

Vol. 88, No. 2

Journals

Search

Partners

Information

September 2001

Previous

Table of contents

Next

The Journal of American History

Table of contents

List journal issues

Home



Printer-friendly
format

How to cite this
electronic article

The Road to Xanadu: Public and Private Pathways on the History Web

Roy Rosenzweig

On August 24, 1965, Theodor Nelson presented a paper to the Association for Computing Machinery national conference in which he introduced the word "hypertext" to refer to "a body of written or pictorial material interconnected in such a complex way that it could not conveniently be presented or represented on paper." Nelson, who had started musing about this sort of associative thinking and linking as a Harvard University graduate student in 1960, viewed "hypertext" as an integral part of an imagined globally interconnected library and publishing system that would "grow indefinitely, gradually including more and more of the world's written knowledge" and "have every feature a novelist or absent-minded professor could want, holding everything he wanted in just the complicated way he wanted it held, and handling notes and manuscripts in as subtle and complex ways as he wanted them handled."¹

1

Two years later, while working at the publisher Harcourt Brace, Nelson—an inveterate coiner of terms whose own Web page lists sixteen words or phrases that he claims to have introduced into general use—started to describe his global library as "Xanadu." "For forty years," Nelson wrote recently, "Project Xanadu has had as its purpose to build a deep-reach electronic literary system for worldwide use and a differently-organized general system of data management."²

Nelson's grand vision of a universal library and publishing system has come in for its share of derision. In 1995, the *Wired* magazine writer Gary Wolf devoted twenty thousand words to detailing what he called "The Curse of Xanadu." "Nelson's Xanadu project," he wrote, "was supposed to be the universal, democratic hypertext library. . . . Instead, it sucked Nelson and his intrepid band of true believers into what became the longest-running vaporware project in the history of computing—a 30-year saga. . . . [an] amazing epic tragedy. . . . [and] an actual symptom of madness." Nelson responded angrily to Wolf's profile, but he has also hinted that he views Xanadu as an impossible dream. He took the word from the imaginary home of Kubla Khan in Samuel Taylor Coleridge's uncompleted poem of the same name; Orson Welles (one of Nelson's heroes) used the same word for Citizen Kane's extravagant, uncompleted mansion.³

And yet, just five years after Wolf's obituary for Xanadu, the dream of a universal hypertextual library seems less like the narcotic imaginings of Samuel Taylor Coleridge or the fantasies of Ted Nelson than a description of a multibillion-dollar industry called the World Wide Web.⁴ Even those of us whose professional calling requires us to think soberly about the distant past need now to consider whether such a contemporary development will reshape the ways we research, teach, and write history. Can professional historians look forward to a future in which they can access all the documentary evidence of the past with the click of a mouse? How far have we already come toward reaching that dream?

Not far enough yet. Even Nelson's 1965 paper on hypertext—quite relevant to anyone interested in the Web, which has hypertext as its most basic protocol—is not yet online. And any reader of this journal could come up with long lists of crucial historical sources only in physical libraries and archives. Still, a startling number of primary and secondary sources important to American historians have suddenly appeared online in the less-than-a-decade history of the World Wide Web. Indeed, so rapid has been the growth of the "history Web," as we will call that virtual world within a virtual world, that it cannot be readily surveyed within a single article. Such topics as how digital history might alter classroom teaching, historical writing, or modes of scholarly discourse, while mentioned here, deserve separate, extended treatment. Instead, I focus on some of the general trends in the growth of the history Web over the past five years, especially its emergence as an extraordinarily rich online archive of primary and secondary sources, a Xanadu, in Nelson's words. What sources are now online? What is the range and quality of this virtual archive? Even more important, *who* has put them there and *who* can use them?

Asking such questions inevitably leads us to wonder about the past, present, and future of one of the Internet's most celebrated qualities—its open and *public* character. As the history Web has grown, it has also become more complex. Many of the most important resources are now "hidden" from view in databases not readily accessible by such Web search engines as Google and AltaVista.⁵ In addition, while many of the creators and owners of Web content still come from what could be broadly called the *public* sector—whether grass-roots enthusiasts, grant-funded university-based projects, or government agencies such as the Library of Congress—*private* corporations (giant information conglomerates selling their wares to libraries, entertainment corporations trying to turn the Web into an advertiser-supported medium, and Internet startups with a range of business plans) are coming to control some of the most valuable real estate on the history Web. Such private control raises questions about what history we will see on our computer screens and who will be able to see it. If the road ahead leads to Xanadu.com rather than Xanadu.edu, what will the future of the past look like?

One, Two, Many History Webs: Surface and Deep, Public and Private

Rapidity of change is a new technology cliché. "The Internet's pace of adoption," observes a United States Department of Commerce report, "eclipses all other technologies that preceded it. Radio was in existence thirty-eight years before fifty million people tuned in; TV took thirteen years to reach that benchmark. . . . Once it was opened to the general public, the Internet crossed that line in four years." In just the past five years, the percentage of the United States population online has more than tripled from 14 to 44 percent. The "Web Characterization Project" of the OCLC (Online Computer Library Center, Inc.) reported 7.1 million unique Web sites in October 2000, a 50 percent increase over the previous year's total and almost a fivefold increase since just 1997. Over that time, the Web has almost entirely displaced other media—especially CD-ROMs—for presenting digital content. Conventional search engines such as Google currently index more than 1.3 billion Web pages. Peter Lyman and Hal R. Varian estimate that in 2000 the World Wide Web consisted of about twenty-one terabytes (a terabyte is 1,000 gigabytes) of static HTML (hypertext markup language) pages and was growing at a rate of 100 percent annually. But increasingly Web "pages" only come into existence as the result of specialized database searches, and those Web-based databases do not turn up in standard Web searches. BrightPlanet Corporation, whose Lexibot software indexes some of the searchable databases not readily accessible by conventional search engines, claims that this "invisible" or "deep" Web (in contrast to the "surface" Web found by the search engines) contains nearly 550 billion individual pages.⁶

8

How much has the history Web changed? No time machine can take us back to the Web of 1995 or 1996 and run comparative searches with today. One imperfect benchmark comes from searches that my colleague Michael O'Malley and I did in the fall of 1996 while writing an article on the history Web for this journal. Running the same searches in the same search engine (AltaVista) returns more than ten times as many "hits" today as four years ago—thereby greatly outpacing the overall growth of the Web and even "Moore's law," which predicts that computing power will double every eighteen months. We had 64 hits on William Graham Sumner, 300 on Eugene Debs, and 700 on Emma Goldman in 1996; the comparable figures for November 2000 were 716, 2,971, and 8,805.⁷

9

The quality of those "hits" improved as well. Four years ago, those looking for Debs on the Web might find some basic biographical information about the socialist leader, but the most interesting insights were how Debs fits into contemporary American life—how different groups (from the Democratic Socialists of America to the National Child Rights Alliance) and individuals (from local activists to Ralph Nader) made use of Debs's past in late-twentieth-century America. Now, however, the Web contains not only up-to-date biographical and historical treatments but also a gallery of images, state-by-state figures on Debs's presidential votes, guides to archival collections, and a substantial body of primary sources—at least a dozen different speeches or articles by Debs and another half dozen contemporary accounts of him.

10

Such raw Web searches do not, however, capture the fullness of the history Web since they do not generally measure the deep Web. For historians, the most notable of such databases are the more than ninety collections gathered under *American Memory*, the online resource compiled by the Library of Congress's National Digital Library Program (NDLP). Four years ago, *American Memory* had some staggering archival riches, but now the collection has grown at least fivefold and includes more than five million items—ranging from 1,305 pieces of African American sheet music to 2,100 early baseball cards. Visitors can examine 117,000 FSA/OWI (Farm Security Administration–Office of War Information) photographs, 422 early motion pictures and sound recordings of the Edison Companies, and 176,000 pages of George Washington's correspondence, letter books, and other papers. Library staff will soon place online another thirty collections,

including such eagerly awaited resources as the thousands of ex-slave narratives of the Federal Writers' Project.⁸

Whereas four or five years ago history materials on the Web were most useful for teaching, the depth of such collections as *American Memory* means that historians can now do serious scholarly research in online collections. With more than 200,000 photographs now available in *American Memory*, anyone studying the history of American photography would need to visit the NDLP. Moreover, the digital format makes possible modes of research that are possible in other media but much more difficult. Take, for example, the old, but still much debated, question of George Washington's religious attitudes. Using the online version of the Washington papers, the historian Peter R. Henriques showed not only that Washington never referred to "Jesus" or "Christ" in his personal correspondence but also that his references to death were invariably "gloomy and pessimistic" with no evidence of "Christian images of judgment, redemption through the sacrifice of Christ, and eternal life for the faithful."⁹

Washington's dark thoughts on death are filed away in the deep Web of such databases as the vast *American Memory* collection not accessible by conventional Web-wide search engines; Henriques's thoughts on Washington (published in print in *Virginia Magazine of History and Biography* but online through Bell & Howell's ProQuest Direct and EBSCO's World History FullTEXT), however, reside in a vast terrain that even BrightPlanet does not fully measure—what we will call the private Web. These are the growing number of online resources only available to paying customers. OCLC's data indicate that the growth of the public Web is slowing at the same time that private, restricted Web sites have gone from 12 to 20 percent of the total Web.¹⁰ Whereas the surface and deep Webs, which together we will call the public Web, contain enormous numbers of primary documents, the private Web abounds in the secondary sources crucial to historical work.

For example, most historians know about JSTOR (Journal Storage: The Scholarly Journal Archive), which includes, in its five-million-page database of 117 academic journals, the full text of fifteen different history journals, most of them running from their inception up to 1995. Many of the nonhistory journals, for example, sociology, economic, and political science journals from the early part of the twentieth century, constitute primary sources of great interest to American historians. Searching JSTOR for Eugene Debs in history journals yields 81 articles, but expanding to other journals gives us another 61 articles, including such significant contemporary sources as John Spargo's "The Influence of Karl Marx on Contemporary Socialism" in the 1910 *American Journal of Sociology*. The word search capabilities of JSTOR also facilitate a kind of intellectual history that cannot be done as easily in print sources. Say you want to trace the changing reputation of Charles Beard in the historical profession; the 191 articles in JSTOR that mention Beard provide an invaluable starting point. Historians of language are already having a field day playing with such massive databases. The librarian and lexicographer Fred Shapiro, for example, has uncovered uses of such phrases as "double standard" (1912), "Native American" (for American Indian, 1931), and "solar energy" (1914) that predate citations in the *Oxford English Dictionary* by decades.¹¹

JSTOR lacks the scholarship of the past five or six years, but online databases from Johns Hopkins University Press's Project Muse and the History Cooperative increasingly provide that as well. Although the History Cooperative, JSTOR, and Muse all restrict access to subscribers, they have emerged under nonprofit auspices. But increasingly important online collections of historical data are in the hands of commercial vendors such as Bell & Howell and the Thomson Corporation, which have vast archives of scholarly publications and primary sources, and Corbis, with its unparalleled archive of historical images. These are the exemplars of the private history Web—a growing realm both under corporate control and accessible only to paying customers.¹²

Everyone a Web Historian: Grass-Roots History Online

Despite the growing significance of the private history Web, the greatest energy over the past decade has actually been in the public Web—public in the sense of both its open access and its control by individuals, nonprofits, or government agencies. Indeed, an astonishing grass-roots movement has fueled its enormous growth. Over the past five years, academic historians, history teachers, and history enthusiasts have created thousands of history Web sites. No one has managed a definitive count of these Web sites, although Yahoo!'s United States history directory includes more than 4,500 sites—a fivefold increase since 1996. My own Center for History and New Media maintains searchable databases of more "serious" history Web sites and has indexed more than 2,100 of them.¹³ Although perhaps one-third of history Web sites have .com addresses (signifying the "commercial" domain in contrast to .edu, .org, or .gov), most of those are actually set up by individuals using free space (albeit festooned with banner and pop-up ads) provided by such companies as AOL (America OnLine), Geocities (a part of Yahoo!), CompuServe (an AOL subsidiary), Lycos, or Prodigy. To a surprising degree, then, history Web sites come from both academics and amateurs who have posted historical material online primarily as a labor of love—the original meaning of amateur.

Civil War enthusiasts, not surprisingly, have brought some of the same passion to presenting history online that they regularly display at Civil War reenactments. "Some days," observes *Choice*, the journal of academic libraries, "it appears that the Internet consists of equal parts Star Trek, stock market reports, soft-core pornography—and Civil War sites." And the historians William G. Thomas and Alice E. Carter have recently filled a two-hundred-page book surveying the Civil War on the Web, "a guide to the very best sites." Although many of these sites come from large institutions such as the Library of Congress, the National Park Service, and the Virginia Center for Digital History (with which Thomas and Carter have been associated), hundreds of passionate and dedicated amateurs have created remarkable sites without any outside financial or institutional support. Thomas R. Fasulo, an entomologist, has, for example, assembled an immense archive on the battle of Olustee (the largest Civil War battle in Florida)—more than forty official reports, fifty firsthand reminiscences in letters, articles, and books, and detailed coverage of all the units participating in the battle. The reenactor Scott McKay has developed an equally massive site on the Tenth Texas Infantry filled with rosters, casualty lists, ordnance records, battle reports,

reminiscences, and personal letters.¹⁴ To be sure, Civil War enthusiasts such as Fasulo and McKay flourished well before the emergence of the Web, but the Internet has made their passions visible and accessible to a much wider audience.

Genealogists have similarly found the Web a welcoming arena for engaging in their passion for the past. The USGenWeb Digital Library has mobilized hundreds of local volunteers to create online transcriptions of census records, marriage bonds, wills, and other public documents. The Family History Library of the Church of Jesus Christ of Latter-day Saints (the Mormon Church) has thrown open its massive genealogical databases, including the Ancestral and Pedigree Resource files (a database of family trees submitted to the Family History Library) and the International Genealogy Index (a name index of records collected by church members)—660 million names in all—the fruits of more than a century of Mormon genealogical work.¹⁵

Family historians have visited such sites in amazing numbers; the Mormon Church's site attracts 129,000 visitors per day, an annual rate of close to 50 million. Online resources have drawn tens of thousands more Americans into the already popular practice of tracing family roots—the most common form of historical research in the United States. Significantly, the Internet's greatest impact may lie in connecting people in common pursuit of their roots, allowing them to share information on common ancestors or to help out fellow genealogists by investigating a local lead. The Mormons alone sponsor 137,000 collaborative e-mail lists to facilitate research exchanges. While the Web has served largely as a publishing and archiving medium for already committed Civil War enthusiasts, it has brought new participants to genealogy by making the sources for family history more readily available. Print authors have even noticed the popularity of Web-based genealogical research; at least a dozen published guides—including *Genealogy Online for Dummies*—offer advice to enthusiasts.¹⁶

The breadth of this grass-roots effort becomes clear when we look at who has posted a random selection of historical documents online. I pulled Diane Ravitch's anthology *The American Reader: Words That Moved a Nation* off my shelf and found online fifteen of the twenty documents (many of them far from mainstream) in her chapter "The Progressive Age." Teachers constituted the largest group of people who have made these documents publicly available—a communications professor at the University of Arkansas posting Elizabeth Cady Stanton's "Solitude of Self," a community college instructor in Ohio providing the Niagara Movement Declaration of Principles, a Hartsdale, New York, high school teacher digitizing M. Carey Thomas's "Higher Education for Women."¹⁷ But many others had little or no academic connection. A black organizer includes W. E. B. Du Bois's "Talented Tenth" essay on his Web site (*Mr. Kenyada's Neighborhood*) because he believes that Du Bois's vision "of our potential capacity to solve problems internally" provides the basis for a new "community-based activism." A German purchasing agent puts Joe Hill's "The Preacher and the Slave" on his *History in Song* Web pages that preserve songs from an American studies course he took at Johannes Gutenberg University a quarter of a century ago. The General Board of Discipleship of the United Methodist Church publishes "Lift Every Voice and Sing," by James Weldon Johnson and J. Rosamond Johnson, with the suggestion that congregations "sing this hymn in worship on a Sunday in February [2000], and celebrate its one hundredth anniversary." The amateur poet Kevin Taylor's Web site includes Alice Duer Miller's pro-suffrage verse "Evolution" because "its message is as important and clear today as it has always been," and Miller "is also the author of *The White Cliffs*—one of my favorite books." The Web takes Carl Becker's vision of "everyman a historian" one step further—every person has become an archivist or a publisher of historical documents.¹⁸

Many of these grass-roots efforts are quite modest, poorly designed Web sites proffering one or two favorite documents with little historical context. But others have grown into massive archives. In early 1995 the graduate student Jim Zwick began posting a few documents on anti-imperialism, the subject of his Syracuse University dissertation, on the Web. Like most historians, Zwick had assembled his own personal archive; he realized that the materials gathered for scholarly research could be made public through the World Wide Web. Five years ago, Zwick was one of the Web history pioneers; now his efforts have expanded well beyond anti-imperialism into such topics as political cartoons and world's fairs and expositions and thousands of historical documents personally digitized by Zwick. The volume of material and the number of users have multiplied more than fivefold. Although Zwick's Web site (now called BoondocksNet.com) remains a one-person operation, its increasing scale has forced him to take ads and sell books in order to support the growing hosting and software costs. Zwick has blazed a path that many future graduate students may (and I think should) follow. Why not take the least visible and most private part of the scholar's work—assembling a body of primary documents—and make it public?¹⁹

The most massive grass-roots Web history effort linked to scholars is, of course, H-Net: Humanities & Social Sciences OnLine. Well known to historians for the more than a hundred specialized discussion lists that it sponsors, H-Net also has a major Web presence, which includes searchable archives of the list discussions. HNet has not been heavily involved in posting historical documents, but its archives are now themselves a significant primary source for the thinking of professional historians, as well as an eclectic reference source to important books and teaching tools. Its most profound impact, however, has been on modes of scholarly communication; since its lists include 60,000 subscribers in ninety countries, it has become an essential way for historians to find out about conferences, grants, jobs, and teaching resources. To some degree, it has also accelerated the pace of scholarly discourse. In 1998, for example, subscribers to H-Amstdy, a part of H-Net, extensively debated Janice Radway's presidential address to the American Studies Association before it had been published in *American Quarterly*. Hundreds of volunteer list editors keep H-Net going, although the energy of Executive Director Mark Kornbluh, who has been very successful in obtaining government grants and university support, has also been

vital to its maintenance and growth. As a result, H-Net remains a free scholarly resource that is also open to interested participants from outside the academy.²⁰

The greatest strength of the grass-roots history Web—its diversity and its links to nonprofessionals—is sometimes its greatest weakness. While academically trained historians such as Zwick and the H-Net community have joined in the bottom-up effort, its overall amateur and eclectic quality obviously poses problems for those committed to professional standards. William Thomas, for example, pronounces Civil War history on the Web "anemic" as well as "healthy." Few sites, he notes, "advance new ideas about the history of the period"; most ignore the scholarly trend toward social history and focus relentlessly on generals and battles. Still worse, "many web sites broadcast old prejudices, ancient theories, and long-disproved arguments about the Civil War," especially the view that the war was fought over tariffs rather than slavery. One site argues, "conditions in northern factories were as bad or worse than those for a majority of slaves" and rejects as "simplistic" the idea that "the Civil War was fought over slavery."²¹

Even amateur sites that stick to presenting primary sources rather than historical interpretations do not always meet professional standards. Reenactors digitizing battle reports or labor organizers posting Joe Hill songs generally do not fuss about proofreading and copy editing. Nor are nonprofessionals inclined to worry about definitive editions, editing, or careful contextualization. There are at least sixteen different online versions of Elizabeth Cady Stanton's well-known speech "Solitude of Self"; they provide conflicting dates on which she gave the speech and different bodies to whom she presented it. Paragraphing and punctuation vary widely, and some excerpt or even edit the speech without indicating the intervention. Only one provides a link to the Library of Congress, which has online a facsimile of a printed pamphlet version of the speech.²²

Some documents found on the Web are, in fact, not "real" documents at all. At least three Web pages promise the "voice" of Eugene Debs, but the recording is actually that of Len Spencer, who recorded one of Debs's speeches around 1905.²³ More than two dozen different Web sites offer versions of what they call the "Willie Lynch speech of 1712," in which a British slave owner from the West Indies allegedly advises Virginia slave owners to control slaves through a strategy of divide and rule. Sometimes the sites add an introduction supposedly written by Frederick Douglass; others falsely describe Lynch as the source of the word "lynching." Despite the sites' repeated assurances about the speech's "authenticity," internal evidence readily betrays its twentieth-century origins. The language incorporates modern syntax, and the content focuses on current-day divisions such as skin color, age, and gender rather than ethnic and national divisions much more important in the early eighteenth century.²⁴

To be sure, a careful search of the Web also turns up evidence of the dubious origins of the Lynch speech. Still, those sites that take the speech entirely at face value overwhelm the Web sources that dispute it. Anyone who simply searched for "Willie Lynch" on the Web would be more than ten times as likely to find evidence of the speech's "authenticity" than information that casts doubt. But the Web is unique in the way it offers entry into the world of information and misinformation in which most people operate and allows us to consider the significance and spread of such urban legends as the Willie Lynch speech, which are orally transmitted at such events as the 1995 Million Man March or the 2001 inaugural protests. The Web itself cannot be blamed for misinformation or misrepresentation; the Lynch speech, in fact, appeared in print as early as 1970. The Web increases our access to documents and information, both spurious and authentic. For both better and worse, the virtual archive of the Web distinguishes itself from traditional libraries and archives by its indiscriminating inclusion of the best—and worst—that has been known and said.²⁵

Despite the abundant misinformation available online, the Internet is—somewhat paradoxically—a superb source for basic factual research, especially when used by those who are careful to determine source quality. My own rendering of the Willie Lynch story comes entirely from research in online sources. Although I have a substantial reference library at home, I now do most of my historical "fact checking" on the Web. I can find correct spellings, birth dates, battle deaths, and election results in online sources more quickly and more accurately than in most standard reference works. The key caveat, of course, is "careful to determine source quality," but most professional historians—and probably most advanced history students or most sophisticated general readers—possess this skill.

Deepening the Public History Web: Universities, Foundations, and the Government

While the largest number of Web sites with historical documents and content have emerged out of this eclectic, grass-roots effort, the largest *volume* of historical documentation exists within the deep Web of online databases and the private Web of materials open only to those who pay. Both efforts share some basic similarities—massive scale and use of databases to organize the materials. But only paying customers can visit the private Web.

Surprisingly, enormous amounts of free online historical material have appeared in the past five years, and much more will appear in the next decade. These sites have primarily benefited from government or foundation funding or, in many cases, both. The most important project, the Library of Congress's National Digital Library, has spent about \$60 million to put more than 5 million historical items online between 1995 and 2000—with three-quarters of the funding coming from private donations. Ameritech, the former Bell telephone company for the Midwest (now owned by SBC Communications), worked with the Library of Congress to provide \$2 million for more than twenty digitization projects at libraries across the country.²⁶ The heavy corporate funding naturally raises the specter of probusiness bias in what gets digitized. The AT&T Foundation, for example, has supported the digitizing of the Alexander Graham Bell Family Papers. The

Reuters America Foundation was probably more likely to support the digitizing of the George Washington Papers than the records of the National Child Labor Committee. Nevertheless, Ameritech has, for example, funded the Chicago Historical Society's efforts to bring its collection of Haymarket affair materials to the Web.

The National Endowment for the Humanities (NEH) has also supported many important projects, particularly favoring those with an educational mission and focus on particular topics. The well-known Valley of the Shadow Project at the University of Virginia brings together a stunning archive of documents about two nearby counties (Augusta County, Virginia, and Franklin County, Pennsylvania) on opposite sides during the Civil War era. Already a major Web destination in 1996, its collection of letters, diaries, newspapers, censuses, and photographs has multiplied tenfold in just the past four years. *The Valley of the Shadow* is remarkable not just for its depth and sophistication but also because it has no physical counterpart. Edward L. Ayers, William G. Thomas, and their collaborators have literally created an archive that did not previously exist by hunting down and digitizing documents found in both public repositories and private hands.²⁷

The New Deal Network (NDN), another NEH-funded project, has similarly created a new, virtual archive, with more than 20,000 photographs, political cartoons, and texts (speeches, letters, and other documents) gathered from multiple sources. Sponsored by the Franklin and Eleanor Roosevelt Institute and led by Tom Thurston, the *New Deal Network* lacks the comprehensiveness of the *Valley of the Shadow*, but it offers a remarkable resource for anyone teaching about the 1930s and 1940s. *History Matters: The U.S. Survey Course on the Web*, the product of my own Center for History and New Media and the American Social History Project and funded by NEH and the Kellogg Foundation, has digitized hundreds of first-person historical documents and contextualized them for use in high school and college classrooms.²⁸

In contrast to the "invented archives" represented by the *Valley*, *NDN*, and *History Matters*, *Documenting the American South* opens up an existing archive—the University of North Carolina at Chapel Hill's unparalleled southern collections—to remote students and scholars. Funded by various grants (from NEH, Ameritech, and the Institute of Museum and Library Services), *Documenting the American South* organizes thousands of documents (largely texts) around such specific topics as "Southern Literature," "First-Person Narratives," "Slave Narratives," "The Southern Homefront, 1861–1865," and "The Church in the Southern Black Community."²⁹ 31

The National Science Foundation (NSF), with a budget thirty times that of NEH, has emerged as an important funder for "digital libraries" as a result of its interest in computing issues rather than in the quality of the content being provided. Whatever the motives, NSF has financed some projects of enormous interest to historians. Michigan State University's National Gallery of the Spoken Word (NGSW) is developing techniques for automatically searching large volumes of spoken materials, including, for example, thousands of hours of nightly TV news broadcasts. Historians may not care about the underlying computer science, but if the NGSW succeeds in creating a "fully searchable digitized database of historical voice recordings that span the 20th century," they will make extensive use of it in their teaching and research.³⁰ 32

Whereas NEH funding has largely supported the creation of digital projects for use in the classroom and NSF has concentrated on the intersection of computing and humanities problems, the Mellon Foundation has focused on library-related issues, especially preservation and storage. It has provided substantial funding to the Cornell and University of Michigan libraries to preserve and then make available a major library of printed materials published between 1850 and 1877 under the rubric of the "Making of America" (MOA). The University of Michigan portion of the collection alone will soon encompass more than 9,600 monographs, 50,000 journal articles, and 3 million pages—a significant portion of the library's imprints from those years.³¹ 33

Like scholars using NDLP, those using MOA can find information previously available in theory but not necessarily in practice. Steven M. Gelber, who was researching the origins of hobbies, notes that he turned up "a treasure trove of data in a matter of a couple of days" that would have taken months to find through traditional research. He calls MOA "the most exciting thing I have seen in research since I first discovered Xerox machines in 1967 and realized I did not have to take notes anymore." This "is what I assumed the future of libraries would be but to be quite honest, I never believed I would live to see so much of the past put online in such an accessible form."³²

Despite the enormous value of the MOA and similar projects, some cautions are in order. Some object that such efforts are a form of burning down the village to save it, since most of the books will ultimately be discarded—both because they are cut up to be scanned and because the storage space is valuable. The novelist Nicholson Baker, for example, has sharply criticized earlier newspaper microfilming projects that have led to the similar destruction of paper copies of the newspapers. As the result of Library of Congress microfilming efforts, for example, libraries across the country dumped their hard copies in the belief that there was now a standard, comprehensive microfilmed version of newspapers that could be reproduced, ordered, and consulted. But Baker argues that the anomalies and holes (missing issues, pages, etc.) in the Library of Congress collection have now become permanent holes in some newspaper records because of the ensuing destruction of holdings in other libraries.³³ Baker and others also note the value of marginalia and other markings that get lost with the disappearance of paper copies as well as the difficulties of fully reproducing images such as nineteenth-century engravings in digital form. Librarians, on the other hand, argue that books and newspapers printed on acidic paper were crumbling and that microfilming or digitizing offers the only practical alternative and the only way to supply "the most content to the most people in a cost-effective manner." While some scholars will bemoan the loss of tangible, historical evidence in the transition from paper to digital images (just as they mourn the disappearance of the card catalog), many others will benefit from their ability much more readily to access the volumes in the MOA collection, many of which are not in a standard university library, and even more the possibility of searching them by words in the text rather than just by title.³⁴

Indeed, the incredible ease of using these newly digitized works may actually pose a problem for future historical work. The MOA collection largely draws from books from Michigan's remote storage that had rarely been borrowed in more than thirty years. Yet the same "obscure" books are now searched more than 500,000 times a month. Will digitization create a new historical research canon in which historians resort much more regularly to works that can be found and searched easily online rather than sought out in more remote repositories? Years ago, the *New York Times* ran an advertisement with the tagline "If it is not in the New York Times Index, maybe it didn't happen." Could we arrive at a future in which, if it is not on the Web, maybe it didn't happen?

Such concerns aside, these grass-roots, government, and nonprofit efforts have begun to deliver, as Gelber observes, "what people have been talking about for ten years—a genuine electronic library, or at least an electronic archive." Historians will spend years examining these digital sources and will not readily exhaust their possibilities. Although the Founding Fathers may be better covered in these resources than labor or feminist militants are, the Web in fact now offers material stretching across the broad range of topics that interest contemporary historians. The always precarious state of the public sphere in contemporary America poses one crucial peril for the continued expansion of this burgeoning free archive. For example, the budget of NEH, the most important funder of humanities work, has declined (in real terms) by about two-thirds in the past twenty years.³⁵ And in the past several years, it has had to fight for its survival. NEH may now face further threats with a Republican president and Congress who traditionally have not been sympathetic to the public sector.

Despite the great success of *American Memory*, which receives 18 million page views per month and has brought primary sources into K–12 classrooms across the country, the Library of Congress seems to be shifting away from its focus on putting its historical collections online. A report by the National Research Council in the summer of 2000 criticized the library for, in effect, paying too much attention to historical sources and not enough to recently created "born digital" materials such as Web sites and electronic journals and books. James O'Donnell, vice provost for Information Systems and Computing at the University of Pennsylvania who chaired the committee producing the report, told the *New York Times*: "Digitizing your analog material is less urgent. . . . [I]f you don't do it this year, it'll still be there in five years, and you could

do it then. Digital information that you're losing is probably lost forever."³⁶ If the Library of Congress turns away from the massive digitizing efforts of the past five years, *American Memory* may turn out to be a forgotten memory from the late twentieth century.

Moreover, most of the government or foundation funding has been significantly enhanced by university support (another part of the endangered public sector) and by substantial infusions of sweat equity from digital pioneers. When the creation of online archives becomes routine, will that university and volunteer support remain available? In other words, is there a stable basis for the continued funding of public sector efforts to create a public, free historical archive?

The continuing erosion of the "public domain" further threatens the public Web. Copyrighted material previously entered this intangible realm of unrestricted use after a twenty-eight-year term renewable once, or a maximum of fifty-six years. In 1976, the copyright law narrowed the public domain by lengthening most existing copyrights to seventy-five years. As a result, the only large bodies of materials for the years after 1923 (the year after which copyright covers most work) are government documents such as the WPA (Works Progress Administration) life histories or the FSA photographs. The Sonny Bono Copyright Term Extension Act of 1998, which extended copyrights for an additional twenty years (in part due to the aggressive lobbying of the Disney Corporation, whose Mickey Mouse was scurrying toward the public domain) means that the copyright line will remain frozen at 1923 until 2018. Thus, Web surfers can easily read F. Scott Fitzgerald's *Tales of the Jazz Age* (1922) but not *The Great Gatsby* (1925), which will not find its way online until 2020. The 1998 copyright extension delivered the single greatest blow to the creation of a free, public historical archive; yet historians were barely at the table when that act passed, crowded out by the high-priced suits from the big media conglomerates. Copyright restrictions are one reason for the persistence of fading digital formats such as CD-ROM. The two United States history CD-ROMs on which I have worked contain copyrighted materials for which we could purchase permission to use in the CD-ROM but not on the Web.³⁷

Selling the Past Online: Information Conglomerates and Internet Startups on the Private History Web

For historians, copyright protection has redlined not only much twentieth-century history but also most secondary literature out of the public Web. But because the problem involves rights and money, one solution similarly involves rights and money: companies that provide copyright digital content, charge for it, and then compensate rights holders out of their revenues. That said, the particular models for selling digital content vary widely as the corporations in the emerging "information business" scramble to evolve the most profitable business model. 41

The most common approach involves high-priced library-based subscriptions to digital content. Individual library subscriptions, which allow the library to provide the materials to all its patrons, generally cost thousands of dollars. The Virtual Library of Virginia (VIVA), which purchases electronic databases for the state's thirty-nine public college and university libraries (a consortium arrangement increasingly common in this environment), currently spends more than \$4 million per year for electronic subscriptions, and individual libraries in the consortium are spending thousands, if not millions, more.³⁸ Annual subscriptions to periodical databases such as ProQuest Direct and Expanded Academic ASAP (EAA) typically run around \$30,000 to \$50,000 for colleges and universities. 42

Other vendors sell digital content on an item-by-item basis—"by the drink"—instead of by subscription. Northern Light, which modestly aspires (in the words of its chief executive officer) "to index and classify all human knowledge to a unified consistent standard and make it available to everyone in the world in a single integrated search," offers more than 700 full-text publications (including a number of history journals) on a per-article basis. You can, for example, get Howard Zinn's article in the *Progressive* on "Eugene V. Debs and the Idea of Socialism" delivered instantly to your Web browser for \$2.95. Contentville, which has more of the feel of a magazine (it was founded by Steven Brill, who made his millions with such publications as *American Lawyer*), offers a smaller selection of articles at similar prices as well as primary 43

source documents such as speeches and legal documents. Prominent academic experts such as Sean Wilentz and Karal Ann Marling recommend the best books on "American Politics since 1787" and "Popular Culture," and contributing editors share their favorite Web sites.³⁹

The vast image library controlled by Corbis, the company owned by the Microsoft founder Bill Gates, offers up the most massive historical database available on the pay-per-drink basis. Corbis has swallowed up many of the world's largest image collections, including the Bettmann Archive and the French photo firm Sygma, and has licensing arrangements with leading photographers and repositories around the globe (from the National Gallery in London to the State Hermitage Museum in St. Petersburg). It also represents another example of the trend toward massive concentration in the digital environment. Increasingly, the world's images are coming under the control of just two giant Seattle-based firms—Corbis and Getty Images, owned by the oil heir Mark Getty. Both aspire to be, as a Corbis ad says, "your single source for an array of diverse images"—"The Place for Pictures Online," in its trademarked phrase. More than two million of Corbis's 65 million images are digitized and available through a fast search engine. Anyone who has done photo research for a book or article will appreciate the ability to sit at home and browse through this incredible collection—seventeen superb photos of Eugene Debs, for example. You can look for free, but using the images (emblazoned with "corbis.com" in the online version and protected with digital watermarks) comes with a price tag that escalates as you move up from a digital image for your personal Web page (\$3), to a glossy print for your wall (starting at \$16.95), to an image that you can publish in a book (generally \$100 or more).⁴⁰

45
 Corbis's charges reflect copyrighted images in many cases, but in others they rest on the company's ownership of an image published widely in the pre-copyright era and available for free if you can get a copy from a less fee-hungry source such as the Library of Congress. You can pay Corbis \$3.00 for a digital image of Walker Evans's photo of the "Interior of a Depression-Era Cabin" or download a higher quality version of the same image in *American Memory* for free. *American Memory* also provides a fuller identification and contextualization of the photo, since its goals are educational and scholarly rather than just pecuniary. Similarly, you can purchase Eugene Debs's 1918 Canton, Ohio, speech, which helped land him in prison for sedition, from Contentville for \$1.95 or you can pick it up for free on at least four different Web sites.

46
 Costs aside, these online databases are already revolutionizing the way historians do their research. Most familiar to historians are the massive bibliographic databases such as America: History and Life and the Arts and Humanities Citation Index. Once upon a time (that is, five or six years ago), historians searched through annual bound volumes to develop bibliographies. Now they typically do these searches quickly and at their own convenience. After assembling a bibliography, historians used to search for and copy articles. But now they can find the full text of a surprisingly wide selection of secondary works online.

47
 The major online sources for full-text journals—Bell & Howell's ProQuest Direct, the Thomson Corporation's Expanded Academic ASAP (EAA), and EBSCO—offer thousands of journals, including dozens of major historical journals, generally from 1989 to the present.⁴¹ Despite some gaps such as most state historical society publications, these databases contain a large percentage of the journal literature of the 1990s that historians would need to consult. Two other nonprofit, but still gated, resources—Project Muse and the History Cooperative—fill in some important gaps in what ProQuest and EAA offer. For still older sources, JSTOR (also available only through hefty library installation charges as well as an annual maintenance fee) provides comprehensive coverage, albeit for a smaller set of journals.

As yet, historical monographs cannot be found in cyberspace as readily as journals can. But perhaps not for long. Questia Media, Inc., backed by \$130 million in venture capital, has created an online liberal arts library of 50,000 scholarly books, which they hope will increase to a quarter million volumes by 2003—what they call the "world's largest digitization project." Taking an approach different from that of ProQuest and EAA, Questia intends to sell subscriptions for \$19.95 per month to "time-crunched" students, who they believe (in the face of some reasonable skepticism) will pay for access to materials that will help write their papers more quickly. At least in history classes, the investment may not pay off: although Questia has more than 9,000 history titles, not a single one of the ten history monographs that United States historians, in a *Journal of American History* survey, listed as "most admired" can be found on the online library's shelves. Its competitors, NetLibrary (with more than \$100 million in venture capital and 25,000 books already online) and Ebrary.com, have still other business models. NetLibrary sells libraries electronic copies of books that can only be accessed by one person at a time; if someone has "checked out" the book, then no one else can "take it out." It markets its 25,000 books in different groupings ranging from the 618-title "business school collection" at an average price of \$40 per volume to 126 volumes on "Countries, Cultures, and Peoples of the World" to 214 volumes of "Cliffs Notes" (the actual literary works are generally thrown in free since they are in NetLibrary's collection of 4,000 public domain books). Ebrary, by contrast, allows users to browse books without charge but requires payment for printing or copying a portion of a book.⁴²

Not all pay services offer copyrighted content. Some serve public domain content but charge in an effort to recoup their digitizing costs. One of the pioneers in this has been HarpWeek, a personal project of John Adler, a retired businessman with an interest in nineteenth-century American history. While most digitizing projects rely on "keyword" searching of the full text, Adler has employed dozens of indexers to read every word in *Harper's Weekly* and examine every illustration and cartoon to create a human index of the full run of the magazine from 1857 to 1912. That labor-intensive indexing means, for example, that *HarpWeek* offers better image searching than many other online sources since the brute power of keyword searching brings much greater rewards in historical texts than in images. Adler has created an extraordinary research resource for nineteenth-century

historians, although an expensive one—the first twenty years, now available, retail for close to \$35,000.⁴³

We can glimpse the outlines of a still more remarkable project—the full text of the *New York Times* for the years 1851 to 1923. The "Universal Library" at Carnegie Mellon University (with aspirations similar to Nelson's Xanadu project and support from Seagate Technology) is scanning the entire public domain era of the *Times*, which it will make available for free online reading. At the same time, it is using optical character recognition to turn the *Times* into searchable text, although the quality of the result remains uncertain at the moment. The Universal Library plans to offer free views of the page images but to charge for access to the searchable text—perhaps \$40 for lifetime subscriptions. At the moment, the vision is more exciting than the implementation—you can't search yet, and the scanned microfilm provided for 1860–1866 includes a number of unreadable pages.⁴⁴

The plan of the university-based Universal Library to charge subscriptions suggests a type of history Web site that sits uneasily between the "public" and "private" categories that we have been using. Like JSTOR and Project Muse—both of them nonprofit ventures that have received substantial support from the Mellon Foundation—it is "public," rather than private, in its ownership, control, and eschewing of profit. Yet, it is (or will be) "private" in its restriction of full access to those who pay. Despite their foundation funding, groups such as JSTOR and Project Muse argue—quite reasonably—that they need income to sustain their operation, to add new journal articles, and to maintain the service. Thus, they charge substantial subscription fees to libraries. Unfortunately, when nonprofits enter the private Web, they not only restrict access but also incur substantial costs; JSTOR and Project Muse spend a considerable part of their income not to create or post content, but to market their services and keep out unauthorized users. Michael Jensen, who helped develop Muse, estimates that "over half of the costs of the online journals project was attributable to systems for *preventing access* to the articles."⁴⁵

50

51

Moreover, even where publication, preservation, or distribution is turned over to a nonprofit such as JSTOR or Project Muse, scholarly authors and journals are still giving up control over presentation and access to a separate entity. The History Cooperative—a partnership of the University of Illinois Press, National Academy Press, the Organization of American Historians, and the American Historical Association—has pioneered the alternative idea of a "cooperative" in which scholars and scholarly organizations will retain a say over these questions.⁴⁶ Historians from these professional societies and their journals felt that this arrangement would allow them, for example, to offer to make their electronic journals as widely available as possible. Hence, while the electronic *Journal of American History* and *American Historical Review* will only be available to subscribers, there is no *additional* subscription charge to individuals or libraries for access. Having a say in a cooperative also makes it easier to experiment with one of the key questions facing scholars—will digital environments allow us to present our scholarship in new—and better—ways?⁴⁷ In the end, the measure of success of scholarly and nonprofit societies is how they improve scholarship and society, not how much revenue they generate.

Some argue that, given these larger social and scholarly goals, scholars should move toward total, free access to the fruits of scholarship, which is, after all, mostly publicly funded in the first place. In 1991 Paul H. Ginsparg, a physicist at the Los Alamos National Laboratory, created *arXiv.org e-Print archive*, which has become an open repository of more than 150,000 "preprints" (non-peer reviewed research papers) in physics, math, and related fields. "E-print" archives in psychology, linguistics, neuroscience, and computer science similarly offer electronic preprints on a free access basis. The Open Archives Initiative advocates expanding these efforts so that they will be "interoperable" (for example, allowing easy searching across multiple archives); include peer-reviewed work; and ultimately form the basis of a "transformed scholarly communication model." The computer scientist Stevan Harnad, one of the most aggressive promoters of such open systems, envisions a future in which "the entire refereed literature will be available to every researcher everywhere at any time for free, and forever."⁴⁸ Thus far, scientists have dominated such open scholarly archive experiments. It remains a question whether they are easily transferable to the humanities, which lack the same preprint traditions and where

speed of publication is much less important. Moreover, the extraordinarily high prices of commercially published science journals have further driven these efforts. No one worries about putting commercial science publishers out of business. But the losers in the demise of the scholarly history journals will be university presses and scholarly societies.

If scholarly societies such as the Organization of American Historians are to survive in a world where all scholarly information is free, they will need to come up with alternative revenue models to support their operations. One promising approach to resolving the contradiction between free public access and continued revenue to support scholarly editing and publication has been pioneered by the Open Book project at the National Academy Press (NAP), which has been led by Michael Jensen, who has also been a key figure in Project Muse and the History Cooperative. NAP, the publishing arm of the National Academy of Sciences, has put its entire front list and much of its backlist online for free in a page image format. Ironically, giving this material away has actually increased NAP's sales because people now order books that they have browsed online but want to own in a hardcopy. Moreover, the book itself—indexed by Web search engines—becomes its best advertisement. Jensen, thus, argues that "free browsing, easy access, and researcher-friendly publication first, and sale second" is "much more in keeping with the role of a noncommercial publisher" and its mission of doing "the most good for society as possible within the constraints of our money."⁴⁹

54

Who Owns the Past Online? Access and Control on the Private History Web

These massive projects, whether public or private, will surely transform historical research and ultimately writing. Those who received their Ph.D.'s before 1990 will probably spend the rest of their careers regaling graduate students with tales of how "in my day, we spent hours turning microfilm readers looking for relevant newspaper articles." Given the enormous gift that commercial digitization is bestowing on the historical profession, it seems a bit churlish to look this particular gift horse in the mouth.

55

Churlish, but surely necessary. Once we get over our excitement 56
 about the digital riches on our screens or the new modes of
 research being opened up, we need to think about the price tag. To
 be sure, in most of the emerging models, libraries rather than
 individual researchers are paying that fee. Still, that money is not
 appearing magically; it is draining other parts of library budgets.
 One part of the budget that is being sucked dry is that for
 purchasing real, not virtual, library books, especially scholarly
 books. To be sure, the main villains in the current crisis in
 scholarly publishing are the commercial vendors who charge
 rapacious prices for science, technology, and medicine journals.
 Libraries that pay \$16,344 annually to subscribe to Reed Elsevier's
Brain Research cannot afford as many history monographs as they
 once purchased—a fact that both scholars and university presses
 are painfully confronting. But electronic resources are also
 squeezing library budgets—they now consume 10 percent of
 library materials budgets, compared to only 25 percent for
 monographs.⁵⁰

The digital library fees also generally flow into the hands of 57
 publishers and especially commercial aggregators rather than
 authors. Freelance writers have sued newspapers and magazines for
 including their work without permission (or compensation) in
 databases marketed by Lexis-Nexis (Reed Elsevier) and Bell &
 Howell. And book publishers have been slow to decide what
 portion of e-book revenues they are going to share with authors.⁵¹

In addition, the appearance of these gated databases poses a 58
 particular problem for independent scholars not affiliated with
 academic institutions. If they happen to live near a major public
 library, they can often access the databases within the walls of that
 library. But they do not have the convenience available to most
 university-based historians of using these resources from their own
 homes.⁵² The same problem faces those affiliated with smaller
 institutions that cannot afford the hefty subscription fees. Some
 scholars, however, now have enhanced access to resources; in
 Virginia, VIVA's statewide subscriptions give historians at
 community colleges and underfunded traditionally black colleges
 access to the same electronic resources as faculty at the well-
 endowed University of Virginia. Nevertheless, signs of an
 academic digital divide loom not only between institutions but also
 within them. For example, law school students and faculty
 generally have access to the complete Lexis-Nexis database (with
 considerable resources for historians), which is generally closed to

other parts of the university. Of course, scholars affiliated with more affluent institutions (and parts of institutions) have always had advantages over their colleagues, and independent scholars have always faced barriers to access.

A more worrisome prospect has to do with the emerging economic structure of the information industry. Previously, publishing was a relatively decentralized and small-scale business with many different publishers, large and small. But online information providers, like many other "new economy" businesses, benefit from a powerful combination of economies of scale and "network effects." In the information business, the fixed costs (for example, software development) are the most important costs; once they are covered, it is not much more expensive to sell to 3,000 libraries than to 30. And "network effects"—the benefits of using a system increases as more people use it since, among other things, they will be familiar with its interface—mean that the biggest players will tend to get bigger. Whereas the factory-based economy favored oligopolies, the information economy is more likely to result in monopolies.⁵³

Not surprisingly, then, the online vending of electronic data has already become concentrated into a very small number of hands. Four gigantic corporations—Reed Elsevier, EBSCO, Bell & Howell, and Thomson—are especially prominent in the provision of electronic content to libraries. Reed Elsevier, which focuses particularly on science journals, is less significant for historians (although it does sell Lexis-Nexis, the online data service vital to anyone writing on the recent past). The privately held EBSCO, which has \$1.4 billion in annual sales, produces nearly 60 proprietary reference databases and full-text versions of more than 2,000 publications. Bell & Howell is a billion-dollar corporation, which acquired UMI (formerly University Microfilms International) in 1985 and Chadwyck-Healey (a leading provider of humanities and social science reference and research publications) in 1999. Its databases include over 20,000 periodical titles, 7,000 newspaper titles, 1.5 million dissertations, 390,000 out-of-print books, 550 research collections, and over 15 million proprietary abstracts. These resources constitute an archive that includes more than 5.5 billion pages of information—all of which is being converted into digital form (though not necessarily searchable text) under the "Digital Vault Initiative," which the company says will create "the world's largest digital archival collection of printed works." ("World's largest" is a popular claim in cyberspace.)

Ultimately, Bell & Howell will offer online the full runs of at least fifty periodicals such as the *New York Times*, *Time*, and the *Wall Street Journal*. (Astonishingly enough, given the scale of the effort involved, Bell & Howell intends to create its own searchable edition of the *New York Times*, and its version will come up to the present rather than stop in 1923.)⁵⁴ The microfilm era in research, which Bell & Howell's UMI launched in 1938, will soon come to an end.

Bell & Howell's even larger rival is the Canadian Thomson Corporation, a "global e-information and solutions company" with close to \$6 billion in annual revenues. Thomson's Gale Group sells thousands of full-text publications (including history journals) to libraries under the "InfoTrac" brand, which includes EAA. It also has extensive reference holdings, including works that historians regularly use (for example, from Macmillan Reference USA and Charles Scribner's Sons). More recently, it has bundled its various products as well as some licensed from other vendors into what it calls its "History Resource Center," billed as "the most comprehensive collection of historical information ever gathered into one source." Designed primarily for undergraduates and to be purchased by college or university libraries, it includes primary documents (from an archive accumulated by Primary Source Media, another Thomson subsidiary), encyclopedia articles, full-text periodicals and journals, maps, photographs and illustrations, overview summaries, a timeline, a bibliography, and annotated links to online special collections. These resources do not come cheap. Prices vary considerably depending on particular arrangements, but an annual license for two simultaneous users can run close to \$12,000.

Bell & Howell and Thomson are involved in a dense web of connections with other online ventures. Thomson, for example, holds the largest stake in WebCT.com, which provides widely used software for placing courses online but bills itself more broadly as an "e-learning hub." WebCT has developed discipline-specific online communities with forums and other resources, including one in history. Part of the reason for Thomson's "strategic investment" is presumably to encourage the selling of custom course materials created by Thomson to students in courses managed through WebCT. Bell & Howell is also eyeing the lucrative textbook (or "courseware") market and has recently launched XanEdu, which repackages the materials that it sells to college libraries as ProQuest and sells them to students as electronic course packs and

61

62

a subscription-based (\$49.90 per year) "elibrary for college students, with targeted content and course-driven pre-selected searches" in such fields as history. For the K–12 and public library markets, Bell & Howell further repackages some of the same resources through BigChalk.com.⁵⁵ Bell & Howell and Thomson, thus, aspire to dominate not only university-based library reference publishing but also textbook publishing and education at all levels. In the new electronic environment, such previously separate enterprises potentially merge together into information "portals" or what XanEdu calls "the ultimate learning destination." Like Ted Nelson from whom they may have borrowed their new corporate moniker, the folks at Bell & Howell dream big, promising that XanEdu will be a "utopia for the mind."

Advertising offers another road to a corporate-owned past. Some believe that the Web will emerge as the primary advertising venue of the future, replacing television and glossy magazines. In that scenario, "free" information would be served up in the same fashion as television offers "free" entertainment. Entrepreneurs and large corporations have launched dozens of Web sites aimed at making money off the provision of historical or educational information and services through advertising or marketing. Some, such as the HistoryChannel.com or Discovery.com, are spin-offs of existing print or cable operations. For example, The HistoryNet.com (billed as "where history lives on the Web") is the online companion to fourteen popular history (mostly military history) magazines, including *Civil War Times*, *Wild West*, and *Aviation History*. In addition to back articles from the magazines, it offers a daily quiz, "This Day in History," recommended Web sites (limited in coverage), online forums (not very active in the fall of 2000), and lists of history-related events and exhibitions—all accompanied by flashing banner ads.

Still other history-related sites are startups created directly for the Web. About.com (formerly the Mining Company), for instance, dubs itself the "Human Internet" and provides human "guides" to more than 700 different subjects, including "Women's History," "Twentieth-Century History," and 10 additional historical subjects. The guides, who generally have an undergraduate history degree, usually offer brief annotated links to Web-based materials, short essays of their own (often with some connection to current events), and online forums. The forums—most of them not especially active—include a homework help feature to which students post queries. (Judging from the answers, I doubt everyone will get an A.)

63

64

Many other Web start-ups have shared About.com's interest in tapping the education "market"—an expansive realm including teachers and students at multiple levels. During the Internet stock fever that raged through most of 1999 and early 2000, education dot-coms sprouted overnight as dreams of IPO (initial public offering) millions danced in the heads of entrepreneurs and venture capitalists. Typical were eCollege, a distance education company that raised \$55 million in an initial public offering in December 1999, and Lightspan, a provider of "curriculum-based educational software and Internet products," including, it promises, lesson plans and source documents in history and other fields.⁵⁶ Lightspan went public at \$11.625 per share in mid-February of 2000 and the stock more than doubled less than a month later.

So far the reality of the sponsored history and education sites has not matched the glittering promises, whether of immense profits or of illuminating content. Generally speaking, the nonprofit sites offer considerably better content. For example, 774 popular history articles available at The HistoryNet.com pale beside the thousands of scholarly articles offered at JSTOR. The richest materials at About.com are those from such sites as *American Memory* and the *New Deal Network*, which are presented framed beneath About.com's banner ads. H-Net and *History Matters* provide considerably more active discussion forums than does The HistoryNet or About.com. The History Channel's list of best history Web sites lists the site of the Eighteenth Louisiana Infantry Regiment but not *Valley of the Shadow* or the Library of Congress's collection of Civil War photographs—presumably because you must sign a partnership agreement with the History Channel and post its banner ad to get listed. One must view skeptically The HistoryNet's claims that it is "the Internet's largest and most content-rich history site" or About.com's boast that "our Guides know their subjects as well as anyone."⁵⁷

Stock prices have been even more inflated than content claims, as the spring 2000 NASDAQ (National Association of Securities Dealers Automated Quotations) crash brutally revealed. About.com lost almost three-quarters of its stock value between March and April 2000; eCollege stock plunged 85 percent, and Lightspan plummeted to just above one dollar a share. "There are a lot more companies in the e-learning space than the education industry needs," acknowledged eCollege's chief executive officer, Oakleigh Thorne. Companies with real rather than virtual sources of revenue also began to wonder whether there really was a pot of gold at the end of the Internet rainbow. In November 2000, the privately held Discovery Communications dropped plans to spin off its Web unit and also dropped most of its Web workers—laying off 40 percent of the regular staff and 150 contract workers. "We cannot achieve near-term profitability from the Internet as a stand-alone business," explained the company president, Michela English. Part of the problem was that none of these sites was ever profitable; they simply lived off venture capital, IPO money, or the largess of wealthy corporate parents. Equally problematic was the drop in Internet advertising rates that accompanied the dive in Internet stocks and the realization by advertisers that few Web surfers (about 0.4 percent) were clicking on banner ads.⁵⁸ The fall in rates was part of a vicious cycle in which dropping stock prices soured advertisers on the Internet and then caused problems for start-ups, which—in a kind of Ponzi scheme—had artificially raised rates in the first place with their own advertising.

The collapse of dot-com stock prices and Internet advertising rates suggests that the future of commercially sponsored history on the Web may not be as rosy as some once believed. The history business has had its share of successes in the "real" world—from *American Heritage* magazine to the History Channel, from the History Book Club to heritage tourism—but it has never been a major American industry.⁵⁹ The past remains a realm in which nonprofits, volunteers, and enthusiasts dominate.

Still, as Susan Smulyan reminds us in her history of the commercialization of American broadcasting, broadcasters and advertisers, as well as listeners, viewed the viability of radio advertising with considerable skepticism. Some day Web advertising may be as "natural" and profitable as television commercials. The drop in Internet advertising *rates*, moreover, has not halted the continuing rise in the overall volume of Internet advertising.⁶⁰ And the bursting of the dot-com stock bubble has not slowed the growth in Internet use or even the increasing importance of the Web as a commercial venue. Whether or not history will turn out to go better with Coke (ads), the selling of digital information (probably largely to libraries rather than individuals) will grow in importance and will be increasingly dominated by a small number of giant corporations. Indeed, we may get a combination of fee-based and advertiser-supported systems. Reed Elsevier's Lexis-Nexis Academic Universe charges substantial subscription fees to libraries but still includes flashing banner ads. (A researcher who is "feeling lucky" can, for example, click a banner and put down some money—perhaps his or her latest research grant—on CybersportsCasino.com's blackjack table.)

To raise an alarm about the capitalist character of the information and publishing business makes little sense since publishing has always been a business. But it has not traditionally been dominated by a few giant corporations. In the fall of 2000 when Reed Elsevier and Thomson jointly purchased the publisher Harcourt (where Ted Nelson thought up the term Xanadu four decades ago) for \$4.4 billion in cash and the assumption of \$1.2 billion of debt, the *New York Times* observed that the price was below what had been expected. "The main reason for the low price," it explained, "is that consolidation in the educational and professional publishing businesses—Harcourt's core—has progressed so far that there are almost no bidders left. Each of Harcourt's main businesses is dominated by just three or four companies, like McGraw-Hill or Pearson. Almost all potential bidders faced antitrust problems or had balance sheets full from recent acquisitions."⁶¹ In a world in which libraries can only buy from one or two vendors, those vendors can easily dictate prices and content. And in a world in which there are only a few publishers, they can also dictate terms to authors as well.

The advertiser-sponsored online world also seems to be heading down the same path of media consolidation augured by the merger of AOL with Time-Warner, Inc. Consider, for example, the history of *Civil War Times* magazine, whose humble origins go back to the 1940s when LeRoy Smith used his army poker winnings to start some history tourism businesses in Gettysburg, Pennsylvania. In 1962, during the Civil War centennial, he and the newspaperman Robert H. Fowler started *Civil War Times*; later they gradually added some other related history publications to what they called Historical Times, Inc. In 1986, Cowles Media purchased Historical Times, Inc., and added still more history magazines, which became part of "Cowles Enthusiast Media" and the basis of The HistoryNet.com, which appeared on the Web in 1996. Two years later, the McClatchy newspaper chain acquired Cowles and then sold off Cowles Enthusiast Media to Primedia—formerly known as K-III Communications, a conglomerate of specialty magazines (for example, *National Hog Farmer* and *Lowrider Bicycle*) pulled together by the leveraged buyout specialists Kohlberg Kravis Roberts back in the go-go 1980s. In fall 2000, Primedia announced plans to purchase About.com for more than half a billion dollars—thereby not only consolidating old media (magazines) and new (Web) but also bringing together under one corporate umbrella two of the main advertiser-sponsored history sites on the Web. A few months later, it purchased half ownership of Brill Media Holdings, the company behind Contentville.com.⁶² 71

Ironically, despite the trend toward online consolidation, one of the greatest frustrations of the historical Xanadu as it exists at the dawn of the new millennium is its myriad divisions. To find what the Internet offers on Eugene V. Debs requires at least a dozen different searches—through a general search engine such as Google; the scholarly article archives at JSTOR, ProQuest, EAA, EBSCO, the History Cooperative, and Project Muse; reference works at the History Resource Center; the popular history writings at The HistoryNet.com; articles and sources at Contentville; the primary sources at *American Memory*; and the image archive at Corbis.com. The capitalist market in information and the limitations of Web search engines have fostered both consolidation and competition. Neither trend is wholly friendly to researchers. 72

Perhaps paradoxically, then, the Web seems to be fostering two contradictory developments. On the one hand, the resources required to publish on the Web are so modest that we have seen an amazing grass-roots publishing effort over the past five years. Yet, on the other hand, the capacity to mount a serious Web-based publishing or information business may be quite limited indeed. Even the Web start-ups such as Questia and NetLibrary are backed by hundreds of millions of dollars in venture capital. To be sure, the nonprofit world also has its giants such as NDLP, but their continuation rests on the shaky base of public sector funding. And Internet-based economies of scale are pushing growing consolidation on a global basis. Will the public history Web survive the onslaught of these mega operations? Will "authority" and "authenticity" reside with the corporate purveyors of the past? And will corporate vendors find scholarly fastidiousness about accuracy and contextualization as appealing as archivists and academics do?

Bell & Howell president James P. Roemer presents his company—notes *Forbes* magazine—"as the guardian of truth in an Internet free-for-all." "There's no guarantee that what you're getting on the Internet is correct or the information you want," he says. The company spokesman Ben Mondloch puts the significance of its Digital Vault Initiative in even broader terms. "We're the only company that could do this," he told a reporter for *Wired News*. "We've become the de facto nation's archive."⁶³

The notion of a privatized and corporatized "national archive" occupies the other end of the continuum from the free and open Xanadu envisioned by Ted Nelson. For a humorous and harrowing glimpse of what that might look like, turn to Neal Stephenson's 1992 cyberpunk novel, *Snow Crash*, in which everything is privately owned, from the FOQNEs (Franchise-Organized Quasi-National Entities) known as Burbclaves, where people live, to the highways run by the competing Fairlanes Inc. and Cruiseways Inc., to the Reverend Wayne's Pearly Gates, which has a monopoly on worship services. The book's protagonist, Hiro Protagonist, is a freelance stringer for the CIC, the Central Intelligence Corporation of Langley, Virginia. The CIC's "database" was, Stephenson writes,

formerly the Library of Congress, but no one calls it that anymore. Most people are not entirely clear on what the word "congress" means. And even the word "library" is getting hazy. It used to be a place full of books, mostly old ones. Then they began to include videotapes, records, and magazines. Then all of the information got converted into machine-readable form, which is to say, ones and zeroes. And as the number of media grew, the material became more up to date, and the methods for searching the Library became more and more sophisticated, it approached the point where there was no substantive difference between the Library of Congress and the Central Intelligence Agency. Fortuitously, this happened just as the government was falling apart anyway. So they merged and kicked out a big fat stock offering.⁶⁴

It is all too easy in the era of cyberspace to get carried away with extravagant visions of the future—whether the utopian dreams of Ted Nelson or the dystopian vision of *Snow Crash*. History tells us that change comes much more slowly and unevenly than most visionaries would like. Still, what is remarkable is how much the practice of researching, teaching, and presenting the past has changed in the short five years since the Web and Internet entered the lives of historians. We have many reasons to celebrate the enormous advances—the vast archive of primary and secondary sources now accessible on our computer screens and available to us as researchers, to our students, and to anyone concerned about the past. But while we celebrate what has been gained, we should be vigilant about what might be lost if the grass-roots energy and the cooperative spirit of enthusiastic amateurs, enterprising librarians, and archivists pursuing personal historical passions and public understanding of the past are squashed by the advance of a corporate juggernaut chasing private profit.

Nevertheless, the power and wealth of the corporate forces should not lead us to assume that we are headed inevitably toward Stephenson's CIC. William Y. Arms, the editor of *D-Lib Magazine*, which focuses on digital libraries, has recently argued that "open access" may, in the end, turn out to dominate the future of information. He observes that whereas ten years ago the percentage of information used in professional work that "was available openly, without payment" was probably 1 percent or less, today most people would say that 5 to 80 percent is available with open access. I can often find historical information more quickly on the public Web (and am thus more likely to use it) than by searching the gated private Web databases that my university provides to me. My library, for example, pays a thousand dollars a year to get the online version of *Books in Print* from the Thomson Corporation, but Amazon.com provides much of the same information for free. Increased computer power, moreover, means that it is increasingly easy to find that information on the vast stretches of the Internet. For Arms, "automated digital libraries combined with open access information on the Internet offer to provide the Model T Ford of information," basic transportation for all.⁶⁵

Historians have a great stake in shaping the roads and cars that will populate the future information superhighways. We need to put our energies into maintaining and enlarging the astonishingly rich public historical Web that has emerged in the past five years. For some, that should mean joining in eclectic but widespread grass-roots efforts to put the past online—whether that involves posting a few documents online for your students or raising funds for more ambitious projects to create free public archives. Just as "open source" code has been the banner of academic computer scientists, "open sources" should be the slogan of academic and popular historians. Academics and enthusiasts created the Web; we should not quickly or quietly cede it to giant corporations. For all of us, shaping the digital future requires a range of political actions—fighting against efforts to slash the budgets of public agencies such as NEH and the Library of Congress that are funding important digital projects; resisting efforts further to narrow the "public domain"; and joining with librarians who have been often alone in raising red flags about the growing power of the information conglomerates.⁶⁶ We may also need to reexamine our own contradictory position as both rights holders and consumers of copyright content. Perhaps we should even insist that the intellectual property we create (often with considerable public funding) should be freely available to all. Unless we act, the digital Xanadu, as Nelson fantasized, may turn out to have everything an "absent-minded professor could want" but only at and for a heavy price.⁶⁷

Notes

Roy Rosenzweig is College of Arts and Sciences Distinguished Professor of History and director of the Center for History and New Media at George Mason University.

Thanks to Steve Brier, Josh Brown, Mary Jane Gormley, Deborah Kaplan, Gary Kornblith, Joanne Meyerowitz, Michael O'Malley, Kelly Schrum, John Summers, Tom Thurston, and members of the *JAH* editorial staff for helpful comments on this article.

Readers may contact Rosenzweig at <rrosenzw@gmu.edu>.

¹ T. H. Nelson, "A File Structure for the Complex, the Changing, and the Indeterminate," *Proceedings of the 20th acm National Conference* (1965), 84–100. Nelson's ideas about hypertext were heavily influenced by Vannevar Bush, "As We May Think" (1945); for a reprint of the article and discussions of its influence, see James M. Nyce and Paul Kahn, ed., *From Memex to Hypertext: Vannevar Bush and the Mind's Machine* (Boston, 1991). Even earlier, in 1938, H. G. Wells talked of creating a "World Encyclopedia" with a true "planetary memory for all mankind": quoted in Michael Lesk, "How Much Information Is There in the World?," unpublished paper, 1997 <<http://www.lesk.com/mlesk/ksg97/ksg.html>>. (Unless otherwise noted, the Web references in this article were rechecked online on May 5, 2001.)

² Theodor Holm Nelson, "Xanalogical Structure, Needed Now More than Ever: Parallel Documents, Deep Links to Content, Deep Versioning, and Deep Re-Use," *ACM Computing Surveys*, 31 (Dec. 1999) <http://www.cs.brown.edu/memex/ACM_HypertextTestbed/papers/60.html>; see also Ted Nelson, "Who I Am: Designer, Generalist, Contrarian Theodor Holm Nelson, 1937–" <<http://www.sfc.keio.ac.jp/~ted/TN/WhoIAm.html>>; and Theodor Holm Nelson, "Opening Hypertext: A Memoir," in *Literacy Online: The Promise (and Peril) of Reading and Writing with Computers*, ed. Myron C. Tuman (Pittsburgh, 1992), 43–57.

³ Gary Wolf, "The Curse of Xanadu," *Wired*, 3 (June 1995) <http://www.wirednews.com/wired/archive/3.06/xanadu_pr.html>; Theodor Holm Nelson, "Errors in 'The Curse of Xanadu,' by Gary Wolf," in Andrew Pam, *Xanadu Australia* <<http://www.xanadu.com.au/ararat>>.

⁴ For a history of the development of the Internet, see John Naughton, *A Brief History of the Future: From Radio Days to Internet Years in a Lifetime* (Woodstock, 2000), 229–63.

⁵ For detailed information on Web search engines, see the materials at *Search Engine Watch* <<http://www.searchenginewatch.com/>>. *Search Engine Watch* and other commentators currently rate Google the best overall Web search tool.

⁶ U.S. Department of Commerce, *The Emerging Digital Economy* (Washington, 1998), quoted in Stephen Segaller, *Nerds 2.01: A Brief History of the Internet* (New York, 1998), 14. "Sizing Up the Web," *New York Times*, Dec. 11, 2000, p. C4. All *New York Times* articles cited here are available online (generally for a per-article fee of \$2.50) at *The New York Times on the Web* <<http://www.nytimes.com>> and (for a library subscription

fee) through *Lexis-Nexis Academic Universe* <<http://web.lexis-nexis.com/universe>>; where a page number is cited, the article was first consulted in the print version of the *Times*; where a specific url (uniform resource locator) is cited, the article is available online for free. Office of Research, oclc (Online Computer Library Center, Inc.), "Web Statistics," in *Web Characterization Project* <<http://wcp.oclc.org/>>. Google <<http://www.google.com>>. Peter Lyman and Hal R. Varian, "How Much Information?," *Journal of Electronic Publishing*, 6 (Dec. 2000) <<http://www.press.umich.edu/jep/06-02/lyman.html>>. BrightPlanet, "The Deep Web: Surfacing Hidden Value," in BrightPlanet.com, *Complete Planet* <<http://www.completeplanet.com/Tutorials/DeepWeb/index.asp>>; Lisa Guernsey, "Mining the 'Deep Web' with Specialized Drills," *New York Times*, Jan. 25, 2001.

⁷ *The Internet Archive* <<http://www.archive.org>> intends "to permanently preserve a record of public material" on the Internet. At the present time, however, use of their archive requires programming skills, and I did not receive a response to the request to use the archive that I submitted in October 2000. For a discussion of the need to archive the Web (and a complaint about lack of response from the *Internet Archive*), see Richard Wiggins, "The Unnoticed Presidential Transition: Whither Whitehouse.gov?," *First Monday*, 6 (Jan. 8, 2001) <http://www.firstmonday.org/issues/issue6_1/wiggins/index.html>. Michael O'Malley and Roy Rosenzweig, "Brave New World or Blind Alley? American History on the World Wide Web," *Journal of American History*, 84 (June 1997), 138.

⁸ See "Collections Currently in Progress," in Library of Congress, *American Memory: Historical Collections for the National Digital Library* <<http://memory.loc.gov/ammem/amfuture.html>>. See, more generally, Committee on an Information Technology Strategy for the Library of Congress of National Research Council, *LC21: A Digital Strategy for the Library of Congress* (Washington, 2000) <<http://books.nap.edu/books/0309071445/html/index.html>>. As of December 2000, the NDLP had 5,772,967 items online, but some *American Memory* materials are available as a result of the Ameritech Program and others as a result of cooperative agreements with other institutions. NDLP Reference Team to Roy Rosenzweig, e-mails, Feb. 15, 2001 (in Rosenzweig's possession).

⁹ Peter R. Henriques, "The Final Struggle between George Washington and the Grim King: Washington's Attitude toward Death and an Afterlife," *Virginia Magazine of History and Biography*, 107 (Winter 1999), 75, 95–96. Henriques discussed his methodology with Rosenzweig on November 6, 2000.

¹⁰ OCLC, "Web Statistics"; Peter B. Hirtle, "Free and Fee: Future Information Discovery and Access *D-Lib Magazine*, 7 (Jan.2001), <<http://www.dlib.org/dlib/january01/01editorial.html>>

¹¹ Kevin M. Guthrie, "Revitalizing Older Published Literature: Preliminary Lessons from the Use of JSTOR," paper presented at the conference "Economics and Usage of Digital Library Collections," Ann Arbor, March 23–24, 2000 <<http://www.jstor.org/about/preliminarylessons.html>>. See also "Editor's Interview: Developing a Digital Preservation Strategy for JSTOR, an interview with Kevin Guthrie," *RLG DigiNews*, 4 (no. 4, 2000) <<http://www.rlg.org/preserv/diginews/diginews4-4.html#feature1>>. John Spargo, "The Influence of Karl Marx on Contemporary Socialism," *American Journal of Sociology*, 16 (July 1910), 21–40. Fred Shapiro's discoveries are discussed in Ethan Bronner, "You Can Look It Up, Hopefully," *New York Times*, Jan. 10, 1999 <<http://www.nytimes.com/library/review/011099language-database-review.html>>.

¹² Barbara Quint, "Gale Group's InfoTrac OneFile Creates Web-Based Periodical Collection for Libraries," *Information Today NewsBreaks*, Oct. 16, 2000 <<http://www.infoday.com/newsbreaks/breaks.htm>>.

¹³ See <<http://dir.yahoo.com/Arts/Humanities/History/>>.

¹⁴ *Choice* quoted in William G. Thomas and Alice E. Carter, *The Civil War on the Web: A Guide to the Very Best Sites* (Wilmington, 2000), xiii; Library of Congress, *American Memory* <<http://memory.loc.gov/>>.

¹⁵ "Facts and Statistics *Family Search*, Church of Jesus Christ of Latter-day Saints <http://www.familysearch.com/Eng/Home/News/frameset_news.asp?PAGE=home_facts.asp>.

¹⁶ *Ibid.* April Leigh Helm and Matthew L. Helm, *Genealogy Online for Dummies* (New York, 1999).

¹⁷ Diane Ravitch, ed., *The American Reader: Words That Moved a Nation* (New York, 1990); Elizabeth Cady Stanton, "The Solitude of Self," in *American Public Address, 1644–1935, University of Arkansas Supplement to Communication 4353*, Bernadette Mink <http://comp.uark.edu/~brmink/stanton.html>; "Niagara Movement Declaration of Principles, 1905" in *American History Class Enhancement Pages*, Thomas Martin <<http://www.sinclair.edu/classenhancements/his101e-tm/civilrt1.htm>>; M. Carey Thomas,

"Higher Education for Women," in *Mrs. Pojer's History Classes' Home Page*, Susan M. Pojer <<http://www.historyteacher.net/USProjects/DBQs2000/APUSH-DBQ-40.htm>>. The last two sites were accessed in October 2000 but were no longer available in May 2001. In the first instance, the material moved to a gated WebCT server.

¹⁸ W. E. Burghart Du Bois, "The Talented Tenth," in *Mr. Kenyada's Neighborhood*, Richard Kenyada <<http://www.kenyada.com/talented.htm>>. Joe Hill, "The Preacher and the Slave," in *History in Song*, Manfred J. Helfert <<http://www.fortunecity.com/tinpan/parton/2/pie.html>>. Dean B. McIntyre, "'Lift Every Voice'—100 Years Old," in *General Board of Discipleship*, United Methodist Church <http://www.gbod.org/worship/default.asp?act=reader&item_id=1786>. Alice Duer Miller, "Evolution," in *poet ch'I*, Kevin Taylor <<http://www.geocities.com/Paris/Bistro/8066/index2.htm>>. Carl Becker, "Everyman His Own Historian," *American Historical Review*, 37 (Jan. 1932), 221–36.

¹⁹ *BoondocksNet.com* <<http://www.BoondocksNet.com>>; Jim Zwick to Rosenzweig, e-mails, Nov. 1, 27, 2000 (in Rosenzweig's possession). Some scholars will face copyright and archival restrictions in placing their research materials online but a surprisingly large percentage of materials that historians use—books, magazines, and newspapers from before 1923 and government documents, for example—are in the public domain.

²⁰ "What Is H-Net?," in *H-Net: Humanities & Social Sciences OnLine*, MATRIX: The Center for Humane Arts, Letters, and Social Sciences OnLine, Michigan State University <<http://www2.h-net.msu.edu/about/>>.

²¹ Thomas and Carter, *Civil War on the Web*, xvi–xix; Golden Ink, *About North Georgia* <<http://ngeorgia.com>>, quoted *ibid.*, xix.

²² Elizabeth Cady Stanton, *Solitude of self: address delivered by Mrs. Stanton before the Committee of the Judiciary of the United States Congress, Monday, January 18, 1892* (Washington, 1915), in Rare and Special Collections Division, Library of Congress, *Votes for Women: Selections from the National American Woman Suffrage Association Collection, 1848–1921* <<http://lcweb2.loc.gov/ammem/naw/nawshom.html>>.

²³ Voice of America, *The Century in Sound: An American's Perspective* <<http://www.voia.gov/century/century.html>>; "Socialist Eugene V. Debs speaks during the presidential campaign of 1904," in *Eyewitness: History through the Eyes of Those Who Lived It*, Ibis Communications, Inc. <<http://>>

www.ibiscom.com/vodebs.htm>; "Eugene V. Debs," in *Pluralism and Unity*, David Bailey, David Halsted, and Michigan State University <<http://www.expo98.msu.edu/sounds/debs.html>>. The voice is correctly identified as that of an actor in Department of History, University at Albany, State University of New York, *U.S. Labor and Industrial History World Wide Web Audio Archive* <<http://www.albany.edu/history/LaborAudio/>>. For discussion of the provenance of the Debs speech, see Roy Rosenzweig and Stephen Brier, *Who Built America? From the Centennial Celebration of 1876 to the Great War of 1914* (cd-rom) (New York, 1993), 352.

²⁴ See, for example, "The Willie Lynch Speech of 1712," in *Shepp's Place*, Will Shepperson <<http://www.eden.rutgers.edu/~wshepp3/lynch.html>>; and Willie Lynch, "How to Control the Black Man for At Least 300 Years," in *KohlBlackTimes.com* <<http://www.kohlblacktimes.com/willie.htm>>. The best online commentary on the Lynch speech is Anne Cleëster Taylor, "The Slave Consultant's Narrative: The Life of an Urban Myth?," in *African Missouri*, Anne Cleëster Taylor <<http://www.umsl.edu/~libweb/blackstudies/narrate.htm>>. See also Mike Adams, "In Search of Willie Lynch," *Baltimore Sun*, Feb. 22, 1998, p. 1 (available online in Lexis-Nexis Academic Universe). Of course, many real documents make points similar to those in the Lynch speech.

²⁵ For a discussion of the inclusiveness of virtual libraries, see James J. O'Donnell, *Avatars of the Word: From Papyrus to Cyberspace* (Cambridge, Mass., 1998), 29–43.

²⁶ Kendra Mayfield, "Library of Congress Goes Digital *Wired News*, Jan. 19, 2001 <<http://www.wired.com/news/print/0,1294,41166,00.html>>. For list of sponsors, see "A Unique Public-Private Partnership Supporting the National Digital Library," in *American Memory*, Library of Congress <<http://memory.loc.gov/ammem/sponsors.html>>. See "Library of Congress/Ameritech National Digital Library Competition," *ibid.* <<http://memory.loc.gov/ammem/award/index.html>>.

²⁷ For an astute discussion of *Valley of the Shadow* <<http://jefferson.village.virginia.edu/vshadow/>>.

²⁸ Franklin and Eleanor Roosevelt Institute and Institute for Learning Technologies, *New Deal Network* <<http://newdeal.feri.org/>>; Center for History and New Media and American Social History Project, *History Matters: The U.S. Survey Course on the Web* <<http://historymatters.gmu.edu>>. *History Matters* also includes annotated lists of history Web sites,

online assignments, interactive exercises on the historian's craft, and teaching forums with leading scholars and teachers.

²⁹ University of North Carolina Libraries, *Documenting the American South* <<http://docsouth.unc.edu/aboutdas.html>>.

³⁰ Special Projects Program in the Information and Intelligent Systems Division of the Directorate for Computer and Information Science Engineering, National Science Foundation, *Digital Libraries Initiative* <<http://www.dli2.nsf.gov/>>.

³¹ Wendy Lougee to Rosenzweig, e-mail, Nov. 3, 2000 (in Rosenzweig's possession); Maria Bonn, project director for MOA, provided helpful information on the project in a phone conversation with Rosenzweig, Nov. 9, 2000.

³² Steven Gelber quoted in Nancy Ross-Flanigan, "The Making of America Michigan Today" (Spring 1998) <<http://www.umich.edu/~newsinfo/MT/98/Spr98/mt15s98.html>>.

³³ "Thoughtful weeding of reformatted material is a necessary element of an overall collection management program in the nation's major research libraries": University of Michigan Digital Library Production Service, "Principles and Considerations for University of Michigan Library Subject Specialists" (Feb. 2000) <<http://www.umdl.umich.edu/policies/digitpolicyfinal.html>>. Nicholson Baker, "Deadline: The Author's Desperate Bid to Save America's Past," *New Yorker*, July 24, 2000, pp. 42–61. See also Nicholson Baker, *Double Fold: Libraries and the Assault on Paper* (New York, 2001).

³⁴ Association of Research Libraries, "Talking Points in Response to Nicholson Baker's Article in the 24 July *New Yorker*" <<http://www.arl.org/scomm/baker.html>>. See also Barbara Quint, "Don't Burn Books! Burn Librarians!! A Review of Nicholson Baker's *Double Fold: Libraries and the Assault on Paper*," *Searcher* 9.6 (June 2001) <<http://www.infotoday.com/searcher/jun01/voice.htm>>. Thanks to Josh Brown for his help with this issue. Searching by the word is only possible where the text has been converted into codes that the computer understands as letters and words. The term "digitizing" can refer confusingly both to scanning an image of a page of text and to converting those images of letters into codes that the computer can understand as letters. It is relatively easy to scan thousands of pages of text as images; it is much harder to get that into machine-readable form. That

requires either retyping or an OCR (optical character recognition) system. MOA uses an automated OCR system, which gives very good but not perfect results.

³⁵ Gelber quoted in Ross-Flanigan, "Making of America." Association of Research Libraries, "Summary of Fiscal Year 1999 Appropriation Request for the National Endowment for the Humanities," in *Association of Research Libraries* <<http://www.arl.org/info/letters/FY1999.html>>; Stanley N. Katz, "Rethinking the Humanities Endowment," *Chronicle of Higher Education*, Jan. 5, 2001, pp. B5–10. All *Chronicle* articles cited here are available online to subscribers at <<http://chronicle.com/weekly/sitesearch.htm>>; where a page number is cited, the article was first consulted in the print version of the *Chronicle*.

³⁶ *LC21*; James O'Donnell quoted in Katie Hafner, "Saving the Nation's Digital Legacy," *New York Times*, July 27, 2000, p. G1. See also Mayfield, "Library of Congress Goes Digital."

³⁷ Daren Fonda, "Copyright's Crusader," *Boston Globe Magazine*, Aug. 29, 1999, quoted in Dennis S. Karjala, *Opposing Copyright Extension* <<http://www.public.asu.edu/~dkarjala/commentary/Fonda8-29-99.html>>. See, for example, *NCC Washington Update*, March 27, 1998 <<http://www2.h-net.msu.edu/~ncc/ncc98/ncc9811mar27.html>>. Rosenzweig and Brier, *Who Built America? From the Centennial of 1876 to the Great War of 1914* (cd-rom); Roy Rosenzweig et al., *Who Built America? From the Great War of 1914 to the Dawn of the Atomic Age in 1946* (CD-ROM) (New York, 2000).

³⁸ Kathy Perry, director of VIVA, provided information to Rosenzweig in several conversations during December 2000 and January 2001.

³⁹ *Contentville* <<http://www.contentville.com/>>.

⁴⁰ Corbis and Getty "have been gobbling up smaller agencies around the world": Gordon Black, "Corbis Courts Online Consumers," *Seattle Times*, Nov. 16, 1999, p. D6. See also Kristi Heim, "Digital Image Is Everything as Gates, Getty Vie for Control of 'Net Art,'" *Denver Post*, March 5, 2000, p. I-03 (both available online through Lexis-Nexis Academic Universe). Corbis Corporation, *Corbis—The Place for Pictures Online* <<http://www.corbis.com>>.

⁴¹ EBSCO's full-text holdings in history do not appear to be as deep as those from ProQuest and EAA. For example, EBSCO does not offer such

standards as *Journal of Women's History*, *Journal of Negro History*, and *Journal of Southern History*, which are in EAA.

⁴² On the electronic book ventures, see Goldie Blumenstyk, "Digital-Library Company Plans to Charge Students a Monthly Fee for Access," *Chronicle of Higher Education*, Nov. 14, 2000; Andrew R. Albanese, "E-Book Gold Rush: Welcome to the Electronic Backlist," *Lingua Franca*, 10 (Sept. 2000) <<http://www.linguafranca.com/print/0009/inside-ebook.html>>; Jennifer Darwin, "Storybook Beginning: Questia Founder Follows Novel Script to Launch Online College Library," *Houston Business Journal*, April 7, 2000 <<http://www.bizjournals.com/houston/stories/2000/04/10/story2.html>>; Lisa Guernsey, "The Library as the Latest Web Venture," *New York Times*, June 15, 2000; LC21, box 1.3; Tom Fowler, "\$90 Million in Funding for Questia," *Houston Chronicle*, Aug. 24, 2000, business p. 1 (available online in Lexis-Nexis Academic Universe); and Kendra Mayfield, "The Quest for E-Knowledge," *Wired News*, Feb. 5, 2001 <<http://www.wired.com/news/print/0,1294,41543,00.html>>. For survey, see David Thelen, "The Practice of American History," *Journal of American History*, 81 (Dec. 1994), 953. History is not particularly well represented in the NetLibrary collection so far. Some other "e-book" vendors concentrate on particular fields, for example, information technology (ITKnowledge) and marketing and finance (Books24x7).

⁴³ See HarpWeek, "Purchase Information," in *HarpWeek* <<http://www.harpweek.com/04Products/products-purchase.htm>>. HarpWeek may also begin levying annual maintenance fees in 2002.

⁴⁴ Robert Thibadeau to Rosenzweig, e-mails, Nov. 1, 2, 2000 (in Rosenzweig's possession); *The Historical New York Times Project* <<http://nyt.ulib.org/>>. For unreadable pages, see, for example, Aug. 6, 1860, and Aug. 6, 1863.

⁴⁵ Michael Jensen, "Mission Possible: Giving It Away While Making It Pay," paper presented at the annual meeting of the Association of American University Presses, Austin, Tex., June 22, 1999 <<http://www.nap.edu/staff/mjensen/aaup99.html>> (emphasis in original).

⁴⁶ On the History Cooperative, see Michael Grossberg, "Devising an Online Future for Journals of History," *Chronicle of Higher Education*, April 21, 2000. *William and Mary Quarterly*, *Western Historical Quarterly*, *History Teacher*, and *Law and History Review* will soon join the *Journal of American History* and the *American Historical Review* in the History Cooperative. (Full disclosure: I was a member of the *Journal of American*

Published by the
Organization of
American Historians

Presented online
in association with
the History Cooperative

© 2001
The Organization of
American Historians

History committee that developed the cooperative project.)

⁴⁷ For an experiment in hypertext publishing, see the articles in Roy Rosenzweig, ed., "Hypertext Text Scholarship and American Studies" <<http://chnm.gmu.edu/aq>>; and Roy Rosenzweig, ed., "Forum on Hypertext Scholarship: aq as Web-Zine: Responses to aq's Experimental Online Issue," *American Quarterly*, 51 (June 1999), 237–82 (available online to subscribers at Project Muse <<http://muse.jhu.edu/>>). See also Roy Rosenzweig, "The Riches of Hypertext for Scholarly Journals," *Chronicle of Higher Education*, March 17, 2000.

⁴⁸ "arXiv Monthly Submission Rate Statistics," <http://arXiv.org/show_monthly_submissions>; Stevan Harnad, "The Future of Scholarly Skywriting," in "i in the Sky: Visions of the Information Future," ed. A. Scammell, Aslib, Nov. 1999 <<http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad99.aslib.html>>. See also Vincent Kiernan, "Open Archives' Project Promises Alternative to Costly Journals," *Chronicle of Higher Education*, Dec. 3, 1999; Herbert Van de Sompel and Carl Lagoze, "The Santa Fe Convention of the Open Archives Initiative," *D-Lib Magazine*, 6 (Feb. 2000) <<http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>>; Stevean Harnad, "Free at Last: The Future of Peer-Reviewed Journals," *D-Lib Magazine*, 5 (Dec. 1999) <<http://www.dlib.org/dlib/december99/12harnad.html>>.

⁴⁹ Jensen, "Mission Possible."

⁵⁰ David D. Kirkpatrick, "Librarians Unite against Cost of Journals," *New York Times*, Dec. 25, 2000, p. C5. Data on library budgets provided by Mary Case of the Association of Research Libraries and published in *ARL Statistics, 1998–99* (Washington, 2000); *ARL Supplementary Statistics, 1998–99* (Washington, 2000). On the crisis in scholarly publishing, see, for example, Sanford G. Thatcher, "Thinking Systematically about the Crisis in Scholarly Communication" and other papers presented at the conference "The Specialized Scholarly Monograph in Crisis; or, How Can I Get Tenure If You Won't Publish My Book?," Washington, Sept. 11–12, 1997 <<http://www.arl.org/scomm/epub/papers/>>; and Roy Rosenzweig, "How Can I Get Tenure If You Won't Publish My Book?," *Organization of American Historians Newsletter*, 29 (Nov. 1997), 5.

⁵¹ Christopher Stern, "Freelancers Get Day in Court," *Washington Post*, Nov. 7, 2000, p. E3. David D. Kirkpatrick, "Publisher Set to Split E-Book Revenue," *New York Times*, Nov. 7, 2000, p. C2.

⁵² The National Coalition of Independent Scholars (NCIS) successfully lobbied the Modern Languages Association to pass two resolutions on access for independent scholars at their December 2000 annual meeting in Washington, D.C. See Margaret Delacy, "A History of NCIS" <<http://www.ncis.org/history.htm>>.

⁵³ On network effects and economies of scale, see Philip E. Agre, "The Market Logic of Information," paper presented at Interface 5, Sept. 2000; Carl Shapiro and Hal Varian, *Information Rules: A Strategic Guide to the Network Economy* (Boston, 1998); and Philip E. Agre, "Notes and Recommendations," *Red Rock Eater Digest*, March 3, 1998 <<http://commons.somewhere.com/rre/1998/notes.and.recommendation2.html>>.

⁵⁴ "State Has Eight Firms on Forbes' List of Biggest 500 Private," Associated Press State & Local Wire, Nov. 16, 2000 (available in Lexis-Nexis Academic Universe); "EBSCO Publishing Corporate Quick Facts," in *EBSCO Publishing Homepage* <<http://www.epnet.com/bground2.html>>. UMI is considering plans to turn the page images into searchable text, potentially a massive project. Paula J. Hane, "UMI Announces Digital Vault Initiative," *Information Today, Newsbreak*, July 13, 1998 <<http://www.infotoday.com/newsbreaks/nb0713-3.htm>>. For a report that digital facsimiles will be provided, see "Times Pages to Be Available on Internet," *New York Times*, Jan. 13, 2001 <<http://www.nytimes.com/2001/01/13/technology/13BELL.html>>. I have heard reports that the pages will ultimately be converted to searchable form through a combination of OCR and retyping of headlines and first paragraphs.

⁵⁵ *BigChalk: The Education Network* <<http://www.bigchalk.com>>.

⁵⁶ On the Internet boom, see Hal R. Varian, "Economic Scene," *New York Times*, Feb. 6, 2001, p. C2. *Lightspan.com* <<http://www.lightspan.com/>>. As of January 2001, most of the links to materials in history said: "We're currently gathering the best educational links for this topic. Soon, you'll have access to expert-selected Web sites, encyclopedia articles, learning activities, lesson plans, and more."

⁵⁷ The list of best Web sites was not officially launched when I viewed it on February 6, 2001, but it already contained a long list of Civil War sites. "The History Channel.Com Network," *The History Channel.com* <<http://network.historychannel.com/index.asp?page=home>>. Cowles History Group, Inc., "The HistoryNet: Advertiser Information," in *The HistoryNet* <<http://www.historychannel.com>>.

thehistorynet.com/forms/adinfo.htm>; "About Us: Our Story," in About.com, *About—The Human Internet* <http://ourstory.about.com/index.htm?PM=59_1100_T>.

⁵⁸ Oakleigh Thorne quoted in Sarah Carr and Goldie Blumenstyk, "The Bubble Bursts for Education Dot-Coms," *Chronicle of Higher Education*, June 30, 2000, pp. A39–40. "Discovery.Com Workers Get Pink Slips," *Washington Post*, Nov. 14, 2000, p. C7. "Online Advertising Rate Card Prices and Ad Dimensions," Aug. 14, 2000, in *AdRelevance*, Jupiter Media Metrix <http://www.adrelevance.com/intelligence/intel_archive.jsp>; Paul F. Nunes, "Wake-up Call for Internet Firms Overly Dependent on Ad Revenues," *BusinessWorld* (Philippines), June 6, 2000 (available in Lexis-Nexis Academic Universe).

⁵⁹ See, for example, Roy Rosenzweig, "Marketing the Past: American Heritage and Popular History in the United States," in *Presenting the Past: Essays on History and the Public*, ed. Susan Porter Benson, Stephen Brier, and Roy Rosenzweig (Philadelphia, 1986), 21–49.

⁶⁰ Susan Smulyan, *Selling Radio: The Commercialization of American Broadcasting, 1920–1934* (Washington, 1994); Stuart Elliott, "Banners' Ineffectiveness Stalls an Up-and-Coming Rival to TV," *New York Times*, Dec. 11, 2000, p. C4; "Dot Coms in the Driver's Seat," Sept. 5, 2000, in *AdRelevance* <http://www.adrelevance.com/intelligence/intel_report_000905.pdf>; "The Failure of New Media," *Economist*, Aug. 19, 2000.

⁶¹ David D. Kirkpatrick, "Media Giants in Joint Deal for Harcourt," *New York Times*, Oct. 28, 2000, p. C1. See also Richard Poynder, "The Debate Heats Up—Are Reed Elsevier and Thomson Corp. Monopolists?," *Information Today Newsbreaks* (30 April 2001) <<http://www.infotoday.com/newsbreaks/nb010430-1.htm>>.

⁶² Brett D. Fromson, "On the Level: Is This a Stock 'Primed' for an Uptick?," *The Street.com*, Dec. 5, 2000 <http://www.thestreet.com/_yahoo/markets/onthelevel/1199748.html>. (The merger was completed March 1, 2001.) "Primedia's Loss Exceeds Expectations, Taking Hit from New-Media Businesses," *WSJ.Com*, Feb. 2, 2001 <<http://public.wsj.com/sn/y/SB981035131440666351.html>>, accessed online Feb. 17, 2001, but not accessible on May 5, 2001.

⁶³ Victoria Murphy, "Unlocking the Vault," *Forbes Magazine*, Nov. 13,

2000 <<http://www.forbes.com/forbes/2000/1113/6613228a.html>> (Forbes now requires that you register to access its articles); Steve Silberman, "Putting History Online," *Wired News*, June 26, 1998 <<http://www.wired.com/news/culture/0,1284,13298,00.html>>. See also Peter Jacso, "With Experience and Content, UMI Is Poised for Conversion Megaproject," *Information Today*, Sept. 8, 1998 <<http://www.infotoday.com/it/sep98/jacso.htm>> and the enhanced version <<http://www.umi.com/hp/News/Reviews/SiteBuilder.html>>; "Bell & Howell's ProQuest Digital Vault Initiative Leaps Forward This Spring," press release, March 22, 2000, in *Bell & Howell's ProQuest Information* <<http://www.proquest.com/division/pr/00/20000322.shtml>>.

⁶⁴ Neal Stephenson, *Snow Crash* (New York, 1992), 22.

⁶⁵ Florence Olsen, "'Open Access' is the Wave of the Information Future, Scholar Says," *Chronicle of Higher Education*, Aug. 18, 2000; William Y. Arms, "Automated Digital Libraries: How Effectively Can Computers Be Used for the Skilled Tasks of Professional Librarianship?," *D-Lib Magazine*, 6 (July-August 2000) <<http://www.dlib.org/dlib/july00/arms/07arms.html>>.

⁶⁶ For a recent effort by librarians and scientists to fight back against the rapacious prices of commercially owned science journals, see Scholarly Publishing & Academic Resources Coalition and Triangle Research Libraries Network, *Declaring Independence: A Guide to Creating Community-Controlled Science Journals* (Washington, 2001) <<http://www.arl.org/sparc/DI/>>.

⁶⁷ Nelson, "A File Structure for the Complex."

Content in the History Cooperative database is intended for personal, noncommercial use only. You may not reproduce, publish, distribute, transmit, participate in the transfer or sale of, modify, create derivative works from, display, or in any way exploit the History Cooperative database in whole or in part without the written permission of the copyright holder.

[September 2001](#)

[Previous](#)

[Table of contents](#)

[Next](#)