Proofs of Genius
Gailey, Amanda

Published by University of Michigan Press

Gailey, Amanda.
Proofs of Genius: Collected Editions from the American Revolution to the Digital Age.

For additional information about this book
https://muse.jhu.edu/book/42621

For content related to this chapter
https://muse.jhu.edu/related_content?type=book&id=1667127
Despite widespread upheavals in literary theory and the academy in general that began in the later half of the twentieth century, editing has largely remained committed to the bedrock principles that have historically supported major collected editions. Through the theory wars and into the digital age, editions organized around authorial genius more or less remain the gold standard—some editions offer glimpses of alternative ways of presenting literature, but entrenched editorial methods, funding structures, and the persistent belief that editing should be an apolitical undertaking have converged as conservative influences on the field, resulting in a state of digital literary editing that does not radically depart from the picture of scholarly editing several decades ago. This chapter examines the state of editing American literature in the final decades of the twentieth century and into the present in order to address some factors that are inhibiting more heterogeneous approaches and to suggest some possible ways forward that more directly speak to a variety of critical interests in texts.

1. “Wimmin” and Literary Study

In the 1970s and 1980s, the rift between the fields of scholarly editing and literary theory seemed nearly insuperable. It was exacerbated by the slow pace of editorial production, which required the continued, devoted application of editorial methodology and goals established sometimes decades earlier, while competing, highly politicized schools of theory developed at lightning speed outside its doors: the evolution of editorial theory and its products in the decades between the emergence of Greg-Bowers and the
advent of the Web is like a time-lapse film of a glacier inching along, as the
days and nights of the theory wars blink by around it. Theory was quickly
changing, fashionable, expressly political; by comparison the slow and fo-
cused work of textual editing seemed hopelessly stodgy. Theory became de-
defined by its assaults on tradition; editing by its adherence to it. By the 1980s,
editors—already seen as pedantic, second-string academics as early as the
early 1960s—were widely viewed as academic support staff.

Among the unfortunate consequences of the marginalization of edit-
ing was the falling out of fashion of bibliographic methods, long a staple
of English studies. Accordingly, as literary scholars became interested in
canon expansion—primarily seeking to include more women and mi-
nority authors in classrooms, textbooks, and scholarship—their work did
not benefit as much from careful editorial and bibliographic work, leaving
much foundational scholarship on newly appreciated texts undone to
this day.1 Literary critics working on texts by canonical authors benefited,
even if only silently, from bibliographies and well-researched editions often
conceived prior to the New Left’s shake-up of American universities, while
work on women and minority authors often required textual recovery at
a time when careful editorial work was unfashionable and institutionally
unrewarded. It was more interesting and institutionally profitable to dis-
cover or assert challenges to the predominantly white and male canon than
to engage in the less flashy work of creating rigorous editions of recovered
authors. Cumulatively, the differences between editions of men and women
or minority writers affirmed long-standing notions about gender, race, cre-
vativity, and intellectual property: male authors who benefited from editions
that foregrounded their solitary genius and their sacrosanct intentions to-
ward their art loomed at the center of the canon, while women authors
stood in a collective mass with picket signs on its boundaries. In Ameri-
can literature, the major exception to this treatment was Emily Dickinson,
whose supposed asexuality and aversion to the literary marketplace made
her a suitable honorary inductee, a woman author whose genius could be
appreciated without inviting in her more combative countrywomen. The
receptions of a few canon-challenging volumes published in the late 1970s
and 1980s demonstrate how controversial editing could be, even in an age
of radical challenges to literary tradition, when editions contested the natu-
ralized ontologies of literary editing.

When the Norton Anthology of Literature by Women was published in
1985 it was meant as an answer to the Norton Anthology of English Litera-
ture, which at the time devoted 96% of its pages to works by male authors. The *Norton Anthology of English Literature*, organized around linguistic and national identity, was also essentially an anthology of literature by men. Yet when Sandra Gilbert and Susan Gubar sought to enrich the canon by collecting and presenting overlooked women authors, not by squeezing a couple more into the hallowed standard *Norton* but by offering an entire, two-thousand-page collection of a variety of texts by women, they were lambasted by several critics for inserting a political thesis into the collection of literature. The most prominent criticism came from Gail Godwin in the *New York Times*, who claimed that the anthology “distorts or undermines the achievements of individual artists” by grouping them all in one volume and thereby implying a singular aesthetic tradition or identity. In the *New Republic*, Denis Donoghue complained about “feminism’s agenda in literary studies” and the reductiveness of the volume even as he conflated the beliefs of many varied international feminists with the anthology’s editors. Donoghue saw the anthology as an assemblage of “documentary evidence to support a case against men—or the world,” and argued that it is a book suited for a sociology class, not a literature class, where inclusion should not be based on historical value or political reform but purely on aesthetic grounds. “It does not even pretend to select its material according to the criteria of literary criticism,” he complained.

In the *National Review*’s sarcastically titled “Wimmin Against Literature,” Jeffrey Hart disparaged the editors, calling them “feminists who refuse to submit themselves to the experience of actually reading, and who subject great and vibrant works to their own grievance machine.” He argued, “These feminist professors do not hesitate to impose their ideology” on us, and chew up literature in “their mincing machine.” Even more measured critics, such as Phyllis Rose writing for the *Atlantic*, faulted the anthology for its political thesis. Rose, who saw the anthology as making an important contribution to the field of literature, nevertheless thought it was “too political for a good anthology of literature.” In one of the most revealing discussions of literary ontology to appear in discussions of editing, Rose writes:

“Literature by women” is in fact a controversial category, not self-evidently valid like “English literature” or “American literature.” Who could imagine refusing to be included in the *Norton Anthology of English Literature*? Who, on the other hand, could take seriously a *Norton Anthology of Literature by Men*?
The editors dodge this important issue and in doing so render the anthology covertly polemical. Of course, the editors are not at all being covertly polemical, given the clear intentions of the anthology and the ire it raised in critics. And some feminist critics would likely have countered at the time that we already have a *Norton Anthology of Literature by Men*, and it is called the *Norton Anthology of English Literature*. If any volume is guilty of covert polemics, it would seem to be the volume claiming to stand for a millennium of literature in the English language that almost entirely excludes texts by women. The covert claim of that anthology is that male authors wrote almost everything worth reading for a thousand years. While Rose finds the categories of “English” and “American” uncontroversial, recent postcolonial challenges to those categories reveal that the lack of controversy was not due to self-evident validity but to naturalized assumptions about national and linguistic histories.

Not long before and after the *Norton Anthology of Literature by Women* was published, two other books intended to expand the canon were published, but their critical receptions were quite different from that of the *Norton*. In 1987, Jean Yellin released her edition of *Incidents in the Life of a Slave Girl*, which was the first edition to credit Harriet Jacobs, and not the pseudonym Linda Brent, as the author. The volume was a major scholarly accomplishment: prior to Yellin’s efforts, critics had neglected this autobiography under the belief that it was written by the white abolitionist (and Jacobs’s benefactor) Lydia Maria Child; even critics who believed Jacobs was the author doubted the veracity of her extraordinary tale of sexual harassment, estrangement from her children, and eight years of confinement in a coffin-like garret. But Yellin’s careful study of the historical record, including a cache of Jacobs’s letters, allowed her to prove that Jacobs was the author and that the book was accurate. *Incidents* quickly became an important addition to American literary history; it offered a rare, woman’s perspective on slavery and complemented texts such as *Narrative of the Life of Frederick Douglass*.

The most critical review of Yellin’s edition was published in the *Washington Post*, where reviewer Theodore Rosengarten grumbled mildly about Yellin’s “feminist convictions,” her tendency to “use big words where little words would do better,” and “the introduction which is weighted down by ideology,” but nevertheless saw the book as important, compelling, and nicely annotated. Wayne Lionel Aponte, writing for the *Nation*, similar-
ly concentrated on the content of Jacobs's narrative, but concluded that it “represents an early attempt to establish an American sisterhood,” a claim he does not acknowledge that Yellin herself makes in the introduction, and just the kind of assertion of a tradition that offended several reviewers of the *Norton Anthology of Literature by Women.* Jacobs could be viewed as part of a feminine tradition, so long as the book seemed to stand alone and the reader drew that conclusion for himself; if it were part of the organization of the volume or asserted in the interstitial editorial content, the volume would be seen as distastefully polemical.

Similarly, in 1979, critical response to a selected edition of Zora Neale Hurston's work was generally positive, though limited. Alice Walker edited *I Love Myself When I Am Laughing . . . And Then Again When I Am Looking Mean and Impressive* through the Feminist Press, which had been responsible for bringing Rebecca Harding Davis's *Life in the Iron Mills* and Charlotte Perkins Gilman's “The Yellow Wall Paper” back into public view. Walker had written about her efforts to recover Hurston, who died in poverty and obscurity, in a piece in *Ms. Magazine,* and *I Love Myself* included selections of Hurston's writings that Walker thought were the best and most representative of the array of Hurston's talents. Despite its overt feminist aim to recover a neglected woman author, the volume failed to attract the controversy that surrounded the *Norton.* It was a *New York Times* recommended book, and a more extensive review in the *Times* praised its editing. Randall Kennedy opened his review by explaining that Hurston was “the leading lady of black American letters between 1920 and 1950,” and then emphasized that the volume is selective: it includes “excerpts” of her autobiography, “selections” from her novels, and is altogether a “representative sample of Zora Neale Hurston's many talents.” Kennedy was writing as an advocate for the study of literature by black authors, but nevertheless felt that a *selected* edition seemed appropriate for such an important figure. He specifically praised Walker for resisting “sociological and political reductionism” in her insistence that “in assessing [Hurston's] place in our literary tradition we should look first to the work itself and only secondarily to the context from which that work emerged.”

Some of the criticism launched at canon-expanding editors may be valid: the editorial commentary in anthologies and other collections sometimes seeks the lowest common denominators among authors and their works in order to assert a tradition or to justify the basis of the collection. However, the common complaints about this kind of editing left few op-
tions for noncontroversial editing of women authors. Some critics claimed it was overly reductive to collect them, because that would assert a common tradition among so many varied voices—though the same critics sometimes also complained that more authors were not included. Sometimes critics grumbled that on the one hand minor women authors were not significant enough to warrant recovery, and on the other hand feminist editors were arrogant in daring to elucidate one of the honorary inductees to an otherwise generally male canon. Single volumes of texts by women, such as *Incidents in the Life of a Slave Girl*, could be interpreted as establishing “an American sisterhood,” but a volume that collected such texts together would be lambasted as polemical and inappropriate. Even more sympathetic critics such as Kennedy seemed to view the “leading lady” of black literature as most suited to a representative sampling, despite the fact that white, male authors were frequently the subjects of multivolume, decades-long collected editions that rigorously examined even their scraps and juvenilia.

Editorial efforts to challenge the canon or to innovate the collection and presentation of literature resemble larger patterns in capitalist relations. Texts constitute a form of cultural capital; they stand in for their authors, and by extension their authors’ social place. When certain texts from a vast, multifaceted literary history are collected and propagated, it suggests not only the value of the kinds of people who wrote those texts—the standouts from a mass of voices—but also affirms and normalizes the ways we think about authorship, intellectual output, and property rights. The picture of editing in the United States in the 1970s and 1980s, a time of continued political tumult on college campuses, shows something happening in our presentation of texts that looks a lot like larger skirmishes over cultural, economic, and political power. Textual representation is concentrated in the hands of a few greats, who enjoy luxurious editions bearing the signs of conspicuous—though increasingly unsustainable—labor, whose collations and textual apparatus, like a gentleman’s clean collar and brushed vest, indicated that the wearer was the sort of person upon whom much labor should be expended. Meanwhile, editors are aggressively dissuaded from organizing the disenfranchised rabble, and collections of women laying claim to cultural power are seen as politically charged as a picket line.

The least controversial alternative for editions of recovered or canon-expanding authors seemed to be editions that only presented selections of a single author, either a single, stand-alone text or limited selections from the corpus—these did not claim any essentializing traditions or similarities,
and they did not claim through comprehensiveness that the author was of extraordinary stature.

All literary editions assert a thesis. For some the thesis is so naturalized that it goes undetected by most readers, while others violate long-held notions of what makes literature important. In the 1970s and 1980s, as the energy behind the Greg-Bowers method was waning, as scholarly editing became a marginalized academic pursuit, and as the theory wars diverted attention to different kinds of questions about texts, editions became a relatively undertheorized influence on how the academy approached texts. Most texts were edited, published, and reviewed without the focus on editorial method or rigor of previous decades, and texts that did attract attention for their controversial editing tended to be attacked and defended by critics who were not particularly interested in the history of editing or in a sophisticated analysis of literary ontology. Consequently, some editions’ theses stood as invisible and unproblematic: (1) established, canonical authors are of such stature that large-scale collections of their every word, including “discarded writings,” are valuable scholarly tools; (2) collections of literature by many different authors in which over 90% of the texts are written by men may stand as representative of a national literature, literature of a period, or literature in its entirety; (3) a selected edition of a minority woman author is ideal for capturing her talents at no loss to the complexity of her art or career. A stand-alone single text is also fine and can be read as an exemplar of a demographic, but the editor should not suggest as much via the volume. However, one thesis is visible and available to challenge: A collection of women authors inserts polemics and sociology into literature and reduces women to representatives of their gender.

2. COLLECTED EDITIONS ON THE EVE OF THE DIGITAL REVOLUTION

Collected editions that were under way in the 1980s were facing their own problems. Some editions were proving to be impracticable, and several projects were abandoned midstream, including editions of Stephen Crane, Hamlin Garland, and Washington Irving. Others, such as the massive collected edition of Whitman begun in 1963, were stretching into their third decade. Unfortunately, these major print undertakings could not easily adjust course, even when new materials were discovered—the Whitman
edition provides a case study for how the lack of flexibility and revision in large-scaled editions could cause editing projects to collapse under their own weight. *The Collected Writings of Walt Whitman* were designed to meet a list of desiderata that the prominent scholar Sculley Bradley had called for—generally, Bradley had hoped for an edition that would allow scholars to trace Whitman's poetic growth, as evidenced through his revision process, across the editions of *Leaves of Grass* (of which he counted nine). His interest in the poet's workshop was compatible with Greg-Bowers, though, since he still viewed the final authorially sanctioned edition as the culmination of authorial intent: previous variants led teleologically to it, and unsanctioned variants were corruptions. Bradley wanted a “true variorum” that would possess four essential characteristics: (1) it should be a faithful copy of the authoritative text; (2) it should track all authorial variants, including manuscript ones, on the same page as the transcription of the principal text; (3) it should include editorial annotations explaining the background and meaning of the text, also on the same page; (4) it should include a rich appendix collecting as much information as is available about sources for the work, the context of the work, important criticism of the work, and commentary on how the variants relate to each other.

This was a very tall order, especially considering that Whitman left behind thousands of manuscripts—though Bradley may not have been aware of the full extent of the manuscript record at the time he envisioned the edition. These goals would prove to be absolutely untenable for a print edition, and would, in fact, not even begin to coalesce until a digital edition was begun over fifty years later. Bradley’s directives were colored by the critical milieu of the midcentury. The textual variants were of interest to him not as alternate performances of the poems—to borrow Cristanne Miller’s phrase in her discussion of Dickinson—but almost exclusively because they were evidence of the poet’s workshop and demonstrated how *Leaves of Grass* evolved into a unified whole by 1881. Consequently, he takes as a given that the deathbed should be considered the authoritative edition, and believes that the record of Whitman’s revisions will show how he “achieved three principle objectives: improved taste by the suppression of indelicacies, solecisms, and coinages; increased force by verbal revision and strengthened imagery; and a greater mastery of rhythm” as well as increasing the “unity of the whole.” In fact, he believed that a variorum that made visible the evolution of *Leaves of Grass* into its unified final edition would vindicate nineteenth-century critics who less than favorably reviewed earlier editions,
since “the work could not possibly have had anything like the compelling wholeness which it has for the modern reader before the edition of 1881.”

The foundational assumption that the deathbed edition should be considered authoritative—an assumption that to some extent persists even today—caused some difficulties for Bradley’s proposed variorum, since he also wanted readers to consider the poems and variants chronologically. Whitman did not organize *Leaves of Grass* chronologically, and any given poem may have undergone any number of revisions over the decades prior to 1891. This complicated chronology and revision history would be accounted for, in Bradley’s plan, through editorial notes, though he did not seem to anticipate how cumbersome such notes would be.

Bradley’s plea for a variorum is important because it articulated “a primary need of American scholarship” that Bradley himself would help remedy over the next twenty-two years. In 1955, New York University Press announced plans to publish a new collected writings of Walt Whitman that would be released under the editorial supervision of Gay Wilson Allen, who had proposed the project to the Press. At the time of its announcement, the *Collected Writings of Walt Whitman* was to include almost everything that Bradley had asked for fourteen years earlier, and some materials that he did not. As Floyd Stovall explained in 1962, the *Collected Writings* were slatted to comprise fourteen volumes: “four volumes of correspondence with annotations, two volumes of major prose works published in Whitman’s lifetime with variant readings, a variorum edition of *Leaves of Grass* in two volumes, a reader’s edition of *Leaves of Grass* in one volume, a bibliography in one volume, the collected notebooks, diaries, and prose fragments in two volumes, the fiction and early verse in one volume, and the journalistic writings in one volume.”

To take on such a daunting task, Allen had enlisted the help of some of the most prominent Whitman scholars in the world and the major collector of Whitman’s manuscripts to serve on the project’s advisory board: Bradley, Roger Asselineau, Harold Blodgett, Charles Feinberg, Clarence Gohdes, Emory Holloway, Rollo G. Silver, and Floyd Stovall. Allen also began recruiting various scholars—some on the advisory board—to take on specific areas of the *Collected Writings*. All of these scholars’ work on the project was paid for only insofar as their home institutions allowed them time for the work—it was a condition of the Press’s involvement in such an ambitious and unwieldy undertaking that it would not fund the editorial work. Given the daunting nature of the project and the lack of direct financial compen-
sation, it was impressive that Allen was able to secure the long-term commitments from leading scholars to edit portions of the project, and spoke to the respectability of editing as a scholarly activity before it fell out of favor.

Unfortunately, almost as soon as work began, all involved were forced to reconsider the feasibility of the project’s high aims. Essentially, the partial failure of the *Collected Writings* resulted from compounding difficulties: first, the editors were initially unaware of the sheer volume of materials that comprised these various aspects of Whitman’s writings. Second, the allocation of different aspects of the project to different editors at different institutions over the decades—necessary to begin to tackle the materials at hand—caused problems in organization and continuity that would eventually prove severe and sooner or later affect almost every subproject of the *Collected Writings*. The most acute organizational crisis occurred within *The Correspondence*, which was originally conceived by its editor, Edwin Haviland Miller, as a two-volume project, organized chronologically. Even as Miller worked, more and more unpublished letters surfaced, rendering his volumes incomplete. The first and second volumes appeared in 1961, making them the first parts of the *Collected Writings* to come off the press. In the mid-1960s, Miller released a third and a fourth volume, and by 1969 he was forced to publish a fifth volume containing an “addenda” of over sixty letters that were not incorporated into the chronology of the first volumes. Earlier, in 1963, Allen wrote of Miller’s editorial problems:

Possibly some letters may be discovered in obscure places even after all the volumes of the *Correspondence* are in print, but the number is not likely to be large, and two thousand letters should make an edition extensive enough for the needs of Whitman students and scholars. But this situation shows how nearly impossible it is to have an absolutely complete edition.18

Allen’s comments show that even only a few years into the *Collected Writings* the editors were forced to reconsider their admirable goals of gathering all of Whitman’s writings, and that Allen himself was willing to consider the *Correspondence* adequate as an *almost* complete correspondence. Unfortunately, Allen underestimated both the number of letters that would continue to surface in coming years and the desires of students and scholars to see all these letters in print—three supplemental volumes were called for, the most recent in 2004. Now divided in a badly broken sequence over seven volumes published over the course of forty-three years, the *Correspondence*
has failed to provide the organized, complete reference that Miller and the other editors had originally hoped.

Similar problems plagued other areas of the *Collected Writings*. Whitman’s journalism, which defies easy and complete collection, since Whitman often wrote anonymously and some newspapers issues are not extant, had seemed to have been abandoned decades into the project. In fact, though, just in 1998 and 2003 the project’s editors released the first two of a projected five volumes, but there is still no sign of the other three. Even if all five volumes materialize, the *Journalism*, like the *Correspondence*, will almost certainly require supplements.

Another confusing aspect of the *Collected Writings* is the division between William White’s three-volume *Daybooks and Notebooks*, published in 1977–78 by NYUP, and Edward Grier’s six-volume *Notebooks and Unpublished Prose Manuscripts*, published in 1984 by NYUP. As Ed Folsom has written, there was simply “no good rationale” for dividing these aspects of the project, and in fact the division resulted from what Folsom describes as “disagreements over editing styles and timetables.” That such disagreements resulted in a nonsensical division of materials and a discrepancy in editorial styles between the two projects is a comment on the general disarray and management problems within the *Collected Writings*.

Perhaps the most interesting editorial issues of the whole project were raised by Bradley’s and Harold Blodgett’s treatment of *Leaves of Grass*. Although Bradley’s first requirement of a proper variorum had initially been that “it must be an exact edition” of the chosen copy-text, the variorum that he and Blodgett edited was no such thing. At the inception of the edition, Bradley and Allen both believed in the tenet popularized by Greg and Bowers that the latest versions of poems should be considered authoritative: in his 1963 overview of the project, Allen asked rhetorically, “Which provides the best text for the definitive *Writings*? Should it not be the 1892 ‘deathbed’ text, which Whitman authorized as his choice, commanding in a preface that henceforth no other be reprinted?” Bradley had also argued for using the Deathbed as the copy-text:

The desire to show the progressive unfolding of Whitman’s mind and art has several times evoked the suggestion that the chronological order be employed in a text of the *Leaves*. But for a definitive edition the purely chronological order must be rejected, first, because of the mechanical difficulty presented by changes of order in the later editions of early texts; secondly, because it would
not result in the publication of any single text as Whitman left it; and thirdly, because it would be an injudicious violation of the author’s specific injunction to follow the text of 1892 in future editions.22

Indeed, Bradley and Blodgett followed this directive in preparing the Comprehensive Reader’s Edition of Leaves of Grass, for which they adhered to final authorial intent, but when preparing the Variorum seemed to violate it in favor of a hodgepodge approach that is difficult to defend. Understandably though unfortunately, Bradley and Blodgett would have to scrap plans to include manuscript drafts in their variorum because the quantity of materials was simply too staggering. Including these variants would likely have exposed the Variorum to the same problems as the Correspondence, and was further complicated by the task of sorting out the complex and sometimes indiscernible relationships among printed texts and manuscript drafts. However, Bradley’s violation of his own proposed ordering scheme is quite mysterious. Rather than producing an “exact edition” of the 1892 text, the Variorum presented the deathbed versions of each poem in the order in which they debuted in Leaves of Grass. First it printed the twelve poems of the 1855 edition, but in their Deathbed form, then the poems new to the 1856 edition, but in their Deathbed form, and so on. The result was a Deathbed-chronological hybrid, which, as Bradley himself explained years earlier, “did not result in the publication of any single text as Whitman left it,” and certainly violated Whitman’s “specific injunction to follow the text of 1892,” except in segments. The inauthenticity is underscored by the editorial apparatus, which notes any changes among the versions leading up to the Deathbed in footnotes.

The volumes comprising the Collected Writings are now spread over five decades of publication and three publishers. They lack a centralized index or table of contents, so that a researcher looking for documents—letters, notebook entries, Leaves of Grass versions—pertaining to any particular poem must first learn what the edition makes available, then search through the volumes for each component project, adjusting for differences in editorial methods. Seen as a rather scattered and inconsistent printing of materials, one of the primary values of the Collected Writings seems to be preservational: while it is severely crippled by organizational problems, at least the edition presents documents, such as Whitman’s letters, that would be difficult for most scholars to access otherwise, and that are vulnerable to damage or loss. In fact, it is precisely this aspect of the Collected Writings
that has opened it to charges that it misrepresents Whitman’s corpus by reproducing the texts that he did not see fit to publish. Folsom has articulated this position strongly by explaining that when reprinted, the sheer bulk of Whitman’s abandoned writings overwhelm his published writings and give the impression that most of what Whitman produced was rubbish. He writes:

The bulk of the Collected Writings will end up containing poor writing or hack writing, really the Discarded Writings . . . . In any case, it’s safe to say that when the project is completed, no other major author will be represented by such a large and luxurious Collected Writings so filled with intrinsically bad and flat writing. Readers will perhaps need to be reminded that such a result is not Whitman’s fault.23

Such concerns echo complaints about collecting “remains” that stretch back a century, such as when the Athenaeum complained in 1903 about “disinterring” abandoned materials in order to publish “books in which a great name vouches for material of little or no absolute value.”24 Certainly this problem is particularly marked in Whitman, who wrote incessantly and in later years seemed constitutionally incapable of throwing anything out. Even if we are to view such volumes as a sort of repository, where materials of questionable value are reprinted for storage should scholars ever find them useful, it is important to note what the Collected Writings did not preserve. Two important bodies of documents in Whitman scholarship—and the ones most in need of curating—his poetry manuscripts and his periodical publications, were omitted from the Variorum. Consequently, the Collected Writings suffers on both counts: it reprints so much discarded writing as to distort Whitman’s work, but leaves out many of the documents that are of most interest to students of Whitman.

The Collected Writings was a colossal undertaking, and the editors approached it with noble goals. While the edition failed on many counts, it has provided not only useful materials for Whitman scholarship, but also hard-learned lessons on editing. In the early days of digital editing John Unsworth argued, “If an electronic scholarly project can’t fail and doesn’t produce new ignorance, then it isn’t worth a damn,” because “reporting and analyzing failure in any research activity, humanistic or scientific” is essential to producing new knowledge and to helping a community of scholars learn from mistakes. Certainly the value of such analysis is not limited to
electronic editions. The *Collected Writings* provided an object lesson for editors in a transitional age: given even widespread, interinstitutional cooperation, impeccable credentials, strict devotion to the work, and decades to complete it in, the editors of this print edition simply could not adequately treat the materials within the constraints of the medium. Print is unforgiving and difficult to fix. When new materials surface, they cannot be seamlessly integrated into a volume already in print. Print also lends itself to singular, linear presentations of materials. If an editor is trying to reconcile all of the demands that Bradley outlined in his list of desiderata, he will, as the *Collected Writings* shows, find print an uncooperative partner. Finally, the *Collected Writings* demonstrates that on-the-fly funding is shaky grounds upon which to build a monumental edition, even if it is the only option.

Early into the project, Allen was encouraged by this “million dollar project without a million dollars,” specifically in the NYUP’s willingness to undertake the publication of the volumes without subsidy. The project’s sprawl and slow pace, though, taxed the NYUP’s commitment to the venture, and editors of later volumes were not assured a publisher—consequently, the edition spans not only decades and the work of dozens of collaborators, but also three different publishers, NYUP, Iowa University Press, and Peter Lang, none of whom maintain sole ownership of the project.

Most importantly, despite the careful attention and labor that went into the *Collected Writings*, it was following a road map created in the critical and technological climate of the midcentury, and by the time the later volumes were released, critical interest in Whitman had shifted, and the editorial commitments of the edition were not optimal for supporting a new generation of scholarship. As literary studies continued to look in different directions—at linguistic sign systems, at the author’s gender, race, and sexual orientation, at the socioeconomic contexts of the work—really, anywhere except at the stand-alone internal coherence of a text or the conscious intentions of its author—scholarly editing and theory seemed increasingly estranged. Anyone worried about the state of scholarly editing at the time would probably have been more concerned by what was not said about it than what was—even the rare controversies were really about how basic editorial decisions aligned with camps in the theory wars rather than careful considerations of how to select, transcribe, and annotate texts.

Then, in 1983, Jerome McGann published *A Critique of Modern Textual Criticism*, which, along with some of his other work, directly addressed ways in which midcentury textual criticism and its attendant editorial ap-
approaches neglected aspects of textual histories that should be of primary interest to critics. The Greg-Bowers method bore an ambivalent relationship with the materiality of texts, depending on intense examination of material textual history only in order to reject it for its corruptions: the goal was to systematically study the material record in order to identify and purge corruptions and recover a purified authorial intent. McGann was also interested in the material history of the text, but instead of rejecting textual changes as a work moved from context to context, he embraced these “fault lines” as precisely what should interest the textual scholar. Rather than seeking a singular intended work, the textual scholar ought to view texts as diachronic, social, embodied things, whose reconfigurations are evidence of their collaborative creation and social meaning. For McGann, the job of the critic was not to ignore or transcend the material record, but to account for it.

McGann’s vision of what an edition might accommodate was largely unrealized until the advent of the Web—before then, print editions that foregrounded the material record and multiple forms of a work typically would have proven infeasibly expensive to produce. But the flexible display, expansibility, and imaging capabilities of the Web made editions with goals in line with McGann’s possible. Perhaps the most successful digital edition of the works of an American author—if not of any author—is the Walt Whitman Archive, begun in 1995 by Kenneth M. Price and Ed Folsom, who continue to edit it in its much-expanded current form.

3. DIGITAL COLLECTED EDITIONS OF AMERICAN LITERATURE

The Web greatly expanded the capacity and improved the organizational flexibility of collected editions, allowing them to fold in multiple publication contexts and accommodate a richer understanding of the author’s work than print collections feasibly could. The Whitman Archive is an instructive case study in digital collection: a mature edition now in its eighteenth year, it has been the U.S. digital editing project most successful at obtaining grant money and creating a sustainable funding model, now with a two-million-dollar endowment. The Whitman Archive was strongly influenced by and, in turn, influenced the early development of standards for the digital editing of American literature. In many ways the Whitman Archive realized the shift in focus that McGann’s 1983 Critique called for—the
Archive is centrally interested in the varying forms that Whitman’s writing took as he revised it for different publication contexts. Whitman himself, who was a sophisticated manipulator of the periodical press, a skilled book designer, and a bit of a hoarder, offers a particularly rich record for such a study—when he died he left thousands of manuscripts, six editions and multiple states of his major work, *Leaves of Grass*, and a decades-long trail of periodical publications.

A rigorous digital edition of Whitman in the 1990s was made possible by the Institute for Advanced Technology in the Humanities (IATH) at the University of Virginia. The Institute was founded in 1993 after the computer science department at UVA decided to use a million dollars donated by IBM to support digital work in the humanities. Planning committees at the university made crucial early decisions about IATH that would make it a powerful supporter of early digital editorial work and also help lock in technological approaches that are still prevalent in digital editing today.

McGann, then a professor at UVA, helped found IATH, making editorial work one of the Institute’s central concerns and consequently one of the oldest and most theorized strains of digital humanities. IATH’s first director, John Unsworth, was committed to making its projects web deliverable, a departure from many early digital humanities approaches that committed much of this early, rigorous digital editorial work to technologies such as TEI (Text Encoding Initiative), which was at the time expressed through SGML (Standard Generalized Markup Language), but would transform into the XML-based methodologies that continue to form the basis for most digital editorial work.

One of the early editorial projects fostered by IATH was the *Whitman Archive*. When Price and Folsom set out to edit the *Whitman Archive* in the 1990s, they began as scholars of Whitman, not students of editorial theory: the impetus behind the *Archive* was to allow readers to understand the vast record of Whitman materials that had been admirably tackled but inadequately treated by the ongoing print collected edition. Most readers in 1995 would have had no way to access some of the editions of *Leaves of Grass*, Whitman’s periodical poems remained uncollected, and Whitman’s manuscript record was not much better organized than how he left it in 1892—dispersed over more than three dozen repositories, idiosyncratically and incompletely cataloged, and inaccessible to many scholars who had an interest in the decaying drafts. So the editors’ primary goal was to collect these materials and make them accessible, enabled by the Web’s capacity
for image display and the possibility of continually growing, inserting, and rearranging materials as their order became clear.

The first goals of the Archive were to collect the different editions of *Leaves of Grass* authorized by Whitman, along with a full record of his manuscripts, his periodical poetry, all known photographs of the poet, contemporaneous criticism, and a bibliography of current scholarship. Each of these efforts was a major aid to readers of Whitman, but the ongoing manuscript collection has been a revelation: prior to digital collection, no single scholar had seen all of these materials, much less had the ability to examine and reexamine them at her leisure. Now hosted by the Center for Digital Research in the Humanities at the University of Nebraska–Lincoln, the *Whitman Archive* continues to build these components of the Archive, but has expanded its parameters to include materials such as translations of Whitman and thousands of letters he penned as an amanuensis for the attorney general after the Civil War. Whitman has become the center, the organizing principle, for a continually expanding digital world—one that powerfully rearticulates the author function as the ontological basis for literary editing.

Other digital editions of American literature are similarly organized. The *Willa Cather Archive* is a more dynamic and extensive digital edition superseding the print collected edition; the *Dickinson Electronic Archives 2*, an overhauled version of the *Dickinson Electronic Archives*, presents transcriptions and scans of Dickinson's manuscripts alongside critical and pedagogical materials in a digital environment that encourages users to think of the materials in “exhibits” to which they can contribute commentary. Less extensive but useful archives are available for a handful of other American authors, including Sarah Orne Jewett, Charles Brockden Brown, and Mark Twain. Some projects focusing on American literature are organized on bases other than authorship, such as *Wright American Fiction*, which offers scans and searchable text for the thousands of books catalogued in Lyle Wright's bibliography of fiction published in the United States between 1851 and 1875. Space constraints preclude a full reckoning of the digital collections of American literature, but what is perhaps most striking about the current landscape of digital American literature collections is how few of them rival midcentury print editions in scope and depth—many projects look at many, many texts, and fewer projects pay close, methodical attention to their editing of the texts, but very few indeed—the *Walt Whitman Archive*, the *Willa Cather Archive*, *Dickinson Electronic Archives 2*, eventual-
ly the *Mark Twain Papers Project*—have undertaken close textual attention to many texts, and these are all organized around authorial genius.

Indeed, for these authors, the genius thesis seems unproblematic—we are interested in literature by Whitman and Dickinson largely because it is by Whitman and Dickinson. However, for many other texts, authorship, authorial intent, and composition processes are not fundamental concerns—scholars may be primarily interested in their similarities to other texts, their material histories, their reception, their performance, and so on. Digital editing offers developed methods for treating the digital equivalent of a collected edition, but few standard ways to rigorously study and present texts whose primary value to readers does not lie in a deep appreciation for the author. This isn’t terribly surprising—as McGann has recently argued, given that the foundational act of textual scholarship has been the establishment of a reliable document, and given that reliability has historically been defined in terms of authorial intent and authority, authorship has primacy in the editorial enterprise. Consequently, digital American literature reproduces the divide that afflicted print editing in the 1970s and 1980s: a handful of canonical authors are the beneficiaries of rigorous, well-funded, long-term editorial labor, while many other authors—disproportionately minorities—are studied as a largely undifferentiated mass in large-scale digitization projects.

The technological underpinnings of digital editions are partially responsible for this. Any serious digital editorial project that hopes to secure conventional grant funding will almost certainly avail itself of the Text Encoding Initiative (TEI), the de facto standard for digital editing in the humanities. As a standard for the careful treatment of literary texts in a digital environment, TEI has no competitor. It has the benefit not only of over twenty years of development by an international community of users, but also speaks to traditions in editing that reach back to Greg-Bowers and even beyond in the European and philological traditions.

To understand how TEI is ideally suited to conventional, author-centered collected editions or editions that study a single work, it is helpful to review how TEI works. TEI is based on a model of textuality known as OHCO, or Ordered Hierarchy of Content Objects. OHCO approaches text as a series of objects with content—say, a paragraph or a sentence—whose order is important—the third paragraph should come after the second—and which are structured hierarchically—sentences are in paragraphs that are in chapters that are in a book. OHCO is a conceptual model describing the structure of texts, and we can consider its merits.
or shortcomings independent of understanding how it is technologically implemented—in many cases, OHCO seems to be a simple and accurate way of describing a text: this book is comprised of chapters, which contain sections, which contain paragraphs, which contain sentences, which contain words. These components are hierarchical, their order is inviolable, and we understand them implicitly as competent readers. Putting aside considerations of textuality that do not lend themselves well to OHCO—which we will consider later on—it is crucial to note that OHCO is born out of and supports a specific textual technology, the markup languages in the SGML (Standard Generalized Markup Language) family. SGML was invented in the late 1960s and 1970s by computer scientists, and has now evolved into XML (Extensible Markup Language), with the related HTML (Hypertext Markup Language), the bread and butter of the Web, as one of its offshoots. It is worth mentioning the origin of SGML when reflecting on its suitability for modeling current critical interests in texts, but for the rest of this discussion I will refer to this language as XML, since that is its present incarnation.

XML puts OHCO into practice. Consider a simple (and very abbreviated) example, in which XML markup—the material in brackets, called tags—describes the components of the text (fig. 10).

```
<book>
  <chapter>
    <subtitle>CHAPTER I.</subtitle>
    <paragraph>Late in the afternoon of a chilly day in February, two gentlemen were sitting alone over their wine, in a well-furnished dining parlor, in the town of P—, in Kentucky. There were no servants present, and the gentlemen, with chairs closely approaching, seemed to be discussing some subject with great earnestness. One of the parties, however, when critically examined, did not seem, strictly speaking, to come under the species. He was a short, thick-set man, with coarse, commonplace features, and that swaggering air of pretension which marks a low man who is trying to elbow his way upward in the world.</paragraph>
  </chapter>
  <chapter>
    <subtitle>CHAPTER II.</subtitle>
    <subtitle>The Mother</subtitle>
    <paragraph>Eliza had been brought up by her mistress, from girlhood, as a petted and indulged favorite.</paragraph>
  </chapter>
</book>
```

Fig 10. {~?~no caption}
The book tag encompasses all of the content. This is one requirement of XML: every XML file must have a single tag, or root element, that contains all of the content. The book contains two chapters; each of those chapters contains a title, subtitle, and paragraphs. It is helpful to think of the tags as nesting Tupperware containers—one large container includes a few smaller ones, which in turn include even smaller ones, and so forth. We can have a few medium ones sitting side by side in the larger one, and the medium ones can have smaller ones sitting side by side within it. But we could not put a large one inside a medium one—a book inside a chapter.

Another requirement of XML involves the lids of these containers. A closing tag matches the name of the element that opened it, but with a slash in front of it: </chapter> closes <chapter>, for example. To continue with the Tupperware analogy, we must close the smaller tags before we put on the lid of the larger container that holds them. You can’t close a medium container and then close the small ones inside it. Similarly, we can’t close a chapter and then close a paragraph within it: this is prohibited by the syntax of XML. So if we try this we will violate the rules of XML and make our file unusable by most programs: <chapter><paragraph></chapter></paragraph>

Fig. 11. Emily Dickinson, [This quiet dust was Gentlemen and Ladies], MS #65, Amherst College Special Collections.
paragraph>. Instead, we must nest the tags, closing content objects before we close their parent content objects. This may seem like a minor point, but it has had profound implications for how we treat texts in a digital environment. For example, consider a text that may interest us both for its literary content and its bibliographic features, such as the manuscript poem by Emily Dickinson shown in figure 11, which spans a recto and verso.

If we try to describe both of these interests in a single XML file, we will quickly run into nesting problems. The recto holds only part of the poem, requiring us to close the recto container before we close its child container, the poem shown in figure 12.

```xml
<leaf>
  <recto>
    <poem>
      <line>This quiet dust was Gentlemen and Ladies</line>
      <line>And Lads and Girls</line>
      <line>Was Laughter and Ability and Sighing</line>
      <line>And Frocks and Curls</line>
      <line>This passive place a summer's nimble mansion</line>
      <line>Where bloom and bees</line>
    </poem>
  </recto>
  <verso>
    <line>Fulfilled their Oriental Circuit</line>
    <line>Then ceased, like These</line>
  </verso>
</leaf>

Fig 12.
```

The squiggly line is where my XML editing program, Oxygen, is calling my attention to the error—I have closed the recto tag before closing the poem tag. The physical bibliography of the document and the literary content of the document are really two different hierarchies, and if we try to blend them in XML, which is incapable of housing conflicting hierarchies, we run into a fatal error. This is a very simple example, but the problem comes up time and again when people consider literary materials with fresh eyes, and are forced to choose one of these hierarchies as the basis for the encoding, wrenching the secondary hierarchy into the file with jury-rigged tagging.

Unlike HTML, which provides tags for document description—so `<p>` means paragraph, `<ol>` means an ordered list, and `<li>` indicates a list item
within it—XML provides no vocabulary, only the syntactical requirements that tags are structured in certain ways, that they nest, and a few other technical necessities for computer processing. But the terms an individual user chooses to tag texts is up to her. This is a great descriptive strength of XML: unlike its diminutive relative HTML, which allows for an easily understood but limited and consistent way of describing document structure for web display, XML empowers different user communities to choose for themselves what data are to be described and using what terms. The possibilities are as vast as human interest—for example, SpacecraftML is an agreed-upon set of tags that an aerospace community uses to describe and exchange data; Music Markup Language and the Music Encoding Initiative each attempt to define an ontology of music notation and data and convey it in XML: in each case, a user community defines tags and where they can be applied. Similarly, the Text Encoding Initiative provides a set of tags for describing texts and rules delineating their use. By using a TEI schema—a file that ensures your XML only uses approved tags and in appropriate places—you agree to describe your texts using TEI’s vocabulary. A major structural division is a div; a person’s name is a persName; a page break is a pb—TEI compliance means that you will use these terms to mean these things, and you will not be able to hierarchically meddle with them (no divs in a persName, for example). Practically speaking, it also recommends to encoders that the types of textual features TEI identifies are the ones that their projects should find salient. While it is possible to build an extension to TEI—a customized tagging set that describes features untreated by TEI—doing so requires much more intellectual and technical labor than using TEI in an orthodox way. This is the strength and the weakness of standards: they make like-minded work much easier to accomplish and communicate, but overall can have a stultifying influence on innovation.

TEI’s tag set has been developed by a community of users—who propose needed encoding, discuss its merits, and publish new releases of the available encoding—for decades, and it reflects an impressive range of scholarly interest in texts. Many TEI tags are designed to describe linguistic phenomenon; others are specialized for specific kinds of literary form, such as drama and poetry. Textual scholars involved with TEI are developing subsets of tags for encoding genetic editions, allowing editors to systematically describe complex compositional histories within manuscripts and sets of manuscripts. A module of tags designed for encoding document structure has also been proposed, which would allow editors to prioritize
the bibliographic structure of a document over its literary content, so that pages, gatherings, and so on are the hierarchy prioritized in the XML, and are not forced into broken, awkward encoding in order to preserve the continuity of the encoding of the literary form.

What has developed is an encoding standard that works quite well for many projects, but in particular those projects that are primarily interested in single authors or single texts. TEI originated and incubated at a time when even open-minded humanities departments had fairly rigid expectations for the presentation of texts, valuing naturalized assumptions about how literature is created, how it should be organized, and its quarantine from contentious politics. Because of both the structure of OHCO/XML and the origins of TEI, the Text Encoding Initiative has developed in a way that offers most support to projects that are interested in single works, single authors, or very formalist markup. These projects, including influential ones hosted by IATH in the 1990s, contributed desiderata in the most formative years of TEI, when the goals of digital editions were primarily to remediate so as to avoid the shortcomings of their print forbears. XML itself, with its prohibition against conflicting hierarchies and nesting structure, lends itself to a view of literature that considers texts (and sometimes works) as discrete, self-contained entities with their own internal coherence. It fundamentally supports a view of literature that says here is a work; here are the texts that comprise it; here are the formal components of the texts; here are interesting sites of composition and revision. At each stage the encoding typically looks inward. In those cases in which editors want to invest in encoding that points out from the individual text—for example, to similar passages in other texts or to continuations in serially printed text—those cross-references can be exceptionally awkward and labor-intensive to pull off. When deciding which of the many TEI tags to implement in the encoding, editors would be wise, in the name of sanity and project management, to develop guidelines that offer tags and encoding strategies for similar features, those formal or compositional attributes that arise in many texts, rather than treating each text idiosyncratically, and this approach will naturally lend itself to looking for habitual features—the structural features and recurring compositional practices found across the collected texts. So the technological requirements, available tag set, and project policies that look for recurrent features in order to minimize workload are all ideally suited to a project that collects the work of a single author with relatively consistent habits that can be viewed as intended or patterned acts, creating
an oeuvre of texts that reflect his genius and evidence self-contained formal elements.

So, although TEI is a rich resource and a testament to the power of collaborative, international work in the humanities, and although it is an invaluable tool for much editorial work, it implicitly supports projects that examine authors and texts as self-contained units more than ones that don’t. Similarly, it supports less controversial, less politicized views of texts, such as examinations of their formal and linguistic structures, over more contentious kinds of interpretive claims, such as content or thematic elements.\textsuperscript{28} These are all understandable and possibly inevitable outcomes of a successful international, interdisciplinary editorial methodology. However, if we are to consider creative new directions for the development of editing, we should bear in mind the influence of the monolithic status of TEI, however excellent it has proven at supporting certain kinds of work.

The limiting influence of TEI has been exacerbated by early hopes that it would allow interoperability: though many involved in TEI have attempted to debunk this myth for several years, TEI has held some allure to organizations who hoped that encoding with TEI would make their XML interoperable with other projects—the dream has been that if we’re all using the same vocabulary to describe texts, we ought to be able, with only minor adjustments, to combine them into new or aggregate collections.\textsuperscript{29} However, the freedom to pick and choose TEI subsets, to apply them idiosyncratically, and to add customized tags, has made interoperability chimeric, except for projects using very light markup—sometimes adhering to a minimalist subset of tagging called TEI Tite—or stripping out much beyond light markup. In those cases TEI loses almost all of its expressive power and becomes HTML on steroids, limited to the lowest common denominator and relegated to describing mostly superficial structural information and metadata. As Google and other private entities have digitized books at a pace most academic undertakings could never compete with, some within the digital humanities community, looking for ways to define a new place for text encoding, have suggested that supporting or joining these mass digitization projects with light markup is the way we should go. During a recent discussion within the TEI, members identified one potential area for growth as claiming a seat in Big Data, and specifically “ensuring that large amounts of lightly but consistently encoded texts (e.g., TEI Tite) are generated and made publicly available, perhaps in a central repository or at least through some centrally coordinated portal.”\textsuperscript{30} It remains unclear whether
this will become a priority for TEI, but if it does, it would mark a significant shift away from what many editors view as a core mission—that our job is careful, expert attention to texts and their salient features, and not the fast and light output of a product. Such a move from artisan to factory text encoding would be an extreme and discomfiting escalation of reductive and homogenizing influences that have perhaps always been present in TEI.

Projects that look at many different texts by different authors may not find such a natural fit in the ontological commitments of XML/TEI/OHCO. Consequently, some such projects, such as Nineteenth-Century African American Women Writers or Wright American Fiction, tend to simply produce lightly encoded texts, including only minimal structural encoding, rightly assuming that keyword searching the content is likelier to assist researchers than in-depth formal encoding. Some such projects have invested in more detailed TEI encoding, but arguably this labor does not support what is most interesting about the materials. Projects focusing on noncanonical texts or unorthodox collections of materials seem to face a choice between light TEI markup, in-depth markup designed to support a different interest in the texts, or designing a better mousetrap. The path of least resistance is to use light, predefined markup. Cumulatively, though, this can subtly reinforce age-old differences in how editors and readers confront texts by canonical—and disproportionately white and male—authors, and texts whose authors are not beheld so reverentially. The former generally receive close, expert attention; the latter are collected in casually edited collections.

4. ALTERNATIVE PRINCIPLES FOR COLLECTION:
A CASE STUDY

For a few years I’ve been working on a digital collection of American literature that has brought some of these issues to the fore. The Tar Baby and the Tomahawk is a digital collection that aims to provide a heavily annotated resource for scholars and students of literature, history, African American studies, visual communication, and education to examine how adults wanted children to think about race during Jim Crow. When complete, The Tar Baby and the Tomahawk will include literature, illustrations, and popular-culture materials featuring characters of different races primarily intended for a juvenile audience roughly between Reconstruction and 1939, when
The Yearling became the first book published for young readers to win a Pulitzer Prize. In some cases, the authorship of this material is collaborative, corporate, or altogether unknown. What binds the materials together is that they all provide evidence of how popular media marketed to children or families during the period of Jim Crow helped to assert, reinforce, and, occasionally, diminish racial inequity.

The materials seem almost defined by their unsuitability for a conventional scholarly editing project. Virtually all of the materials are noncanonical or decanonized texts; their authorship is frequently slippery or of little interest—many of the texts are derivative works, sometimes by one or more uncredited authors; and their afterlives in unauthorized or appropriated forms are often more significant to our study than their pristine origins. The conventional ways of conceptualizing a scholarly edition or digital archive, along with the methods and technologies developed around conventional editions, while eminently reasonable for certain types of materials, have proven unwieldy and inappropriate for ours.

Take, for instance, the works of Joel Chandler Harris, which comprise a large portion of the materials we’ve collected. If editorial work moved faster than glaciers, there would likely be a sprawling, multivolume print edition of Joel Chandler Harris’s works in every American research library. In the mid-1950s, when editors undertook the preparation of modern editions of so many American authors, Harris may have seemed like a contender for such work. At the time of his death in 1908 he was one of the most popular American authors, and Theodore Roosevelt published a letter mourning the loss of a national treasure, declaring Harris’s fiction the most likely of American works to endure. In the 1920s, over a decade after Harris’s death, a survey of U.S. high school and college teachers showed that Harris was considered one of the five most important authors in the United States.

Harris published dozens of novels and collections of short stories over his literary career, which stretched from 1881 to 1908, but his most popular works were his Uncle Remus books, in which a loyal former slave tells folk stories to an unnamed white child. Today, Harris’s phonetic spellings of nineteenth-century middle Georgian African American dialect, which strike many readers as difficult or offensive, and his paternalistic approach to Remus and implicit nostalgia for the antebellum days, have helped boot him from the canon. Perhaps most damaging to Harris’s reputation, however, was the appearance of Song of the South. Produced by Walt Disney and released by RKO Radio Pictures in 1947, the film trades on the worst aspects
of the Harris tales: the more complex Remus character of Harris’s books is caricatured into a hyperbolically loyal and happy servant to white children, and today, over sixty years after its premiere, Disney views the film as an embarrassment and refuses to rerelease it.

Though Harris himself has fallen into obscurity, and though we no longer view his works as unambiguously good, his texts and their reception suggest interesting and instructive patterns about late nineteenth-century American attitudes toward race and culture. To study these patterns, though, requires a perspective very different from the view afforded by an author- or work-centered edition. Harris’s texts are entertaining to read and offer fascinating glimpses into U.S. racial history, but certainly the best treatment of them is not presenting Harris as he may have once been viewed: a highly canonical genius whose compositional process elicits scholarly curiosity or admiration. It is the reception of Harris’s works, how they were pirated, appropriated into popular culture, and generally diffused into American racial consciousness that is of interest, and studying these is not well supported by methods developed for author-centered editions.

The project’s interest in Harris’s Uncle Remus tales is in many ways similar to ongoing scholarship on Harriet Beecher Stowe’s *Uncle Tom’s Cabin*, which in recent years has enjoyed attention from scholars who examine its cultural significance through the many sympathetic and hostile appropriations it spurred: parodies, homages, minstrel shows, and so on. Stowe’s and Harris’s central characters followed a very similar path through American and international culture. Both Uncle Tom and Uncle Remus were born in the pages of American periodicals—Stowe’s in the *National Era* and Harris’s in the *Atlanta Journal Constitution*. Both Harris and Stowe wrote well-intended but sentimental and paternalistic depictions of black characters in an effort to effect social change. Stowe, of course, sought the end of slavery, and Harris, writing at the height of Jim Crow in the American South, hoped to humanize African Americans to his white readers in an effort to end the epidemic of lynchings. Stowe’s and Harris’s political goals, however flawed they now may seem in execution, were progressive in their times and elicited both admiration and hostility.

Stowe’s and Harris’s moral earnestness made them easy targets for parodists and hacks. Just as Stowe’s characters were quickly subsumed into consumer culture and the minstrel stage, Uncle Remus was featured in pirated publications, abridgments, household decorations, advertisements, coloring books, menus, postcards, and corporate logos. The cultural reach
of Harris’s characters far exceeded his grasp. It is this reach that is of most interest about Harris, though the author’s intentions and direct creations delimit the scope of most editorial undertakings. One of the few digital projects to trace the cultural reconfigurations of an American literary text is Stephen Railton’s *Uncle Tom’s Cabin in American Culture*, which is full of rich material and editorial insights. Railton’s site includes images of several editions of the book; Stowe’s own *Key to Uncle Tom’s Cabin* and adaptation of the novel for the stage; “pretexts” that illustrate the culture into which Stowe was publishing her work; numerous reviews of the book; adaptations of the book for children; 3D manipulable images of memorabilia based on *Uncle Tom’s Cabin*; and images of the book’s eventual transformation for stage and screen. All of this traces the text’s trajectory from its roots in mid-nineteenth-century abolitionism to its transformation into a twentieth-century industry of racial degradation and caricature.

The problem with the site is that few developed methods available to digital literary scholarship support this kind of approach to texts. For example, Railton is rightfully interested in the covers of early editions of *Uncle Tom’s Cabin*. From the golden age of conspicuous literary consumption, these covers speak to what publishers and purchasers found important and beautiful about the text. They worked as the marketable face of the book, and if the ubiquitous still-uncut pages of nineteenth-century gift editions tell us anything, it is that the face of the book most frequently held the owner’s interest. Yet TEI does not have a single tag, much less a developed module, for describing the outside of a book—something that is relevant to many literary projects. The *Walt Whitman Archive*, for example, omits cover information from its encoding, since it is unsupported by TEI. However, the 1856 edition of *Leaves of Grass* was widely known at the time for the adulatory Emerson quote—“I greet you at the beginning of a great career”—that Whitman brazenly reprinted on the spine. If the *Archive* wanted to include this, scholars would have to develop an ad hoc TEI extension. One of *The Tar Baby and the Tomahawk*’s customizations to TEI was to create a simple tag for the illustrator of a book, a person as important to much of children’s literature as the author. The orthodox TEI approach to treating illustrators is, bizarrely, as a specialized kind of editing—that is, TEI recommends encoding the illustrator this way

<editor role=“illustrator”>A. B. Frost</editor>, 
as though the illustrator were merely another corrupting or altering influence on a pristine text.

The lack of an <illustrator> tag in TEI reflects a bias in the way the standard describes texts. The literary structures of a text, but not its relationship to other texts, nor its collaborative aspects, are robustly supported by the TEI tag set. A project designed around a core of stand-alone texts written by one author is much more suited to this than a thematically oriented project, a collection of collaborative works or texts by different authors, or one that examines textual transmission and appropriation. In the case of the author-centered archive, the design of the project matches up with the nesting structures of TEI: the identity of the author contains individual texts that are comprised of chapters, which hold paragraphs, and so on. But if a project hopes to examine the kinds of connections and cultural dispersions that Railton's work addresses, for example, TEI begins to seem like a hindrance: a significant investment of time and labor into tagging that supports little of the intellectual interest of the project. Yet when Railton's site on *Uncle Tom's Cabin* was updated to conform to technological standards, the focus was on migrating the data into TEI-compliant XML. Given that compliance with TEI is a de facto requirement for serious—that is, rigorous and funded—digital editorial projects, the author-centered model is not only implicitly encouraged by the current granting system but is also clearly the path of least resistance for anyone with an interest in digitally editing American literature. For many projects, though, it seems that we lack good editorial standards for describing what is of most value: how, in Joseph Grigely's words, those “post-textual reconfigurations of a work tell us something about the personality of a culture.”

**5. Alternative Methodologies**

TEI is indispensable for many digital editing tasks, but was simply not intended to note the relationships among ephemera such as a children's menu, a do-it-yourself comic, or the other many merchandise tie-ins and other cultural goods generated by the publication of the Uncle Remus tales. Similarly, it is not suited to capture what is of interest in the array of materials pertaining to *Uncle Tom's Cabin* that Railton has collected, or the larger patterns of co-opting Uncle Tom and Uncle Remus, which ranged from
friendly retellings to degrading parodies. We have a markup vocabulary for noting intricate structures within a single text, but lack a graceful way of noting patterns or relationships to which a text belongs.

Consider Harris’s first story collection, *Uncle Remus: His Songs and Sayings*, from 1880. This collection was arguably his most influential, and its story about Brer Rabbit’s encounter with a tar baby would be Harris’s most widely recognizable tale. *The Tar Baby and the Tomahawk* has collected a variety of materials, both texts and nontextual objects, that were directly or indirectly based on Harris’s Uncle Remus tales and that demonstrate the tales’ cultural influence. In some cases an item was clearly influenced by a particular book or story, as with “Tar-Baby Nails.” Other items make use of characters that recur in several of his collections. Figure 13 illustrates the relationships among many materials we have encountered while working on Harris’s texts.36

Viewed this way, it is easy to see how this now fairly obscure text functioned as a cultural vector. However, the diagram in figure 13 does not capture many details of the relationships that are of interest to literary scholarship. A few details that would likely be of wide interest include the following:

1. What kind of thing is the appropriation?
2. Was the appropriation authorized by Harris?
3. Is the appropriation sympathetic to or critical of the aims of the source text?
4. What aspects of the source text were appropriated: characters, illustrations, plot, direct language?

There are more questions we would want to ask, of course, but these few may serve as examples. A model that could support these kinds of queries would need to be ontologically nuanced; that is, it would need to be a carefully constructed formal system that describes the entities, their properties, and their relationships to each other. Figure 14 attempts to represent how a character in one object (a book) is appropriated into an illustration in another object (a menu). I have kept the figure simple and only included here some representative entities and attributes.

A few technologies could allow us to implement this model. Relational databases appeal to some literary archives that attempt to foreground relational structures,37 though commonly used database technologies such as Structured Query Language (SQL) are too inelastic to form the basis of a distributed, flexible system for describing textual entities in different
Figure 13. Diagram of appropriations of *Uncle Remus: His Songs and Sayings*.

Figure 14. Diagram of textual transmission.
archive and editions. Semantic web technologies such as RDF (Resource Description Framework) and OWL (Web Ontology Language) are proving to be viable, flexible alternatives to relational databases and can accommodate an expanding and diverse set of claims about entities and their properties and relationships. While admittedly difficult to learn and technically implement in comparison to inline markup or limited relational database technologies, the formal specifications of the semantic web, expressible in XML, allow projects to represent particular entities and relationships among them within sophisticated ontologies. Semantic web technologies provide ontological rules and a syntax for expressing them. For example, if we were to describe a set of relationships involved in textual transference in OWL, we could claim the following:

1. Uncle Remus is a character in “The Wonderful Tar Baby Story”
2. Joel Chandler Harris wrote “The Wonderful Tar Baby Story”
3. A toy advertisement features Uncle Remus.

The inferential rules we can describe through OWL would later allow us to derive from this information that this toy advertisement features a character that Joel Chandler Harris created, even though we never directly stated that. This is a simple example, but because the technologies allow us to continually provide the ontology with new entities, it can grow complex very quickly and ultimately allow us to derive sophisticated information about entities inferred from a multitude of single, separately entered statements about them. Essentially, the data-interchange standards of the semantic web are much more expansible than brittle relational databases.

Semantic web technologies, combined with a constrained vocabulary tailored to the purpose, could help literary scholars and cultural historians track many different kinds of textual content and describe intertextual relationships. Further, the flexibility and expansibility of semantic web technologies could allow for aggregation: individual projects describing historically or thematically similar materials could combine records and allow for the mapping or graphing of relationships among materials across projects. But currently, the lack of a developed method for tracking cultural transmission is a significant lacuna in digital literary scholarship. As we seek to build upon the rich tradition of the collected edition, an ontological framework for describing intertextual relationships could prove fruitful.

Digital humanists have already been exploring the use of semantic web technologies to enable other ways of studying American literature. In 2003,
NINES (Networked Interface for Nineteenth-century Electronic Scholarship) was established as both an organization comprised of transatlantic scholars who evaluate materials related to the study of the British and American long nineteenth centuries, as well as a “federated collective” of the peer-reviewed materials that the organization vets. NINES tries to combat the siloing of digital scholarship that is encouraged by many grant programs, institutional structures, and the XML-based technologies that lend themselves to self-contained (and often author-centric) scholarship. Its primary tool to resist editorial isolationism is Collex, an open-source application that Bethany Nowviskie, one of its developers, described as “facilitating primary interpretive gestures of exploration and explanation in a broad and socially networked manner.” Collex was designed to use “facets,” in which loose, nonhierarchical claims about relationships, such as creation date, author, and so on are housed in the metadata, and folksonomic tagging, in which users assign free-form descriptive labels to digital objects, in order to connect and exhibit materials that may never have cohabited the same scholarly archive. This semantic web approach was meant to allow NINES to present materials in much more flexible ways than is enabled by strict, prescriptive metadata, and to more closely track the varied interpretive and associative interests of readers. Collex has the potential to offer innovative approaches to collection by using RDF (but not OWL) in combination with a large body of digital materials. However, in practice, the RDF metadata supplied by individual projects tend to record orthodox information—adding rich, innovative metadata for Collex, metadata that break with the organizational principles of the digital project, has typically been an afterthought. By the time a contributing project has been built and is ready to be peer reviewed, the metadata that would make Collex realize its potential are understandably not a priority for editors who designed the resource along different and more conventional lines. The progress of Collex suggests that once a project is designed as a silo there is little reward for the labor of reorganization, and that innovative principles of collection ought to be part of the initial design of individual digital archives and editions.

In the last couple of years, new developments in digital collection portend exciting new ways we might think about collecting and presenting our cultural inheritance. The Digital Public Library of America gathers digital resources into one portal, where users can create exhibitions or build apps. Another digital undertaking, Viral Texts: Mapping Networks of Reprinting in 19th-Century Newspapers and Magazines, which is still in early stages of development, seeks to examine and theorize how some nineteenth-
century texts get widely reprinted. Projects like these evidence a growing interest in diversifying the ways in which we gather and represent American literature.

The digital representation of American literature continues and complicates a trajectory we could trace at least as far back as the origins of the country itself, when collected editions became a way to showcase individual genius and the cultural potential of an entire community or nation. Through the centuries, collected editions of American authors have proven robust and flexible vehicles, capable of helping form a national literary identity, representing emerging conceptions of intellectual labor and property, serving as enduring memorials for amateur talent, erecting public monuments to greatness, and motivating an academic industry. The collected edition is a venerable genre, an aspirational form for authors and editors, both the bedrock and apex of the canon. Its success is evident in how undetectably it has tended to work. The collected edition has clearly buoyed a certain view of authorship, promoting the image of the genius in his workshop directly communicating to the individual reader. This view of authorship has historically favored—and continues to favor—authors whose race, class, and gender conform to long-held notions of what venerable geniuses look like. Since the mid-twentieth century, the rigorous, developed methods for presenting carefully edited texts have been based on the gold standard of the collected edition. Through the latter half of twentieth century and into the twenty-first, these standards seem to have come at the expense of the development of other methods for careful collection, description, and presentation of literature that would better express critical interests that have developed over the last fifty years. The modern academy understands the importance of historicizing the texts we read, yet even as curricular models and critical studies move away from single-author models, and as digital collections spring up around any number of critical interests, editorial methodologies to express even simple interpretive questions, narratological elements, and intertextual phenomena remain largely undeveloped. The collected edition will surely remain an important genre for organizing texts, but as digital scholarship and readerly interests look beyond siloed collections, we need to develop alternative methods for rigorously representing American texts.