

# Literary Letters Lost and Found in Cyberspace

**Matthew G. Kirschenbaum**

**University of Maryland**

The following text originally appeared on [Matthew G. Kirschenbaum's blog](#).  
It is reprinted here with permission from the author.

Good afternoon. I'm neither a librarian nor an archivist, but as someone who makes the study of texts and their attendant technologies their professional business I want to make a few points about the true nature of electronic documents; and thereby contribute my perspective to today's discussions. I said a few points, but really it's just variations on a theme which I'll give to you up front: for every problem electronic documents create problems for preservation, problems for access, problems for cataloging and classification and discovery and delivery there are equal and potentially enormous opportunities.

That's a point that may seem breathtakingly obvious to those of us in this room today, but it's still not a point, I think, grasped by the public at large, or even sub-sets of the public whose interests overlap with our own. "[Literary Letters Lost in Cyberspace](#)," laments the *New York Times Book Review* in an essay a few weeks ago by Rachel Donadio, one of the paper's writers and editors (September 4, 2005). The gist of the piece is that as more and more correspondence between authors and their publishers shift to email, an important body of scholarly primary source material crucial to literary criticism and biography, textual editing, and historical study is jeopardized.

The article gets some important things right. It points out that with the rise of email the volume of correspondence between authors and their editors has not diminished but rather accelerated dramatically. This presents more, not fewer opportunities for scholars and biographers, though the essay does assert that most email is written in a careless, ephemeral style, a point which I'd dispute because I write email too, a lot of it in fact, and while some of it is careless and ephemeral much of it carefully thought out and edited, particularly when I send it to *my* publishers. More to the point, the article repeatedly frames the issue in terms of a dearth of personal or organizational protocol for archiving email, rather than technical obstinacy and hardcore digital preservation issues. Donadio mentions the case of Deborah Treisman, the *New Yorker's* fiction editor: "Unfortunately, since I haven't discovered any convenient way to electronically archive e-mail correspondence, I don't usually save it, and it gets erased from our server after a few months," Treisman said. 'If there's a particularly entertaining or illuminating back-and-forth with a writer over the editing process, though, I do sometimes print and file the

e-mails.' The fiction department files eventually go to the New York Public Library, she said, 'so conceivably someone could, in the distant future, dig all of this up.'"

Here we see preservation framed as a fundamentally social rather than a technological challenge. This is right I think. But the article fails to draw the obvious conclusion, which in this case is simply that Ms. Treisman should *start* systematically saving her email and that her systems people should start backing it up every night rather than wiping it every few months. Likewise, when novelist Zadie Smith laments, "I have a normal Yahoo account that saves e-mails instantly, but not to the hard drive. I've e-mailed Yahoo and asked how you can save all your own e-mails onto disk or whatever, but I get no reply," the solution is to switch to a different service provider, one who will allow her to download all of her 12,000 messages to and from agents, editors, and other literati to her own personal files, which she can take ownership of. Smith also points out that she no longer has any drafts or early versions of her works in progress since she simply executes a save command that overwrites any earlier copy of a file. The truth is those earlier versions no doubt do exist somewhere in the flotsam and jetsam of her file system, but she need not resort to high-end recovery tools to get at them. She just needs to start saving her versions. In fact, as storage costs continue to plummet and personal hard drives begin to edge up to the terabyte threshold, saving every state of every file is likely to become routine, the default. It will cost more to take the time to search a file system, locate a file, and then overwrite it than it will to simply keep it somewhere on a hard disk whose aerial density is at something 10,000 tracks and sectors per square inch.

There's no question that certain things are lost when documents are prepared and transmitted in electronic formats. The texture, heft, even smell of the paper, the coffee cup's stain, the crinkled edges and dog-eared pages, the physical abrasions of marks and erasures. Let's think for a moment about what's gained though. By opening my word processor's Properties window I can ascertain, to the date- and millisecond, when the file was first created and when it was last edited. I can count the number of words and characters, but more interestingly the number of minutes spent editing the document. This is the kind of information scholars and editors of the literary classics would weep to have. How long was Coleridge *really* at work on "Kubla Kahn" before he was interrupted by the man from Porlock? The point here is that electronic objects are self-documenting to a remarkable degree, and this is a phenomenon that can and should be exploited as new social and technological practices evolve to preserve them.

Since the debut of Word 97, users have had the ability to "track changes" in their documents, the software automatically logging each and every addition and deletion, as well as changes in formatting all date- and time-stamped. In fact, one common workplace gaffe is to send a client or correspondent an electronic copy of a document with the *view* of the changes turned off, but the changes themselves not yet accepted or rejected thereby displaying the word-by-word processing of the text to its intended recipient as soon as the view of the changes is turned back on in their own copy of Word. Wikis, notably the Wikipedia, offers similar systems, where the version and revision history of the document is transparent to all who access it. I can find out much more about [who contributed what to a Wikipedia article when](#) than I ever could with a printed

reference work often authored anonymously or by committee. Fixing such version histories is a trivial computational undertaking, not least because at the center of a CPU you'll find not teeming rows of ones and zeros but a crystal clock.

In terms of challenges to future historians, Donadio cites Steven Kellman who has just written a new biography of Henry Roth; he suggests, rather indisputably, that "Our understanding of the Constitution . . . would be quite different if the thoughts about it exchanged by Jefferson, Madison and Monroe had vanished into the electronic ether." True enough. But there's nothing inherent in the technology that makes email especially susceptible to vanishing into the electronic ether. On the contrary, as Oliver North and other malefactors have found out, the stuff is remarkably pesky and hard to expunge. A single email message may leave traces of itself inscribed on a dozen different servers as it makes its way across the network, a potential for proliferation that is further exacerbated by backup services at each site. While I don't mean to minimize the very real technical challenges in the realm of digital preservation, it's worth remembering that email and other textual forms have it easier than with other media since often we're dealing with ASCII and XML rather than binaries and proprietary formats.

Consider what a treasure trove if Jefferson, Madison and Monroe had been corresponding via email and their messages had been systematically archived. This is a feature routinely associated with listserv technology and other electronic mail systems. In the case of the [William Blake Archive](#), one of the scholarly text and image encoding projects begun in the mid-90s at the [Institute for Advanced Technology in the Humanities](#) at the University of Virginia, we have ten years and around 10,000 messages worth of correspondence between the editors and project staff on every facet of the project's development, large and small, all date- and time-stamped, threaded, hyperlinked, and network accessible to the project participants.

The plight of novelists like Zadie Smith who fear that computer use is keeping them from saving earlier versions of their work is particularly poignant. The irony is that versioning is a hallmark of electronic textual culture. Not only in a principled or potential way, but as a thriving industry of document management systems, file comparison utilities, and so-called version control or concurrent versions systems, the archetype of which, CVS, originated in the mid-1980s when a Dutch computer scientist crafted a repository structure that would allow himself and his students to work on a large distributed programming project without overwriting one another's code. While most prevalent in software development, especially open source projects with their frequently globally dispersed contributors, a version control system like CVS (or its heir apparent, Subversion) is capable of managing any kind of data, textual or otherwise. A version control system retains every state of every file checked in or out of the repository, and is capable of stepping back through the decision trees that result to access the data at any point in its revision history. In essence, what these systems represent are temporal extrusions of the immediate documentary event on a user's screen.

I've just embarked on a [new project](#) that asks poets and fictions writers to contribute original material to just such a CVS repository, checking their work in and out of the

repository structure each and every time they wish to edit or compose. The result will be a Web-accessible archive of the creative process, with the full text of each and every version of a writer's text available for reading, much as though one were looking over the writer's shoulder as he or she worked. I've sent invitations to a number of established contemporary writers in an effort to get this off the ground. Want to see how Robert Pinsky writes a poem? You'll be able to. Assuming, that is, he answers my email.