

2007

Building a Spenser Archive - One Scan at a Time

David Lee Miller

University of South Carolina - Columbia

Follow this and additional works at: https://scholarcommons.sc.edu/engl_facpub



Part of the [English Language and Literature Commons](#)

Publication Info

Published in *Duke University Libraries*, ed. B. Ilene Nelson, Volume 20, Issue 2/3, 2007, pages 14-19.
Miller, D. L. (2007). Building a Spenser archive - One scan at a time. *Duke University Libraries*, 20(2/3), 14-19.

© Duke University Libraries, 2007, Duke University

This Article is brought to you by the English Language and Literatures, Department of at Scholar Commons. It has been accepted for inclusion in Faculty Publications by an authorized administrator of Scholar Commons. For more information, please contact digres@mailbox.sc.edu.



Editor's Note: David Lee Miller, professor of English and Comparative Literature at the University of South Carolina, spent several days in February at Duke's Rare Book, Manuscript, and Special Collections Library, examining the Library's 1609 edition of Edmund Spenser's *The Faerie Queene*. Miller was also at Duke to attend a conference, "Producing the Renaissance Text: Current Technologies of Editing—In Theory and Practice." What follows is a slightly revised version of the paper Professor Miller presented at the conference.

THE IDEAL COPY
TEXT FOR ANY
EDITION WOULD
BE ONE CONTAIN-
ING THE FINAL,
CORRECTED STATE
FOR EVERY FORME.
BUT IN THE EARLY
DAYS OF PRINT-
ING, PROOFING
WAS OFTEN DONE
ON THE FLY,
WITH CORRECTED
AND UNCOR-
RECTED SHEETS
COMBINED INDIS-
CRIMINATELY IN
ANY GIVEN COPY
SENT TO THE
BINDER.

BUILDING A SPENSER ARCHIVE—ONE SCAN AT A TIME

DAVID LEE MILLER

In the late 1990s, a team of American researchers persuaded Oxford University Press that the time had come for a new scholarly edition of the works of Edmund Spenser. The players were Joseph Loewenstein (Washington University), Patrick Cheney (Penn State), Elizabeth Fowler (University of Virginia), and me. From the beginning we imagined our goal as a digital archive from which various physical texts might be derived: a hardcover library edition, a classroom text, a paperback of the *View of the Present State of Ireland*, and perhaps others. The matrix from which these books are generated will be an open-access digital archive built to serve everyone from beginning students to the geekiest of bibliographers.

So the first principle I'm here to offer is that in the new age of editing, hard copy texts will be captures from an electronic database. Many things follow from this principle, most of which I can't tell you about because we're learning as we go and the field is changing fast. But here are a few conclusions we've drawn

so far, gathered under six headings: digital copy text, digital collation, hypertext commentary, collaboration, teaching, and the immaterial text.

DIGITAL COPY TEXT

Given what we know about early modern printing practices, there's really no reason for any single copy of a given edition to serve as copy text. The crucial unit of analysis for textual editors is not the book. Nor is it the page. It's the "forme"—that layout of pages set up together and locked within a chase to be printed on a sheet, which will then be folded and cut. The ideal copy text for any edition would be one containing the final, corrected state for every forme. But in the early days of printing, proofing was often done on the fly, with corrected and uncorrected sheets combined indiscriminately in any given copy sent to the binder.

The result is that the ideal copy may or may not exist on a shelf somewhere between one set of covers. Charleton Hinman created a facsimile of the Shakespeare first folio by cherry-picking the images of corrected (and

well-inked) pages from various existing copies; his example takes on a new interest now that we can store high-resolution digital scans of existing copies on a server. Why not follow Hinman's lead by recombining scans to create a virtual copy text consisting entirely of corrected formes?

The biggest obstacle is to get enough copies scanned—the process can be quite expensive—but it does seem reasonable to expect that over time most copies of most early witnesses will be digitized. Our goal for Spenser is to collect TIFF scans of as many copies as we can. This will cost a lot and take a long time, but sooner or later it will happen—and long before it does, we will have witnesses enough to compose our virtual copy text.

DIGITAL COLLATION

This goal of collecting scans will have other advantages as well. One of the purely practical obstacles to editing a book like *The Faerie Queene* has always been the difficulty of collating multiple copies. Over a hundred copies of the 1590 edition are thought to survive, but they are scattered all over the world, and each copy takes three or four days to collate. Until recently no one, not even the editors of the Johns Hopkins Variorum edition, had ever collated more than three or four copies. The team of Japanese scholars who prepared the text for the recent Longman edition were able to collate a dozen, but to do it they had to work from microfilm and photocopies. This method carries inevitable limitations—for instance, it's difficult to recognize where a copy may have been "sophisticated" along the way.

Take, for example, the description of Satyrane in Book I, canto vi of *The Faerie Queene*. All copies in 1590 say that among the beasts he compelled with

iron yokes was the "Wolfe both swift and cruell" (I.vi.26.5). This is a problem because the previous line lists the "Tigre cruell," with both cruells in the rhyming position. Sure enough, in the *Faults Escaped* that accompanies most copies of the 1590 printing, we find that "swift and cruell" should read "fierce and fell." Yamashita et al. list this as a press variant in 1590 because they think that Malone 615, housed in the Bodleian, contains the corrected reading. It's always a good idea to be suspicious of copies that incorporate corrections from the *Faults Escaped* list; I've found other instances in which a copy was "improved" by some earlier owner or seller taking a hint from that source. But you can't tell this sort of thing from microfilm. You have to go into the Bodleian and look at page 85 of Malone 615, in which case you will see that the correction has actually been pasted in over the uncorrected state, which can still be seen if you lift the flap of paper on which the correction has been printed.

Even very high-resolution scans will never completely replace the occasional need for first-hand examination of the physical evidence. They will, however, reduce that need, since they capture so much more data than any other kind of image. And, what may prove most valuable in the long run, they hold out the possibility of making such first-hand examination more efficient by telling us where to look.

Optical character recognition may someday be sophisticated enough to do preliminary collations of early modern books, but unless Google knows something we don't (and they may), that's nowhere near achievable for the present. What OCR can do, though, is identify what counts as a character or as the space between characters. Computer science students working with Joseph Loewenstein and Keith Bennett at

EVEN VERY HIGH-
RESOLUTION
SCANS WILL NEVER
COMPLETELY
REPLACE THE
OCCASIONAL NEED
FOR FIRST-HAND
EXAMINATION OF
THE PHYSICAL
EVIDENCE. THEY
WILL, HOWEVER,
REDUCE THAT
NEED, SINCE THEY
CAPTURE SO MUCH
MORE DATA THAN
ANY OTHER KIND
OF IMAGE.



THE HYPERTEXT ENVIRONMENT NOT ONLY OFFERS A LARGER QUANTITY AND VARIETY OF ANNOTATION AVAILABLE WITH A MOUSE-CLICK, IT ALSO OFFERS THE PROSPECT THAT OUR CONFERENCE ORGANIZERS REFER TO AS THE “CONTINUOUSLY REVISED ONLINE EDITION.”

Washington University have developed a program that works with TIFF files, using OCR to locate characters on the page, but then switching to a direct comparison of pixel patterns to detect significant variation. This program, currently in its beta stage and slated for further testing, is known as “Digicoll.” Digicoll isn’t smart enough to do the collating for us, but it is patient enough to cull through as many copies as we can scan in order to flag discrepancies and say to a human editor, “Here, come have a look at this, will you?”

Operating on a substantial archive of scans, such a program should enable us to collate many more copies than have ever been collated before, and to do it with a higher degree of accuracy.

HYPERTEXT COMMENTARY, OR “OH WHAT AN ENDLESSE WORKE HAVE I IN HAND!”

This topic may quickly provoke the reflection that sometimes limits are a good thing, since they force an editor to be both selective and concise. This is one reason—one of many—that it’s good to have the interplay between digital editions and hard copy derivatives: the economics of the book require distillation where those of the internet solicit a *jouissance* of proliferation. Still, the hypertext environment not only offers a larger quantity and variety of annotation available with a mouse-click, it also offers the prospect that our conference organizers refer to as the “continuously revised online edition.” Such continuous revision needn’t always entail expansion, but it will certainly invite editors to imagine their texts as a set of portals leading into a virtual encyclopedia of contexts and specialized studies. Indeed, if there’s going to

be a *Spenser Encyclopedia*—a superb reference work—why shouldn’t its entries be placed online and linked to a hypertext edition?

Of course that’s only the beginning. Can we get an audiofile of Seamus Heaney reading his favorite passages from *The Faerie Queene*? What about specialized studies of everything Elizabethan, from architecture to zoology? And if Google is going to put the entire public domain online, why shouldn’t we be able to create a digital simulacrum of Spenserian intertextuality, with direct links from a given passage in *The Faerie Queene* to its tributaries in Virgil, Ovid, Chaucer, Ariosto, Tasso, and the Bible?

COLLABORATION

One of the more attractive features of digital projects as a form of scholarship is that they require extensive collaboration: sociologically they are the antithesis of the monograph. They push us to build partnerships across disciplines, forcing humanists and computer scientists to explain themselves to each other and to work with the library and the school of library science. And they regularly give rise to new possibilities for collaboration, since every obstacle is an opportunity to involve another specialist. The Spenser Project has formed mutually beneficial working relationships with Early English Books Online and with the Wordhoard project at Northwestern, and it has brought different schools and departments at Washington University and at the University of South Carolina into collaboration on specific tasks.

Most recently, I was discussing with Joseph Loewenstein how to annotate certain lines of *The Faerie Queene*, and it emerged from the

discussion that we have different notions of how Spenser’s syntax works. I consulted with a specialist in our linguistics program, and the next thing I knew we were drawing up a grant proposal and designing a curriculum that would enable graduate students to pursue advanced study in literature and linguistics aimed at the formal analysis of syntax in *The Faerie Queene*. Add the advances in theoretical linguistics over the last few decades to the kinds of flexible and sophisticated concordance made possible by programs like Wordhoard, and you can see how new studies of early modern syntax might be created to extend and educate our intuitions as editors and close readers. Syntactic analysis can also be used to create a tag set and add to our textual transcriptions a markup layer that will flag significant features, providing a basis for further study and a useful model for corpus-based linguistic analysis.

TEACHING

In various ways, the kinds of collaboration I’ve been describing can be extended into the classroom. Joseph Loewenstein started a few years ago talking about the “bench humanities,” and with the help of our new project director Amanda Gailey, also at Washington University, he has followed through by creating a Spenser course with a lab component. Students in the lab worked on XML markup of various texts, studying the markup language and the TEI guidelines, debating the kinds of questions that come up when you try to design a tag set, and in the end successfully encoding substantial chunks of the transcriptions provided to us through our working arrangement with Early English Books Online. Another XML workshop is planned for this

summer at Washington University, which will in turn provide the model for a course next year in the honors college at South Carolina.

Meanwhile I’ve been experimenting with editorial commentary as a way of teaching *The Faerie Queene*. Exercises in preparing commentary on a specific passage give first-time students a chance to think directly about a fundamental question: what and how much do they need to know in order to read the poem? Students preparing commentary have to look closely at the language of a selected passage, think seriously about whether mythological references are decorative or functional, ponder the importance of historical references and literary allusions, and figure out for themselves and each other what exactly counts as “comprehension” with a text as complex as Spenser’s. Instead of writing individual term papers, they work in small teams to construct their own commentaries on various episodes complete with a critical introduction explaining their editorial decisions, and they present their work to their peers at the end of the semester. I think this procedure sometimes works better than a more conventional combination of essays and exams to give undergraduate English majors a sharp and memorable sense of Spenser as a writer.

THE IMMATERIAL TEXT

It’s a commonplace of the new bibliography to emphasize the ways in which the printed text itself was always a collaborative product, not an immaculate conception of the authorial mind for which print is merely a necessary evil, an imperfect, accident-prone source of what editors sometimes still call “corruption.” But there’s nothing commonplace about



EXERCISES IN PREPARING COMMENTARY ON A SPECIFIC PASSAGE GIVE FIRST-TIME STUDENTS A CHANCE TO THINK DIRECTLY ABOUT A FUNDAMENTAL QUESTION: WHAT AND HOW MUCH DO THEY NEED TO KNOW IN ORDER TO READ THE POEM?



AFTER NOTICING THIS DISCREPANCY, I STARTED GOING AROUND THE BORDER WITH A MAGNIFYING GLASS TO LOCATE THE BREAKS BETWEEN THE PIECES IT'S MADE OF. AND WHILE I WAS DOING THAT, SOMETHING ELSE ENTIRELY LEAPT INTO VIEW.



Mark M. Zupan

the endless particularity of the material text, and about all the ways it can call attention to the circumstances of its making and its circulation.

My first experience collating a copy of the 1590 *Faerie Queene* took place at the Ransom Humanities Center in Austin, Texas. I got very excited the first time I found a previously unrecorded variant. It was so . . . factual. One such variant I found on signature X4 of the Pforzheimer copy. This variant was unrecorded in part, I'm sure, because it doesn't occur in the text at all: look at the upper left-hand corner of the ornamental box that frames the "argument" to canto x. See the difference between the image from the Pforzheimer copy, on the left side of your handout, and the one from the Stark copy, on the right? These ornamental boxes are made up of separate pieces fitted together; in the Pforzheimer copy, one of these pieces is turned the wrong way. If you look even more closely, you can see that the piece forming the entire left-hand side of the box has been replaced.

This is a fact. What it means, I can't yet tell you. I don't imagine the "furniture"—the wood blocks and wedges that hold the type in place—was loose, because I haven't found evidence of other movement on the page. I assume, then, that for some reason



Mark M. Zupan

the chase must have been opened, whereupon pieces fell out or were removed, and one of them was put back wrong. To figure out why, you have to look at what else in the forme has been changed, and if you find any changes you have to see whether they coincide with this one—that is, whether they occur in all the same copies.

I'm still collating, still gathering my data, so I'm not ready to say what it means. Instead, let me tell you something else. After noticing this discrepancy, I started going around the border with a magnifying glass to locate the breaks between the pieces it's made of. And while I was doing that, something else entirely leapt into view.

It was a hair. A single strand of hair, as white as the page itself, rooted in the weave of the paper and spiraling up into view as if it had sprouted there. It had been invisible to the naked eye, but loomed so large in the magnifying glass that I felt a small, momentary shock, and pulled back. I had been reading Philip Gaskell's account of how sheets of paper were made by pouring a

paste of pounded rags over a fine mesh screen and pressing the water out, but now suddenly the details became real to me in a completely different way. This *happened*. More than four hundred years ago, an actual person (Giles the paper-maker?) pressed the sheet from which this page was folded and cut. Maybe he scratched his beard, and the hair is his, or maybe it was there in the rags, left over from some former owner with a more obscure itch. But there it was, and there it had probably been for the last 413 years, not the least bit allegorical until I and my amazement happened along to seize upon it—figuratively speaking, of course—and subject it to bemused scrutiny.

I have spent many long hours since then, whole days in fact, staring with fascination at the variously smudged and discolored surfaces of page after page in copies of *The Faerie Queene*, and I'll be doing it again next week right in the rare book reading room down the hall. What makes this looking so fascinating is not, however, just the minute particularity of each single page. In fact, it's only now and then that I look directly *at* a single page. Most of the time I'm staring into a mirror, and this mirror is angled toward a second mirror which is angled toward the open book. That's with my left eye; my right eye, meanwhile, is trained on a computer screen displaying a high-resolution digital image of the same page from my control text; or I might be using a printout of the scan. This is a variation on the technique known as optical collation, developed by Randall McCloud of the University of Toronto. The set of mirrors I use was developed by Carter Hailey of the University of Virginia.

What I see at such moments is a highly detailed image, including the

smudged outlines of the letters, bits of foreign matter embedded in the paper, water stains, the texture of the weave, the tears and scraped places. But for all its magnification of physical detail, this image is wholly immaterial: it exists neither on the page of the book to my left nor on the computer screen to my right. Its location is the visual cortex, where the images from my binocular vision are stereoscopic (or "collated") with such precision that even small discrepancies seem to float up off the page, occupying a different depth of field. It's a very useful thing for editing, but it's also a visionary experience. I see both the material object and the ways in which it differs from itself, for of course the whole purpose of collation is to take into account the fact that there is not one material text but many, no two of them quite identical.

I guess I'm telling you this because even though so much of the value and the interest of editing, these days, come from new technologies, new forms of collaboration, and new ways of construing the physical object, there's still a part of the process that is quite personal, indeed almost incommunicable, involving no technology more sophisticated than a pair of mirrors on lamp stands and no collaboration more extensive than that between your right and left eye. It is, as I said, visionary. In one way you're a bit like Arthur after he wakes from his dream, staring at the "pressed grass" where Gloriana lay beside him—I never realized that this could be an allegory of the printing press. But in another way you're like Arthur before he wakes, peering intently into your own mind to behold there the likeness of *The Faerie Queene*. That's a stereoscopic effect technology can't explain, but for me it's still the reason to edit the text. ☸



THERE'S STILL A PART OF THE PROCESS THAT IS QUITE PERSONAL, INDEED ALMOST INCOMMUNICABLE, INVOLVING NO TECHNOLOGY MORE SOPHISTICATED THAN A PAIR OF MIRRORS ON LAMP STANDS AND NO COLLABORATION MORE EXTENSIVE THAN THAT BETWEEN YOUR RIGHT AND LEFT EYE. IT IS, AS I SAID, VISIONARY.