it accessible to a highly specialized readership with a decidedly academic interest in the text only (Haehnert 10). The physical limitations of the print book render a transparent and accurate representation of texts' complex diachronic features impossible (ibid. 10; cf. McGann, Dahlström 2, Andrews 65, Fraistat 331).

Hypertext, on the other hand, can easily accommodate complex textual relationships and "dramatically reorganiz[e] text in networked ways, delinearizing and interlinking the text both within its own boundaries and in relation to other such texts" (Fitzpatrick 96). Roberta Capelli, writing about mediaeval manuscripts, puts it this way:

> A single page from a manuscript may provide a series of multi-ordered information (text and paratext), multilevel information (palimpsestic writing), and multi-layered information (variants and copying activity from multiple sources). The tangible and intangible pluridimensionality of the manuscript . . . requires synchronization of all the data and metadata which they are capable of providing and which is impossible to realize in the static and linear form of traditional printing, but which is congenial to the non-linear form of the hypertext. (6)

Publishing critical editions in digital form thus promises scholars to do away with the practical, materially based disadvantages of the print book, including limitations on the extent of the apparatus and the number of variant readings that can reasonably be represented.

One of the most vivid visions of the enhanced functionality digital critical editions could afford scholars comes from a 1996 article by Charles L. Ross, in which he proposes a multilayer arrangement:

> There could be *both* 'default' texts . . . *and* means for the reader to create further texts according to other assumptions. For instance, the reader would access both the paradigmatic (vertical) and syntagmatic (horizontal) levels of the text, add or

delete various classes of transmissive factors, choose among variants, and thus create *a* (never *the*) text." (230, original emphases)[3]

The "default" texts Ross is referring to here correspond, essentially, to the clear reading text eclectic editors create, but he also imagines functionalities that allow the reader-user of the edition to toggle on and off other iterations of, and interventions into, the text according to her specific research interests.

Peter Robinson, in 2003, echoes Ross's aspiration and also expresses a desire for "co-operative [sic] and distributed editions," editions that are the result of work contributed by an open-ended number of scholars and readers, editions that can be amended and extended over time. Boot and van Zundert, similarly, argue for open-ended, networked editions that can be enhanced by many contributors over time in 2011 (143–44). They stress the need for versioning control so that readers will always be able to access an edition's previous states (150). Robinson even sees "obvious pedagogical opportunities" in a sort of gamification of scholarly editions that demand a high level of engagement from their readership (Robinson$_{2003}$).[4]

This is only a short list of the various expectations for digital critical editions; since almost none have been fulfilled up to now, it shall suffice to illustrate the large gap between aspiration and reality. Despite the promises of the digital environment that scholars recognized early on (e.g., McGann, Ross), there is still, today, a dearth of digital critical editions whose functionality goes beyond what we are familiar with from the print paradigm (e.g., Robinson$_{2003}$, Dahlström, Schmidt$_{2014}$ 11, Andrews 65).

Realizing the Promise of the Digital—with Markup

Work on implementing these functionalities has been going on for well over twenty years now. What almost all digital editions created to date have in common is their being based on semantic markup languages, predominantly the markup guidelines recommended by the Text Encoding

---

[3] See ibid. 228–29 for an elaboration of this vision.
[4] Incidentally, IVANHOE, a "playspace" and "pedagogical environment for interpreting textual and other cultural materials" released in 2006 (and headed, among others, by Jerome McGann), is defunct nowadays; the news item announcing its "temporary" unavailability was posted in January 2012 (IVANHOE).

Initiative (TEI). The following section will discuss the rationale for markup and the development of markup languages in more detail before outlining the considerable disadvantages of generalized markup for work on texts of the humanities.

Many non-technological factors contribute to the current wanting state of digital critical editions: lack or insufficient duration of institutional funding, lack of programming expertise, the perceived prestige gap between print and online editions, uncertainty about digital editions' sustainability, maybe even insufficient interest from the scholarly communications community overall, etc. (cf. Fischer 89, Price). While I am not prepared, then, to argue that the deficiencies of semantic markup are the sole cause for the dearth of fully realized digital critical editions, the following will at least show that encoding texts with markup may not be the be-all and end-all of their development.

First, some preliminary remarks: I am concerned here with semantic (or descriptive or generalized) markup that describes portions of text in terms of their function.[5] Semantic markup allows computers to distinguish between ambiguous forms, which humans—with their natural speech processing capabilities—have no difficulty telling apart. While presentational markup can indicate to the processor that a word is to be displayed in italics, for example, semantic markup clarifies if this word is a foreign term, a citation, or an emphasized word. This allows for carrying out complex search operations; for example, markup makes it possible to retrieve all citations (or references to places and persons, to give just two more examples) in a given text.[6]

Further, semantic markup thus implies a reasonably clear distinction between transcription (the content between tags) and interpretation (the tags we choose to assign to specific content). Thus, `<rubbish>Achievement of your happiness is the only moral purpose of your life</rubbish>` clearly signals both an act of transcription and an act of interpretation by the

---

[5] By contrast, presentational (or prescriptive or procedural or typographic) markup like HTML describes what processing is to be carried out on particular portions of text; it describes how they are to be displayed (Pichler and Bruvik 187).

[6] Project Gutenberg, founded in 1971, should definitely be mentioned as a successful example of a text digitization effort. It presents texts in markup-free ASCII-only format to make them available to the broadest possible reading public (Fitzpatrick 101). Since the texts are not marked up semantically, however, they do not allow for a properly "critical" treatment (ibid. 101, cf. also Sperberg-McQueen 60 n. 6).

encoder. Evidently, common markup schemes advocate the use of more "neutral" and "objective" tags like `<line>` and `<para>`, but the point still holds, and its significance will become clear in time.

Computers in Editorial Projects Before the Widespread Adoption of Markup

Before encoding texts with markup proliferated, literary scholars and editors saw the advantages of making use of computers.[7] Early on, they were mainly interested in (and limited to) speeding up the time-intensive work of creating concordances and collations; they were concerned, overall, with enhancing the preparatory work that goes into the creation of an edition. For example, in a 1962 article, Ephim G. Fogel writes admiringly of several successful projects creating concordances with punch cards and "data-processing machines" such as the IBM 705 (15).[8] Corpus linguistics adopted computer technology early on: in 1963, W. Nelson Francis and Henry Kučera chose to make use of an encoding system developed by the United States Patent Office in 1959 for compiling their Brown University Standard Corpus of Present-Day American English, the first machine-readable corpus of modern English (Francis and Kučera).[9] Each entry was preceded by an abbreviated form of a citation reference indicating specific sections and line numbers in the source document (Hockey[2000] 27).

Other early notational editorial practices anticipated or paralleled the advent of what we now recognize as markup.[10] In 1979, Wilhelm Ott describes a collation program, TVGL, used at the University of Tübingen, that, after input of magnetic tape containing two text versions, would output text annotated with characters like =, −, and + to indicate replacement, deletion, and

---

[7] The journal *Computers and the Humanities* was founded in 1966 (Hockey[2004a]).

[8] The following passage from Fogel's paper is too precious not to replicate here (and it provides a potent reminder of how speculative forecasts of future technological developments are, then as now): "'The latest [computers],' writes Ritchie Calder, 'are a thousand times faster than those of three years ago and a million times faster than those of ten years ago,' and he reports that in June, 1959, in Paris, at an International Conference on Information Processing, scientists seriously discussed 'machines which would memorize all the knowledge in the world.' One's mind reels and retreats to somewhat less staggering fantasies in which the C. W. Wallaces and Leslie Hotsons [contemporary American scholars of Elizabethan literature] of the twenty-first century, working in American repositories, ask computers to search magnetic tapes of British archives for all occurrences of names with, say, the components *Sh*, *k*, *sp*, *r* or *M*, *r*, *l*. A daydream high fantastical, perhaps; yet the photoduplication during World War II of a vast number of British documents, now available at the Library of Congress on microfilm, provides a notable precedent for the internationalization of archive." (18–19)

[9] Cf. Hockey[2001] for a discussion of computer programs for corpus linguistics.

[10] Cf. Ore, who argues that forms of markup has been used in Western scholarly work since antiquity.

addition of text, respectively (33–35). For his groundbreaking synoptic edition of Joyce's *Ulysses* (1984), Hans Walter Gabler employed a computer program, TuStep (Tübingen System of Text Processing Programs), to help him sequence the different manuscript versions (Hockey$_{2000}$ 126), and it is not far-fetched to regard his diacritical marks as a form of text markup.[11]

The Development of Markup

One of the best-known early markup schemes for citations and locations within a text is COCOA (word COunt and COncordance Atlas), created in the 1960s. A short example illustrates its features well:

```
<W Shakespeare>
<T Merchant of Venice>
<A 2>
<S 6>
<C Graziano>
This is the penthouse under which Lorenzo
Desired us to make stand
<C Salerio>
His hour is almost past
<C Graziano>
And it is marvel he outdwells his hour
For lovers ever run before the clock.
```

(qtd. in Hockey$_{2000}$ 27–28)

It is apparent that this markup is flat; there are no end tags as they are familiar to us now from XML, and the end of a particular element can only be signaled by the beginning of a new element of the same class (Hockey$_{2004b}$ 368); in the example above, the markup signals the end of Graziano's speech only by the beginning of Salerio's speech. Because COCOA cannot handle embedding one element into another, then, the syntax is not suitable for encoding text functions with a high degree of granularity or encoding more complex text relations such as additions, deletions, or annotations, and thus "attempts to use it for any kind of critical apparatus have failed because its structures are not powerful enough" (Hockey$_{2004b}$ 368).[12] The same is true for GML, Generalized Markup Language,

---

[11] Charles L. Ross argues that Gabler's *Ulysses* "reveals the printed book yearning for its electronic transformation" (227).

[12] On the other hand, COCOA's structure does offer some flexibility in terms of encoding overlapping hierarchies, a feature that "outstrips [those] of almost all modern markup schemes" (Hockey$_{2004a}$), as I will discuss at a later point.

developed in the early 1970s by Charles Goldfarb, Edward Mosher, and Raymond Lorie for IBM to handle documents intended for print (Schmidt[2010] 339).

In 1986, ISO, the International Organization for Standardization, standardized SGML syntax (Standard Generalized Markup Language), which was derived from GML specifications but is much more complex. It saw sustained and widespread use in industry and commerce. Since the syntax was non-proprietary and well documented, users were free to develop their own markup nomenclatures based on SGML syntax "and were not restricted to use software from any particular vendor" (Pichler and Bruvik 165). `<s><name>Bach</name> was a <emp>great</great> composer.</s>` is an example of a sentence encoded with SGML syntax (ibid. 164), which would prove to become the most firmly established syntax for future developments (indeed, it is a direct precursor to XML) since it can be applied to many different types of text and complex text relations.

Independent of the standardization and proliferation of SGML, the [Ludwig] Wittgenstein Archives Project at the University of Bergen developed a markup syntax, MECS (Multi-Element Code System), in 1991 (Pichler and Bruvik 186) that allows for a detailed description of text elements and overlapping hierarchies. An encoded version of `The` `trees` *`are green`* *`with white`* `flowers` in MECS looks like this: `<us1/trees <i/are green/us1> with white/i> flowers` (Pichler and Bruvik 189–90). However, MECS "can be processed only by way of software written at Bergen . . . specific to the requirements of the Wittgenstein Archives Project" (Hockey[2004b] 369) and has, therefore, not found much use outside Wittgenstein philology (Hrachovek and Köhler).

The Text Encoding Initiative

As humanities scholars started to make use of SGML syntax, different—incompatible— nomenclatures proliferated. In 1987, Nancy Ide of Vassar College in Poughkeepsie convened a meeting of likeminded scholars "to examine the possibility of creating a standard encoding scheme for humanities electronic texts," and they agreed on a set of principles to guide this endeavor (Hockey [2004a]). Out of this event grew the Text Encoding Initiative (TEI), whose first version of the *Guidelines for Electronic Text Encoding and Interchange of Machine-Readable Texts* was published in 1994 (Pichler and Bruvik 175). The first three versions of the Guidelines were based on SGML

syntax; in a two-year effort, TEI rewrote the *Guidelines* to use XML syntax and published its XML-based version P4 in 2004 (Text Encoding Initiative). The most recent version, P5, was initially released in 2007 and is being continually revised and updated (TEI Consortium x, i).

Figure 1 shows a portion of Henry David Thoreau's *Walden* encoded according to TEI guidelines for the *Digital Thoreau*'s digital "fluid-text edition":

```
 1  ▼                    <seg n="2a" change="#wc_0a">
 2  ▼                        <note type="header" resp="#clapper">
 3                           <ab type="string">Economy 2a written: A; rewritten: C, C.</ab>
 4  ▼                        <ab type="parsed">
 5                               <seg type="chapterNum">1</seg>
 6                               <seg type="chapterTitle">Economy</seg>
 7                               <seg type="paragraph">2a</seg>
 8                               <seg type="written">A</seg>
 9                               <seg type="rewritten">C, C</seg>
10  ╚                        </ab>
11  ▼                    </note> I should not <app xml:id="walc01-app-0004">
12                           <lem>obtrude my affairs so much on the notice of my readers</lem>
13  ▼                        <rdg wit="#wc_0a">presume to talk so much about myself and my affairs as I
14                               shall in this <del>lecture</del>
15  ▼                           <del>
16                                   <add>book</add>
17  ╚                           </del>
18  ▼                           <del>
19                                   <add>work</add>
20  ╚                           </del>
21                               <add>book</add>
22  ╚                        </rdg>
23  ▼                        <rdg wit="#wc_0c">obtrude <del>myself and</del> my affairs so much on the
24                               notice of my readers<anchor type="hn" xml:id="hn-01-010"/>
25  ╚                        </rdg>
26  ╚                    </app> if very <app xml:id="walc01-app-0005">
```

Figure 1. TEI markup of Henry David Thoreau's *Walden* (*Digital Thoreau*).

Lines 1 through 9 describe the chunk of *Walden* that is being marked up. The text portions not contained by descriptive tags (such as `I should not` on line 11) are the base text the *Digital Thoreau* uses and upon which it builds its encoding of variant readings taken from different documents. The totality of variant readings are enclosed in `<app>` tags (e.g., lines 11–26). Embedded in those, the base text's version is described with `<lem>` tags (line 12) and its attendant variants with `<rdg>` tags. Those are further identified by `wit` attributes specifying the individual variant (lines 13 and 23). Inside those `<rdg>` tags, the type of text interventions are noted with, for example, `<del>` (line 14) or `<add>` (line 16) tags. Editorial notes can be inserted via `<note>` tags.

The markup in Figure 1 is relatively straightforward; the current version of the TEI *Guidelines* specifies around 550 different tags to encode humanities texts with markup, annotations, and metadata (Schmidt$_{2012}$ 4). Today, TEI markup is "the *de facto* representation standard for textual scholarship" (Andrews 63).

Criticism of Generalized Markup

Criticism of TEI markup usually focuses on two problematic areas: its non-interoperability and its inability to efficiently and correctly encode overlapping hierarchies.

Desmond Schmidt, perhaps the most vocal critic of the use of TEI in digital editions, locates the inadequacies of TEI markup (and, in fact, any kind of generalized markup) for use in the humanities in its origin as an industrial standard for marking up documents intended to be printed in a certain way (Schmidt$_{2010}$ 339, 341–42). The markup most widely used for creating digital editions is, indeed, as I hope has become clear in my discussion of the development of markup above, adapted from the same medium born-digital editions are ostensibly seeking to supersede.

Pichler and Bruvik maintain that "it is only through the disciplines and technologies of markup that one can carry out [the separation between transcription and presentation] on all levels" (185). Yet Schmidt, as well as Peter Shillingsburg, bring up the larger question of whether it is acceptable to embed markup (which is based on acts of historically situated interpretation) in the text repository proper (Schmidt$_{2010}$, Shillingsburg 164–67).

I will synthesize those views in the following.

Overlapping Hierarchies

SGML, XML, and the resulting TEI were "devised on the assumptions that a document is a logical construct" with a distinctly hierarchical tree structure (Vanhoutten; cf. Schmidt$_{2014}$ 5). They specify that elements must all be well formed, that is, properly nested; if <a> precedes , then must precede . Otherwise, the parser will return a validation error; it will not know how to output elements that are not nested. Yet, humanities texts not uncommonly contain *overlapping* hierarchies (Barnard et al. first discussed this problem in SGML in 1988). We may include in this category textual variation, which involves an overlap of markup and content, and concurrently existing hierarchies within texts that are impossible to encode (they would involve an overlap of markup, if not content).

Figure 1 showed a successful encoding of textual variation. However, an operation like this becomes more and more difficult the more variants one wants to encode. Complex variant

structures cannot be encoded without copying much of the base text over and over, making transcription redundant and error-prone. Schmidt gives the example of joining two paragraphs from one version together in another version. The string `</p><p>` varies over the two versions in this situation. "Such variation can't be specified without copying the entire content of the two paragraphs as one alternative, and the other merged paragraph as the other, if well-formedness is to be preserved" (Schmidt[2010] 347).

Furthermore, it also becomes increasingly difficult "to record precisely what is a variant of what" (Schmidt and Colomb 502). How, for example, could one precisely indicate a simple transposition of two words from one version to the other? Suppose the base version of the work reads, `Ben likes Holly`, and another version reads, `Holly likes Ben`. We could choose to encode `Holly likes Ben` as a variant of `Ben likes Holly`, or we could encode `Holly` as a variant of `Ben`, and `Ben` as a variant of `Holly`. While the second solution is less redundant and more precise than the first, it still does not indicate the dynamic nature of the transposition-as-transposition. On the basis of this very simple example, the pitfalls of encoding textual variation within the same document become apparent.

The most prominent example cited for the second kind of overlap, concurrent hierarchies within texts, involves the markup of dramatic works, where the transcriber might want to encode speeches, lines, and verses; not uncommonly, lines and verses run over from one to another character's speech. Another example would be trying to encode a metaphor in poetry that does not neatly conform to line boundaries (cf. Robinson[2003]).

The TEI *Guidelines* address the problem of overlap explicitly:

> Non-nesting information poses fundamental problems for any XML-based encoding scheme, and it must be stated at the outset that no current solution combines all the desirable attributes of formal simplicity, capacity to represent all occurring or imaginable kinds of structures, suitability for formal or mechanical validation. The representation of non-hierarchical information is thus necessarily a matter of trade-offs among various sets of advantages and disadvantages. (TEI Consortium 636)

The ability to encode overlapping hierarchies in a way that is faithful to the source documents of the work in question would be an essential prerequisite to realizing Charles L. Ross's vision of the

multi-layer, toggle-able digital critical edition. Yet overlapping hierarchies can only be approximated by generalized markup, "and this approximation gets worse the more detail there is to record" (Schmidt$_{2010}$ 348).

Lack of Interoperability

Data is "interoperable" when it can be loaded into a variety of software without requiring human modification of that data. If the data needs to be modified before it can be processed by software other than the one in whose framework it was created (which often means that the underlying information is damaged in the process), it is merely "interchangeable" (cf. Schmidt$_{2014}$ 7). As I mentioned above, the TEI *Guidelines* were established with the intent of creating a standard encoding scheme for the humanities. Yet in scholarly practice, Tara L. Andrews maintains, the *Guidelines* are "routinely customized for each new project . . . exception becomes the rule" (63). Whereas standardization "would allow for true progress toward digital critical editions," she laments that the widespread practice of customizing the TEI nomenclature for individual projects renders interoperability virtually impossible (ibid. 63).[13]

This lack of interoperability goes beyond the level of markup, however. Online environments, too, are often created for specific projects. As soon as the grant money for those projects runs out, their resources run the risk of deteriorating and becoming obsolete, and the underlying encoded text becoming inaccessible (cf. Schmidt$_{2010}$ 348).[14]

It is evident that non-interoperability complicates the realization of Robinson's and Boot and van Zundert's desire for cooperative and distributed editions significantly.

Embedding Transcription and Markup in the Same Document

I argued before that markup allows for a clear distinction between transcription and interpretation. This is true as long as one is actually able to view the TEI source file and make out tags as opposed to content. Yet Shillingsburg (164–67) and Schmidt question whether it is a good idea in the first place to record both transcription and interpretation in the same source file. In fact, Schmidt argues

---

[13] Cf. Morrisey for a discussion of interoperability problems as concerns SGML and XML.
[14] The demise of IVANHOE described in footnote 4 of this paper is but one example of this.

that texts should be transcribed in plain-text format ASCII so as not to "embed the technology and interpretations of today into texts that will be archived for the future" (Schmidt[2010] 349). Instead of embedding markup in transcriptions directly, he suggests a data architecture that layers interpretative documents over sustainable plain-text transcription documents, with the interpretative layers pointing to the transcription, and not the other way around.[15]

I need to mention here the online edition of St. Patrick's *Confessio*, funded by the Royal Irish Academy. The creators chose to tokenize every single word of their base text (that is, they tagged every word with a specific ID attribute), which allows them to augment the text with very precise annotations and variant notations, referring to specific words or ranges of words (Cream). While their markup still suffers from all the disadvantages of TEI described above, such practice—which I have observed in no other digital edition I surveyed in the course of my research—does, at least, show scholarly editors implicitly questioning the common modus operandi in the digital realm.

Coda: The State of the Digital Critical Edition

Peter Robinson, in his insightful 2013 article "Towards a Theory of Digital Editions," analyzes current practice that privileges the display and meticulous transcription of documents over a functionally richer experience of the *work* they are a part of: swept away by the data storage facilities the online environment affords (including innumerable high-resolution full-color images would have been impossible in print form), creators of digital editions run the risk of turning them into mere "data silos," as Neil Fraistat puts it (331). "Should this model of the digital edition prevail, we will see a flood of facsimile editions in digital form" (Robinson[2013] 127), but only few that offer ways *into* the documents they display (ibid. 126). Yuri Cowan identifies as positive the prevailing tendency of editors "not so much 'editing' as . . . creating editions, and . . . making archives rather than forming eclectic texts" (226), but vast archives of images and transcription are hardly useful for the reader who does not approach the work they represent with an in-depth understanding in

---

[15] I am going to enthusiastically refer the reader to Schmidt[2010] at this point; exploring his fascinating suggested data architecture is beyond the scope of this paper.

the first place. Not that such archives and digital facsimile editions are useless—on the contrary[16]—but they are hardly *critical*. "Attending to infrastructure emerges as an applied form of scholarship, with the boundaries of [the digital humanities] expanded beyond discursive modes of interpretation and into the realm of critical making" (Coble et al. 5). At the least, for the reader-user not to drown in the wealth of information digital editions can provide, the editor has to augment her traditional skillset with those of an information architect who can point out several ways for the reader to navigate the edition and fruitfully interact with it (cf. Moran et al., Price). And if this is difficult to do on the basis of inadequate semantically encoded transcriptions, then, perhaps—even if it may be a discomforting prospect—it is time to think about alternative ways of going about it.[17] If, however, providing curated and critically negotiated reading experiences is not in actuality the objective of most digital bibliography projects, then their creators ought to be more upfront about the fact that their praxis is, essentially, redefining the meaning of "critical edition," or label their projects differently altogether.

Conclusion

> Principles may define practice. But practice may become so accepted, so ingrained,
> that principles are determined by practice, and not the other way about.
> (Robinson[2013] 126)

This paper illustrated the gap between aspirations for what the digital critical edition *could* be and what it *is* in the second decade of the twenty-first century. It maintains that online editorial praxis is still very much indebted to models belonging to the print paradigm. Specifically, it outlined the development of semantic markup languages, which underlie all online editions today, and highlighted arguments according to which their inadequacy for creating digital critical editions can be traced back to their origin in print-based technologies. The fact that most articulations of digital critical editions' functionality remain aspirational can at least in part be attributed to the

---

[16] Cf. Cowan 224–26 for a discussion of the revival of material bibliography that digital facsimile editions make possible.
[17] Schmidt[2014] and Wittern both describe alternatives to or augmentations of generalized markup that could annul its evident disadvantages; exploring their ideas in further detail is material for another paper, however.

limitations of semantic markup. Semantic markup certainly is a powerful and necessary tool, but it will not be the only tool needed for making real progress toward realizing the promise of the digital for critical editions.

Works Cited

Andrews, Tara L. "The Third Way: Philology and Critical Edition in the Digital Age." *Variants* 10 (2013): 61–76. Print.

Barnard, David, et al. "SGML-Based Markup for Literary Texts: Two Problems and Some Solutions." *Computers and the Humanities* 22.4 (1988): 265–276. Print.

Boot, Peter, and Joris van Zundert. "The Digital Edition 2.0 and the Digital Library: Services, not Resources." *Bibliothek und Wissenschaft* 44 (2011): 141–52. Print.

Bryant, John. *The Fluid Text: A Theory of Revision and Editing for Book and Screen*. Ann Arbor: U of Michigan P, 2002.

Capelli, Roberta. "Practical and Theoretical Implications of Digitizing the Middle Ages." *CLCWeb: Comparative Literature and Culture* 15.3 (2013): <http://dx.doi.org/10.7771/1481-4374.2248>. PDF.

Coble, Zach, et al. "Process as Product: Scholarly Communication Experiments in the Digital Humanities." *Journal of Librarianship and Scholarly Communication* 2.3 (2014): 1–11. PDF.

Cowan, Yuri. "Reading Material Bibliography and Digital Editions." *Editors, Scholars, and the Social Text*, ed. Darcy Cullen. Toronto: U of Toronto P, 2012. 223–38. Print.

Cream, Randall. "The Technologies of the HyperStack." *St. Patrick's Confessio*. Royal Irish Academy, 2011. Web. 6 May 2015. <http://www.confessio.ie/about/technologies>.

Dahlström, Mats. "Digital Incunables: Versionality and Versatility in Digital Scholarly Editions." *Electronic Publishing in the Third Millennium. Proceedings of an ICCC/IFIP Conference Held at the Kaliningrad State University, Kaliningrad/Svetlogorsk, Russia, 17–19 August, 2000*. Ed. Peter Linde, John W.T. Smith, and Elena Emelianova. Washington: ICCC, 2000. 224–34. <http://elpub.architexturez.net/doc/oai-elpub.id-0024>. PDF [page references are to PDF copy].

Digital Thoreau. Paul Schacht, dir. Dev. and maintained by the State University of New York at Geneseo, the Thoreau Society, and the Walden Woods Project. Geneseo: SUNY, 2014. 26 Nov. 2014. <http://www.digitalthoreau.org>.

Finneran, Richard J., ed. *The Literary Text in the Digital Age*. Ann Arbor: U of Michigan P, 1996. Print.

Fischer, Franz. "All Texts Are Equal, But ... Textual Plurality and the Critical Text in Digital Scholarly Editions." *Variants* 10 (2013): 77–92. Print.

Fitzpatrick, Kathleen. *Planned Obsolescence: Publishing, Technology, and the Future of the Academy*. New York: New York UP, 2011. Print.

Fogel, Ephim G. "Electronic Computers and Elizabethan Texts." *Studies in Bibliography* 15 (1962): 15–31. Print.

Fraistat, Neil. "Textual Addressability and the Future of Editing." *European Romantic Review* 23.3 (2012): 329–333. Print.

Francis, W. Nelson, and Henry Kučera. *Brown Corpus Manual*, 1964 [revised edition, 1979]. Repr. by Uni Research (Anne Lindebjerg), 11 Sep. 1997. Web. 11 May 2015. <http://clu.uni.no/ icame/brown/bcm.html>.

Haehnert, Alexandra. "Digital Critical Editions: Early Prospects and Current Realities." Seminar paper, Portland State University. Dec. 2014. <http://issuu.com/ahhnrt/docs/digital_ critical_editions>. PDF.

Hockey, Susan. "Concordance Programs for Corpus Linguistics." *Corpus Linguistics in North America: Selections from the 1999 Symposium*, ed. Rita C. Simpson, and John M. Swales. Ann Arbor: U of Michigan P, 2001. 76–97. Print.

–––. *Electronic Texts in the Humanities: Principles and Practice*. Oxford: Oxford UP, 2000. Print.

–––. "The History of Humanities Computing." *A Companion to Digital Humanities*, ed. Susan Schreibman et al. Oxford: Blackwell, 2004. John Unsworth, 10 Dec. 2007. Web. 11 May 2015. <http://digitalhumanities.org:3030/companion/view?docId=blackwell/9781405103213/ 9781405103213.xml&chunk.id=ss1-2-1>. [Referred to as Hockey$_{2004a}$ throughout the paper]

–––. "The Reality of Electronic Editions." *Voice, Text, Hypertext*: *Emerging Practices in Textual Studies*, ed. Raimonda Modiano et al. Seattle: U of Washington P, 2004. 361–77. Print. [Referred to as Hockey~2004b~ throughout the paper]

Hrachovek, Herbert, and Dieter Köhler. "Digital Hermeneutics: A New Approach to Wittgenstein's *Nachlass*." *Wittgenstein und die Zukunft der Philologie: Eine neubewertung nach 50 Jahren*, ed. Rudolf Haller, and Klaus Puhl. Vienna: öbv, 2002. Repr. *Herbert Hrachovec*. Herbert Hrachovec, n.d. Web. 11 May 2015. <http://hrachovec.philo.at/Kirchberg2_ 010202.htm>.

IVANHOE. *Ivanhoe – A Ludic Playspace*. NINES, 20 Jan. 2012. Web. 11 May 2015. <http://www.ivanhoegame.org/>.

Joyce, James. *Ulysses: A Critical and Synoptic Edition*. Ed. Hans Walter Gabler. 3 vol. New York: Garland, 1984. Print.

Maron, Nancy, et al. "Publarians and Lubishers: Role Bending in the New Scholarly Communications Ecosystem." Proceedings of the Charleston Library Conference. 2013. PDF. 11 May 2015. <http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1533& context=charleston>.

McGann, Jerome J. "The Rationale of HyperText." *Jerome McGann*. U of Virginia, 6 May 1995. Web. 11 May 2015. <http://www2.iath.virginia.edu/public/jjm2f/rationale.html>.

Morrissey, Sheila M. "'More What You'd Call "Guidelines" Than Actual Rules': Variation in the Use of Standards." *Journal of Electronic Publishing* 14.1 (2011). Web. 11 May 2015. <http://dx.doi.org/10.3998/3336451.0014.104>.

Ore, Espen S. "Document Markup: Why? How?" *Historical Social Research / Historische Sozialforschung* 37.3 (2012): 106–24. PDF [page references are to PDF copy]. 6 May 2015. <http://www.cceh.uni-koeln.de/files/Ore_final.pdf>.

Ott, Wilhelm. "A Text Processing System for the Preparation of Critical Editions." *Computers and the Humanities* 13.1 (1979): 29–35. Print.

Pichler, Alois, and Tone Merete Bruvik. "Digital Critical Editing: Separating Encoding from

    Presentation." *Digital Critical Editions*, ed. Daniel Apollon, et al. Champaign: U of Illinois P,

    2014. 179–99.

Price, Kenneth M. "Electronic Scholarly Editions." *A Companion to Digital Literary Studies*, ed.

    Susan Schreibman and Ray Siemens. Oxford: Blackwell, 2009. Repr. *Walt Whitman Archive*.

    Kenneth M. Price, n.d. Web. 11 May 2015. <http://www.whitmanarchive.org/

    about/articles/anc.00267.html>.

Robinson, Peter. "Making Electronic Editions and the Fascination of What Is Difficult."

    Linguistica Computazionale 20 (2004): 415–38. PDF [page references are to PDF copy]. 6 May

    2015. <http://dho.ie/sites/default/files/Fascination_of_what_is_difficult.pdf>.

———. "Towards a Theory of Digital Editions." *Variants* 10 (2013): 105–31. Print.

———. "Where We Are with Electronic Scholarly Editions, and Where We Want to Be." *Jahrbuch*

    für Computerphilologie 5 (2003). Ludwig Maximilian U of Munich, 2003. Web. 6 May 2015.

    <http://computerphilologie.uni-muenchen.de/jg03/robinson.html>.

Ross, Charles L. "The Electronic Text and the Death of the Critical Edition." Finneran 225–31.

Schmidt, Desmond. "The Inadequacy of Embedded Markup for Cultural Heritage Texts."

    Literary and Linguistic Computing 25.3 (2010): 337–56. Print.

———. "The Role of Markup in the Digital Humanities." *Historical Social Research / Historische*

    Sozialforschung 37.3 (2012): 125–46. PDF [page references are to PDF copy]. 6 May 2015.

    <http://www.cceh.uni-koeln.de/files/Schmidt_final.pdf>.

———. "Towards an Interoperable Digital Scholarly Edition." *Journal of the Text Encoding*

    Initiative 7 (2014). PDF [references are to numbered paragraphs]. 6 May 2015.

    <http://jtei.revues.org/979>.

Schmidt, Desmond, and Robert Colomb. "A Data Structure for Representing Multi-Version

    Texts Online." *International Journal of Human-Computer Studies* 67.6 (2009): 497–514. Print.

Shillingsburg, Peter L. "From Physical to Digital Textuality: Loss and Gain in Literary Projects."

    CEA Critic 76.2 (2014): 158–68. Print.

Sperberg-McQueen, C.M. "Textual Criticism and the Text Encoding Initiative." Finneran 37–61.

TEI Consortium. "TEI P5: Guidelines for Electronic Text Encoding and Interchange."
Version 2.8.0. Text Encoding Initiative, 6 Apr. 2015. Web. 6 May 2015. <http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf>.

Text Encoding Initiative. "TEI Task Force on SGML to XML Migration." Text Encoding
Initiative, June 2004. 6 May 2015. <http://www.tei-c.org/Activities/Workgroups/MI/>.

Vanhoutte, Edward. "Electronic Textual Editing: Prose Fiction and Modern Manuscripts:
Limitations and Possibilities of Text-Encoding for Electronic Editions." *Electronic Textual Editing*, ed. John Unsworth et al. TEI Consortium, 31 Oct. 2007. Web. 11 May 2015.
<http://www.tei-c.org/About/Archive_new/ETE/Preview/vanhoutte.xml>.

Wittern, Christian. "Beyond TEI: Returning the Text to the Reader." *Journal of the Text Encoding Initiative* 4 (2013). PDF. 6 May 2015. <http://jtei.revues.org/691>.