Traces of the Old, Uses of the New

Earhart, Amy E.

Published by University of Michigan Press

Earhart, Amy E.
Traces of the Old, Uses of the New: The Emergence of Digital Literary Studies.

For additional information about this book
https://muse.jhu.edu/book/43167

For content related to this chapter
https://muse.jhu.edu/related_content?type=book&id=2792691
Cyberspace is an environment comprised entirely of 0’s and 1’s: simple binary switches that are either off or on. No in-between. No halfway. No shades of gray. All too often, when it comes to virtual culture, the subject of race seems to be one of those binary switches: either it’s completely “off” (i.e., race is an invisible concept because it’s simultaneously unmarked and undiscussed), or it’s completely “on” (i.e., it’s a controversial flashpoint for angry debate and overheated rhetoric). While there are similar patterns of silence about race when it comes to interpersonal interaction in “the real world,” the presence of visual and aural markers of race (no matter how inaccurate those may be) means that race is rarely (if ever) as invisible offline as it is in cyberspace.

—Beth E. Kolko, Lisa Nakamura, and Gilbert B. Rodman, Introduction to Race in Cyberspace, Routledge, 2000, 1

In 2010, I located Sharon Harris’s Early American Women Writers website. By 2011, the site had been removed.

As the previous chapters chronicle, scholars working in the early period of digital literary production experimented with delivery, forms of access, interfaces, and representations of materiality. The digital edition and digital archive models were successfully
replicated leading to a period of rapid digital production. Working within these models, a subset of scholars focused on what I dub “digital recovery projects,” archives and editions that used digitization to expand what such scholars saw as an outmoded new critical literary canon that excluded work by women, people of color, queers, and others. Digital recovery projects emerged out of activist cultural studies communities and were tied to evolving understandings of the emergent Internet.

In the 1970s and 1980s, canon expansion was the literary holy grail. Groundbreaking scholarship, such as *The Madwoman in the Attic* by Gilbert and Gubar, argued for a broader canon, heralding a new generation of scholars that would begin to edit and publish previously excluded writers. But scholarly presses had limited budgets and numerous texts remained unpublished or out of print, rejected because they were deemed noncommercially viable. Further, scholars of African American, Asian American, Native American, and Latino/a literatures charged that the white feminist movement had not sufficiently attended to writers of color. Among the most influential texts that attracted criticism was the Gilbert and Gubar *Norton Anthology of Literature by Women* (1985), condemned for its lack of women writers of color, divergent classes, and pre-1800 women’s writing. Scholars charged that the *Norton Anthology* had created an alternative canon that was as problematic as the previous canon that had excluded women’s writing. Margaret Ezell’s influential *Writing Women’s Literary History* (1993), for example, articulates how “structures used to shape our narrative of women’s literary history may have unconsciously continued the existence of the restrictive ideologies that initially erased the vast majority of women’s writings from literary history and teaching texts.” One remedy to what was seen as a continued exclusionary canon was the launch of activist presses such as the Feminist Press (1970) and the Kitchen Table: Women of Color Press (1980), dedicated to publishing materials that mainstream and even scholarly presses were not printing.
Despite valiant efforts by activist presses and the increased publication of “lost” texts by mainstream and scholarly presses, it was clear that such efforts alone would not solve the canonical omissions. Dissatisfied, scholars and activists began to look for other solutions to solve the canon problem.

In the 1990s, proponents of the Internet began to tout the developing technological infrastructure as a tool to democratize knowledge. Popular culture portrayed the early 1990s Internet as an idealized, democratic, and free space that needed protection from the corporate market forces intent upon invading the open space theretofore dominated by scientists, hackers, and geeks. Advocates of the free web championed three ideas: “1) Access to computers should be unlimited and total; 2) All information should be free; (and) 3) Mistrust authority and promote decentralization,” all ideas designed to allow “bubbles” of information to rise from the bottom, sowing “seeds of revolutionary change.”

Scholars, too, began to understand the net as a space that altered power structures. As Paul Delany reports,

The Internet has thus mutated into an unforeseen and unplanned information space. Its virtues can all be attributed to its collegial political economy: in a word, its openness. Internet’s most important features are its relatively small hardware investment, a weak (but not ineffective) central administration, little censorship, and an absence of specifiable “bottom-line” objectives. Its explosive growth in the last few years confirms the dynamism of a collegial cyberspace culture in which millions of users exchange information, collaborate on creative projects, and have their say on any subject they care about.

For scholars interested in challenging the traditional canon, the technological possibilities were a boon. It was imagined that the web would allow those previously cut off from intellectual capi-
tal to gain knowledge that might be leveraged to change their social position and would allow those who had been silenced to have a voice. Hypertext theorist Jay David Bolter, for example, promoted the freeing power of the web-based environment as a space that encouraged “the abandonment of the ideal of high culture (literature, music, the fine arts) as a unifying force. If there is no single culture, but only a network of interest groups, then there is no single favored literature or music.” The belief in the low cost of Internet-based publication would encourage scholars to embrace the Internet as a vehicle for disseminating cultural materials. In 1996, Peter Shillingsburg anticipated that it “eventually will cost less to produce and therefore, one assumes, to purchase a compact disk than it cost to produce and purchase Hans Gabler’s edition of Ulysses.” In hindsight, Shillingsburg and the larger scholarly community have rejected this view as naïve and unrealistic, yet it was a common refrain in the early, heady days of digital recovery, where the faith in the Internet’s ability to shift hierarchical structures became part of the narrative that in no small part drove the proliferation of digital literary recovery projects.

From the mid-1990s through the mid-2000s there was an explosion of digital sites that championed work previously excluded from the canon, what I label digital literary recovery projects. Scholars worked individually or as small collectives to produce bibliographies or to publish primary texts. Simple HTML projects, such as The 19th Century American Women Writers Web (19CWWW), Voices from the Gaps, Early American Women Writers, The Black Poetry Page, The Online Archive of Nineteenth-Century U.S. Women’s Writings, and American Women Writers 1890 to 1939—Modernism and Mythology, were developed by scholars without the support of a digital humanities center, technological collaborators, or external funding. The activist recovery projects of this period were positioned to critique and respond to perceived weaknesses in the existing canon. Projects during this pe-
period were created using HTML, HyperText Markup Language, in large part because HTML was a fairly simple to learn tagging system that became even easier to use with the development of early HTML editors, including Claris Home Page, Mozilla Composer, and Adobe PageMill, all of which allowed novices to easily publish their findings. Sites built during this period were largely labors of love launched by scholars interested in disseminating materials that they found buried in difficult to access rare book rooms, crumbling newspapers, and unknown journals.

Such work envisages digital literary scholarship as a tool that might be utilized to meet the theoretical demands of scholarly work that reinserted women, queers, and people of color into the canon. For example, groundbreaking feminist scholar Judith Fetterley, whose 1978 *The Resisting Reader: A Feminist Approach to American Fiction* did much to challenge the traditional literary canon, nurtured a small online digital project to bring scholarly attention to the huge numbers of nineteenth-century women’s texts that deserve critical scrutiny. Fetterley’s personal website at the University of Albany, from which she is now retired, includes several pages of bibliographies and chronologies of what were then little known literary texts by American women.10 Titled the 19th-Century Bibliography Project, the site includes bibliography entries focused on women writers from the 1820s and 1830s arranged chronologically and alphabetically (see fig. 3.1).

Developed by students in a graduate course that Fetterley taught in 1995 and then compiled and published on the web by her student, Annie (formerly Lois Dellert) Raskin, Fetterley viewed the project as both pedagogical and activist. Articulating similar concerns about the canon as Ezell, Fetterley explains that:

Underlying this project was the desire to make clear to students that the texts available for such a course as this represent only a fragment of the total work of women writers during the period, and that we are as much in danger of
developing a skewed version of American literary history if we look only at the texts by women that have been reprinted in the last twenty years as we are if we look only at the work of male writers. In other words, the category of the “marginalized” can seem quite privileged when set in the context of a category that might well be called “the not-yet-marginalized.”

The online bibliography was a means of enacting a broader canon than that represented in print and is also indicative of the many such early digital projects where self-publication was designed to distribute work in progress, expand the literary and historical record, and spur new interest in the recovered texts. Unlike a print bibliography, which purports to be an authoritative, complete, and static overview of a particular author or topic, the 19th-Century Bibliography Project is understood to be incomplete, in effect “an introduction to the various bibliographic resources
available for research in this field.” The project also views the digital as a means to initiate collective action and, accordingly, invites revisions: “We offer these bibliographies in the hope that they will be useful to others and we welcome comments, corrections and discussion.”

The collaborative approach to developing and sharing the materials, with Fetterley utilizing the classroom as a textual cottage industry, is typical of this period, and other projects such as The Charles Chesnutt Archive also follows the pedagogical model to bring “lost” texts online. Unfortunately, though, the 19th-Century Bibliography Project is one of the many digital projects built during this period that are in danger of disappearance. The project is fragile not because of its digital form but because there is no preservation strategy nor community surrounding the project. While participation and revision is invited, there is no indication that the materials found in the bibliography had more than perhaps a revision or two. Published in 1996, the pages were last updated in 1997 and are now static and even decaying, an emblem of the larger state of digital recovery projects from this period.

In hindsight the limitations of recovery projects are clear, but for scholars interested in reworking the canon in the late 1990s, the web seemed a space of possibility. Scholars believed that the digital environment prevented presses and editors from limiting the types of work published. “What is new in the twenty-first century,” writes Susan Fraiman, “is that now the guest list of history-making women is electronic—and there are always more seats at the table.”

The belief in the Internet as a means to break the canon was likewise espoused by the editors of Romantic Circles: “One of the strengths of Web publishing is that it facilitates—even favors—the production of editions of texts and resources of so-called non-canonical authors and works.” Like Fraiman, the editors of Romantic Circles laud web publishing’s ability to allow the scholar to work outside traditional publication mechanisms and to create expansive canons. The editors go on to
identify the simplicity of digital publication as a key component of the success of the web. Success “is in part a function of the relative simplicity of HTML (and all of the simpler document-type-descriptions of SGML) and of ‘workstation publishing’ in general when compared to traditional commercial or academic letterpress production and distribution methods.” The ease of publication identified by the editors was what allowed the broad range of small-scale recovery projects to explode in this early period, as most of the projects were produced using simple hand coding of HTML or HTML editing programs. Digital recovery project URLs reveal that creators of projects often “published” their materials on personal webspace allotted by colleges and universities to their faculty and students. Technologically inexperienced users could create a simple textual project using low cost and low expertise methods, an approach contradictory to much of contemporary digital project development that tends to be conducted with the more complicated TEI/XML, databases, tools, and even datamining and algorithmic approaches. The low threshold at this moment of digital literary studies resulted in a broader set of textual materials being created and a democratic digital development.

Tracking the rise of the queercore, riot grrrl, and straight edge punk movements in the late 1980s and early 1990s provides one lens through which to read the digital literary studies movement. Emerging from 1970s punk, queercore, riot grrrl, and straight edge punk movements adopted self-publication as activist interventions. Amy Spencer describes the riot grrrl movement as “about using anything you can get your hands on to shape your own cultural entity; your own version of whatever is missing in mainstream culture,” an ideology replicated in digital recovery projects from the mid to late 1990s. As did those interested in the digital recovery of texts, activist oriented music refused to rely on traditional methods of artistic dispersal, instead creating a self-publishing movement that allowed participants to create
a broader canon of artistic expression. Riot grrl self-publication gave voice to the excluded:

BECAUSE us girls crave records and books and fanzines that speak to US that WE feel included in and can understand in our own ways. BECAUSE we wanna make it easier for girls to see/hear each other’s work so that we can share strategies and criticize-applaud each other. BECAUSE we must take over the means of production in order to create our own moanings. BECAUSE viewing our work as being connected to our girlfriends-politics-real lives is essential if we are gonna figure out how we are doing impacts, reflects, perpetuates, or DISRUPTS the status quo.¹⁸

Embracing a personal is political approach, both groups made visible excluded women, people of color, and queers—whether the exclusion was at a punk show, in print culture, in the classroom, or on a music label. The notion of such work as “underground” was vitally important, as the groups positioned their work against that of the dominant society. If mainstream record companies wouldn’t sign queer singers, then singers would produce and distribute their work directly to fans. If scholarly publishers refused to print a relocated women’s text, then scholars would create and publish the materials on the web for use by those interested.

Often localized and idiosyncratic, both movements were able to use new models of technology to promote inclusion and share understandings of voice and power through do-it-yourself (DIY) production of materials. Moore and Mitchel argue, “We see the DIY mechanism as both a relational mechanism because it both ‘alters connections between people, groups, and networks,’ which creates solidarities, and as a ‘cognitive mechanism,’ because it ‘operates through alterations of individual and collective perception’ by providing a foundation for the creation of imagined communities via taste and aesthetic choice.”¹⁹ A hallmark of the
DIY literary recovery projects is the development of a virtual community through various linking strategies. Projects often referenced other like-minded recovery projects through a list of links, creating an ever-expanding virtual network. For example, the Native Web included a list of authors, writers, and biographies (see fig. 3.2).

Native Web is activist in construction, community-focused in orientation, with a primary goal “to foster communication among peoples engaged in the present,” a common thread in such digital recovery work. The importance of what I call “curated hyperlinked” sites has been unremarked upon by digital humanities scholars, a remnant of late 1990s web culture that now seem simplistic and out of date. Yet such work was pivotal to the formation of digital literary culture. Influential curated hyperlinked sites of the period include Alan Liu’s The Voice of the Shuttle and Randy Bass’s The American Studies Crossroads Project. Donna Campbell’s American Writers project provides an excellent example of curated hyperlinked projects used to construct a community of scholars interested in revising canons (see figs. 3.2 and 3.3).

Launched in 1997, Campbell created the website to support a literature class. Housed on her personal webspace at Gonzaga University, the site grouped materials by author, time period, and literary movement. Historical, cultural, and biographical materials are also included as are links to primary texts. Campbell’s curated hyperlink site and others like it were developed to renegotiate canonical boundaries. Hence, the American Writers author list is far more diverse than contemporary anthologies, including little studied women, Native American, Latino/a, African American, and Asian American writers. Rather than devaluing such sites as technologically simplistic, we must resituate curated hypertext sites as crucial genres in the evolution of digital literary studies.

Open access, the use of digital technologies to distribute scholarly materials without cost or restriction, is also forecast in the 1990s digital recovery movement. For example, Kim Wells's
Fig. 3.2. Native Web screenshot, Authors, Writers & Biographies.

Fig. 3.3. Zitkala Sa page, American Writers curated hyperlinked site.
Domestic Goddess website, which focused on canonically excluded women writers, positioned open access and education as central goals of digital recovery projects:

I think it is our duty as teachers not to ignore the possibilities of making research easily available on the Internet. If educators do not provide the information, who will? Do students have to pay for it, as a lot of encyclopedia companies are requiring now? This makes information only commercial, and even if they buy a subscription, sometimes the information they get is incomplete, and encyclopedic, rather than critical. These sites are great—but I want to teach for free.

The desire to educate and freely disperse information is apparent in other early sites, including Kristin Mapel-Bloomberg’s American Women Writers 1890–1939, Modernism and Mythology, and Shari Benstock’s Women of the Left Bank, all of which publish supplemental scholarly materials to their accompanying print books. Such work, tied to DIY activist movements that bled into scholarship, is among the earliest representation of open source scholarly publishing and, like curated hyperlinked texts, needs to be reconfigured within the history of digital literary studies. Though often dismissed as simplistic or unimportant, open source distribution of materials brought attention to writers that had not been represented in the canon and helped to launch our current understanding of public scholarship.

As individual scholars began to share recovered texts online, libraries began to explore how they might use the digital to bring their collections to a larger audience. Etext centers were started at Rutgers University and Princeton University (Center for Electronic Texts in the Humanities [CETH]), the University Library of the University of North Carolina at Chapel Hill (Documenting the American South), the University of Virginia (The Electronic
Text Center [Etext]), and other universities. Some of the centers emphasized out-of-print and recovery texts, as was the case with the University of North Carolina’s *Documenting the American South*, but most were focused on the digitization of canonical texts. The UVA Etext Center is arguably the most important center during this period due to the quality and quantity of the projects they produced, as well as the students trained within the center who include notable contemporary digital scholars Stephen Ramsay, David Gants, Lisa Spiro, Tanya Clement, Matthew Kirschenbaum, and Amanda French. The Center digitized a broad assortment of texts, recovered and traditional; noncanonical texts by Native American, African Americans, and women writers coexisted with Founding Father documents by Thomas Jefferson and Alexander Hamilton. While the bulk of the texts that the Center digitized were selected by Etext staff from the UVA library stacks, a type of pre-Google digitization project, individual scholars would also suggest targeted texts of interest, including many of the noncanonical texts, for digitization. Virginia’s Etext Center was unusual in its digitization of noncanonical texts as most other centers, according to Mandell and Gamer, were traditional in their approach: “the web—at least in its first stages—did reproduce canonical biases long inscribed in Romantic poetry. This was particularly true of ‘early’ electronic text collections, like those compiled at Oxford, Toronto, Berkeley, and Carnegie Mellon, to name a few.” Those centers that produced noncanonical texts shared a common focus on faculty and graduate student input in text selection. Scholarly participation, therefore, was a central factor in the selection of texts digitized during this period.

One successful recovery collaboration between an etext center, the University of Virginia, and an individual scholar, Jean Lee Cole, was the *Winnifred Eaton Digital Archive (WEDA)* (see fig. 3.4).

Immediately after publishing her 2002 book on Eaton, *The Literary Voices of Winnifred Eaton: Redefining Ethnicity and Au-
What's In and What's Out?

• 75

thentricity (Rutgers University Press), Cole uncovered numerous “lost” Eaton stories and novels. Unable to secure a publisher for the little known author’s work (publication of Eaton’s work was not commercially viable, according to the presses she contacted), Cole explored other options for publication. When she attended the Rare Book School at the University of Virginia, Cole was put in touch with the Etex Center to discuss the possible digital publication of the materials. Cole created a digital archive housed on the UVA Etex Center servers that included an introduction to Eaton, twenty-three short stories, two novels, and thirteen nonfiction pieces. The WEDA archive is a particularly interesting example from this period as it is representative of the fluid boundaries between print and digital. The digital archive began as a way to add materials to the print monograph. Once the archive was complete, Cole recognized that scholarly credit would not be extended to her digital work, so she published a
critical article, checklist of materials, and a reprint of one short story in the journal *Legacy*. Cole’s use of the digital as a means of increasing access to “lost” works is indicative of the period where the digital was seen as a tool for extending the canon.

As we moved out of the 1990s, the flurry of digital recovery projects and the promise of activist work to restructure the canon began to fade. Digital recovery projects had made a huge impact on the texts available online, but such work seemed to contrast with the growing professionalization of digital literary studies. While digital editions and archives had coexisted with digital recovery work in this earlier period, the increasing institutionalization of digital humanities shifted the focus away from such technologically simplistic work and toward more robust and complex projects. The decline of digital recovery projects will produce what Kenneth Price calls a “newly emerging digital canon of American literature.” Unfortunately, the canon we now have was created “partly by design and partly by chance,” a canon much more narrow than that of the contemporary print canon.

The late 2000s marked a period of rapid decline in recovery work. By 2008, most of the etext centers including the Oxford Text Archive, the UVA Etext Center, CETH, and the Carnegie Mellon center had been shuttered or absorbed by their institutions’ libraries, resulting in a rapid decrease of new digital recovery projects. The proliferation of early activist projects by scholars interested in recovery dropped precipitately as literary scholarship shifted its focus away from canonicity (see fig. 3.5).

JSTOR data provides a snapshot of the larger literary field’s interest in particular subjects. Collecting data from journals, books, and primary sources, JSTOR data provides one context for understanding trends in scholarship. An analysis of JSTOR data from 1950 to 2014 reveals that scholarship focused on canon concerns peaked in 1999 and has been declining since that period. It is not surprising that the same pattern of interest and decline is apparent in the digital record of textual recovery.
In addition to a decrease in new projects, a large number of digital recovery projects produced during the heyday of such work have been or are in danger of being lost, of grave concern as it is possible that we are creating a digital canon that is more representative of the literary canon of 1950 than that of the literary canon of 2014. Analysis of Alan Liu’s *Voice of the Shuttle* (*VOS*), a curated hyperlinked project that compiled a broad range of primary and secondary humanities materials from 1994 to 2009, provides a means of measuring the loss of early digital recovery projects. Subjects range from the technology of writing, politics and government, photography, dance, anthropology, and cultural studies among other categories. Links to primary texts are included in multiple categories, including Literature (in English), Literatures (Other than English), Minority Studies, Cultural Studies, and Gender and Sexuality Studies. Because Liu designed the *VOS* as comprehensive and also relied on an early form of crowd sourcing, a network of scholars that con-
tributed links and updates, the site provides an excellent measure of the state of digital projects during the years the site was active. While the site has been static since 2009, the Wayback Machine, part of The Internet Archive, has archived earlier versions of the project and related links that though incomplete provides one of the best tools for scholars exploring the state of humanities digital projects of the period. Examination of digital texts curated by the VOS reveals that a tremendous number of early digital recovery projects have been lost. Not only are links dead but detailed searches of the web suggest that materials have been completely removed. For example, correspondence with Sharon Harris, creator of the *Early American Women’s Website (EAWW)*, reveals that the site was removed in 2011. According to Harris, there are no available backup files. The loss of the *EAWW* is a familiar occurrence. Examination of projects included in the VOS reveals a great deal of loss. For example, contemporary writer Maya Angelou is included in the Contemporary American Authors and African American Literature categories with a total of seven sites related to her work. Only two of the seven projects remain online. Of the five original projects, we are left with scant details that we might use to locate the original documents, such as original URL, author, site creator, and hosting university. This unfortunate loss is replicated time and time again, leaving the digital literary canon smaller with each loss.

Tracking the production and subsequent loss of recovery projects reveals the haphazard history of the preservation of such projects. *The Winnifred Eaton Digital Archive* is one such example. Hosted by the University of Virginia’s Etext Center as a separate and distinct digital archive from 2003 to 2010, the archive has been absorbed into the library’s digital collections by the department of Digital Curation Services at the University of Virginia, and, as part of the transition, most Etext Center materials have been cleaned and migrated to stable repositories. All Eaton texts, then, are preserved and available through a standard search of
VIRGO, the University of Virginia Library database. The preservation of individual recovery texts is rare and those involved in such work at the University of Virginia show great forethought. However, preservation decisions have created a new form, perhaps a new edition, of Eaton’s work rather than a copy of the original archive. UVA has broken the original digital archive into individual items, in effect mimicking the original print versions of Eaton’s work. In addition, the original archive includes images from magazines that accompanied the texts. These images are no longer available in the VIRGO versions. Also removed from the current Eaton texts are files that include information regarding the production of the digital version of the text, such as original creator of the materials, conversion metadata, and editing history. The elimination of such material erases the bibliographical markers of production that scholars are interested in viewing when they evaluate the reliability of a text. Such losses are exacerbated by the dissolution of the archive itself. The stand-alone digital archive of Eaton’s works reveals the volume and breadth of her writing and gives weight to the importance of resituating Eaton within the American literary tradition. The transfer of the body of work into single entries obscures the largess of Eaton’s work, and Eaton’s legacy recedes into the database, leaving her recovered body of work obscured once more.

The Winnifred Eaton Digital Archive is just one of many early digital recovery projects that are unstable. Some recovery projects contain texts that are unavailable as open source online materials but are found in subscription databases. Other projects have been fragmentally archived, as is the case with the Eaton materials. Still other projects have disappeared completely and are available only in the original print form from which they were recovered. All losses impact the shape of our current digital canon. Scholars spent years digitizing and recovering texts, providing open access versions of work by authors who had been understudied. Every text that has disappeared has erased crucial knowledge that will
now need to be recovered or, worse, undertaken once more. The loss of a digital presence of an often reproduced canonical author may not hurt the awareness and study of the canonical author, but the loss of the single digital instance of a recovered text by a little known author will, in effect, send the author back into the hidden archives that scholars have worked to expose. Additionally, such built projects provide important cultural and historical data that reveal understandings of criticism and culture at a particular historical moment. It matters that scholars and fans chose to dedicate scholarly sites to certain authors and not others. The loss of digital recovery projects leaves gaping holes in our understanding of late twentieth-century literary scholarship. Digital humanists are fond of talking about sustainability as a problem for current and future works, but it is clear that we already have sustained a good deal of loss within the broadly defined digital canon.

While a good many of the early small-scale digital projects have been displaced or lost from our current digital canon, a few have managed not only to survive but to thrive and offer models of best practices for recovery and sustainability. Projects created in connection with an institutional structure, a university or etext center, have had a much better chance of surviving, in part because institutions understand standards and preservation approaches, have a greater possibility of staff dedicated to such work, and see the projects as part of past investments of the institution and, as such, economically and intellectually important. For example, the Victorian Women Writers Project was launched in 1995 and quickly became known as an exemplary digital recovery project. By mid-2000, however, the project was faded. The markup was old and incomplete, search capabilities were limited, and the project looked dated. In 2010 the Indiana University Library team, headed by Michelle Dalmau, refurbished the project. Texts were recoded to meet TEI P5 standards and additional materials, including “newly encoded texts and related contextual materials like criti-
cal introductions, biographical sketches, and annotations,” were added. Harnessing the expertise of digital librarians at Indiana University, graduate and undergraduate students contributed to the revision.42 The same support is usually not offered to the one-off, stand-alone DIY projects outlined at the beginning of this chapter, which suffer from their position outside the institutional structure. Other projects have found institutional support from digital organizations, such as NINES (Nineteenth-Century Scholarship Online). The project 19: Interdisciplinary Studies in the Nineteenth Century was initially formed as a simple HTML journal. Scholars affiliated with the project participated in a 2005 NINES summer workshop during which they learned to encode their documents with the international standard of TEI/XML. Once the project was re-marked with TEI, it was brought into the NINES federated collection of nineteenth-century materials, which expanded its user base, allowed scholars to use different tools to manipulate and examine the materials, and increased the project’s chance of long-term sustainability.

The attention to infrastructure is central to what might be the most successful early recovery project, the Women Writers Project. The project is designed “to bring texts by pre-Victorian women writers out of the archive and make them accessible to a wide audience of teachers, students, scholars, and the general reader,”43 a goal shared by many of the early digital DIY projects. Begun in the department of English at Brown University, later moved to Computing and Information Services and, finally, the Brown library, the WWP has always demonstrated its connection to the mainstream of academia. The founders of the project, Susanne Woods and Stuart Curran, brought legitimacy to the project since they were known scholars in the subject area who could command respect for their work, even when that work was being disseminated through digital means. The advisory board and initial editors were or have become leaders in their respective fields. In addition, the strategic alignment of the WWP with a highly
respected press, Oxford, and university, Brown, suggests that the effort to position the project as inside the academy not only aids legitimacy of the project but increases its ability to be sustained. Cleverly tying the new with the old, Woods suggests that harnessing traditional values and approaches to scholarship aid the sustainability of the WWP. As Woods describes the project, “The long-term future of the WWP, then, lies in its position as a bridge between old ways of reading and new.” Infrastructures of support, then, are a necessity in digital project preservation.

Infrastructure support extends beyond institutions to individuals who work in collective communities. Active communities are the difference between preservation and loss. Curated hyperlink projects, for example, are structured to take advantage of active communities that produce webpages. As communities lose interest, projects disappear leading to the death of links, a process that eventually renders the curated hyperlink project unusable. An active and interested community, however, is incredibly powerful as demonstrated by viral archiving examples. Viral archiving is the replication of a text for use or distribution. Perhaps the most studied text preserved through viral archiving is Aggripa (A Book of the Dead), a born digital art collaboration of Dennis Ashbaugh, William Gibson, and Kevin Begos Jr. Matthew Kirschenbaum and Alan Liu’s exemplary work on Aggripa reveals that “the text of the poem as it circulates online today is not a digital copy whose bits were lifted from one of the project’s diskettes, but rather the result of a manual transcription of a video cassette . . .” The hackers who chose to preserve the poem were responding to the unique challenge of a software text that was designed to be unarchivable, and though we don’t have an exact replica of the original poem, a surrogate for study has remained through the communities’ efforts. Similar efforts include the reproduction of a poignant letter written in 1795 by Judith Cocks, a slave in Connecticut, to her master. Originally recovered by the Early American Women Writers website, the letter was replicated in an entry
on the Women of Color in Accounting Facebook page. The reproduction of the letter on a nonacademic page is not uncommon; “The Marriage of Okiku-San,” a short story found within the *Winnifred Eaton Digital Archive* and reprinted by *Legacy*, is preserved in various sites including a genealogy site and various Chinese and Japanese sites. While viral archiving successfully preserves texts, the happenstance approach to such preservation is by no means an appropriate long-term strategy if we are interested in producing a comprehensive humanities data set.

Crucial to establishment of an inclusive set of digital literary texts is an understanding of the role of technological standards in preservation and canonization. Digital humanists are well aware of the importance of updating technologies for preservation, and successful projects such as the *Walt Whitman Archive* have made the transition from HTML, to SGML, to TEI/XML metadata structures. However, technology standards have impacted the ways that materials are utilized, leaving some digital recovery projects unused and excluded from what is developing as technological standards that are elemental to the formation of what is and is not included in the digital literary canon. Kenneth Price alludes to the centrality of technology standards in canon selection when he argues that “people ready to embrace high quality work wherever it is found hold in highest regard digital work that features a rigorous editorial process and adheres to international standards (for example, TEI/XML).” The notion of technological standardization as a marker of high quality digital work provides another clue to the displacement of digital recovery texts. An MLA bibliography search includes almost one thousand digital literary projects, yet the vast majority of the small-scale, older recovery projects included by MLA are invisible in current digital literary scholarship. Glynis Carr’s *The Online Archive of Nineteenth-Century U.S. Women’s Writing*, active from 1997 to 2001, is one such project listed in the MLA bibliography. A number of the texts included on the site, such as *Aunt*
Lindy by former slave Victoria Earle Matthews, are not available in digital form elsewhere. Though Matthews's text would be very useful for those examining nineteenth-century issues of religion and race in postbellum America, the text is not considered part of the current digital literary studies canon in large part because of its technological infrastructure. The project shares characteristics with other dying or lost digital recovery projects. Glynis Carr produced the project with a small number of students, without a digital humanities center or library support, and published the project on her personal staff page at the university at which she is employed. Carr is also not active within the digital humanities community. All of these reasons may contribute to the lack of attention given to the site, but perhaps most damaging is that the site was produced in HTML, a form resistant to preservation strategies and noninteroperable with current datamining approaches. A common strategy in early digital recovery projects, the use of simple technologies for digital project creation, now mark projects as “low quality,” a stigma that creators of the projects are well aware. Interviews with Jean Lee Cole, creator of the Winnifred Eaton Digital Archive; Sharon Harris, creator of the Early American Women’s Website; and Donna Campbell, creator of American Authors reveal that all three scholars believe that current technological standards have stymied project preservation. Harris, for example, reports that she removed her EAWW project because it was “not in a sophisticated platform.” Cole and Campbell both view their projects as excluded from digital literary scholarship primarily due to technological approaches. Some digital scholars see the overdetermination of technology as a disturbing trend that leads to a narrow digital literature canon. Amanda Gailey and Andrew Jewell call this trend “a ‘hipster ethos’” that makes “absent” “celebrations of content-rich digital humanities projects.” Gailey and Jewell are concerned with current production of digital editions, but the emphasis on technological standards has an equally damaging impact on the inclusion of a diverse canon.
Barriers to digital recovery projects are also found in the current ways that we theorize what is an appropriate “text” for digitization and analysis. Margaret Ezell cautions that we have not revised the way in which we understand texts and because of this elision certain texts, particularly noncanonical texts, are not being digitized. She argues that “while we increasingly have the ability to digitalize any text we please . . . editors do not please to select certain types of material and this is in part because perhaps we are not yet changing some of the basic assumptions about what an ‘edition’ does, or in Hunter’s terms, what is ‘appropriate.’” Clearly the question of canon is still in flux, regardless of the belief that such concerns were resolved by the culture wars of the last century. We might also have a historical problem in the very emergence of digital humanities that contributes to the selection of materials for digitization. Martha Nell Smith contends that digital humanities developed as a space to which practitioners fled from the shifts in the profession that arose out of the cultural studies movement. In “The Human Touch: Software of the Highest Order, Revisiting Editing as Interpretation,” Smith highlights the digital humanities’ retreat into modes of analytics, objective approaches as “safe” alternatives to the messy fluidities found in literary studies. She notes, “It was as if these matters of objective and hard science provided an oasis for folks who did not want to clutter sharp, disciplined, methodical philosophy with considerations of the gender-, race-, and class-determined facts of life . . . Humanities computing seemed to offer a space free from all this messiness and a return to objective questions of representation.” If Smith is correct, then it should be no surprise that the recovery of messy lost texts has not been a priority for the field.

Additional reasons for the exclusion of particular texts are economic. Citing a number of press directors, John Willinsky points to the continued cost of scholarly publication due to the continued need for press staff. As Kenneth Price has so adroitly stated, digital materials might be free for those who use the project,”But
free stuff comes from somewhere, and it is rarely, if ever, free to produce.” Price, who has successfully guided the *Walt Whitman Archive* through numerous grant applications, emphasizes that “the creation of digital editions is expensive, and the demand for external and internal grant support always exceeds the money available. We need to reflect, then, on what gets funded and what does not and to take care not to institute an even more narrowly conceived canon than in the past.” NEH grants, which fund a majority of the digital literary projects, are often judged by impact and impact is most recognized by numbers of hits to the site. We know that canonical writers have a greater chance of a large following than little known writers. So, it follows that a good number of the archive and edition projects are focused on canonical writers, such as Rossetti, Melville, or Whitman. “What is marketable in terms of digitalization projects are . . . editorial projects covering great vast expanses of materials of varying natures,” states Margaret Ezell, “and that digitalization in this instances ‘sells’ because of its ability to include ‘everything’ and link it in a comprehensible searchable and sustainable system.” This system tends to exclude outlying texts, including the “single, uniquely existing manuscript.” Such structural concerns need to remain at the fore of our work with digitization, since without careful attention to issues of the canon, we risk creating a digital canon that looks more like a New Critical canon than a contemporary, diverse body of literature.

What I am arguing seems to move against current digital humanities trends regarding preservation. As Kathleen Fitzpatrick and Matthew Kirschenbaum, among others, have argued, digital forms are far more stable than most critics presume. Forensic computing is able to recover data etched on hard drives and viral preservation has preserved games, texts like *Agrippa*, and even some of the recovered texts this chapter has reviewed. However, viral preservation or even planned obsolescence are not appropriate for certain kinds of materials, such as recovered texts. While
we might bemoan the loss of Bethany Nowviskie’s early digital project, *The John Keats Hypermedia Archive*, we can easily access Keats materials in print or even online. The majority of the early digital recovery work discussed in this chapter, however, is not easily accessible. Few of these texts are in print. A few more are available on for-profit databases or on microfilm, but most are available only with a return to the one or two libraries that own the original physical copy of the book, journal, or newspaper. The targeted preservation of such materials is not technical, as Matthew Kirschenbaum has taught us, but social, and as a field we need to put such materials at the front of the preservation queue.

Central to our efforts is the need to build a community invested in the consideration of the digital literary canon. We need an initial targeted triage focused on digital recovery projects that reproduce materials not accessible in print or open access digital form. Successful models exist. We eventually need a thoughtful, targeted approach to recovery and preservation, but until we have resolved the larger digital preservation problem we might think in short-term increments. In “The Future of Preserving the Past,” Dan Cohen writes, “Worrying too much about the long-term fate of digital materials in many ways puts the cart ahead of the horse . . . Instead of worrying about long-term preservation, most of us should focus on acquiring the materials in jeopardy in the first place and on shorter-term preservation horizons, 5 to 10 years, through well-known and effective techniques such as frequent backups stored in multiple locations and transferring files regularly to new storage media, such as from aging floppy discs to DVD-ROMs. If we do not have the artifacts to begin with, we will never be able to transfer them to one of the more permanent digital archives being created by the technologists.” Cohen’s suggested approach won’t resolve our preservation problems, but he reminds us that simple, short-term solutions might provide a stopgap measure for the hemorrhaging of recovery projects.
Existing organizations, including digital humanities centers, collectives, and existing projects, have preserved recovery projects. For example, the University of Nebraska’s Center for Digital Research has provided support to projects that otherwise might have disappeared, such as Marta Werner’s *Radical Scatters*. NINES and its sister organizations under the Advanced Resource Consortium (ARC) have established networks that foster and support various forms of digital work including recovery. Another avenue for increasing preservation is to incorporate legacy projects into current digital projects. Given the collaborative, extensible, interoperable nature of digital work, there is no reason why contemporary projects might not be connected to legacy projects. In effect, communities of interest are best able to engage with discussions of and experiments to preserve crucial early digital recovery texts. The communities of interest must also engage in the shifting technological and economic requirements of digital text recovery. Consortia might be built to target the quickly disappearing early digital recovery projects. Infrastructures might be proposed to preserve instances of the various projects. Such approaches are best built into existing societies or organizations, particularly those organizations with an activist bent. For example, the College Language Association has long focused on African diaspora texts. Should the organization choose to act as a hub for recovery and preservation efforts, the effort would serve as an extension of work already underway.

One of the powerful things about the early period of digital literary studies is the DIY approach that many scholars embraced, the sheer joy and freedom of bringing important texts to the larger scholarly community. As we move from simple HTML sites to TEI and visualization projects, as we move from individual or small collective projects to larger team projects, from nonbudgeted projects to large, externally funded projects, we see fewer scholars working with digital textual recovery. This should concern digital humanists, and we should, accordingly, begin to
strategize how we might reverse this trend. Small steps are underway. We need to examine the canon that we, as digital humanists, are constructing, a canon that skews toward traditional texts and excludes crucial work by women, people of color, and the GLBTQ community. We need to reinvigorate the spirit of previous scholars who believed that textual recovery was crucial to their work, who saw the digital as a way to enact changes in the canon. If, as Jerome McGann suggests, “the entirety of our cultural inheritance will be transformed and reedited in digital forms,”61 then we must ensure that our representation of culture does not exclude work by those writers previously excluded.