Back to the future: what digital editors can learn from print editorial practice

Daniel Paul O'Donnell
Department of English, University of Lethbridge, Canada

Abstract

This article revisits the question of the intellectual adequacy of the print critical edition. Contemporary theory and current digital practice have encouraged editors and users of editions to dismiss various aspects of the print critical edition–particularly the reading text and the critical apparatus–as artifacts of an obsolete technology. Using database theory, the author shows how many of these basic elements in fact represent the most intellectually efficient possible way of organizing information about texts and the readings of their underlying witnesses. By recognizing the inherent sophistication of the classical model, digital editors can improve of print practice by exploiting features of the new medium that make it easier to present such data in interactive ways.

Correspondence:
Daniel Paul O'Donnell,
Department of English,
University of Lethbridge,
4401 University Drive W,
Lethbridge AB,
Canada T1K 3M4.
E-mail:
daniel.odonnell@uleth.ca

This essay was written in 1997. It has not been significantly updated but reflects that particular moment in the debate on electronic text.

1 Introduction

The last decade or so has proven to be a heady time for editors of digital editions. With the maturation of the digital medium and its application to an ever increasing variety of cultural objects, digital scholars have been led to consider their theory and practice in fundamental terms. The questions they have asked have ranged from the nature of the editorial enterprise to issues of academic economics and politics; from problems of textual theory to questions of *mise-en-page* and navigation: what is an Edition? What kinds of objects can it contain? How should it be used? Must it be critical? Must it have a reading text? How should it be organized and displayed? Can intellectual responsibility be shared among editors and users? Can it be shared across generations of editors and users? While some of these questions clearly are related to earlier debates in print theory and practice, others involve aspects of the production of editions not relevant to or largely taken for granted by previous generations of print-based editors.

The answers that have developed to these questions at times have involved radical departures from earlier norms (though see Karlsson and Malm, 2004 for an opposing view). The flexibility inherent to the electronic medium, for example, has encouraged editors to produce editions that users can manipulate interactively, displaying or suppressing different types of readings, annotation, and editorial approaches, or even navigate in rudimentary 3D virtual reality. The relatively low production, storage, and publication costs associated with digital publication, similarly, have encouraged the development of the archive as the *de facto* standard of the genre: users of digital editions now expect to have access to all the evidence used by the editors in the construction of their texts (assuming, indeed, that editors actually have provided some kind of mediated text): full-text transcriptions, high-quality facsimiles of all known witnesses, and tools for building alternate views of the underlying data.

Literary and Linguistic Computing, Vol. 24, No. 1, 2009.
doi:10.1093/llc/fqn039

There have been experiments in editing non-textual objects (Reed Kline, 2001; Foys, 2003), in producing image-based editions of textual objects (Kiernan, 1999/2003), and in recreating digitally the aspects of the sensual experience users might have had in consulting the original objects (British Library; for a discussion, see O'Donnell, 2005b). There have been editions that radically decenter the reading text (e.g. Robinson, 1996), and editions that force users to consult their material using an editorially imposed conceit (Reed Kline, 2001). Even elements carried over from traditional print practice have come in for experimentation and redesign: the representation of annotation, glossaries, or textual variation, for example, are rarely the same in any two electronic editions, even in editions published by the same press (O'Donnell, 2005b, § 5).[1]

Much of the impetus behind this theoretical and practical experimentation has come from developments in the wider field of textual and editorial scholarship, particularly work of the book historians, new philologists, and social textual critics who came into prominence in the decade preceding the publication of the earliest modern digital editorial projects (e.g. McGann, 1983/1992; McKenzie, 1984/1999; Cerquiglini, 1989; Nicols, 1990; for a review see Greetham, 1994, pp. 339–343). Despite significant differences in emphasis and detail, these approaches are united by two main characteristics: a broad interest in the editorial representation of variance as a fundamental feature of textual production, transmission, and reception; and opposition to earlier, intentionalist, approaches that privileged the reconstruction of a hypothetical, usually single, authorial text over the many actual texts used and developed by historical authors, scribes, publishers, readers, and scholars. Working largely before the revolution in Humanities Computing brought on by the development of structural markup languages and popularity of the Internet (O'Donnell, 2004), these scholars nevertheless often expressed themselves in technological terms, calling for changes in the way editions were printed and organized (see, for example, the call for a loose leaf edition of Chaucer in Pearsall, 1985) or pointing to the then largely incipient promise of the new digital media for representing texts as multi-forms (e.g. McGann, 1995; Shillingsburg, 1996).

A second, complementary, impetus for this experimentation has been the sense that the digital editorial practice is, or ought to be, fundamentally different from and even opposed to that of print. This view is found to a greater or lesser extent in both early speculative accounts of the coming revolution (e.g. McGann, 1995; the essays collected in Finneran, 1996 and Landow and Delaney, 1993) and subsequent, more sober and experienced discussions of whether digital practice has lived up to its initial promise (e.g. Robinson, 2004, 2005, 2006; Karlsson and Malm, 2004). It is characterized both by a sense that many intellectual conventions found in print editions are at their root primarily technological in origin, and that the new digital media offer what is in effect a *tabula rasa* upon which digital editors can develop new and better editorial approaches and conventions to accommodate the problems raised by textual theorists of the 1980s and 1990s.

Of course in some cases, this sense that digital practice is different from print is justified. Organizational models, such as the Intellectual Commons or Wiki have no easy equivalent in print publication. Technological advances in our ability to produce, manipulate, and store images cheaply, likewise, have significantly changed what editors and users expect editions to tell them about the primary sources. The ability to present research interactively has opened up rhetorical possibilities for the representation of textual scholarship difficult or impossible to realize in the printed codex.

But the sense that digital practice is fundamentally different from print has been also at times more reactionary than revolutionary. If digital theorists have been quick to recognize the ways in which some aspects of print editorial theory and practice have been influenced by the technological limitations of the printed page, they have been also at times too quick to see other, more intellectually significant aspects of print practice as technological quirks. Textual criticism in its modern form has a history that is now nearly 450 years old (Greetham, 1994, p. 313); in its general desire to produce 'better' texts (however 'better' is defined at the moment in question), it is 'the most ancient of

scholarly activities in the West' with a history stretching back to the end of the sixth century BCE (Greetham, 1994, p. 297). The development of the critical edition over this period has been as much an intellectual as a technological process. While the limitations of the printed page have undoubtedly dictated the form of many features of the traditional critical edition, centuries of refinement—by trial-and-error as well as outright invention—also have produced conventions that transcend the specific medium for which they were developed. In such cases, digital editors may be able to improve upon these conventions by recognizing the (often unexpressed) underlying theory and taking advantage of the superior flexibility and interactivity of the digital medium to improve their representation. But in order to improve on the work of our print predecessors, we need first to understand what they were up to.

2 Expert knowledge and the critical text

Perhaps no area of traditional print editorial practice has come in for more practical and theoretical criticism than the provision of synthetic, stereotypically eclectic, reading texts.[2] Of course this criticism is not solely the result of developments in the digital medium: suspicion of claims to definitiveness and privilege is, after all, perhaps the most characteristic feature of post-structuralist literary theory. It is the case, however, that digital editors have taken to avoiding the critical text with a gusto that far outstrips that of their print colleagues. It is still not unusual to find a print edition with some kind of critical text; the provision of similarly critical texts in digital editions is far more unusual. While most digital projects do provide some kind of top-level reading text, few make any strong claims about this text's definitiveness. More commonly, as in the early, ground breaking editions of the *Canterbury Tales Project* (CTP), the intention of the guide text is, at best, to provide readers with some way of organizing the diversity, without any direct claim to authority (Robinson n.d.):

> We began . . . work [on the CTP] with the intention of trying to recreate a better reading text of the *Canterbury Tales*. As the work progressed, our aims have changed. Rather than trying to create a better reading text, we now see our aim as helping readers to read these many texts. Thus from what we provide, readers can read the transcripts, examine the manuscripts behind the transcripts, see what different readings are available at any one word, and determine the significance of a particular reading occurring in a particular group of manuscripts. Perhaps this aim is less grand than making a definitive text; but it may also be more useful.

There are some exceptions to this general tendency—both in the form of digital editions that are focused around the provision of editorially mediated critical texts (e.g. McGillivray, 1997) and projects, such as the Piers Plowman Electronic Archive (PPEA) that hope ultimately to derive such texts from material collected in their archives. But even here, I think it is fair to say that the provision of a synthetic, critical, text is not what most digital editors consider to be the really interesting thing about their projects. What distinguishes the computer from the codex and makes digital editing such an exciting enterprise is precisely the ability the new medium gives us for collecting, cataloguing, and navigating massive amounts of *raw* information: transcriptions of *every* witness, collations of *every* textual difference, facsimiles of *every* page of *every* primary source. Even when the ultimate goal is the production of a critically mediated text, the ability to archive remains distracting.[3]

In some areas of study, this emphasis on collection over synthesis is perhaps not a bad thing. Texts like *Piers Plowman*, the *Canterbury Tales*, and, in print, *Hamlet* and *Het achterhuis* (The Diary of Anne Frank) have such complex textual histories that they rarely have been archived in any form useful to the average scholar; in such cases, indeed, the historical tendency—seen from our post-structuralist perspective—has been towards *over*-synthesis. In these cases, the most popular previous print editions were put together by editors with strong ideas about the nature of the textual history and/or authorial intentions of the works in question. Their textual histories, too, have

tended to be too complex for easy presentation in print format (e.g. Manly and Rickert, 1940). Readers with only a passing interest in the textual history of these texts have been encouraged implicitly or explicitly to leave the question in the hands of experts.

The area in which I work—Old English textual studies—has not suffered from this tendency in recent memory. Editions of Old English texts historically have tended to be under- rather than over-determined—even in print. In most cases, this is excused by the paucity of surviving witnesses. Most Old English poems (about 97% of the known canon), for example, survive in unique manuscripts (Sisam, 1953; Jabbour, 1968; O'Donnell, 1996a). Even when there is more primary material, Anglo-Saxon editors work in a culture that resists attempts at textual synthesis or interpretation, preferring parallel-text or single-witness manuscript editions whenever feasible, and limiting editorial interpretation to the expansion of abbreviations, word division, and metrical layout—or, in a student edition, perhaps, the normalization of certain unusual linguistic and orthographic features. One result of this is that print practice in Anglo-Saxon studies over the last century or so has anticipated to a great extent many of the aspects that in other periods distinguish digital editions from their print predecessors.

The scholarly history of *Cædmon's Hymn*, a text I have recently edited for the Society of Early English and Norse Electronic Texts series (O'Donnell, 2005a), is a perfect example of how this tendency manifests itself in Old English studies. *Cædmon's Hymn* is the most textually complicated poem of the Anglo-Saxon period, and, for a variety of historical, literary, and scholarly reasons, among the most important: it is probably the first recorded example of sustained poetry in any Germanic language; it is the only Old English poem for which any detailed account of its contemporary reception survives; it is found in four recensions and twenty-one medieval manuscripts—a textual history which can be matched in numbers, but not complexity, by only one other vernacular Anglo-Saxon poem, *Bede's Death Song* (the most recent discussion of these issues is O'Donnell, 2005a).

The poem has also been well studied. Since the 1930s, published semi-diplomatic transcriptions have existed of all known witnesses (Dobbie, 1937).[4] Facsimiles of the earliest manuscripts of the poem (dating from the mid-eighth century) have been available from various sources since the beginning of the twentieth century (e.g. Dobiache-Rojdestvensky, 1928), and were supplemented in the early 1990s by a complete collection of high-quality black-and-white photos collected in Fred C. Robinson and E.G. Stanley's *Old English Poems from Many Sources* (1991). Articles and books on the poem's transmission and textual history have appeared quite regularly for over 100 years and the poem has been at the centre of most debates about the nature of textual transmission in Anglo-Saxon England since at least the 1950s. Taken together, the result of this activity has been the production of what, in everything except medium of production and distribution, can be understood as a fairly *avant garde* digital edition *avant la lettre*: a lightly mediated, witness- and facsimile-based archive that, constructed over a number of generations by independent groups of scholars, even anticipates several recent calls for the development of a new model of collective, multi-project and multi-generational digital editorial scholarship (e.g. Ore, 2004; Robinson, 2005).

The poem's print scholarly history anticipates contemporary digital practice in another way as well: until recently, *Cædmon's Hymn* had never been the subject of a modern critical textual edition. The last century has seen the publication of a couple of student editions of the poem and some specialized reconstructions of one of the more corrupt recensions (Wuest 1906; Smith, 1938/1978; O'Donnell, 1996b; Cavill, 2000)—but no critical works that have attempted to encapsulate and transmit in textual form what is actually known about the poem's transmission and recensional history. The closest thing to a standard edition for most of the last 100 years, two versions of the poem edited by Elliot Van Kirk Dobbie for the Anglo-Saxon Poetic Records (ASPR) (Dobbie, 1942), actually ignored the editor's own earlier work on the poem's textual history. Thus, in his earlier, never renounced, study of the poem, *The Manuscripts of*

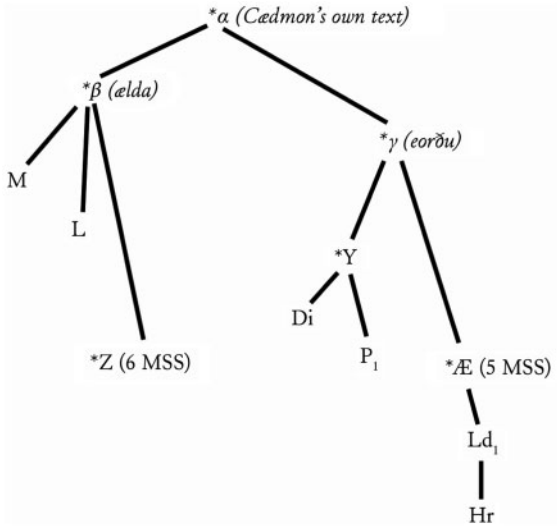

Fig. 1 Reproduction of stemma of Cædmon's Hymn from Dobbie (1937)

Cædmon's Hymn (Dobbie, 1937), Dobbie argued that the *Hymn* developed along something like the following lines, with an early Northumbrian division between versions of the poem reading *aelda* or *eordu* in line 5b that was subsequently reproduced in later West-Saxon manuscripts (Fig. 1).[5]

In his actual edition (Dobbie, 1942), however, Dobbie reproduced two hypothetical reconstructions using witnesses grouped by dialect: a Northumbrian text based on the combined readings of manuscripts M, L, Di, and P_1, and a West-Saxon text based on readings from witnesses to the *Z and *Æ recensions and the manuscripts Ld_1 and Hr. In doing so, Dobbie created texts substantively different from those found in any surviving witness and obscured the readings he himself considered to be fundamental to the poem's recensional division.

The problem with this approach—and with the history of *Cædmon's Hymn* in Anglo-Saxon scholarship—should be clear enough. On the one hand the poem's textual history is, by Anglo-Saxon standards, quite complex and the subject of intense debate by professional textual scholars. On the other, the failure until recently to provide any kind of critical text representing the various positions in the debate has all but hidden the significance of this research—and its implications for work on other aspects of the *Hymn*—from the general reader. Instead of being able to take advantage of the expert knowledge acquired by editors and textual scholars of the poem over the last 100 years of modern *Cædmon's Hymn* scholarship, readers instead have been forced either to go back to the raw materials and construct their own texts over and over again or rely on a standard edition that misrepresents its own editor's considered views of the poem's textual history.

This is not an efficient use of these readers' time. As Kevin Kiernan has argued, the textual history of *Cædmon's Hymn* is not a spectacle for casual observers (Kiernan, 1990), and most people who come to study *Cædmon's Hymn* are not interested in collating transcriptions, deciphering facsimiles, and weighing options for grouping the surviving witnesses: they want to study its sources and analogues, its composition and reception, its prosody, language, place in the canon, significance in the development of Anglo-Saxon Christianity, or usefulness as an index in discussions of the position of women in Anglo-Saxon society—in other words, all the *other* things we do with texts when we are not studying their transmission. What these readers want—and certainly what I want when I consult an edition of a work I am studying for reasons other than its textual history—is a text that is accurate, readable, and hopefully based on clearly defined and well-explained criteria. They want, in other words, to be able to take advantage of the expert knowledge of those responsible for putting together the text they are consulting. If they don't like what they see, or if the approach taken is not what they need for their research, then they may try to find an edition that is better suited to their particular needs. But they will not—except in extreme cases I suspect—actually want to *duplicate* the effort required to put together a top-quality edition.

3 Print convention and database theory

The failure of the print editors of *Cædmon's Hymn* over the last 100 years to provide a critical editorial

account of their actual knowledge of their poem is very much the exception that proves the rule: for the provision of this kind of expert knowledge is an area in which digital editors generally are much poorer than their print predecessors. This is ironic both because the encoding and dissemination of expert knowledge is in other fields of computer science a principal preoccupation, and because traditional print editorial conventions already provide a robust and sophisticated model for how this knowledge could be organized. As digital editors, we can improve on this practice both because it has been developed in the past for the most part by trial-and-error rather than explicit intellectual modelling, and because computers actually are better suited than codices for implementing many aspects of the print model. In coming to define our work for a large part in reaction to print editions, however, we have failed to recognize some of the most promising areas opened for us by the new media. Digital editing as a discipline will only truly come of age when it recognizes the solidity of the foundation it has been given to build upon.

Perhaps the most important thing to recognize about traditional print editorial practice is its intellectual efficiency. At a conceptual level, print editorial conventions developed over the last several 100 years form an almost textbook example of the parsimonious organization of information about texts and witnesses. While there are some technological and conventional limitations to the way this information is used and presented in its codex form, it is difficult if not impossible to come up with a theoretically more flexible and efficient rhetorical organization.

Demonstrating this requires us to make a brief excursion into questions of database theory and design. Most contemporary databases are built on a relational data model.[6] The goal of relational database design is to generate a set of relationship schemas that allows us to store information without unnecessary redundancy, yet also allows us to retrieve information easily (Silberschatz *et al.* 2006, p. 263). The relational model organizes information into 2D tables, each row of which represents a relationship among associated bits of information. Complex data commonly require the use of more than one set of relations or tables. The goal is to avoid complex redundancies: in a well-designed relational database, no piece of information that logically follows from any other appears more than once.[7]

The process used to eliminate redundancies and dependencies in relational database design is known as normalization. When data has been organized so that it is free of all such inefficiencies, it is usually in third normal form. How this works can be best seen through an example. The following invoice is from a hypothetical book store (adapted from Krishna, 1992, p. 32):

Invoice Number	JJSmith0001
Customer ID:	JJS01
Name:	Jane J. Smith
Address:	323 Fifteenth Street S., Lethbridge, Alberta T1K 5X3.

ISBN	Author	Title	Price	Quantity	Item Total
0-670-03151-8	Pinker, Stephen	The Blank Slate: The Modern Denial of Human Nature	$35.00	1	$35.00
0-8122-3745-5	Burrus, Virginia	The Sex Lives of Saints: An Erotics of Ancient Hagiography	$25.00	2	$50.00
0-7136-0389-5	Dix, Dom Gregory	The Shape of the Liturgy	$55.00	1	$55.00
Grand total					$140.00

Describing this information in relational terms is in this case a three-step process. The first step involves identifying what it is that is to be included in the data model by providing explicit names for information implicit in the document's structure (parentheses are used to indicate information that can occur more than once on the invoice):

> *Invoice*: invoice_number, customer_id, customer_name, customer_address, (ISBN, author, title, price, quantity, item_total), grand_total

Repeating information in this document about the actual books sold (ISBN, author, title, price, quantity, item_total) is then removed and placed in a second table, whose connection to the first is indicated by the value of the invoice_number key:

Invoice: invoice_number, customer_id, customer_name, customer_address, grand_total
Invoice_Item: invoice_number, ISBN, author, title, price, quantity, item_total

The final step involves removing functional dependencies and redundancies within these two tables. In this database, for example, information about a book's author, title and item_price are functionally dependent on its ISBN: for each ISBN, there is only one possible author, title, and item_price. Likewise customer_id is associated with only one customer_name and customer_address. These dependencies are eliminated by placing the dependent material in two new tables, *Customer* and *Book*, which are linked to rest of the data by the customer_id and ISBN keys. This leaves us with four sets of relations, none of which can be broken down any further, at which point, the data are said to be in *third normal form*:

Invoice: invoice_number, customer_id, grand_total
Invoice_Item: invoice_number, ISBN, quantity, item_total
Customer: customer_id, customer_name, customer_address
Book: ISBN, author, title, price

The normalization process becomes interesting when one applies it to the type of information editors commonly collect about textual witnesses. The following, for example, is a simplified version of a sheet I used to record basic information about each manuscript witness to *Cædmon's Hymn*:

The sheet has what are essentially fields for the manuscript sigil, date, scribe, location, and, of course, the text itself—something that can be seen, on analogy with our book store invoice, as a repeating set of categories: lines, words, morphemes, and the like. Additional information used in editing medieval texts, though not typically included on an informal sheet like this, includes the relationship between individual readings in this manuscript to some kind of canonical reference system, glossary and grammatical information, and collations.

As with our hypothetical invoice, it is possible to place this data in normal form. The first step, once again, is to extract the relevant relations from the manuscript sheet and, in this case, the unstated, expert knowledge an editor typically brings to his or her task. This leads at the very least to the following set of relations:[8]

Manuscript: shelf_mark, date, scribe, location, (ms_instance, canonical_reading, dictionary_form, grammatical_information, translation)

Extracting the repeating information about individual readings, leaves us with two tables linked by the key shelf_mark:

Manuscript: shelf_mark, date, scribe, location
Text: shelf_mark, ms_instance, canonical_reading, dictionary_form, grammatical_information, translation

And, placing the material in third normal form, at least one more:

Manuscript: shelf_mark, date, scribe, location
Text: shelf_mark, ms_instance, canonical_reading
Glossary: canonical_reading, dictionary_form, grammatical_information, translation

The significant thing about this process is the resemblance the final normalized arrangement

Shelf-Mark:	B1 Cambridge, Corpus Christi College 41
Date:	s. xi-1
Scribe:	Second scribe of the main Old English text.
Location:	Copied as part of the main text of the Old English translation of the *Historia ecclesiastica* (p. 332 [f. 161v]. line 6)
Recension:	West-Saxon *eorðan* recension
Text:	Nuweherigan sculon heofonrices weard metodes mihte &hismod ge Þanc weorc wuldor godes [etc]

bears to the stereotypical structure of a traditional print-based critical edition: a section with bibliographic (and other) information about the text and associated witnesses, a section relating manuscript readings to editorially privileged forms, and a section containing abstract lexical and grammatical information about words in the abstract text. Moreover, although familiarity and the use of narrative descriptions can obscure this fact in practice, much of the information contained in these sections in traditional print editions actually is tabular in form: in structural terms, glossaries are best understood as highly structured lists in which each item contains similar material in a similar order: lemma, grammatical information, translation, instances; not surprisingly dictionaries were among the first genres to be modelled in structural markup languages like (S)GML (see Elliot, 2000 for the *Oxford English Dictionary*). Bibliographical discussions, too, often consist of what are in effect, structurally organized lists: shelf-mark, bibliography, provenance, etc.[9]

4 The critical text as database view

There is one place, however, in which this presentation of the data model underlying traditional print practice is misleading: the representation of the critical text itself. For while it is possible to see the how the other sections of a print critical edition (including the *apparatus variorum*) might be rendered in tabular form, the critical text itself—the place where editors present the synthetic reading text that is the result of their editorial efforts—is never presented in anything resembling the non-hierarchical, tabular form a relational model would lead us to expect. Indeed the essential point of the editorial text—and the source of post-structural criticism of the approach—is the elimination of non-hierarchical choice: in constructing synthetic texts, print editors impose order on the textual evidence by privileging a single reading at each collation point in their editions, relegating the rest—and even then only a sample—to appearance in small type at the bottom of the page in the critical apparatus. Although it is the defining feature of the print critical edition, the critical text itself would appear to be the only part that is not directly part of the underlying, and extremely efficient, data model developed through the centuries.

This does not invalidate my larger argument, however. Because we build databases precisely in order to acquire this ability to select and organize raw data. If the critical text in a print edition is not actually a database itself, it is a database *view*—that is to say a 'window on the database through which data required for a particular user or application can be accessed' (Krishna, 1992, p. 210). In computer database management systems, views are built by querying the underlying data and building new relations containing one or more answers from the results. In print editorial practice, the critical text is built by the editor according to more or less explicit selection criteria and processing instructions that are designed to produce a single value for each query. Hence, a typical student edition of a medieval or classical text might be understood as a selection of readings that match paradigmatic forms in the standard grammars and reflect statistically common syntactic patterns. A modern spelling edition of Shakespeare is a view built on some underlying principle that has been processed to replace Renaissance spellings with their modern equivalents. An edition like the Kane-Donaldson *Piers Plowman* is built on a far more complex set of selection criteria and processing based on the editors' research into Middle English scribal practice. Editorial emendations are, in this sense, simply an additional column in the unstated, underlying table: editors build them as alternatives to readings from surviving witnesses and select them whenever no other reading in the underlying database meets the selection requirements.[10]

5 Improving on print practice

If this understanding of the critical text and its relationship to the data model underlying print critical practice is correct, then one obvious place where digital editors perhaps could improve on print conventions might seem to lie in formalizing and

automating the process by which print editors process and query the data upon which their editions are based. Such an approach, indeed, would have two main advantages: it would allow us to test others' editorial approaches by modelling them programmatically; and it would allows us to take advantage of the inherent flexibility of the digital medium by providing users with access to limitless critical texts of the same work. Where, for economic and technological reasons, print editions tend to offer readers only a single critical approach and text, digital editions could now offer readings a series of possible approaches and texts built according to various selection criteria. In this approach, users would read texts either by building their own textual queries, or by selecting pre-made queries that build views by dynamically modelling the decisions of others—a Kane-Donaldson view of *Pier Plowman*, perhaps, or a Gabler reading text view of *Ulysses*.

In actual practice, we are a long way from being able to build anything but the simplest of texts in this manner. Certain processes can, of course, be automated and even improved upon electronically—we can collate readings from different witnesses, derive manuscript stemma, automatically normalize punctuation and spelling, even model scribal performance (Ciula, 2005; O'Donnell, 2005c). And it is easy to see how it we might be able to build databases and queries so that we can model human editorial decisions in relatively simple cases—reproducing the flawed dialectal texts of *Cædmon's Hymn* discussed above, perhaps, or building simple student editions of small poems.

But such conceptually simple tasks are at the extreme outer limits of what it is currently possible, let alone economically reasonable, to do. Going beyond this—learning to automate higher level critical decisions involving cultural, historical, or literary distinctions—is beyond the realm of current database design and artificial intelligence even for people working in fields vastly better funded than textual scholarship. Thus, while it would be a fairly trivial process to generate a reading text based on a single witness from an underlying relational database, building automatically a best text edition—that is to say, an edition in which a single witness is singled out automatically for reproduction on the basis of some higher level criteria—is beyond our current capabilities. Other distinctions of the type made every day by human editors—distinguishing between good and bad scribes, assessing *difficilior* versus *facilior* readings, or weighing competing evidence of authorial authorization—belong to the realm of science fiction.[11]

This suggests on the one hand that we are far away from being able to truly automate our digital textual editions, and on the other that we need to find some way of incorporating expert knowledge into the editions that we currently can build that is commensurate with their complexity. The more evidence we cram into our digital editions, the harder it becomes for uninitiated readers to make anything of them. No two witnesses to any text are equally reliable, authentic, or useful for all purposes at all times. In the absence of a system that can build custom editions in response to naïve queries—'build me a general interest text of Don Juan', 'eliminate unreliable scribes', or even 'build me a student edition'—editors will need to provide readers with explicit expert guidance as to how the at times conflicting data in their editions is to be assessed. In some cases, it is possible to use hierarchical and object-oriented data models to encode these human judgements so that they can be generated dynamically (see note 11 above). In other cases, digital editors, like their print predecessors, will simply have to build critical texts of their editions the old fashioned way—by hand—or run the risk or failing to pass on the expert knowledge they have built up over years of scholarly engagement with the primary sources.

It is here, moreover, that digital editors can improve theoretically and practically the most on traditional print practice. For if critical reading texts are, conceptually understood, the equivalent of query-derived database views, then there is no reason why readers of critical editions should not be able to entertain multiple views of the underlying data. Critical texts, in other words—as post-structuralist theory has told us all along—really are neither right nor wrong: they are simply views of a textual history constructed according to different,

more or less explicit, selection criteria. In the print world, economic necessity and technological rigidity imposed constraints on the number of different views editors could reasonably present to their readers—and often encouraged them to see the production of a single definitive critical text as the primary purpose of their editions. Digital editors, on the other hand, have the advantage of a medium that allows the inclusion much more easily of multiple critical views, a technology in which the relationship between views and data is widely known and accepted, and a theoretical climate that encourages an attention to variance. If we are still far from being at the stage in which we can produce critical views of our data using dynamic searches, we are able even now to hard-code such views into our editions in unobtrusive and user-friendly ways.[12]

6 Conclusion

And so in the end, the future of digital editing may lie more in our past than we commonly like to consider. While digital editorial theory has tended to define its project largely in reaction to previous print practice, this approach both underestimates the strength of the foundation we have been given to build upon, and the true significance of our new medium. For the exciting thing about digital editing is not that it can do everything differently, but rather that it can do some very important things better. Over the course of the last half millennium, print editorial practice has evolved an extremely efficient intellectual model for the organization of information about texts and witnesses. As digital editors, we can greatly improve upon this model both by recognizing and formalizing its intellectual strength, and by implementing it far more fully and flexibly than print editors themselves could ever imagine. The question we need to answer, then, is not whether we can do things differently, but how doing things differently can improve on current practice. But we will not be able to answer this question until we recognize what current practice already does very well.

References

Bart, P. R. (2006). Controlled experimental markup in a TEI-conformant setting. *Digital Medievalist* 2.1. http://www.digitalmedievalist.org/journal/2.1/bart/.

British Library. Turning the Pages. http://www.bl.uk/onlinegallery/ttp/ttpbooks.html.

Cavill, P. (2000). The manuscripts of Cædmon's Hymn. *Anglia*, **118**: 499–530.

Cerquiglini, B. (1989). *Éloge de la variante: Histoire critique de la philologie*. Paris, Éditions de Seuil.

Ciula, A. (2005). Digital palaeography: using the digital representation of medieval script to support palaeographic analysis. *Digital Medievalist* 1.1 http://www.digitalmedievalist.org/article.cfm?RecID=2.

Dobbie, E. V. K. (1937). *The Manuscripts of Cædmon's Hymn and Bede's Death Song with a Critical Text of the Epistola Cuthberti de obitu Bedæ*. Columbia University Studies in English and Comparative Literature, 128. New York, Columbia University Press.

Dobbie, E. V. K. (ed.) (1942). *The Anglo-Saxon Minor Poems*. The Anglo-Saxon Poetic Records, a Collective Edition, 6. New York: Columbia University Press.

Dobiache-Rojdestvensky, O. (1928). Un manuscrit de Bède à Léningrad. *Speculum*, **3**: 314–21.

Elliot, L. (2000). How the *Oxford English Dictionary* went online. *Ariadne* 24. http://www.ariadne.ac.uk/issue24/oed-tech.

Finneran, R. J. (ed.) (1996). *The Literary Text in the Digital Age*. Ann Arbor: University of Michigan Press.

Foys, M. K. (ed.) (2003). *The Bayeux Tapestry: Digital Edition*. Leicester: SDE.

Greetham, D. C. (1994). *Textual Scholarship*. New York, Garland.

Jabbour, A. A. (1968). *The Memorial Transmission of Old English Poetry: A Study of the Extant Parallel Texts*. Unpublished PhD dissertation, Duke University.

Karlsson, L. and Malm, L. (2004). Revolution or remediation? A study of electronic scholarly editions on the web. *HumanIT*, **7**: 1–46.

Kiernan, K. S. (1990). Reading Cædmon's Hymn with someone else's glosses. *Representations*, **32**: 157–74.

Kiernan, K. S. (ed.) (1999/2003). *The Electronic Beowulf*, 2nd edn. London: British Library.

Krishna, S. (1992). *Introduction to Database and Knowledge-base Systems*. Singapore, World Scientific.

Landow, G. P. and Delaney, P. (eds) (1993). *The Digital Word: Text-based Computing in the Humanities*. Cambridge, MA, MIT Press.

Manly, J. M. and Rickert, E. (1940). *The Text of the Canterbury Tales*. Chicago, University of Chicago Press.

McGann, J. J. (1983/1992). *A Critique of Modern Textual Criticism*. Charlottesville: University of Virginia Press.

McGann, J. J. (1995). Rationale of the Hypertext. http://www/iath.virginia.edu/public/jjm2f/rationale.htm

McGillivray, M. (ed.) (1997). *Geoffrey Chaucer's Book of the Duchess: A Hypertext Edition*. Calgary, University of Calgary Press.

McKenzie, D. F. (1984/1999). *Bibliography and the Sociology of Texts*. Cambridge: Cambridge University Press.

Nicols, S. G. Jr (ed.) (1990). *Speculum* 65.

O'Donnell, D. P. (1996a). *Manuscript Variation in Multiple-Recension Old English Poetic Texts: The Technical Problem and Poetical Art*. Unpublished PhD dissertation, Yale University.

O'Donnell, D. P. (1996b). A Northumbrian version of 'Cædmon's Hymn' (eordu recension) in Brussels, Bibliothèque Royale MS 8245-57 ff. 62r2-v1: Identification, edition, and filiation. In Houwen, L. A. J. R. and MacDonald, A. A. (eds), *Beda Venerabilis: Historian, Monk and Northumbrian*. Mediaevalia Groningana, 19. Groningen: Egbert Forsten, pp.139–65.

O'Donnell, D. P. (2005a). *Cædmon's Hymn: A Multimedia Study, Edition, and Archive*. SEENET A.8. Cambridge, D.S. Brewer.

O'Donnell, D. P. (2005b). O Captain! My Captain! Using Technology to Guide Readers Through an Electronic Edition. *Heroic Age* 8. http://www.heroicage.org/issues/8/em.html

O'Donnell, D. P. (2005c). The ghost in the machine: revisiting an old model for the dynamic generation of digital editions. *HumanIT*, **8** (2005): 51–71.

Ore, E. S. (2004). Monkey business—or what is an edition? *Literary and Linguistic Computing*, **19**: 35–44.

Pearsall, D. (1985). Editing medieval texts. In McGann, J. J. (ed.), *Textual Criticism and Literary Interpretation*. Chicago, U Chicago, pp. 92–106.

Reed Kline, N. (ed.) (2001). *A Wheel of Memory: The Hereford Mappamundi*. Ann Arbor, University of Michigan Press.

Robinson, F. C. and Stanley, E. G. (eds) (1991). *Old English Verse Texts from Many Sources: A Comprehensive Collection*. Early English Manuscripts in Facsimile, 23. Copenhagen, Rosenkilde & Bagger.

Robinson, P. (nd). New Methods of Editing, Exploring, and Reading the *Canterbury Tales*. http://www.cta.dmu.ac.uk/projects/ctp/desc2.html.

Robinson, P. (ed.) (1996). *The Wife of Bath's Prologue on CD-ROM*. Cambridge, Cambridge University Press.

Robinson, P. (2004). Where are We with Electronic Scholarly Editions, and Where to We Want to be? *Jahrbuch für Computerphilologie* Online at <http://computerphilologie.uni-muenchen.de/ejournal.html>. Also available in print: *Jahrbuch für Computerphilologie*. pp. 123–143.

Robinson, P. (2005). Current issues in making digital editions of medieval texts—or, do electronic scholarly editions have a future? *Digital Medievalist* 1.1 http://www.digitalmedievalist.org/article.cfm?RecID=6.

Robinson, P. (2006). The Canterbury Tales and other medieval texts. In Unsworth, J., Burnard, L. and Katherine O'Brian O'Keefe (eds), *Electronic Textual Editing*. New York: Modern Language Association of America.

Shillingsburg, P. L. (1996). Electronic editions. *Scholarly Editing in the Computer Age: Theory and Practice*, 3rd edn.

Silberschatz, A., Korth, H., and Sudarshan, S. (2006). *Database System Concepts*. New York, McGraw-Hill.

Sisam, K. (1953). *Studies in the History of Old English Literature*. Oxford, Clarendon Press.

Smith, A. H. (ed.) (1938/1978). *Three Northumbrian Poems: Cædmon's Hymn, Bede's Death Song, and the Leiden Riddle*. With a bibliography compiled by M. J. Swanton. Revised edn *Exeter Medieval English Texts*. Exeter: University of Exeter Press.

Sutherland, K. (ed.) (1997). *Electronic Text: Investigations in Method and Theory*. Oxford: Clarendon.

Wuest, P. (1906). Zwei neue Handschriften von Cædmons Hymnus. *ZfdA*, **48**: 205–26.

Notes

1 As this list suggests, my primary experience with actual practice is with digital editions of medieval texts. Recent theoretical and practical discussions, however, suggest that little difference is to be found in electronic texts covering other periods.

2 *Synthetic* here is not quite synonymous with eclectic as used to describe the approach of the Gregg-Bower's school of textual criticism. Traditionally, an eclectic text is a single, hypothetical, textual reconstruction (usually of the presumed Authorial text) based on assumption of divided authority. In this approach, a copy text is used to supply accidental details of spelling and punctuation and (usually) to serve as a default source for substantive readings that affect the meaning of the abstract artistic work. Readings from this copy text are then corrected by emendation or, preferably, from forms found in other historical witnesses. In this essay, *synthetic* is used to refer to a critical text that attempts to summarize in textual form an editorial position about an abstract work's development at some point in its textual history. All eclectic texts are therefore synthetic, but not all synthetic texts are eclectic: a best text (single witness) edition is also synthetic if, as the name implies, an editorial claim is being made about the particular reliability, historical importance, or interest of the text as represented in the chosen witness—as opposed to a diplomatic transcription, in which the focus is on reporting the details of a given witness as accurately as possible.

3 It is indeed significant that the PPEA—the most ambitious digital critical edition of a medieval text that I am aware of—is at this stage in its development publishing primarily as an *archive*: the development of critical texts of the A-, B-, and C-text traditions has been deferred until after the publication of individual edition/facsimiles of the known witnesses (Bart, 2006).

4 Transcriptions, editions, facsimiles, and studies mentioned in this paragraph in many cases have been superseded by subsequent work; readers interested in the current state of *Cædmon's Hymn* should begin with the bibliography in O'Donnell (2005a).

5 While there is reason to doubt the details of Dobbie's recensional division, his fundamental conclusion, that dialect did not play a crucial role in the poem's textual development remains undisputed. For recent (competing) discussions of the *Hymn's* transmission, see O'Donnell (2005a) and Cavill (2000).

6 There are other types of databases, some of which are at times more suited to representation of information encoded in structural markup languages such as XML, and to the type of manipulation common in textual critical studies (see below, note 11). None of these other models, however, expresses information as parsimoniously as does the relational model (Silberschatz *et al*. 2006, p. 362–5).

7 This is a rough rather than a formal definition. Formally, a well-designed relational database normally should be in either third normal form or Boyce–Codd normal form (BCNF). A relation is said to be in third normal form when (1) the domains of all attributes are atomic, (2) all non-key attributes are fully dependent on the key attributes (Krishna, 1992, p. 37). A relation is said to be in BCNF if whenever a non-trivial functional dependency $\rightarrow$ A holds in R, X is a superkey for R (Krishna, 1992, p. 38). Other normal forms exist for special kinds of dependencies (Silbertschatz *et al*., 2006, p. 293–98).

8 In actual practice, the model would be far more complex and include multiple levels of repeating information (words within lines, for example). This example also assumes that the word is the basic unit of collation; while this works well for most Old English poetry, it may not for other types of literature.

9 Of course, critical editions typically contain far more than bibliographic, textual, and lexical/grammatical information. This too can be modelled relationally, however, although it would be quixotic to attempt to account for the infinite range of possible material one might include in a critical edition in this essay. Thus cultural information about a given text or witnesses is functionally dependent on the specific text or witness in question. Interestingly, the more complex the argumentation becomes, the less complex the underlying data model appears to be: a biographical essay on a text's author, for example, might take up but a single cell in one of our hypothetical tables.

10 The critical apparatus in most print and many digital editions is itself also usually a view of an implicit textual database, rather than the database itself. Although it usually is presented in quasi-tabular form, it rarely contains a complete accounting for every form in the text's witness base.

11 This is not to say that it is *impossible* to use data modelling to account for these distinctions—simply that we are far from being able to derive them arbitrarily from 2D relational databases, however complex. Other data models, such as hierarchical or object-oriented databases can be used to build such distinctions into the data itself, though this by definition involves the application of expert knowledge. In O'Donnell (2005a), for example, the textual apparatus as a hierarchical database. This allows readers to in effect query the database, searching for relations predefined as significant, substantive, or orthographic by the editor. See O'Donnell (2005a, §§ ii.7, ii.19, 7.2–9).

12 In the case of my edition of *Cædmon's Hymn*, this takes the form of multiple critical texts and apparatus: several reconstructions of the poem's archetypal form, and various critical views of the poem's five main

recensions and collations. The criteria used to construct these views is indicated explicitly in the title of each page and explained in detail in the editorial introductions. The individual editions were extracted from an SGML-encoded text using style sheets—in essence hard-wired database queries reflecting higher level editorial decisions—but presented to the reader as a series of progressively abstract views. In keeping with the developing standard for digital textual editions, the edition also allows users direct access to the underlying transcriptions and facsimiles upon which it is based. The result is an edition that attempts to combine the best of the digital and print worlds: the archiving function common to most electronic editions (and traditionally the focus of *Cædmon's Hymn* textual research in print), with the emphasis on the presentation of expert knowledge characteristic of traditional print editorial practice.