Chapter Title: Can Information Be Unfettered? Race and the New Digital Humanities Canon

Chapter Author(s): AMY E. EARHART

Book Title: Debates in the Digital Humanities

Book Editor(s): Matthew K. Gold

Published by: University of Minnesota Press

Stable URL: https://www.jstor.org/stable/10.5749/j.ctttv8hq.21

Can Information Be Unfettered? Race and the New Digital Humanities Canon

AMY E. EARHART

In the 1990s, the rallying cry of proponents of the Internet was the democratization of knowledge made possible by the developing technological infrastructure. Lost or excluded texts began to be published on the net, some developed by scholars, others by fans, and still others by libraries and museums. I remember the possibilities that these materials offered for the literary scholar. I could create a website for students that linked the recovered e-text of Harriet Wilson's *Our Nig*, period images of slaves, and the variety of African American cultural and historical documents found on the then-fledgling Schomburg Research Center website. The seemingly expansive materials for use on the web were far more complete than materials found in print anthologies or other such course materials. For scholars interested in reinserting writers of color into critical discussions, the recovery efforts were a boon. We imagined that the free access to materials on the web would allow those previously cut off from intellectual capital to gain materials and knowledge that might be leveraged to change the social position of people of color. The new space of the Internet would allow those who had been silenced to have a voice. Hypertext theorist Jay David Bolter 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" (233). As the 1990s drew to a close, and the number of digitally recovered texts seemed to grow each day, Bolter's prediction seemed correct. However, a review of digitized materials production and the current treatment of race in the digital canon suggests that Bolter's hopes have not been realized.

I want to focus my discussion by examining a subset of the digital humanities, digital texts.[1] I'm interested in the digital work being produced by those associated with academia and those with strong connections to traditional humanities fields including history, literature, classics, art history, and archeology, among others. My

focus includes pay-walled scholarly production, such as the excellent *Clotel* project published by Virginia's Rotunda Press, and open-access materials but excludes large-scale digital projects produced by for-profit publishers, such as Gale-Cengage, or nonscholarly produced projects, such as Google Books. I am also most interested in projects that make something. Here I would like to echo Stephen Ramsay's recent argument that "Digital Humanities is about building things." While Ramsay has come under fire for his insistence on the applied nature of digital humanities, the history of digital humanities reveals the centrality of building to the field. In fact, scholars invested in early work on race in digital humanities insisted on building editions and digital texts as an activist intervention in the closed canon. While we should continue to explore tool building, visualization, and data mining as crucial areas within digital humanities, the narrow digital canon should remind us why we cannot stop digital edition work.

While those invested in digital text production should continue to flesh out the digital canon, other areas of digital humanities, such as tool building and visualization, should also be invested in the investigation of canon on their work. For example, the Metadata Offer New Knowledge (MONK) project has harnessed materials from Documenting the American South, Early American Fiction, Early English Books Online (EEBO), Eighteenth Century Collections Online (ECCO), Nineteenth-Century Fiction, Shakespeare, and Wright American Fiction 1850–75. While the purpose of MONK is not text recovery but visual analysis, a broad understanding of the literature of this period is only as good as the data from which the analysis draws. In the case of MONK, a quick search reveals that texts by Sojourner Truth, Sui Sin Far, and Maria Christina Mena—authors of color included in most standard anthologies of American literature—are absent.[2] Add to this MONK's claim that "for users of public domain materials, MONK provides quite good coverage of 19th century American fiction," and we are reminded that a more direct analysis of the position of race in digital humanities work is necessary (Unsworth and Muller, 2). We shouldn't be surprised at the lack of certain texts used in the MONK project, as the digitized humanities corpora is scant and the project's primary goal was to develop a prototype of data-mining work rather than an inclusive data set. However, the emphasis on "good coverage" should concern digital humanists. As a field where collaborative teams are able to produce better results than that of the lone scholar, it is important to include project participants that can help speak to the importance of cultural criticism. Without careful and systematic analysis of our digital canons, we not only reproduce antiquated understandings of the canon but also reify them through our technological imprimatur.

Unlike related fields, digital humanities has historically deemphasized theoretical examination of the digital utilizing cultural studies frameworks. Those working within rhetoric, media, and communication, and particularly those working in game studies, have constructed a body of scholarly work that interrogates the theoretical implications of race construction within technology.[3] With superb work being

produced by scholars such as Lisa Nakamura, Beth Kolko, and Tara McPherson, these fields have begun the difficult work of theorizing the way in which technology impacts the digital object. In digital humanities, however, we have much theoretical work to do in the selection of materials and application of digital tools to them.

To understand the current position of race and digital humanities work, we must turn to the emergence of the World Wide Web. As the web began to gain popularity in the 1990s, it was portrayed as an idealized, democratic, and free space. *Time* magazine's 1994 story "Battle for the Soul of the Internet" is indicative of the understanding of the newly popular space, uncolonized and free. The original users of the web, scientists, computer geeks, and hackers, according to the story, were battling against the corporate market intent on invading their open space. Advocates of the free web were interested in three ideas: "1) Access to computers should be unlimited and total; 2) All information should be free; 3) Mistrust authority and promote decentralization," all designed to allow "bubbles" of information to rise from the bottom, sowing "seeds of revolutionary change" ("Battle for the Soul of the Internet"). Scholars, too, began to see the net as a space that altered power structures. As Paul Delany notes,

> 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. (Childers and Delaney)

The revolutionary power of the net was based on the belief that the open digital environment was unpoliced and unregulated, open to all who wanted to participate. The low cost, the "small hardwave investment" that Delany points to, was also crucial, as it allowed scholars to produce their materials more cheaply, increasing the types and numbers of texts available. In his 1996 essay "Principles for Electronics Archives, Scholarly Editions, and Tutorials," Peter Shillingsburg concurs it "eventually will cost less to produce and therefore, one assumes, to purchase a compact disk than it costs to produce and purchase Hans Gabler's edition of *Ulysses*" (25). While, in hindsight, we have rejected this view as too simplistic, it was a common refrain in the early, heady days of digital recovery, and the decentralization and shifting power structures, as related by these statements, became part of the narrative and the mythology that in no small part drove the digital literary studies recovery projects that grew in the 1990s.

For scholars interested in reworking the canon, the web seemed an unfettered space that allowed the scholar direct control of what would be digitized and how

it would be presented. Susan Fraiman, director of the Orlando Project, lauded the expansive digital environment, remarking that "what is new in the twenty-first century, however, is that now the guest list of history-making women is electronic—and there are always more seats at the table" (143). The belief in the web as a space in which the canon might be broken 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. This 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" (Fraistat, Jones, and Stahmer). The ease of publication, identified by the editors, was what allowed the broad range of small-scale recovery projects to explode around the web in this early period. The insistence on the web's ability to build new canons, of an applied approach to the digital tied to a theoretical model, is a hallmark of past and current digital humanities work.

The importance of building tools for digital work, common to and controversial in contemporary digital humanities, has deep roots in this early recovery work. The applied/theoretical model voiced by digital humanists also fits historically into the work of cultural studies scholars prior to the mainstream Internet. For example, Jean Fagan Yellin's work on Harriet Jacobs's *Incidents in the Life of a Slave Girl* took various trajectories of inquiry, but her 1987 Harvard University Press edition, the edition that brought the important volume to the center of the African American literary tradition, was an applied piece set within her larger scholarly body of work. In recovering the text and conducting the painstaking research required to do so, Yellin has followed a model familiar to those working with digital projects—application driven scholarship. This model of building, grown from the cultural studies recovery tradition, would expand as additional scholars began to explore the digital as a tool by which to recover texts by writers of color.

During the 1990s and early 2000s, in many ways the most productive age of digital recovery to date, projects fell into two distinctive scholarly groups. One group was the small-scale project in which scholars worked individually or as small collectives. These projects, including *The Charles Chesnutt Archive, Voices from the Gaps,* and *The Online Archive of Nineteenth-Century U.S. Women's Writings*, were mostly unfunded and produced outside of digital humanities centers or libraries. The second type of project was produced by e-text centers, digital humanities centers, or libraries and museums. Individual scholarly participation was less central to such work. These projects include the *Emory Women Writers Resource Project*; the various projects produced by the Virginia Center for Digital History, including the *Valley of the Shadow* and *Race and Place: An African-American Community in the Jim Crow South*; and the digital *Schomburg Library of Nineteenth-Century Black Women Writers.*

Of these two types of projects, the dominance of small-scale recovery efforts nurtured by an individual scholar who wanted to bring lost texts to scholarly and public attention is surprising. Simple HTML projects such as *The Charles Chesnutt Archive*, by Stephanie Browner; *Voices from the Gaps*, by Toni McNaron and Carol Miller; and *The Online Archive of Nineteenth-Century U.S. Women's Writings*, by Glynis Carr, were developed without the support of a digital humanities center, technological collaborators, or external funding. Some projects, such as *The Charles Chesnutt Archive*, were positioned as undergraduate teaching and learning tools, with undergraduate student partners in the recovery process.

Alan Liu's *Voice of the Shuttle* provides a good measure of the huge number of early recovery projects focused on literature and history written by and about people of color. A quick perusal of "The Minority Studies" section, however, reveals that a tremendous number of the projects have become lost. For example, of the six sites listed in "General Resources in Minority Literature," half cannot be located, suggesting that they have been removed or lost. The same trend is found with other projects listed on the site. While only 50 percent of the projects in the "General Resources in Chicano/Latino Literature" section are still online, other areas, such as Asian American literature, have a higher percentage of active projects.[4] 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.

Cocurrent to the DIY projects were institutional initiatives focused on bringing lost texts to view. Much of this work occurred within e-text centers, such as Rutgers University's Center for Electronic Texts in the Humanities (CETH); fledgling digital humanities centers, such as the University of Virginia's Institute for Advanced Technology in the Humanities (IATH); and museums and libraries, including the New York Public Library. With limited exceptions, a majority of the early projects reinforced canonical bias. Catherine Decker argues that the canon crops up in these projects because of their funding and institutional affiliations: "The reasons for the canonicity of the bulk of the electronic texts available on the web are hardly elusive: most of the large textbases are located at and funded by major universities (or grants to university scholars that include a cut for the university at which the project is situated)." Martha Nell Smith extends this contention and argues that digital humanities developed as a space to which practitioners hoped to flee 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" (4). If Smith is correct, then we not only have a selection problem in digital humanities but also have a historical structural problem that might be more difficult to reverse.

One only needs to review the current work in digital literary studies to see that we have not escaped the traditional canon by turning to new methods of publication. The proliferation of early projects I have cataloged remain but a trace in the current digital canon. A search of websites referenced by the Modern Language Association (MLA) bibliography reveals almost one thousand individual sites, yet very few of these projects are represented in MLA presentations, leaving many new to the field to assume that there are a small number of digital projects, often focused around a core group of digital humanities practitioners. The perception of limited projects and practitioners is what has driven the recent controversy of the digital humanities star system, highlighted by William Pannapacker's post-MLA 2011 blog post, "Digital Humanities Triumphant?" While I, along with many others, reject Pannapacker's representation of key digital humanities scholars as the cool kids at the high-school lunch table, the perception of the exclusionary world of digital humanities is reinforced by a perception of limited projects. Impacting the perception of digital humanities as exclusive, in both practice and product, is the granting model. Examination of funded projects reveals that the shift toward innovation has focused on technological innovation, not on innovative restructuring of the canon through recovery. The National Endowment of Humanities (NEH) awarded 141 Digital Humanities Start-Up Grants from 2007 through 2010. Of those grants, only twenty-nine were focused on diverse communities and sixteen on the preservation or recovery of diverse community texts. It is striking to examine the authors and historical figures individually cited in the list of funding: Shakespeare, Petrach, Melville, Jefferson, David Livingstone,[5] and Whitman. While there are grants to support work on indigenous populations, African and African American materials, and Asian American materials, in addition to others, the funding of named great men of history deserves scrutiny and even, perhaps, a specific funding program to encourage recovery efforts (NEH).

There may be many reasons for the lack of attention to noncanonical texts. 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'" ("Editing Early Modern Women's Manuscripts," 107). Additional reasons for exclusion are structural, such as the cost of production. Ken Price has discussed "the strong correlation between the 'significance' on which successful grant writing depends and the traditional canon" (281). Susan Belasco, for example, believes "that the traditional

standards for tenure and promotion are, in fact, more entrenched than ever and worse—more restrictive and un-imaginative than they were for an earlier generation" (333). Or, as Martha Nell Smith has argued, the digital humanities community might be adverse to the expansion of the canon and the work that has been reinserted into the mix. All of these possible explanations deserve critical attention if the digital humanities community wants to promote a broader 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. *19: Interdisciplinary Studies in the Nineteenth Century* is one such project. Begun as a simple HTML journal, scholars affiliated with the project participated in a Nineteenth-Century Scholarship Online (NINES) summer workshop during which they learned to encode 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, helping expand its user base and take an important step toward long-term sustainability. It provides a positive example of how we might take the institutional structures that have developed in relation to the digital humanities canon and leverage them to support small-scale projects. *19: Interdisciplinary Studies in the Nineteenth Century* also provides a pivotal clue that might explain why certain projects are currently excluded from the digital canon and others are not. The archive demonstrates that project value is created by editorial principles, content, *and* technological infrastructure. Projects like *19* have been revitalized by alignment with an institution, whether a collective, like NINES, a digital humanities center, like the Maryland Institute for Technology in the Humanities (MITH), or a library, like the University of Nebraska. The case of *19* suggests that standards and institution have become a core part of project success and sustainability, crucial to the canonization of digital work. Ken Price alludes to the new canon criteria 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)" (275). If institutional affiliation and technological standards are necessary components to the success of a project, then digital humanists must investigate how we might provide both to DIY scholars. Groups like NINES, which provide workshops and outreach, have modeled potential ways by which to generate such parameters. Additional efforts must follow.

If, indeed, we are beginning to construct a digital canon that weighs content and technological choices equally, then it is crucial for digital humanists to theorize the technological with the same rigor as we theorize the content. Alan Liu has more broadly seen the problem as an absence of cultural criticism, noting that "rarely do we extend the issues involved into the register of society, economics, politics, or culture" into our digital work. If we do not theorize our technological approaches with a mind toward cultural constructions, we will continue to exclude certain materials from digitization. One possible model is found in the partnership of Timothy Powell and the Ojibwe elders, who have created the *Gibagadinamaagoom* archive.

Powell has written extensively on the crucial impact of technological application to indigenous cultural materials and argues that current work needs to be revamped so that "digital technology can more accurately and artistically represent the indigenous origins and spiritual story lines of expressive culture on these [the Americas] continents" (Powell and Aitken, 253). Part of Powell's response to cultural inclusion was to construct a partnership that shifted ownership of cultural materials from that of the scholar to the Ojibwe elders, in effect creating a mechanism by which the tribe might control their own cultural materials.

My digital project, *The Nineteenth-Century Digital Concord Archive*, is similarly invested in exploring how to appropriately apply technological standards to shifting constructions of race represented in textual materials. Our current challenge is how we represent varying representations of blackness found in the census in a database. How do we represent, technologically, the identification of the same person as West Indian, Mulatto, or black? Amanda Gailey's recent article, "A Case of Heavy Editing: The Example of Race and Children's Literature in the Gilded Age," reveals the depth of theoretical inquiry that Gailey has invested in applying TEI appropriately to complex texts that bear markers of postbellum racial construction, particularly in the decision to utilize TEI to facilitate searching. Gailey notes that "we will use the <orig> and <reg> combination instead of the <sic> and <cor> combination (meaning *sic* and "correction"), as the former pair makes no claim about the rightness or wrongness of the readings, only how standardized their spellings are" (136). The TEI tag selection is in keeping with current cultural criticism regarding race and language, which rejects the superiority of standard English. The choice of <orig> and <reg> reveals that Gailey refuses to value one language usage over the other, as would be implied through the choice of the <sic> and <cor> tags. These examples provide a helpful way to imagine the next steps involved in digital humanities work and the treatment of race. While we need to continue to consider how to invigorate a robust digital recovery, we also have a good bit of theoretical work to do in the selection, editing, and technological manipulation of our materials.

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 under way. 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" (72), then we must ensure that our representation of culture does not exclude work by people of color.

NOTES

1. I use the term "digital text" to include digital edition and digital text. I see these as two distinctive types of textual production. I define a digital edition as a project that emphasizes textual variants using traditional bibliographical methods. A digital text is one version of a text that has been brought digital.

2. For example, all of these authors are included in *The Bedford Anthology of American Literature* and *The Heath Anthology of American Literature.*

3. I recognize that there is an ongoing discussion of how to view the distinction between media studies and digital humanities. My point is that the larger body of digital humanities work has been less concerned with cultural issues than the work produced by those who self-identify as media, rhetoric, communication, and film studies scholars.

4. I utilized a general Google search to locate the projects. If they were not found through this search, I located the institution or scholar associated with the materials to see if materials had been moved elsewhere.

5. David Livingstone is a Victorian explorer of Africa. The grant was awarded to create an online scholarly edition of his Nwangwe field diary.

BIBLIOGRAPHY

"Battle for the Soul of the Internet." *Time,* July 25, 1994. http://www.time.com/time/magazine/article/0,9171,981132,00.html.

Belasco, Susan. "The Responsibility Is Ours: The Failure of Infrastructure and the Limits of Scholarship." *Legacy: A Journal of American Women Writers* 26, no. 2 (2009): 329–36.

Bobley, Brett. "New from ODH: Summary Findings of NEH Digital Humanities Start-Up Grants (2007–2010)." National Endowment for the Humanities, Office of Digital Humanities, September 7, 2010. http://www.neh.gov/ODH/ODHUpdate/tabid/108/EntryId/144/New-from-ODH-Summary-Findings-of-NEH-Digital-Humanities-Start-Up-Grants-2007-2010.aspx.

Bolter, J. D. *Writing Space: The Computer, Hypertext, and the History of Writing.* Boston: Houghton Mifflin, 1991.

Childers, Peter, and Paul Delany. "Introduction: Two Versions of Cyberspace." *Works and Days* 23, no. 4 (1994): 61–78. http://www.sfu.ca/delany/wkndays.htm.

Decker, Catherine. "Crossing Old Barriers: The WorldWideWeb, Academia, and the Romantic Novel." *Romanticism on the Net* 10 (1998). http://www.erudit.org/revue/ron/1998/v/n10/005794ar.html.

Ezell, Margaret J. M. "Editing Early Modern Women's Manuscripts: Theory, Electronic Editions, and the Accidental Copy-Text." *Literature Compass* 7, no. 2 (2010): 102–9. http://onlinelibrary.wiley.com/doi/10.1111/j.1741-4113.2009.00682.x/full.

———. "The Myth of Judith Shakespeare: Creating the Canon of Women's Literature." *New Literary History* 21, no. 3 (1990): 579–92.

———. *Writing Women's Literary History*. Baltimore: Johns Hopkins University Press, 1993.

Fraiman, Susan. "In Search of Our Mother's Gardens–With Help from a New Digital Resource for Literary Scholars." *Modern Philology* 106, no. 1 (2008): 142–48.

Fraistat, Neil, Steven E. Jones, and Carl Stahmer. "The Canon, the Web, and the Digitization of Romanticism." *Romanticism on the Net* 10 (1998). http://www.erudit.org/revue/ron/1998/v/n10/005801ar.html

Gailey, Amanda. "A Case for Heavy Editing: The Example of Race and Children's Literature in the Gilded Age." In *The American Literature Scholar in the Digital Age,* edited by Amy E. Earhart and Andrew Jewell, 125–44. Ann Arbor: University of Michigan Press, 2010.

Kirschenbaum, Matthew. "What Is Digital Humanities and What's It Doing in English Departments?" *ADE Bulletin* 150 (2010): 55–61. http://mkirschenbaum.files.wordpress.com/2011/01/kirschenbaum_ade150.pdf. Reprinted in this volume.

Liu, Alan. "Where Is Cultural Criticism in the Digital Humanities." *Alan Liu*. January 2011. http://liu.english.ucsb.edu/where-is-cultural-criticism-in-the-digital-humanities/.

McGann, Jerome. "Culture and Technology: The Way We Live Now, What Is To Be Done?" *New Literary History* 36, no. 1 (2005): 71–82. http://muse.jhu.edu/journals/new_literary_history/v036/36.1mcgann.html.

National Endowment for the Humanities. "Digital Humanities Start-up Grants." August 6, 2010. http://www.neh.gov/grants/guidelines/digitalhumanitiesstartup.html.

Omeka. Roy Rosenzweig Center for History and New Media. George Mason University. http://omeka.org/

Powell, Timothy B., and Larry P. Aitken. "Encoding Culture: Building a Digital Archive Based on Traditional Ojibwe Teachings." In *The American Literature Scholar in the Digital Age,* edited by Amy E. Earhart and Andrew Jewell, 250–74. Ann Arbor: University of Michigan Press, 2010.

Price, Ken. "Digital Scholarship, Economics, and the American Literary Canon." *Literature Compass* 6, no. 2 (2009): 274–90. http://onlinelibrary.wiley.com/doi/10.1111/j.1741-4113.2009.00622.x/full.

Ramsay, Stephen. "Who's in and Who's Out." Stephen Ramsay Blog. January 8, 2011. http://lenz.unl.edu/papers/2011/01/08/whos-in-and-whos-out.html.

Shillingsburg, Peter L. "Principles for Electronic Archives, Scholarly Editions, and Tutorials." In *The Literary Text in the Digital Age*, edited by Richard J. Finneran, 23–35. Ann Arbor: University of Michigan Press, 1996.

Smith, Martha Nell. "The Human Touch: Software of the Highest Order: Revisiting Editing as Interpretation." *Textual Cultures* 2, no. 1 (2007): 1–15.

Spencer, Amy. *DIY: The Rise of Lo-FI Culture*. London: Marion Boyars, 2005.

Unsworth, John, and Martin Muller. *The MONK Project Final Report*. 2009. http://www.monkproject.org/MONKProjectFinalReport.pdf.

The Social Contract of Scholarly Publishing

DANIEL J. COHEN

When Roy Rosenzweig and I finished writing a full draft of our book *Digital History*, we sat down at a table and looked at the stack of printouts.

"So, what now?" I said to Roy naively. "Couldn't we just publish what we have on the web with the click of a button? What value does the gap between this stack and the finished product have? Isn't it 95 percent done? What's the last five percent for?"

We stared at the stack some more.

Roy finally broke the silence, explaining the magic of the last stage of scholarly production between the final draft and the published book: "What happens now is the creation of the *social contract* between the authors and the readers. We agree to spend considerable time ridding the manuscript of minor errors, and the press spends additional time on other corrections and layout, and readers respond to these signals—a lack of typos, nicely formatted footnotes, a bibliography, specialized fonts, and a high-quality physical presentation—by agreeing to give the book a serious read."

I have frequently replayed that conversation in my mind, wondering about the constitution of this social contract in scholarly publishing, which is deeply related to questions of academic value and reward.

For the ease of conversation, let's call the two sides of the social contract of scholarly publishing the *supply side* and the *demand side*. The supply side is the creation of scholarly works, including writing, peer review, editing, and the form of publication. The demand side is much more elusive—the mental state of the audience that leads them to "buy" what the supply side has produced. In order for the social contract to work, for engaged reading to happen, and for credit to be given to the author (or editor of a scholarly collection), both sides need to be aligned properly.

The social contract of the book is profoundly entrenched and powerful—almost mythological—especially in the humanities. As John Updike put it in his diatribe against the digital[1] (and most humanities scholars and tenure committees would still agree), "The printed, bound and paid-for book was—still is, for the

moment—more exacting, more demanding, of its producer and consumer both. It is the site of an encounter, in silence, of two minds, one following in the other's steps but invited to imagine, to argue, to concur on a level of reflection beyond that of personal encounter, with all its merely social conventions, its merciful padding of blather and mutual forgiveness."

As academic projects have experimented with the web over the past two decades, we have seen intense thinking about the supply side. Robust academic work has been reenvisioned in many ways: as topical portals, interactive maps, deep textual databases, new kinds of presses, primary source collections, and even software. Most of these projects strive to reproduce the magic of the traditional social contract of the book, even as they experiment with form.

The demand side, however, has languished. Far fewer efforts have been made to influence the mental state of the scholarly audience. The unspoken assumption is that the reader is more or less unchangeable in this respect, only able to respond to and validate works that have the traditional marks of the social contract: having survived a strong filtering process, near-perfect copyediting, the imprimatur of a press.

We need to work much more on the demand side if we want to move the social contract forward into the digital age. Despite Updike's ode to the book, there *are* social conventions surrounding print that are worth challenging. Much of the reputational analysis that occurs in the professional humanities relies on cues beyond the scholarly content itself. The act of scanning a CV is fraught with these conventions.

Can we change the views of humanities scholars so that they may accept, as some legal scholars already do, the great blog post as being as influential as the great law review article? Can we get humanities faculty, as many tenured economists already do, to publish more in open access journals? Can we accomplish the humanities equivalent of FiveThirtyEight.com, which provides as good, if not better, in-depth political analysis than most newspapers, earning the grudging respect of journalists and political theorists? Can we get our colleagues to recognize outstanding academic work wherever and however it is published?

I believe that to do so, we may have to think less like humanities scholars and more like social scientists. Behavioral economists know that although the perception of value can come from the intrinsic worth of the good itself (e.g., the quality of a wine, already rather subjective), it is often influenced by many other factors, such as price and packaging (the wine bottle, how the wine is presented for tasting). These elements trigger a reaction based on stereotypes—if it's expensive and looks well wrapped, it must be valuable. The book and article have an abundance of these value triggers from generations of use, but we are just beginning to understand equivalent value triggers online—thus the critical importance of web design and why the logo of a trusted institution or a university press can still matter greatly, even if it appears on a website rather than a book.

Social psychologists have also thought deeply about the potent grip of these idols of our tribe. They are aware of how cultural norms establish and propagate

themselves and tell us how the imposition of limits creates hierarchies of recognition. Thinking in their way, along with the way the web works, one potential solution on the demand side might come not from the scarcity of production, as it did in a print world, but from the scarcity of attention. That is, value will be perceived in any community-accepted process that narrows the seemingly limitless texts to read or websites to view. *Curation* becomes more important than publication once publication ceases to be limited.

NOTES

This chapter originally appeared as "The Social Contract of Scholarly Publishing" (http://www.dancohen.org/2010/03/05/the-social-contract-of-scholarly-publishing/).

1. http://www.nytimes.com/2006/06/25/books/review/25updike.html.

Introducing Digital Humanities Now

DANIEL J. COHEN

Do the digital humanities need journals? Although I'm very supportive of the new journals that have launched in the last year and although I plan to write for them from time to time, there's something discordant about a nascent field—one so steeped in new technology and new methods of scholarly communication—adopting a format that is struggling in the face of digital media.

I often say to nondigital humanists that every Friday at five I know all of the most important books, articles, projects, and news of the week—without the benefit of a journal, a newsletter, or indeed any kind of formal publication by a scholarly society. I pick up this knowledge by osmosis from the people I follow online.

I subscribe to the blogs of everyone working centrally or tangentially to digital humanities. As I have argued[1] from the start[2] and against the skeptics and traditionalists who think blogs can only be narcissistic, half-baked diaries, these outlets are just publishing platforms by another name, and in my area there are an incredible number of substantive ones.

More recently, social media such as Twitter has provided a surprisingly good set of pointers toward worthy materials I should be reading or exploring. (And as happened with blogs five years ago, the critics are now dismissing Twitter as unscholarly, missing the filtering function it somehow generates among so many unfiltered tweets.) I follow as many digital humanists as I can on Twitter and created a comprehensive list of people in digital humanities.[3] (You can follow me @dancohen.)

For a while I've been trying to figure out a way to show this distilled "Friday at five" view of digital humanities to those new to the field or those who don't have time to read many blogs or tweets. This week I saw a tweet from Tom Scheinfeldt[4] (who in turn saw a tweet from James Neal[5]) about a new service called Twittertim.es,[6] which creates a real-time publication consisting of articles highlighted by people you follow on Twitter. I had a thought: what if I combined the activities of several hundred digital humanities scholars with Twittertim.es?

Digital Humanities Now (*DHN*)[7] is a new web publication that is the experimental result of this thought. It aggregates thousands of tweets and the hundreds of

articles and projects those tweets point to and boils everything down to the most-discussed items, with commentary from Twitter. A slightly longer discussion of how the publication was created can be found on the *DHN* "About" page.[8]

Does the process behind *DHN* work? From the early returns, the algorithms have done fairly well, putting on the front page articles on grading in a digital age and bringing high-speed networking to liberal arts colleges, Google's law archive search, and (appropriately enough) a talk on how to deal with streams of content given limited attention. Perhaps *Digital Humanities Now* will show a need for the light touch of a discerning editor. This could certainly be added on top of the raw feed of all interest items[9] (about fifty a day, out of which only two or three make it into *DHN*), but I like the automated simplicity of *DHN* 1.0.

Despite what I'm sure will be some early hiccups, my gut is that some version of this idea could serve as a rather decent new form of publication that focuses the attention of those in a particular field on important new developments and scholarly products. I'm not holding my breath that someday scholars will put an appearance in *DHN* on their CVs. But as I recently told an audience of executive directors of scholarly societies at an American Council of Learned Societies meeting, if you don't do something like this, someone else will.

I suppose *DHN* is a prod to them and others to think about new forms of scholarly validation and attention beyond the journal. Ultimately, journals will need the digital humanities more than we need them.

NOTES

This chapter originally appeared as "Introducing Digital Humanities Now" (http://www.dancohen.org/2009/11/18/introducing-digital-humanities-now/).

1. http://www.dancohen.org/2008/12/05/leave-the-blogging-to-us/.
2. http://www.dancohen.org/2005/12/16/creating-a-blog-from-scratch-part-1-what-is-a-blog-anyway/.
3. http://twitter.com/dancohen/digitalhumanities/members.
4. http://www.foundhistory.org/, http://twitter.com/#!/foundhistory.
5. https://twitter.com/#!/james3neal.
6. http://tweetedtimes.com/.
7. http://digitalhumanitiesnow.org/.
8. http://digitalhumanitiesnow.org/about/.
9. http://feeds.feedburner.com/DigitalHumanitiesNow.

Text: A Massively Addressable Object

MICHAEL WITMORE

At the Working Group for Digital Inquiry at Wisconsin, we've just begun our first experiment with a new order of magnitude of texts. Jonathan Hope and I started working with thirty-six items about six years ago when we began to study Shakespeare's First Folio plays (Witmore and Hope). Last year, we expanded to three-hundred and twenty items with the help of Martin Mueller at Northwestern, exploring the field of early modern drama. Now that the University of Wisconsin has negotiated a license with the University of Michigan to begin working with the files from the Text Creation Partnership (TCP), which contains over twenty-seven thousand items from early modern print, we can up the number again. By January, we will have begun our first one-thousand item experiment, spanning items printed in Britain and North America from 1530 through 1809. Robin Valenza and I, along with our colleagues in computer sciences and the library, will begin working up the data in the spring. Stay tuned for results.

New experiments provide opportunities for thought that precede the results. What does it mean to collect, tag, and store an array of texts at this level of generality? What does it mean to be an "item" or "computational object" within this collection? What is such a collection? In this post, I want to think further about the nature of the text objects and populations of texts we are working with.

What is the distinguishing feature of the digitized text—that ideal object of analysis considered in all its hypothetical relations with other ideal objects? The question itself goes against the grain of recent materialist criticism, which focuses on the physical existence of books and practices involved in making and circulating them. Unlike someone buying an early modern book in the bookstalls around St. Paul's four hundred years ago, we encounter our TCP texts as computational objects. That doesn't mean that they are immaterial, however. Human labor has transformed them from microfilm facsimiles of real pages into diplomatic quality digital transcripts, marked up in TEI so that different formatting features can be distinguished. That labor is as real as any other.

What distinguishes this text object from others? I would argue that a text is a text because it is *massively addressable at different levels of scale*. Addressable here means that one can query a position within the text at a certain level of abstraction. In an earlier post, for example, I argued that a text might be thought of as a vector through a metatable of all possible words (Witmore). Why is it possible to think of a text in this fashion? Because a text can be queried at the level of single words and then related to other texts at the same level of abstraction: the table of all possible words could be defined as the aggregate of points of address at a given level of abstraction (the word, as in Google's new Ngram corpus). Now, we are discussing ideal objects here; addressability implies different levels of abstraction (character, word, phrase, line, etc.), which are stipulative or nominal: such levels are not material properties of texts or Pythagorean ideals; they are, rather, conventions.

Here's the twist. We have physical manifestations of ideal objects (the ideal *1 Henry VI*, for example), but these manifestations are only provisional realizations of that ideal. (I am using the word manifestation in the sense advanced in the Online Computer Library Center's Functional Requirements for Bibliographic Records [FRBR] hierarchy.[1]) The book or physical instance, then, is *one of many levels of address*. Backing out into a larger population, we might take a genre of works to be the relevant level of address. Or we could talk about individual lines of print, all the nouns in every line, every third character in every third line. All this variation implies massive flexibility in levels of address. And more provocatively, when we create a digitized population of texts, our modes of address become more and more abstract: all concrete nouns in all the items in the collection, for example, or every item identified as a "History" by Heminges and Condell in the First Folio. Every level is a provisional unity: stable for the purposes of address but also stable because it is the object of address. Books are such provisional unities. So are all the proper names in the phone book.

The ontological status of the individual text is the same as that of the population of texts: both are massively addressable, and when they are stored electronically we are able to act on this flexibility in more immediate ways through iterative searches and comparisons. At first glance, this might seem like a Galilean insight, similar to his discipline-collapsing claim that the laws that apply to heavens (astronomy) are identical with the ones that apply to the sublunar realm (physics). But it is not.

Physical texts were *already* massively addressable before they were ever digitized, and this variation in address was and is registered at the level of the page, chapter, the binding of quires, and the like. When we encounter an index or marginal note in a printed text—for example, a marginal inscription linking a given passage of a text to some other in a different text—we are seeing an act of address. Indeed, the very existence of such notes and indexes implies just this flexibility of address.

What makes a text a text—its susceptibility to varying levels of address—is a feature of book culture and the flexibility of the textual imagination. We address ourselves to this level, in this work, and think about its relation to some other.

"Oh, this passage in *Hamlet* points to a verse in the Geneva bible," we say. To have this thought is to dispose relevant elements in the data set in much the same way a spreadsheet aggregates a text in ways that allow for layered access. A reader is a maker of such a momentary *dispositif* or device, and reading might be described as the continual redisposition of levels of address in this manner. We need a phenomenology of these acts, one that would allow us to link quantitative work on a culture's "built environment" of words to the kinesthetic and imaginative dimensions of life at a given moment.

A physical text or manifestation is a provisional unity. There exists a potentially infinite array of such unities, some of which are already lost to us in history: what was a relevant level of address for a thirteenth-century monk reading a manuscript? Other provisional unities can be operationalized now, as we are doing in our experiment at Wisconsin, gathering one thousand texts and then counting them in different ways. Grammar, as we understand it now, affords us a level of abstraction at which texts can be stabilized: we lemmatize texts algorithmically before modernizing them, and this lemmatization implies provisional unities in the form of grammatical objects of address.

One hundred years from now, the available computational objects may be related to one another in new ways. I can only imagine what these are: every fourth word in every fourth document, assuming one could stabilize something like "word length" in any real sense. (The idea of a word is itself an artifact of manuscript culture, one that could be perpetuated in print through the affordances of moveable type.) What makes such thought experiments possible is, once again, the addressability of texts as such. Like a phone book, they aggregate elements and make these elements available in multiple ways. You could even think of such an aggregation as the substance of another aggregation, for example, "all the phone numbers belonging to people whose last name begins with A." But unlike a phonebook, the digitized text can be reconfigured almost instantly into various layers of arbitrarily defined abstraction (characters, words, lines, works, genres). The mode of storage or virtualization is precisely what allows the object to be addressed in multiple ways.

Textuality *is* massive addressability. This condition of texts is realized in various manifestations, supported by different historical practices of reading and printing. The material affordances of a given medium put constraints on such practices: the practice of "discontinuous reading" described by Peter Stallybrass, for example, develops alongside the fingerable discrete leaves of a codex. But addressability as such: *this* is a condition rather than a technology, action, or event. And its limits cannot be exhausted at a given moment. We cannot, in a Borgesian mood, query all the possible data sets that will appear in the fullness of time. And we cannot import future query types into the present. But we can and do approximate such future searches when we automate our modes of address in unsupervised multivariate statistical analysis—for example, factor analysis or Principle Component Analysis (PCA). We want all the phonebooks. And we can simulate some of them now.

NOTES

This chapter originally appeared as "Text: A Massively Addressable Object" (http://winedarksea.org/?p=926).

1. http://www.oclc.org/research/publications/library/2003/lavoie_frbr.pdf.

BIBLIOGRAPHY

Hope, Jonathan, and Michael Witmore. "The Hundredth Psalm to the Tune of 'Green Sleeves': Digital Approaches to Shakespeare's Language of Genre." *Shakespeare Quarterly* 61, no. 3 (2010): 357– 90.

Witmore, Michael. "Texts as Objects II: Object Oriented Philosophy. And Criticism?" *Wine Dark Sea*. September 17, 2009. http://winedarksea.org/?p=381.

The Ancestral Text

MICHAEL WITMORE

In this post I want to understand the consequences of "massive addressability" (Witmore) for "philosophies of access"—philosophies that assert that all beings exist only as correlates of our own consciousness. The term "philosophy of access" is used by members of the speculative realist school: it seems to have been coined largely as a means of rejecting everything the term names. Members of this school dismiss the idea that speculative analysis of the nature of beings can be replaced by an apparently more basic inquiry into how we access the world, an access obtained through either language or consciousness. The major turn to "access" occurs with Kant, but the move is continued in an explicitly linguistic register by Heidegger, Wittgenstein, Derrida, and a range of poststructuralists.

One reason for jettisoning the priority of access, according to Ray Brassier, is that it violates "the basic materialist requirement that being, though perfectly intelligible, remain irreducible to thought" ("The Enigma of Realism"). As will become clear, I am sympathetic to this materialist requirement and more broadly to the speculative realist project of dethroning language as our *one and only* mode of access to the world. (There are plenty of ways of appreciating the power and complexity of language without making it the wellspring of Being, as some interpreters of Heidegger have insisted.) Our quantitative work with texts adds an unexpected twist to these debates: as objects of massive and variable address, texts are "handled" in precisely the ways usually reserved for nonlinguistic entities. When taken as objects of quantitative description, texts possess qualities that—at some point in the future—could be said to have existed in the present, *regardless* of our knowledge of them. There is thus a temporal asymmetry surrounding quantitative statements about texts: if one accepts the initial choices about what gets counted, such statements can be "true" now even if they can only be produced and recognized later. Does this asymmetry, then, mean that language itself, "though perfectly intelligible, remain[s] irreducible to thought" (Brassier)? Do iterative methods allow us to satisfy Brassier's materialist requirement in the realm of language itself?

Let us begin with the question of addressability and access. The research described on this blog involves the creation of digitized corpora of texts and the mathematical description of elements within that corpus. These descriptions obtain at varying degrees of abstractions (nouns describing sensible objects, past verb forms with an auxiliary, etc.). If we say that we know something quantitatively about a given corpus, then we are saying that we know it on the basis of a set of relations among elements that we have provisionally decided to treat as countable unities. Our work is willfully abstract in the sense that, at crucial moments of the analysis, we foreground relations as such, relations that will then be reunited with experience. When I say that objects of the following kind—"Shakespearean texts identified as comedies in the First Folio"—contain more of a certain type of thing (first- and second-person singular pronouns) than objects of a different kind (Shakespeare's tragedies, histories), I am making a claim about a relation between groups and what they contain. These groupings and the types of things that we use to sort them are provisional unities: the circle we draw around a subset of texts in a population could be drawn another way if we had chosen to count other things. And so, we must recognize several reasons why claims about these relations might always be revised.

Every decision about what to count offers a caricature of the corpus and the modes of access this corpus allows. A caricature is essentially a narrowing of address: it allows us to make contact with an object in some of the ways Graham Harman has described in his work on vicarious causation. One can argue, for example, that the unity "Shakespeare's Folio comedies" is really a subset of a larger grouping or that the group can itself be subdivided into smaller groups. Similarly, one might say that the individual plays in a given group aren't really discrete entities and so cannot be accurately counted in or out of that group. There are certain words that *Hamlet* may or may not contain, for example, because print variants and multiple sources have made *Hamlet* a leaky unity. (Accommodating such leaky unities is one of the major challenges of digital text curation.) Finally, I could argue that addressing these texts on the level of grammar—counting first- and second-person singular pronouns—is just one of many modes of address. Perhaps we will discover that these pronouns are fundamentally linked to semantic patterns that we haven't yet decided to study but should. All of these alternatives demonstrate the provisional nature of any decision to count and categorize things: such decisions are interpretive, which is why iterative criticism is not going to put humanities professors out of business. But such counting decisions are not—and this point is crucial—simply another metaphoric reduction of the world. Principal component analysis (PCA), cluster analysis, and the other techniques we use are clearly *inhuman* in the number of comparisons they are able to make. The detour through mathematics is a detour *away* from consciousness, even if that detour produces findings that ultimately converge with consciousness (i.e., groupings produced by human reading).

Once the counting decisions are made, our claims to know something *in a statistical sense* about texts boils down to a claim that a particular set of relations

pertains among entities in the corpus. Indeed, considered mathematically, the things we call texts, genres, or styles simply *are* such sets of relations—the mathematical reduction being one of many possible caricatures. But counting is a very interesting caricature: it yields what is there now—a real set of relations—but is nevertheless impossible to contemplate at present. Once claims about texts become mathematical descriptions of relations, such statements possess what the philosopher Quentin Meillassoux calls ancestrality, a quality he associates primarily with statements about the *natural* world. Criticizing the ascendance of what he calls the Kantian dogma of correlationism—the assumption that everything that can be said "to be" exists only as correlate of consciousness—Meillassoux argues that the idealist or critical turn in Continental philosophy has impoverished our ability to think about anything that exceeds the correlation between mind and world. This "Great Outdoors," he goes on to suggest, is a preserve that an explicitly speculative philosophy must now rediscover, one that Meillassoux believes becomes available to us through mathematics. So, for example, Meillassoux would agree with the statement, "the earth existed 4.5 billion years ago," precisely because it can be formulated mathematically using measured decay rates of carbon isotopes. The statement itself may be ideal, but the reality it points to is not. What places the Great Outdoors out of doors, then, is its indifference to our existence or presence as an observer. Indeed, for Meillassoux, it is *only* those things that are "mathematically conceivable" that exceed the post-Kantian idealist correlation. For Meillassoux, "all those aspects of the object that can be formulated in mathematical terms can be meaningfully conceived as properties of the object in itself." Clearly such a statement is a goad for those who place mind or natural language at the center of philosophy. But the statement is also a philosophical rallying cry: be curious about objects or entities that do not reference human correlates! I find this maxim appealing in the wake of the "language is everything" strain of contemporary theory, which is itself a caricature of the work of Wittgenstein, Derrida, and others. Such exaggerations have been damaging to those of us working in the humanities, not least because they suggest that our colleagues in the sciences do *nothing but* work with words. By making language everything—and, not accidentally, making literary studies the gatekeeper of all disciplines—this line of thought amounts to a new kind of species narcissism. Meillassoux and others are finding ways to not talk about language all the time, which seems like a good thing to me.

But would Meillassoux, Harman, and other speculative realists consider texts to be part of the Great Outdoors? Wouldn't they have to? After all, statements about groupings in the corpus can be true now even when there is no human being to recognize that truth as a correlate of thought. Precisely *because* texts are susceptible to address and analysis on a potentially infinite variety of levels we can be confident that a future scholar will find a way of counting things that turns up a new but as yet unrecognized grouping. Human reading turned up such a thing when scholars in the late nineteenth century "discovered" the genre of Shakespeare's late

romances. (Jonathan Hope and I have, moreover, redescribed these groupings statistically [Witmore and Hope].) Like our future mathematical sleuth might do a century from now, nineteenth-century scholars were arguing that romance was already a *real* feature of the Shakespearean corpus, albeit one that no one had yet recognized. They had, in effect, picked out a new object by emphasizing a new set of relations among elements in a collection of words. Couldn't we expect another genre to emerge from this sort of analysis—a genre X, let's say—given sufficient time and resources? Would we accept such a genre if derived through iterative means?

I can imagine a day, one hundred years from now, when we have different dictionaries that address the text on levels we have not thought to explore at present. What if someone creates a dictionary that allows me to use differences in a word's linguistic origin (Latinate, Anglo-Saxon, etc.) to relate the contents of one text to another? What if a statistical procedure is developed that allows us to "see" groupings we *could* recognize today but simply have not developed the mathematics to expose? When you pair the condition of massive addressability with (1) the possibility of new tokenizations (new elements or strata of address) or (2) the possibility that all token counts past and future can be subjected to new mathematical procedures, you arrive at a situation in which something that is arguably true now about a collection of texts can only be known in the future.

And if something can be true about an object now without itself being a correlate of human consciousness, isn't that something part of the *natural world,* the one that is supposed to be excluded from the charmed circle of the correlation? Does this make texts more like objects in nature, or objects in nature more like texts? Either way, the Great Outdoors has become larger.

NOTE

This chapter originally appeared as "The Ancesteral Text" (http://winedarksea.org/?p=979).

BIBLIOGRAPHY

Brassier, Ray. "The Enigma of Realism: On Quentin Meillassoux's After Finitude." In *Collapse: Philosophical Research and Development*. Vol. 2. 2007.

Meillassoux, Quentin. *After Finitude: An Essay on the Necessity of Contingency*. New York: Continuum, 2008.

Witmore, Michael. "Text: A Massively Addressable Object." In *Wine Dark Sea*. December 31, 2010. http://winedarksea.org/?p=926.

Witmore, Michael, and Jonathan Hope. "Shakespeare by the Numbers: On the Linguistic Texture of the Late Plays." In *Early Modern Tragicomedy*, edited by Subha Mukherji and Raphael Lyne, 133–53. Woodbridge, Suffolk, UK: D. S. Brewer, 2007.

This page intentionally left blank