CENTER for HISTORY and NEW MEDIA

- **chnm home**
- **projects**
- **tools**
- **resources**
- **news**

print this article

Resources

- **History Departments Around the World**
  - **Essays on History & New Media**
    - **Syllabus Finder**
      - **H-Bot**
    - **Tools Center**
      - **Services**

Essays

- **Designing for the Web**
  - **History on the Web**
- **Teaching Digital History**
  - **CD-ROMs**
- **History of the Internet**

Scarcity or Abundance? Preserving the Past in a Digital Era

Roy Rosenzweig

ON OCTOBER 11, 2001, THE SATIRIC Bert Is Evil web site, which displayed photographs of the furry Muppet in Zelig-like proximity to villains such as Adolf Hitler (see Figure 1), disappeared from the web—a bit of collateral damage from the September 11th attacks. Following the strange career of Bert Is Evil shows us possible futures of the past in a digital era—futures that historians need to contemplate more carefully than they have done so far.

In 1996, Dino Ignacio, a twenty-two-year-old Filipino web designer, created Bert Is Evil ("brought to you by the letter H and the CIA"), which became a cult favorite among early tourists on the World Wide Web. Two years later, Bert Is Evil won a "Webby" as the "best weird site." Fan and "mirror" sites appeared with some embellishing on the "Bert Is Evil" theme. After the bombing of the U.S. embassies in Kenya and Tanzania in 1998, sites in the Netherlands and Canada paired Bert with Osama bin Laden. [1]

This image made a further global leap after September 11. When Mostafa Kamal, the production manager of a print shop in Dhaka, Bangladesh, needed some images of bin Laden for anti-American posters, he apparently entered the phrase "Osama bin Laden" in Google's image search engine. The Osama and Bert duo was among the top hits. "Sesame Street" being less popular in Bangladesh than in the Philippines, Kamal thought the picture a nice addition to an Osama collage. But when this transnational circuit of imagery made its way back to more Sesame Street–friendly parts of the world via a Reuters photo of anti-American demonstrators (see Figure 2), a storm of indignation erupted. Children's Television Workshop, the show's producers, threatened legal action. On October 11, 2001, a nervous Ignacio pushed the delete key, imploring "all fans [sic] and mirror site hosts of 'Bert is Evil' to stop the spread of this site too."[2]

Ignacio's sudden deletion of Bert should capture our interest as historians since it dramatically illustrates the fragility of evidence in the digital era. If Ignacio had published his satire in a book or magazine, it would sit on thousands of library shelves rather than having a more fugitive existence as magnetic impulses on a web server. Although some historians might object that the Bert Is Evil web site is of little historical significance, even traditional historians should worry about what the digital era might mean for the historical record. U.S. government records, for example, are being lost on a daily basis. Although most government agencies started using e-mail and word processing in the mid-1980s, the National Archives still does not require that digital records be retained in that form, and governmental employees profess confusion over whether they should be preserving electronic files.[3] Future historians may be unable to ascertain not only whether Bert is evil, but also which

undersecretaries of defense were evil, or at least favored the concepts of the "evil empire" or the "axis of evil." Not only are ephemera like "Bert" and government records made vulnerable by digitization, but so are traditional works—books, journals, and film—that are increasingly being born digitally. As yet, no one has figured out how to ensure that the digital present will be available to the future's historians.

But, as we shall see, tentative efforts are afoot to preserve our digital cultural heritage. If they succeed, historians will face a second, profound challenge—what would it be like to write history when faced by an essentially complete historical record? In fact, the Bert Is Evil story could be used to tell a very different tale about the promiscuity and even persistence of digital materials. After all, despite Ignacio's pleas and Children's Television Workshop's threats, a number of Bert "mirror" sites persist. Even more remarkably, the Internet Archive—a private organization that began archiving the web in 1996—has copies of Bert Is Evil going back to March 30, 1997. To be sure, this extraordinary archive is considerably more fragile than one would like. The continued existence of the Internet Archive rests largely on the interest and energy of a single individual, and its collecting of copyrighted material is on even shakier legal ground. It has put the future of the past—traditionally seen as a public patrimony—in private hands.

Still, the astonishingly rapid accumulation of digital data—obvious to anyone who uses the Google search engine and gets 300,000 hits—should make us consider that future historians may face information overload. Digital information is mounting at a particularly daunting rate in science and government. Digital sky surveys, for example, access over 2 billion images. Even a dozen years ago, NASA already had 1.2 million magnetic tapes (many of them poorly maintained and documented) with space data. Similarly, the Clinton White House, by one estimate, churned out 6 million e-mail messages per year. And NARA is contemplating archiving military intelligence records that include more than "1 billion electronic messages, reports, cables, and memorandums."[4]

Thus historians need to be thinking simultaneously about how to research, write, and teach in a world of unheard-of historical abundance and how to avoid a future of record scarcity. Although these prospects have occasioned enormous commentary among librarians, archivists, and computer scientists, historians have almost entirely ignored them. In part, our detachment stems from the assumption that these are "technical" problems, which are outside the purview of scholars in the humanities and social sciences. Yet the more important and difficult issues about digital preservation are social, cultural, economic, political, and legal—issues that humanists should excel at. The "system" for preserving the past that has evolved over centuries is in crisis, and historians need to take hand in building a new system for the coming century. Historians also tend to assume a professional division of responsibility, leaving these matters to archivists. But the split of archivists from historians is a relatively recent one. In the early twentieth century, historians saw themselves as having a responsibility for preserving as well as researching the past. At that time, the vision and membership of the American Historical Association—embracing archivists, local historians, and "amateurs" as well as university scholars—was considerably broader than it later became.[5]

Ironically, the disruption to historical practice (to what Thomas Kuhn called "normal science") brought by digital technology may lead us "back to the future." The struggle to incorporate the possibilities of new technology into the ancient practice of history has led, most importantly, to questioning the basic goals and methods of our craft. For example, the Internet has dramatically expanded and, hence, blurred our audiences. A scholarly journal like this one is suddenly much more accessible to high school students and history enthusiasts. And the work of history buffs is similarly more visible and accessible to scholars. We are forced, as a result, to rethink who our audiences really are. Similarly, the

capaciousness of digital media means that the page limits of journals like this one are no longer fixed by paper and ink costs. As a result, we are led to question the nature and purpose of the scholarly journals—why do they publish articles with particular lengths and structures? Why do they publish particular types of articles? The simultaneous fragility and promiscuity of digital data requires yet more rethinking—about whether we should be trying to save everything, who is "responsible" for preserving the past, and how we find and define historical evidence.

Historians, in fact, may be facing a fundamental paradigm shift from a culture of scarcity to a culture of abundance. Not so long ago, we worried about the small numbers of people we could reach, pages of scholarship we could publish, primary sources we could introduce to our students, and documents that had survived from the past. At least potentially, digital technology has removed many of these limits: over the Internet, it costs no more to deliver the AHR to 15 million people than 15,000 people; it costs less for our students to have access to literally millions of primary sources than a handful in a published anthology. And we may be able to both save and quickly search through all of the products of our culture. But will abundance bring better or more thoughtful history?[6]

Historians are not unaware of these challenges to the ways that we work. Yet, paradoxically, these fundamental questions are often relegated to more marginal professional spaces—to casual lunchtime conversations or brief articles in association newsletters. But in this time of rapid and perplexing changes, we need to engage with issues about access to scholarship, the nature of scholarship, the audience for scholarship, the sources for scholarship, and the nature of scholarly training in the central places where we practice our craft—scholarly journals, scholarly meetings, and graduate classrooms. That scholarly engagement should also lead us, I believe, to public action to advocate the preservation of the past as a public responsibility—one that historians share. But I hope to persuade even those who do not share my particular political stance that professional historians need to shift at least some of their attention from the past to the present and future and reclaim the broad professional vision that was more prevalent a century ago. The stakes are too profound for historians to ignore the futures of the past.

ALTHOUGH HISTORIANS HAVE MOSTLY BEEN SILENT, archivists, librarians, public officials, and others have loudly warned about the threatened loss of digital records and publications for at least two decades. Words such as "disaster" and "crisis" echo through their reports and conference proceedings. As early as 1985, the Committee on the Records of Government declared, "the United States is in danger of losing its memory."[7] More than a dozen years later, a project called "Time and Bits: Managing Digital Continuity" brought together archivists, librarians, and computer scientists to address the problem once again. Conferees watched the Terry Sanders film Into the Future: On the Preservation of Knowledge in the Electronic Age, and some likened it to Rachel Carson's Silent Spring and themselves to the environmentalists of the 1960s and 1970s. A Time and Bits web site assembled conference materials and promoted "ongoing digital dialogue." But, as if to prove the conference's point, the site disappeared in less than a year. Computer scientist Jeff Rothenberg may have been over-optimistic when he quipped, "Digital documents last forever—or five years, whichever comes first."[8]

Those worried about a problem like digital preservation that lacks public attention are prone to exaggerate. Probably the greatest distortion has been the implicit suggestion that we have somehow fallen from a golden age of preservation in which everything of importance was saved. But much— really, most—of the record of previous historical eras has disappeared. "The members of prehistoric societies did not think they lived in prehistoric times," Washington Post writer Joel Achenbach observes.

"They merely lacked a good preservation medium." And non-digital records that have survived into this century—from Greek and Chinese antiquities to New Guinean folk traditions to Hollywood films—are also seriously threatened.[9]

Another exaggeration involves stories about the grievous losses that never occurred. One widely repeated story is that computers can no longer read the data tapes from the 1960 U.S. Census. In truth, as Margaret Adams and Thomas Brown from the National Archives have shown, the Census Bureau had by 1979 successfully copied almost all the records to newer "industry-compatible tapes." Yet, even in debunking one of the persistent myths of the digital age, Adams and Brown reveal some of the key problems. In just a decade and a half, migrating the census tapes to a readable format "represented a major engineering challenge"—hardly something we expect to face with historical records originating from within our own lifetimes. And although "only 1,575 records . . . could not be copied because of deterioration," the absolute nature of digital corrosion is sobering.[10] Print books and records decline slowly and unevenly—faded ink or a broken-off corner of a page. But digital records fail completely—a single damaged bit can render an entire document unreadable. Here is the key difference from the paper era: we need to take action now because digital items very quickly become unreadable, or recoverable only at great expense.

This has already happened—albeit not as much as sometimes suggested. "Ten to twenty percent of vital data tapes from the Viking Mars mission," notes Deanna Marcum, the president of the Council on Library Information Resources, "have significant errors because magnetic tape is too susceptible to degradation to serve as an archival storage medium." Often, records lack sufficient information about their organization and coding to make them usable. According to Kenneth Thibodeau, director of the National Archive and Record Administration's Electronic Records Archives program, NARA lacked adequate documentation to make sense of several hundred reels of computer tapes from the Department of Health and Human Resources and data files from the National Commission on Marijuana and Drug Abuse. Some records could be recovered by future digital archaeologists but sometimes only through an unaffordable "major engineering challenge."[11] The greatest concern is not over what has already been lost but what historians in fifty years may find that they can't read.

Many believe—incorrectly—the central problem to be that we are storing information on media with surprisingly short life spans. To be sure, acid-free paper and microfilm last a hundred to five hundred years, whereas digital and magnetic media deteriorate in ten to thirty years. But the medium is far from the weakest link in the digital preservation chain. Well before most digital media degrade, they are likely to become unreadable because of changes in hardware (the disk or tape drives become obsolete) or software (the data are organized in a format destined for an application program that no longer works). The life expectancy of digital media may be as little as ten years, but very few hardware platforms or software programs last that long. Indeed, Microsoft only supports its software for about five years.[12]

The most vexing problems of digital media are the flipside of their greatest virtues. Because digital data are in the simple lingua franca of bits, of ones and zeros, they can be embodied in magnetic impulses that require almost no physical space, be transmitted over long distances, and represent very different objects (for instance, words, pictures, or sounds as well as text). But the ones and zeros lack intrinsic meaning without software and hardware, which constantly change because of technological innovation and competitive market forces. Thus this lingua franca requires translators in every computer application, which, in turn, operate only on specific hardware platforms. Compounding the difficulty is

that the languages being translated keep changing every few years.

The problem is still worse because of the ability of digital media to create and represent complex, dynamic, and interactive objects—another of their great virtues. Even relatively simple documents that appear to have direct print analogs turn out to be more complex. Printing out e-mail messages makes rapid searches of them impossible and often jettisons crucial links to related messages and attachments. In addition, multimedia programs, which generally rely on complicated combinations of hardware and software, quickly become obsolete. Nor is there any good way to preserve interactive and experiential digital creations. That is most obviously true of computer games and digital art, but even a large number of ordinary web pages are generated out of databases, which means that the specific page you view is your own "creation" and the system can create an infinite number of pages. Preserving hypertextually linked web pages poses the further problem that to save a single page in its full complexity could ultimately require you to preserve the entire web, because virtually every web page is linked to every other. And the dynamic nature of databases destabilizes mundane business and governmental records since they are often embedded in systems that automatically replace old data with new—a changeability that, notes archival educator Richard Cox, threatens "the records of any modern day politician, civic leader, businessperson, military officer, or leader."[13]

WHILE THESE TECHNICAL DIFFICULTIES ARE IMMENSE, the social, economic, legal, and organizational problems are worse. Digital documents—precisely because they are in a new medium—have disrupted long-evolved systems of trust and authenticity, ownership, and preservation. Reestablishing those systems or inventing new ones is more difficult than coming up with a long-lived storage mechanism.

How, for example, do we ensure the "authenticity" of preserved digital information and "trust" in the repository? Paper documents and records also face questions about authenticity, and forgeries are hardly unknown in traditional archives. The science of "diplomatics," in fact, emerged in the seventeenth century as a way to authenticate documents when scholars confronted rampant forgeries in medieval documents. But digital information—because it is so easily altered and copied, lacks physical marks of its origins, and, indeed, even the clear notion of an "original"—cannot be authenticated as physical documents and objects can. We have, for example, no way of knowing that forwarded e-mail messages we receive daily have not been altered. In fact, the public archive of Usenet discussion groups contains hundreds of deliberately and falsely attributed messages. "Fakery," write David Bearman and Jennifer Trant, "has not been a major issue for most researchers in the past, both because of the technical barriers to making plausible forgeries, and because of the difficulty with which such fakes entered an authoritative information stream."[14] Digital media, tools, and networks have altered the balance.

"It took centuries for users of print materials to develop the web of trust that now undergirds our current system of publication, dissemination, and preservation," notes Abby Smith, a leading figure in library and preservation circles. Digital documents are disrupting that carefully wrought system by undercutting our expectations of what constitutes a trusted and authentic document and repository. But to make the transition to a new system requires not just technical measures (such as digital signatures and "watermarks") but, as Clifford Lynch, the executive director of the Coalition for Networked Information, observes, also figuring out responsibility for guaranteeing claims of authorship and financing for a system of "authentication and integrity management."[15]

Such questions are particularly hard to answer since digitization also undercuts our sense of who owns

such materials and, thus, who has the right and responsibility to preserve them. Consumers (including libraries) have traditionally purchased books and magazines under the "first sale" doctrine, which gives those who buy something the right to make any use of it, including lending or selling it to others. But most digital goods are licensed rather than sold. Because contract law governs licenses, vendors of digital content can set any restrictions they choose—they can say that the contents may not be copied or cannot be viewed by more than one person at a time. Adobe's eBook reader even includes a warning that a book may not be read aloud.[16]

But if libraries don't own digital content, how can they preserve it? The problem will become even worse if publishers widely adopt copy protection schemes as they are seriously considering doing for electronic books. Even a library that had the legal right to preserve the content would have no reason to assume that it would be able to do so; meanwhile, the publisher would have little incentive to keep the protection system functioning in a new software environment. In general, digital rights management systems and other forms of "trusted computing" undercut preservation efforts by embedding centralized control in proprietary systems. "If Microsoft, or the U.S. government, does not like what you said in a document you wrote," speculates Free Software advocate Richard Stallman, "they could post new instructions telling all computers to refuse to let anyone read that document."[17]

Licensed and centrally controlled digital content not only erodes the ability of libraries to preserve the past, it also undercuts their responsibility. Why should a library worry about the long-term preservation of something it does not own? But then, who will? Publishers have not traditionally assumed preservation responsibility since there is no obvious profit to be made in ensuring that something will be available or readable in a hundred years when it is in the public domain and can't be sold or licensed.[18]

The digital era has not only unsettled questions of ownership and preservation for traditional copyrighted material, it has also introduced a new, vast category of what could be called semi-published works, which lack a clear preservation path. The free content available on the web is protected by copyright even though it has not been formally registered with the Library of Congress Copyright Office or sold by a publisher. That means that a library that decided to save a collection of web pages—say, those posted by abortion rights organizations—would technically be violating copyright.[19] The absence of this "process" is the most fundamental problem facing digital preservation. Over centuries, a complex (and imperfect) system for preserving the past has emerged. Digitization has unsettled that system of responsibility for preservation, and an alternative system has not yet emerged. In the meantime, cultural and historical objects are being permanently lost.

Four different systems generally preserve cultural and historical documents and objects. Research libraries take responsibility for books, magazines, and other published cultural works, including moving images and recorded sound. Government records fall under the jurisdiction of the National Archives and a network of state and local archives.[20] Systems for maintaining other cultural and historical materials are less formal or centralized. "Records" and "papers" from businesses, voluntary associations, and individuals have found their way into local historical societies, specialized archives, and university special collections. Finally, the semi-published body of material we have called "ephemera" has been most often saved by enthusiastic individuals—for example, postcard and comic book collectors—who might later deposit their hoard in a permanent repository.[21]

While research libraries have tried to save relatively complete sets of published works, other historical

sources have generally only been preserved in a highly selective and sometimes capricious fashion—what archivists call "preservation through neglect." Materials that lasted fifty or one hundred years found their way into an archive, library, or museum. Although this inexact system has resulted in many grievous losses to the historical record, it has also given us many rich collections or personal and organizational papers and ephemera.[22]

But this "system" will not work in the digital era because preservation cannot begin twenty-five years after the fact. What might happen, for example, to the records of a writer active in the 1980s who dies in 2003 after a long illness? Her heirs will find a pile of unreadable 5 1/4" floppy disks with copies of letters and poems written in WordStar for the CP/M operating system or one of the more than fifty now-forgotten word-processing programs used in the late 1980s.[23]

Government archives similarly continue to rely on the unwarranted assumption that records can be appraised and accessioned many years after their creation. A recent study, "Current Recordkeeping Practices within the Federal Government," which surveyed more than forty federal agencies, found widespread confusion about "policies and procedures for managing, storing, and disposing of electronic records and systems." "Government employees," it concluded, "do not know how to solve the problem of electronic records—whether the electronic information they create constitutes records and, if so, what to do with the records. Electronic files that qualify as records—particularly in the form of e-mail, and also word processing and spreadsheet documents—are not being kept at all as records in many cases."[24]

This uncertainty and disarray would not be so serious if we could assume that it could be simply sorted out in another thirty years. But if we hope to preserve the present for the future, then the technical problems facing digital preservation as well as the social and political questions about authenticity, ownership, and preservation policy need to be confronted now.

AT LEAST INITIALLY, archivists and librarians tended to assume that a technical change—the rise of digital media—required a technical solution. The simplest technical solution has been to translate digital information into something more familiar and reassuring like paper or microfilm. But, as Rothenberg points out, this is a "rear-guard action" that destroys "unique functionality (such as dynamic interaction, nonlinearity, and integration)" and "core digital attributes (perfect copying, access, distribution, and so forth)" and sacrifices the "original form, which may be of unique historical, contextual, or evidential interest."[25]

Another backward-looking solution is to preserve the original equipment. If you have files created on an Apple II, then why not keep one in case you need it? Well, sooner or later, a disk drive breaks or a chip fails, and unless you have a computer junkyard handy and a talent for computer repair, you are out of luck. "Technological preservation," moreover, requires intervention before it is too late to save not just the files but also the original equipment. The same can be said of what is probably the most widely accepted current method of digital preservation—"data migration," or moving the documents from a medium, format, or computer technology that is becoming obsolete to one that is becoming more common.[26] When the National Archives saved the 1960 U.S. Census tapes, they used migration, and large organizations use this strategy all the time—moving from one accounting system to another. Because we have lots of experience migrating data, we also know that it is time consuming and expensive. One estimate is that data migration is equivalent to photocopying all the books in a library every five years.[27]

Some like Rothenberg also worry, for example, about the loss of functionality in migrating digital files. Moreover, the process can't be automated because "migration requires a unique new solution for each new format or paradigm and each type of document that is to be converted into that new form." Rothenberg is also derisive about the practice of translating documents into standardized formats and then re-translating as new formats emerge, which he finds "analogous to translating Homer into modern English by way of every intervening language that has existed during the past 2,500 years."[28]

Rothenberg's favored alternative is "emulation"—developing a system that works on later generations of hardware and software but mimics the original. In principle, a single emulation solution could preserve a vast store of digital documents. In addition, it holds the greatest promise for preserving interactive and multimedia digital creations. But critics of emulation tellingly note that it is only a theoretical solution. Probably the best strategy is to reject the all-or-nothing, magic-bullet approaches implicit in the proposals of the most passionate advocates of any particular strategy—whether creating hard copies, preserving old equipment, migrating formats, or emulating hardware and software. Margaret Hedstrom, one of the leading figures in digital preservation research, argues persuasively that "the search for the Holy Grail of digital archiving is premature, unrealistic, and possibly counter-productive." Instead, we need to develop "solutions that are appropriate, effective, affordable and acceptable to different classes of digital objects that live in different technological and organizational contexts."[29]

But even the most calibrated mix of technical solutions will not save the past for the future because, as we have seen, the problems are much more than technical and involve difficult social, political, and organizational questions of authenticity, ownership, and responsibility. Multiple experiments and practices are under way—more than can be discussed here. But I want to focus on some widely discussed approaches or experiments as illustrative of some of the possibilities and continuing problems.

ONE OF THE EARLIEST AND MOST INFLUENTIAL approaches to digital preservation (and digital authenticity) was what archivists call the "Pitt Project," a three-year (1993–1996) research effort funded by the National Historical Publications and Records Commission (NHPRC) and centered at the University of Pittsburgh School of Information and Library Studies. For historians, what is most interesting (and sometimes puzzling) about the Pitt Project approach is the way that it simultaneously narrows and broadens the role of archives and archivists through its focus on "records as evidence" rather than "information." "Records," David Bearman and Jennifer Trant explain, "are that which was created in the conduct of business" and provide "evidence of transactions." Data or information, by contrast, Bearman "dismisses as non-archival and unworthy of the archivist's attention."[30] From this point of view, the government's record of your Social Security account is vital but not the "information" contained in letters that you and others might have written complaining about the idea of privatizing Social Security.

The Pitt Project produced a pathbreaking set of "functional requirements for evidence in electronic record keeping"—in effect, strategies and tactics to ensure that electronic records produce legally or organizationally acceptable evidence of their transactions. Such a focus responds particularly well to worries about the "authenticity" of electronic records. But for historians (and for some archivists), the focus on records as evidence rather than records as sources of information, history, or memory seems disappointingly narrow. Moreover, as Canadian archivist Terry Cook points out, the emphasis on "redesigning computer systems' functional requirements to preserve the integrity and reliability of

records" and assigning "long-term custodial control . . . to the creator of archival records" privileges "the powerful, relatively stable, and continuing creators of records capable of such reengineering" and ignores artists, activists, and "marginalized and weaker members of society" who have neither the resources nor inclination to produce "business acceptable communications."[31]

While the Pitt Project emphasizes archival professionalism, a narrowing of the definition of recordkeeping, a rejection of the custodial tradition in archives, and planning for more careful collecting in the future rather than action in the present, the Internet Archive has taken precisely the opposite approach. It represents a grass-roots, immediate, enthusiast response to the crisis of digital preservation that both expands and further centralizes archival responsibility in ways that were previously unimaginable. Starting in September 1996, Brewster Kahle and a small staff sent "crawlers" out to capture the web by moving link-by-link and completing a full snapshot every two months. Although in part a philanthropic venture funded by Kahle, the Internet Archive also has a commercial side. Kahle's for-profit web navigation service, Alexa Internet (bought by Amazon in 1999 for $300 million), is what actually gathers the web snapshots, which it uses to analyze patterns of web use, and then donates them to the Internet Archive.[32]

By February 2002, the Internet Archive (IA) had gathered a monumental collection of more than 100 terabytes of web data—about 10 billion web pages or five times all the books in the Library of Congress—and was gobbling up 12 terabytes more each month. That same fall, it began offering public access to most of the collection through what Kahle called the "Wayback Machine"—a wry reference to the device used by the time-traveling Mr. Peabody in the Rocky and Bullwinkle cartoons of the 1960s. Astonishingly, a single individual with a very small staff has created the world's largest database and library in just five years.[33]

In December 2001, shortly after the Wayback Machine became public, the search engine company Google unveiled "Google Groups," another massive digital archive—this one under purely commercial auspices. Google Groups provides access to more than 650 million messages posted over the past two decades to "Usenet," the online discussion forums that predate even the Internet. Although "ownership" seems like a dubious concept in relation to a public discussion forum, Google purchased the archive from Deja.com, which had brought the groups to the web but then collapsed in the Internet bust. Despite Deja.com's failure, Google sees the Usenet Archive as another attractive feature in its stable of online information resources and tools.[34]

Both IA and Google Groups are libraries organized on principles that are more familiar to computer scientists than to librarians, as Peter Lyman, who knows both worlds as the head of the University of California at Berkeley library and as a member of the IA board, points out. The library community has focused on developing "sophisticated cataloging strategies." But computer scientists, including Kahle, have been more interested in developing sophisticated search engines that operate directly on the data we see (the web pages) rather than on the metadata (the cataloging information). Whereas archival and library projects focus on "high-quality collections built around select themes" and make the unit of cataloging the web page, the computer science paradigm "allows for archiving the entire Web as it changes over time, then uses search engines to retrieve the necessary information."[35]

Projects designed by librarians and archivists generally have the advantages of precision and standardization. They favor careful protocols and standards such as the Dublin Core, the OAIS (Open Archival Information System), and the EAD (Encoded Archival Description). But the expense and

difficulty of the protocols and procedures mean that less well funded and staffed archives and libraries often ignore them. Responding to presentations by advocates of standards at a conference, computer scientist Jim Miller warned that if archivists push for too much cataloging metadata "they might end up with none."[36]

The Internet Archive, which is the child of the search engines and the computer scientists, is an extraordinarily valuable resource. Most historians will not be interested now, but in twenty-five or fifty years they will delight in searching it. A typical college history assignment in 2050 might be to compare web depictions of Muslim Americans in 1998 and 2008. But any appreciation of the IA must acknowledge its limitations. For example, large numbers of web pages do not exist as "static" HTML pages; rather, they are stored in databases, and the pages are generated "on the fly" by search queries. As a result, the IA's crawlers do not capture much of the so-called "deep web" that is stored in databases. Multimedia files—streaming media and flash—also do not seem to be captured. In addition, the Internet Archive's crawls cannot go on forever; at some point, they stop, since, as one of the computer scientists who manages them acknowledges, "the Web is essentially infinite in size." Anyone who browses the IA regularly encounters such messages as "Not in Archive" and "File Location Error" or even "closed for maintenance."[37]

Some pages are missing for legal and economic as well as technical reasons. Private, gated sites are off-limits to the Internet Archive's crawlers. And many ungated sites also discourage the crawlers. The New York Times allows free access to its current contents, but charges for articles more than one week old. If the IA gathered up and preserved the Times's content, there would be no reason for anyone to pay the Times for access to its proprietary archive. As a result, the Times includes a "robots exclusion" file on its site, which the IA respects. Even those sites without the robots exclusion file and without any formal copyright are still covered by copyright law and could challenge the IA's archiving of their content. To avoid trouble, the IA simply purges the pages of anyone who complains. It is as if Julie Nixon could write to the National Archives and tell them to delete her father's tapes or an author could withdraw an early novel from circulation.[38]

Thus the Internet Archive is very far from the complete solution to the problem of digital preservation. It does not deal with the digital records that vex the National Archives and other repositories because they lack the public accessibility and minimal standardization in HTML of web pages. Nor does it include much formally published literature—e-books and journals—which is sold and hence gated from view. And even for what it has gathered, it has not yet hatched a long-term preservation plan, which would have to incorporate a strategy for continuing access to digital data that are in particular (and time-bound) formats. Even more troubling, it has no plan for how it will sustain itself into the future. Will Kahle continue to fund it indefinitely?[39] What if Amazon and Alexa no longer find it worthwhile to gather the data, especially since acquisition costs are doubling every year?

Similar questions could be raised about "Google Groups." What if the company decides that there is no prospect of gaining adequate advertising revenue by making old newsgroup messages available (as, indeed, Deja.com previously determined)? While appreciating Google's entrepreneurial energy in preserving and making available an enormous body of historical documents, we should also look carefully at the way private corporations have suddenly entered into a realm—archives—that was previously part of the public sector—a reflection of the privatization sweeping across the global economy. At least so far, our most important, and most imaginatively constructed, digital collections are in private hands.[40]

GIVEN THAT THE PRESERVATION OF cultural heritage and national history are arguably social goods, why shouldn't the government take the lead in such efforts? One reason is that at least some key aspects of the digital present—the Bert story, for example—do not follow national boundaries and, indeed, erode them. If national archives were part of the projects of state-building and nationalism, then why should states support post-national digital archives? The declining significance of state-based national archives may mirror the decline of the contemporary national state. So far, the Smithsonian Institution and the Library of Congress have worked with the Internet Archive only where they needed its help in documenting some particularly national stories—the elections of 1996 and 2000 and the September 11th attacks.

Another reason for the limited government role is that the digital preservation crisis emerged most dramatically during the anti-statist Reagan revolution of the 1980s. In the 1970s, for example, the electronic records program of the National Archives made a modest, promising start. But, as archivist Thomas E. Brown writes, it went into "a near total collapse in the 1980s." The staff dropped to seven people by 1983, and, amazingly, this beleaguered group charged with guarding the nation's electronic records had no access to computer facilities. Things began to improve in the early 1990s, but, after 1993, the electronic records program suffered from further cutbacks in the federal work force. An underfunded and understaffed National Archives was hardly in a position to develop a solution to the daunting and mounting problem of electronic federal records.[41]

The Library of Congress also initially eschewed a leading role in preserving digital materials, as the National Research Council later complained. Here, too, one could detect the weakening influence of the state. The library's high-profile effort in the digital realm was "American Memory," which digitized millions of items from its collections and placed them online. Teachers, students, and researchers love American Memory, but it did nothing to preserve the growing number of "born digital" objects. Not coincidentally, American Memory was a project that could attract large numbers of private and corporate donors, who often saw sponsorship as good advertising and who paid for three-quarters of the project.[42]

Better developed state-centered approaches to digital preservation have, not surprisingly, emerged outside the United States—in Australia and Scandinavia, for example. Norway requires that digital materials be legally deposited with the national library in return for copyright protection.[43] One of the key ways that the Library of Congress could help preserve the future of digital materials would be to aggressively assert its copyright deposit claims, which would finesse some of the legal and ownership issues troubling the Internet Archive.[44]

Nevertheless, the National Archives and the Library of Congress have very recently begun—prodded by outside critics and supported belatedly by Congress—to take a more aggressive approach on digital preservation. The archives is proposing a "Redesign of Federal Records Management" to respond to the reality that "a large majority of electronic record series of continuing value are not coming into archival custody." It is also working closely with the San Diego Supercomputing Center on developing "persistent object preservation" (POP), which creates a description of a digital object (and groups of digital objects) in simple tags and schemas that will be understandable in the future; the records would be "self-describing" and, hence, independent of specific hardware and software. The computer scientists maintain that records in this format will last for three hundred to four hundred years.[45]

In December 2000, the Library of Congress launched the most important initiative, the National Digital Information Infrastructure Program (NDIIP). Even this massive and important federal initiative bore the marks of the anti-statist, privatization politics of the 1980s. Congress gave the library $5 million for planning and promised another $20 million when it approved the plan. But the final $75 million will only be distributed as a match against an equal amount in private funds.[46]

Although the future of the digital present remains perilous, these recent initiatives suggest some encouraging strategies for preserving the range of digital materials. A combination of technical and organizational approaches promises the greatest chance of success, but privatization poses grave dangers for the future of the past. Advocates of digital preservation need to mobilize state funding and state power (such as the assertion of eminent domain over copyright materials) but infuse it with the experimental and ad hoc spirit of the Internet Archive. And we need to recognize that, for many digital materials (especially the web), the imperfect computer-science paradigm probably has more to recommend it than the more careful and systematic approach of the librarians and archivists. What is often said of military strategy seems to apply to digital preservation: "the greatest enemy of a good plan is the dream of a perfect plan."[47] We have never preserved everything; we need to start preserving something.

GIVEN THE ENORMOUS BARRIERS to saving digital records and information, it comes as something of a surprise that many continue to insist that a perfect plan—or at least a pretty good plan—will eventually emerge. Techno-optimists such as Brewster Kahle dream most vividly of the perfect plan and its startling consequences. "For the second time in history," Kahle writes with two collaborators, "people are laying plans to collect all information—the first time involved the Greeks which culminated in the Library of Alexandria . . . Now . . . many [are] once again to take steps in building libraries that hold complete collections." Digital technology, they explain, has "gotten to the point where scanning all books, digitizing all audio recordings, downloading all websites, and recording the output of all TV and radio stations is not only feasible but less costly than buying and storing the physical versions."[48] Librarians and archivists remain skeptical of such predictions, pointing out the enormous costs of cataloging and making available what has been preserved, and that we have never saved more than a fraction of our cultural output. But, whatever our degree of skepticism, it is still worth thinking seriously about what a world in which everything was saved might look like.

Most obviously, archives, libraries, and other record repositories would suddenly be freed from the tyranny of shelf space that has always shadowed their work. Digitization also removes other long-term scourges of historical memory such as fire and war. The 1921 fire that destroyed the 1890 census records provided a crucial spark that finally led to the creation of the National Archives. But what if there had been multiple copies of the census? The ease—almost inevitability—of the copying of digital files means that it is considerably less likely today that things exist in only a single copy.[49]

What would a new, virtual, and universal Alexandria library look like? Kahle and his colleagues have forcefully articulated an expansive democratic vision of a past that includes all voices and is open to all. "There are about ten to fifteen million people's voices evident on the Web," he told a reporter. "The Net is a people's medium: the good, the bad and the ugly. The interesting, the picayune and the profane. It's all there." Advocates of the new universal library and archive wax even more eloquently about democratizing access to the historical record. "The opportunity of our time is to offer universal access to all of human knowledge," said Kahle.[50]

Kahle's vision of cultural and historical abundance merges the traditional democratic vision of the public library with the resources of the research library and the national archive. Previously, few had the opportunity to come to Washington to watch early Thomas Edison films at the Library of Congress. And the library could not have served them if they had. Democratized access is the real payoff in electronic records and materials. It may be harder to preserve and organize digital materials than it is paper records, but, once that is accomplished, they can be made accessible to vastly greater numbers of people. To open up the archives and libraries in this way democratizes historical work. Already, people who had never had direct access to archives and libraries can now enter. High school students are suddenly doing primary source research; genealogy has exploded in popularity because you no longer have to travel to distant archives.

This vision of democratic access also promises direct and unmediated access to the past. Electronic commerce enthusiasts tout "disintermediation"—which is the elimination of the insurance and real estate broker and other intermediaries—and the emergence of one like eBay made up of only buyers and sellers. In theory, the universal digital library might bring a similar cultural disintermediation in which people interested in history make direct contact with the documents and artifacts of the past without the mediation of cultural brokers like librarians, archivists, and historians. Sociologist Mike Featherstone speculates on the emergence of a "new culture of memory" in which the existing "hierarchical controls" over access would disappear. This "direct access to cultural records and resources from those outside cultural institutions" could "lead to a decline in intellectual and academic power" in which the historian, for example, no longer stands between people and their pasts.[51] The "Wayback Machine" encapsulates this vision of disintermediation by suggesting that everyone, like Mr. Peabody and his boy Sherman, can jump in a time machine and find out what Columbus or Edison was "really" like. Of course, most historians would argue that, while digital collections may put "the novice in the archive,"[52] he or she is not so likely to know what to do there. Still, the balance of power may shift. Ask any travel agent how the widespread access to information undercuts professional control.

Most historians have not embraced this vision in which everyone becomes his or her own historian. Nor have they enthusiastically endorsed the vision of a universal library that contains all voices and all records. In my informal polling, most historians recoil at the thought that they would need to write history with even more sources.[53] Historians are not particularly hostile to new technology, but they are not ready to welcome fundamental changes to their cultural position or their modes of work. Having lived our professional careers in a culture of scarcity, historians find that a world of abundance can be unsettling.

Abundance, after all, can be overwhelming. How do we find the forest when there are so many damned trees? Psychologist Aleksandr Luria made this point in his famous study of a Russian journalist, "S" (S. V. Shereshevskii), who had an amazingly photographic memory; he could reproduce complex tables of numbers and long lists of words that had been shown to him years earlier. But this "gift" turned out to be a curse. He could not recognize people because he remembered their faces so precisely; a slightly different expression would register as a different person. Grasping the larger point of a passage or abstract idea "became a tortuous . . . struggle against images that kept rising to the surface in his mind." He lacked, as psychologist Jerome Bruner notes, "the capacity to convert encounters with the particular into instances of the general."[54]

If historians are to set themselves "against forgetting" (in Milan Kundera's resonant phrase), then they may need to figure out new ways to sort their way through the potentially overwhelming digital record

of the past. Contemporary historians are already groaning under the weight of their sources. Robert Caro has spent twenty-six years working his way through just the documents on Lyndon B. Johnson's pre-vice-presidential years—including 2,082 boxes of Senate papers. Surely, the injunction of traditional historians to look at "everything" cannot survive in a digital era in which "everything" has survived.[55]

The historical narratives that future historians write may not actually look much different from those that are crafted today, but the methodologies they use may need to change radically. If we have, for example, a complete record of everything said in 2010, can we offer generalizations about the nature of discourse on a topic simply by "reading around"? Wouldn't we need to engage in some more methodical sampling in the manner of, say, sociology? Would this revive the social-scientific approaches with which historians flirted briefly in the 1970s? Wouldn't historians need to learn to write complex searches and algorithms that would allow them to sort through this overwhelming record in creative, but systematic, ways? The future gurus of historical research methodology may be the computer scientists at Google who have figured out how to search the equivalent of a 100-mile-high pile of paper in half a second. "To be able to find things with high accuracy and high reliability has an incredible impact on the world"—and, one might add, future historians. Future graduate programs will probably have to teach such social-scientific and quantitative methods as well as such other skills as "digital archaeology"(the ability to "read" arcane computer formats), "digital diplomatics" (the modern version of the old science of authenticating documents), and data mining (the ability to find the historical needle in the digital hay).[56] In the coming years, "contemporary historians" may need more specialized research and "language" skills than medievalists do.

HISTORIANS HAVE TIME TO THINK ABOUT changing their methods to meet the challenge of a cornucopia of historical sources. But they need to act more immediately on preserving the digital present or that reconsideration will be moot; they will be struggling with a scarcity, not an overabundance, of sources. Surprisingly, however, historians themselves have been scarce on this issue.[57] Archivists and librarians have intensely debated and discussed digitization and digital presentation for more than a decade. They have written hundreds of articles and reports, undertaken research projects, and organized conferences and workshops. Academic and teaching historians have taken almost no part in these conferences and have contributed almost nothing to this burgeoning literature. Historical journals have published nothing on the topic.[58]

Part of the reason is that preserving the born-digital materials for future historians seems like a theoretical and technical issue, tomorrow's problem or at least someone else's problem. Another reason for this disinterest is the divorce of archival concerns from the historical profession—a part of the general narrowing of the concerns of professional historians over the past century. In the late nineteenth and early twentieth centuries, historians and archivists were closely aligned. Perhaps the most important committee of the American Historical Association in the 1890s was the Historical Manuscripts Commission, which led to the AHA's influential Public Archives Commission. Archival concerns found a regular place in the AHA's Annual Meeting, the American Historical Review, and especially the voluminous AHA annual reports. Most important, the AHA led the fight to establish the National Archives. But in 1936 (in the midst of an earlier technological upheaval that came with the emergence of microfilm), the Conference of Archivists left the AHA to create the Society of American Archivists. The professions charged with writing about the past and preserving the records of the past have sharply diverged in the past seven decades. Today, only 82 of the 14,000 members of the AHA identify themselves as archivists.[59]

But historians ignore the future of digital data at their own peril. What, for example, about the long-term preservation of scholarship that is—increasingly—originating in digital form? Not only do historians need to ensure the future of their own scholarship, but linking directly from footnotes to electronic texts —an exciting prospect for scholars—will only be possible if a stable archiving system emerges.[60] For the foreseeable future, librarians and archivists will be making decisions about priorities in digital preservation. Historians should be at the table when those decisions are made. Do they wish to endorse, for example, the Pitt Project's emphasis on preserving records of business transactions rather than "information" more broadly?

One of the most vexing and interesting features of the digital era is the way that it unsettles traditional arrangements and forces us to ask basic questions that have been there all along. Some are about the relationship between historians and archival work. Should the work of collecting, organizing, editing, and preserving of primary sources receive the same kind of recognition and respect that it did in earlier days of the profession? Others are about whose overall responsibility it is to preserve the past. For example, should the National Archives expand its role in preservation beyond official records? For many years, historians have taken a hands-off approach to archival questions. With the unsettling of the status quo, they should move back more actively into this realm. If the web page is the unit of analysis for the digital librarian and the link the unit of analysis for the computer scientists, what is the appropriate unit of analysis for historians? What would a digital archival system designed by historians look like? And how might we alter and enhance our methodologies in a digital realm? For example, in a world where all sources were digitized and universally accessible, arguments could be more rigorously tested. Currently, many arguments lack such scrutiny because so few scholars have access to the original sources—a problem that has arisen especially sharply in the recent controversies over Michael A. Bellesiles' *Arming America: The Origins of a National Gun Culture* (2000). In a new digital world, would historians then be held to the same standard of "reproducible" results as scientists?[61]

Of course, when historians get to the preservation table, they will discover a cultural and professional clash between their own impulses, which are to save everything, and those of librarians and archivists who believe that selection, whether passive or active, is inevitable. The National Archives, for example, only permanently accessions 2 percent of government records.[62] This conflict surfaced in the 1980s and 1990s, when librarians tried to bring in scholars to discuss priorities in preserving books that were deteriorating because of acidic paper. Librarians found the discussion "frustrating." "Many scholars," recalls Deanna Marcum, declared "that everything had to be saved and they could not make choices." Not surprisingly, scholars have responded very differently to Nicholson Baker's sharp attack on the microfilming and disposal of aging books and newspapers in Double Fold than have archivists and librarians. Whereas many scholars have shared Baker's outrage that books and newspapers have been destroyed, archivists and librarians have responded in outrage to what they see as his failure to understand the pressures that make it impossible to save everything. Whereas historians with their gaze fixed on the past worry about information scarcity (the missing letter or diary), archivists and librarians recognize that we now live in a world of overwhelming information abundance.a href="#N_63_" name="ref63" id="ref63">63 If historians are going to join in preservation discussions, they will have to make themselves better informed about the simultaneous abundance of historical sources and scarcity of financial resources that lead archivists and librarians to respond with exasperation to scholars' blithe insistence that everything must be saved.

Preservation of the past is, in the end, often a matter of allocating adequate resources. Perhaps the

largest problem facing the preservation of electronic government records has nothing to do with technology; it is, as various reports have noted, "the low priority traditionally given to federal records management." In the absence of new resources, the costs of preservation will come from the money that our society, in the aggregate, allocates for history and culture. Richard Cox, for example, has argued that a greater portion of the budget of the National Historical Publications and Records Commission (NHPRC) should go to electronic records preservation and management and correspondingly less money should go to the letterpress Documentary Editions that the commission also funds, since "most of the records represented by the documentary editions are not immediately threatened." This stance does not endear him to documentary editors, who are much better represented among professional historians than are archivists.[64]

The alternative to squabbling over inadequate resources that are appropriated for these purposes is joint action to secure further funds. When Shirley Baker, president of the Association of Research Libraries, challenged historian Robert Darnton's favorable review of Baker's book and noted "choices have always had to be made" in the absence of "greater public commitment to the preservation of the historical record," Darnton responded by urging the establishment of "a new kind of national library dedicated to the preservation of cultural artifacts" (including disappearing digital records) and funded by income generated by the sale or rental of bandwidth.[65] Such state-based solutions return us to the kind of alliance between historians and archivists that led to building of the National Archives in the 1930s, an era of growing rather than waning confidence in the nation-state. Historians need to join in lobbying actively for adequate funding for both current historical work and preservation of future resources. They should also argue forcefully for the democratized access to the historical record that digital media make possible. And they must add their voices to those calling for expanding copyright deposit—and opposing copyright extension, for that matter—of digital materials so as to remove some of the legal clouds hanging over efforts like the Internet Archive and to halt the ongoing privatization of historical resources. Even in the absence of state action, historians should take steps individually and within their professional organizations to embrace the culture of abundance made possible by digital media and expand the public space of scholarship—for example, making their own work available for free on the web, cross-referencing other digital scholarship, and perhaps depositing their sources online for other scholars to use. A vigorous public domain today is a prerequisite for a healthy historical record.[66]

More than a century ago, Justin Winsor, the third president of the AHA, concluded his Presidential Address—focused on a topic that would be considered odd today, that of preserving manuscript sources for the study of history—with a plea to the AHA "to convince the National Legislature" to support a scheme "before it is too late" to preserve and make known "what there is still left to us of the historical manuscripts of the country." For founders of the historical profession such as Winsor, the need to engage with history broadly defined—not just how it was researched but also how it was taught in the schools or preserved in archives—came naturally; it was part of creating a historical profession.[67] In the early twenty-first century, we are likely to be faced with recreating the historical profession, and we will be well served by such a broad vision of our mission. If the past is to have an abundant future, if the story of Bert Is Evil and hundreds of other stories are to be fully told, then historians need to act in the present.

Footnotes:

This article has benefited greatly from the generous and astute comments of a number of friends and colleagues: Joshua Brown, Michael Grossberg, Deborah Kaplan, Gary Kornblith, Michael O'Malley, Kelly

[1] Greg Miller, "Cyberculture: The Scene/The Webby Awards," *Los Angeles Times* (March 9, 1998): D3. On Ignacio, see the interview "Dino Ignacio: Evil Incarnate," in Philippine Web Designers Network, Philweavers, http://www.philweavers.net/profiles/dinoginacio.html; Buck Wolf, "Osama bin Muppet," *ABC News*, http://www.abcnews.go.com/sections/us/WolfFiles/wolffiles190.html; "Media Killed Bert Is Evil," http://plaza.powersurfr.com/bert/, viewed online April 15, 2002, but unavailable as of July 4, 2002; Peter Hartlaub, "Bert and bin Laden Poster Tied to S.F. Student," *San Francisco Chronicle* (October 12, 2001): A12; Gina Davidson, "Bert and Bin: How the Joke Went Too Far," *The Scotsman* (October 14, 2001): 3.

[2] "Bert Is Evil!" in Snopes.com, http://www.snopes2.com/rumors/bert.htm; "Bert Is Evil—Proof in the Most Unlikely Places," in *HermAphroditeZine*, http://www.pinktink3.250x.com/hmm/bert.htm; Josh Grossberg, "The Bert-Bin Laden Connection?" in *E! Online News*, October 10, 2001, http://www.eonline.com/News/Items/0,1,8950,00.html; Joey G. Alarilla, "Infotech Pinoy Webmaster Closes Site after 'Bert-Bin Laden' Link," Philippine Daily Inquirer (October 22, 2001): 17; Dino Ignacio, "Good-bye Bert," in *Fractal Cow*, http://www.fractalcow.com/bert/bert.htm. See also Michael Y. Park, "Bin Laden's Felt-Skinned Henchman?" *Fox News* (October 14, 2001), http://www.foxnews.com/story/0,2933,36218,00.html; Declan McCullagh, "Osama Has a New Friend," *Wired News* (October 10, 2001), http://www.wired.com/news/conflict/0,2100,47450,00.html; "Sesame Street Character Depicted with bin Laden on Protest Poster," AP Worldstream (October 11, 2001). Nikke Lindqvist, N!kke, http://www.lindqvist.com/art.php?incl=bert.php&lang=eng, provides an excellent chronicle of the unfolding story. Significantly, many of the links on this site, which I first viewed in February 2002, were no longer working in March 2003.

[3] Jeffrey Benner, "Is U.S. History Becoming History?" *Wired News* (April 9, 2001), http://www.wired.com/news/print/0,1294,42725,00.html.

[4] Arcot Rajasekar, Richard Marciano, and Reagan Moore, "Collection-Based Persistent Archives," http://www.sdsc.edu/NARA/Publications/OTHER/Persistent/Persistent.html; U.S. Congress, House Committee on Government Operations, Taking a Byte out of History: The Archival Presentation of Federal Computer Records, HR 101–987 (Washington, D.C., 1990); National Academy of Public Administration, *The Effects of Electronic Recordkeeping on the Historical Record of the U.S. Government* (Washington, D.C., 1989), 8, 29; Joel Achenbach, "The Too-Much-Information Age," Washington Post (March 12, 1999): A01; General Accounting Office (hereafter, GAO), *Information Management: Challenges in Managing and Preserving Electronic Records* (Washington, D.C., 2002), 11, 66. See also Alexander Stille, *The Future of the Past* (New York, 2002), 306; Richard Harvey Brown and Beth Davis-Brown, "The Making of Memory: The Politics of Archives, Libraries, and Museums in the Construction of National Consciousness," *History of the Human Sciences* 11, no. 4 (1998): 17–32; Deanna Marcum, "Washington Post Publishes Letter from Deanna Marcum," CLIR Issues, no. 2 (March/April 1998), http://www.clir.org/pubs/issues/issues02.html#post.

[5] John Higham, History: Professional Scholarship in America (1965; rpt. edn., Baltimore, 1983), 16–20. See also American Historical Association Committee on Graduate Education, *The Education of Historians in the 21st Century* (Urbana, Ill., forthcoming 2004). To observe this broader vision is not to deny the very different historical circumstances (such as the disorganization of archives), the obvious blindness of the early professional historians on many matters (such as race and gender), and the early tensions between "amateurs" and professionals.

[6] For interesting observations on "abundance" in two different realms of historical work, see James O'Toole, "Do Not Fold, Spindle, or Mutilate: Double Fold and the Assault on Libraries," *American Archivist* 64 (Fall/Winter 2001): 385–93; John McClymer, "Inquiry and Archive in a U.S. Women's History Course," *Works and Days* 16, nos. 1–2 (Spring/Fall 1998): 223. For a sweeping statement

about political and cultural implications of "digital information that moves frictionlessly through the network and has zero marginal cost per copy," see Eben Moglen, "Anarchism Triumphant: Free Software and the Death of Copyright," *First Monday* 4, no. 8 (August 1999), http://www.firstmonday.dk/issues/issue4_8/moglen/index.html.

[7.] Committee on the Records of Government, Report (Washington, D.C., 1985), 9 (the committee was created by the American Council of Learned Societies, the Council on Library Resources, and the Social Science Research Council with funding from the Mellon, Rockefeller, and Sloan foundations); John Garrett and Donald Waters, *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information* (Washington, D.C., 1996); Paul Conway, *Preservation in the Digital World* (Washington, D.C., 1996), http://www.clir.org/pubs/reports/conway2/index.html. For other reports with similar conclusions, see, for example, the 1989 report of the National Association of Government Archives and Records Administrators, cited in Margaret Hedstrom, "Understanding Electronic Incunabula: A Framework for Research on Electronic Records," *American Archivist* 54 (Summer 1991): 334–54; House Committee on Government Operations, *Taking a Byte out of History; Committee on an Information Technology Strategy for the Library of Congress, Computer Science and Telecommunications Board, Commission on Physical Sciences, Mathematics, and Applications, and the National Research Council, LC21: A Digital Strategy for the Library of Congress* (Washington, D.C., 2000), http://books.nap.edu/html/lc21/index.html; GAO, *Information Management; NHPRC Electronic Records Agenda Final Report (Draft)* (St. Paul, Minn., 2002).

[8.] Margaret MacLean and Ben H. Davis, eds., Time and Bits: Managing Digital Continuity (Los Angeles, 1998), 11, 6; Jeff Rothenberg, Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation (Washington, D.C., 1998), http://www.clir.org/pubs/reports/rothenberg/contents.html. The 1997 conference "Documenting the Digital Age" has also disappeared from the web, nor is it available in the Internet Archive. The Sanders film is available from the Council on Library and Information Resources, http://www.clir.org/pubs/film/future/order.html.

[9.] Achenbach, "Too-Much-Information Age." See also Stille, *Future of the Past; Council on Library and Information Resources* (hereafter, CLIR), *The Evidence in Hand: Report of the Task Force on the Artifact in Library Collections* (Washington, D.C., 2001), http://www.clir.org/pubs/reports/pub103/contents.html.

[10.] Margaret O. Adams and Thomas E. Brown, "Myths and Realities about the 1960 Census," *Prologue: Quarterly of the National Archives and Records Administration* 32, no. 4 (Winter 2000), http://www.archives.gov/publications/prologue/winter_2000_1960_census.html. See also letter of August 15, 1990, from Kenneth Thibodeau, which says that recovering the records took "substantial efforts" by the Bureau of the Census, quoted in House Committee on Government Operations, Taking a Byte out of History, 3. According to Timothy Lenoir, it is now too expensive to rescue the computer tapes that represent Douglas Englebart's pioneering hypermedia-groupware system called NLS (for oNLine System)—the basis of many of the features of personal computers. Timothy Lenoir, "Lost in the Digital Dark Ages" (paper delivered at "The New Web of History: Crafting History of Science Online," Cambridge, Mass., March 28, 2003).

[11.] Marcia Stepanek, "From Digits to Dust," Business Week (April 20, 1998); House Committee on Government Operations, Taking a Byte out of History, 16; Jeff Rothenberg, "Ensuring the Longevity of Digital Documents," *Scientific American* (January 1995): 42–47. See also Garrett and Waters, Preserving Digital Information. Many Vietnam records are stored in a database system that is no longer supported and can only be translated with difficulty. As a result, the Agent Orange Task Force could not use important herbicide records. Stille, *Future of the Past*, 305.

[12.] Most Microsoft software moves into what the company calls the "non-supported phase" after just four or five years, although it offers a more limited "extended support phase" that lasts up to seven years. After that, you are out of luck. Microsoft, "Windows Desktop Product Life Cycle Support and Availability Policies for Businesses," October 15, 2002, http://www.microsoft.com/windows/lifecycle.mspx; Lori Moore, "Q&A: Microsoft Standardizes Support Lifecycle," Press Pass: Information for

Journalists (October 15, 2002), http://www.microsoft.com/presspass/features/2002/Oct02/10-15support.asp. On media longevity, see Rothenberg, Avoiding Technological Quicksand; MacLean and Davis, Time and Bits; Margaret Hedstrom, "Digital Preservation: A Time Bomb for Digital Libraries" (paper delivered at the NSF Workshop on Data Archiving and Information Preservation, March 26–27, 1999), http://www.uky.edu/~kiernan/DL/hedstrom.html; Frederick J. Stielow, "Archival Theory and the Preservation of Electronic Media: Opportunities and Standards below the Cutting Edge," *American Archivist* 55 (Spring 1992): 332–43; Charles M. Dollar, *Archival Theory and Information Technology: The Impact of Information Technologies on Archival Principles and Methods* (Ancona, Italy, 1992), 27–32; GAO, *Information Management*, 50–52.

[13.]Richard J. Cox, "Messrs. Washington, Jefferson, and Gates: Quarrelling about the Preservation of the Documentary Heritage of the United States," *First Monday* 2, no. 8 (August 1997), http://firstmonday.org/issues/issue2_8/cox/. See also Peter Lyman and Brewster Kahle, "Archiving Digital Cultural Artifacts: Organizing an Agenda for Action," *D-Lib Magazine* 4, nos. 7–8 (July/August 1998), http://www.dlib.org/dlib/july98/07lyman.html. Voyager's CD-ROM explicating Beethoven's Ninth Symphony—a landmark work in multimedia—no longer operates, in part, because Apple changed a CD-ROM driver that the program relied on. Robert Winter, *Ludwig Van Beethoven Symphony No. 9* (Santa Monica, Calif., 1991). Digital art presents particularly difficult problems; see, for example, Scott Carlson, "Museums Seek New Methods for Preserving Digital Art," *Chronicle of Higher Education* (August 16, 2002).

[14.]Margaret Hedstrom, "How Do We Make Electronic Archives Usable and Accessible?" (paper delivered at "Documenting the Digital Age," San Francisco, February 10–12, 1997); Luciana Duranti, "Diplomatics: New Uses for an Old Science," *Archivaria* 28 (Summer 1989): 7–27; Peter B. Hirtle, "Archival Authenticity in a Digital Age," in *Council on Library and Information Resources, Authenticity in a Digital Environment* (Washington, D.C., 2000), http://www.clir.org/pubs/reports/pub92/contents.html; CLIR, Evidence in Hand; Susan Stellin, "Google's Revival of a Usenet Archive Opens Up a Wealth of Possibilities But Also Raises Some Privacy Issues," New York Times (May 7, 2001): C4; David Bearman and Jennifer Trant, "Authenticity of Digital Resources: Towards a Statement of Requirements in the Research Process," *D-Lib Magazine* 4, no. 6 (June 1998), http://www.dlib.org/dlib/june98/06bearman.html.

[15.]Abby Smith, "Authenticity in Perspective," in CLIR, Authenticity in a Digital Environment (Washington, D.C., 2000), http://www.clir.org/pubs/reports/pub92/smith.html; Clifford Lynch, "Authenticity and Integrity in the Digital Environment: An Exploratory Analysis of the Central Role of Trust," in *Authenticity in a Digital Environment*, http://www.clir.org/pubs/reports/pub92/contents.html. See also M. T. Clanchy, *From Memory to Written Record: England 1066–1307*, 2d edn. (Oxford, 1993); see also *Research Libraries Group, Attributes of a Trusted Digital Repository: Meeting the Needs of Research Resources; An RLC-OCLC Report* (Mountain View, Calif., 2001), http://www.rlg.org/longterm/attributes01.pdf.

[16.]Brewster Kahle, Rick Prelinger, and Mary E. Jackson, "Public Access to Digital Materials" (white paper delivered at the Association of Research Libraries and Internet Archive Colloquium "Research in the 'Born-Digital' Domain," San Francisco, March 4, 2001), available at http://www.dlib.org/dlib/october01/kahle/10kahle.html.

[17.]Committee on Intellectual Property Rights in the Emerging Information Infrastructure, National Research Council, et al., *The Digital Dilemma: Intellectual Property in the Information Age* (Washington, D.C., 1999), http://books.nap.edu/html/digital_dilemma/; Richard Stallman, "Can You Trust Your Computer?" *Newsforge* (October 21, 2002), http://newsforge.com/newsforge/02/10/21/1449250.shtml?tid=19. The Digital Millennium Copyright Act makes it illegal to circumvent technical protection services. See Peter Lyman, "Archiving the World Wide Web," in *CLIR, Building a National Strategy for Digital Preservation: Issues in Digital Media Archiving* (Washington, D.C., 2002), http://www.clir.org/pubs/reports/pub106/web.html.

[18.]CLIR, Evidence in Hand.

[19.]Committee on Intellectual Property Rights in the Emerging Information Infrastructure, National Research Council, Digital Dilemma.

[20.]As with our network of research libraries, this system is a modern invention. The first public governmental archive came with the French Revolution; the British Public Record Office opened in 1838, and the National Archives is of startlingly recent vintage: the legislation establishing it did not come until 1934. Donald R. McCoy, "The Struggle to Establish a National Archives in the United States," in *Guardian of Heritage: Essays on the History of the National Archives*, Timothy Walch, ed. (Washington, D.C., 1985), 1–15.

[21.]Don Waters, "Wrap Up" (paper delivered at the DAI Institute, "The State of Digital Preservation: An International Perspective," Washington, D.C., April 25, 2002), available at http://www.clir.org/pubs/reports/pub107/contents.html; Dale Flecker, "Preserving Digital Periodicals," in *CLIR, Building a National Strategy for Digital Preservation*.

[22.]Michael L. Miller, "Assessing the Need: What Information and Activities Should We Preserve?" (paper delivered at "Documenting the Digital Age," San Francisco, February 10–12, 1997), copy in possession of author. To be sure, it has been biased toward the preservation of the records of the rich and powerful, although in more recent years energetic, "activist archivists" have sought out more diverse sets of materials. Ian Johnston, "Whose History Is It Anyway?" *Journal of the Society of Archivists* 22, no. 2 (2001): 213–29.

[23.]See Adrian Cunningham, "Waiting for the Ghost Train: Strategies for Managing Electronic Personal Records before It Is Too Late" (paper delivered at the Society of American Archivists Annual Meeting, Pittsburgh, August 23–29, 1999), available at http://www.rbarry.com/cunningham-waiting2.htm. For numbers of commercial word-processing programs, see House Committee on Government Operations, Taking a Byte out of History, 15.

[24.]*SRA International, Report on Current Recordkeeping Practices within the Federal Government* (Arlington, Va., 2001), http://www.archives.gov/records_management/pdf/report_on_recordkeeping_practices.pdf. This report responded to an earlier GAO report: U.S. Government Accounting Office, National Archives: Preserving Electronic Records in an Era of Rapidly Changing Technology (Washington, D.C., 1999). Archival consultant Rick Barry reports that four-fifths of e-mail creators he surveyed "do not have a clue" whether their e-mail was an official record and that most are "largely unaware" of official e-mail policies. Quoted in David A. Wallace, "Recordkeeping and Electronic Mail Policy: The State of Thought and the State of the Practice" (paper delivered at the Annual Meeting of the Society of American Archivists, Orlando, Florida, September 3, 1998), http://www.rbarry.com/wallace.html.

[25.]Rothenberg, *Avoiding Technological Quicksand*. For the long controversy over NARA and the printing of e-mail, see Bill Miller, "Court Backs Archivist's Rule: U.S. Agencies May Be Allowed to Delete E-Mail," Washington Post (August 7, 1999): A02; Wallace, "Recordkeeping and Electronic Mail Policy"; GAO, *Information Management*, 57–65.

[26.]See Stewart Granger, "Emulation as a Digital Preservation Strategy," *D-Lib Magazine* 6, no. 10 (October 2000), http://www.dlib.org/dlib/october00/granger/10granger.html, on this as the "dominant" approach. An even earlier intervention version of "migration" is to move digital objects to "standardized" formats immediately or as quickly as possible, to put them in non-proprietary, open-source, commonly accepted formats (for instance, ASCII for text, .tiff for images, etc.) that are likely to be around for a long time. Of course, popular standards are no guarantee of longevity; in 1990, NARA was arguing that spreadsheets formatted for Lotus 1–2–3 were not a preservation problem since the program was so "widespread." House Committee on Government Operations, Taking a Byte out of History, 12.

[27.]Warwick Cathro, Colin Webb, and Julie Whiting, "Archiving the Web: The PANDORA Archive at the National Library of Australia" (paper delivered at "Preserving the Present for the Future Web Archiving," Copenhagen, June 18–19, 2001). See also Diane Vogt-O'Connor, "Is the Record of the 20th Century at Risk?" *CRM: Cultural Resource Management* 22, no. 2 (1999): 21–24.

[28.]Rothenberg, Avoiding Technological Quicksand.

[29.]Margaret Hedstrom, "Digital Preservation: Matching Problems, Requirements and Solutions" (paper delivered at the NSF Workshop on Data Archiving and Information Preservation, March 26–27, 1999), NSFWorkshop/hedpp.html">http://cecssrv1.cecs.missouri.edu/NSFWorkshop/hedpp.html (accessed March 2002 but unavailable in May 2003). See also Margaret Hedstrom, "Research Issues in Digital Archiving" (paper delivered at the DAI Institute, "The State of Digital Preservation: An International Perspective, Washington, D.C., April 25, 2002, available at http://www.clir.org/pubs/reports/pub107/contents.html). Rothenberg himself is currently undertaking research on emulation, and other emulation research is going on at the University of Michigan and Leeds University and at IBM's Almaden Research Center in San Jose, California. Daniel Greenstein and Abby Smith, "Digital Preservation in the United States: Survey of Current Research, Practice, and Common Understandings" (paper delivered at "Preserving History on the Web: Ensuring Long-Term Access to Web-Based Documents," Washington, D. C., April 23, 2002), copy in possession of author. More recently, Rothenberg has apparently tempered his position on emulation versus migration.

[30.]David Bearman and Jennifer Trant, "Electronic Records Research Working Meeting, May 28–30, 1997: A Report from the Archives Community," *D-Lib Magazine* 3, nos. 7–8 (July/August 1997), http://www.dlib.org/dlib/july97/07bearman.html; Terry Cook, "The Impact of David Bearman on Modern Archival Thinking: An Essay of Personal Reflection and Critique," *Archives and Museum Informatics* 11 (1997): 23. See further Margaret Hedstrom, "Building Record-Keeping Systems: Archivists Are Not Alone on the Wild Frontier," Archivaria 44 (Fall 1997): 46–48. See also David Bearman and Ken Sochats, "Metadata Requirements for Evidence," in University of Pittsburgh, School of Information Sciences, the Pittsburgh Project, NHPRC/">http://www.archimuse.com/papers/NHPRC/. (Many parts of this site have disappeared, but this undated paper is available at NHPRC/BACartic.html">http://www.archimuse.com/papers/NHPRC/BACartic.html.) David Bearman, "An Indefensible Bastion: Archives as Repositories in the Electronic Age," in Bearman, ed., Archival Management of Electronic Records (Pittsburgh, 1991), 14–24; Margaret Hedstrom, "Archives as Repositories—A Commentary," in *ibid*.

[31.]Cook, "Impact of David Bearman on Modern Archival Thinking," 15–37. From another perspective, the Pitt Project broadened, rather than narrowed, the concerns of electronic archivists, since previously the focus had been on statistical databases. In one effort to join the emphasis on records as evidence with a broader social cultural focus, Margaret Hedstrom argues that "to benefit fully from the synergy between business needs and preservation requirements, cultural heritage concerns should be linked to equally critical social goals, such as monitoring global environment change, locating nuclear waste sites, and establishing property rights, all of which also depend on long-term access to reliable, electronic evidence." Quoted in Richard J. Cox, "Searching for Authority: Archivists and Electronic Records in the New World at the Fin-de-Siècle," First Monday 5, no. 1 (January 3, 2000), http://firstmonday.org/issues/issue5_1/cox/index.html. The Pitt Project has been the subject of enormous discussion and significant debate among archivists; a full and nuanced treatment of the subject is beyond the scope of this article. Whereas Cook offers serious criticism of Bearman, the leader of the project along with Richard Cox, he also celebrates Bearman as "the leading archival thinker of the late twentieth century." Linda Henry offers a sweeping attack on Bearman and other advocates of a "new paradigm" in electronic records management in "Schellenberg in Cyberspace," *American Archivist* 61 (Fall 1998): 309–27. A more recent critique is Mark A. Greene, "The Power of Meaning: The Archival Mission in the Postmodern Age," *American Archivist* 65, no. 1 (Spring/Summer 2002): 42–55. Terry Cook puts the story in historical perspective (but from his particular perspective) in "What Is Past Is Prologue: A History of Archival Ideas since 1898, and the Future Paradigm Shift," *Archivaria* 43 (Spring 1997), available at http://www.rbarry.com/cookt-pastprologue-ar43fnl.htm. The project "Preservation of the Integrity of Electronic Records" (called the UBC Project because it was carried out at the University of British Columbia) and the InterPARES project (International Research on Permanent

Authentic Records in Electronic Systems), which built on the UBC Project, have taken a different approach, but they share the Pitt Project's emphasis on the problem of "authenticity" and on "records" rather than the broader array of sources that generally interest historians. Luciana Duranti, *The Long-Term Preservation of Authentic Electronic Records: Findings of the InterPARES Project* (Vancouver, 2002), http://www.interpares.org/book/index.htm. The December 2002 draft of the NHPRC Electronic Records Agenda Final Report suggests that the consensus among archivists is moving toward a broader definition of records. My understanding of these issues has been greatly aided by attending the December 8–9, 2002, meeting convened to discuss that agenda and by conversations with Robert Horton of the Minnesota Historical Society, who is the leader of that effort.

[32.] Carolyn Said, "Archiving the Internet: Brewster Kahle Makes Digital Snapshots of Web," *San Francisco Chronicle* (May 7, 1998): B3; Brewster Kahle, "Preserving the Internet," Scientific American (March 1997), http://www.sciamdigital.com; Kendra Mayfield, "Wayback Goes Way Back on Web," *Wired News* (October 29, 2001), http://www.wired.com/news/print/0,1294,47894,00.html; Mike Burner, "The Internet Archive Robot," e-mail to Robots Mailing List, September 5, 1996, http://www. robotstxt.org/wc/mailing-list/1258.html. On Alexa, see Rajiv Chandrasekaran, "Seeing the Sites on a Custom Tour: New Internet Search Tool Takes Selective Approach," *Washington Post* (September 4, 1997): E01; Tim Jackson, "Archive Holds Wealth of Data," *Financial Times* (London) (November 24, 1997): 15; Laurie J. Flynn, "Alexa's Crusade Continues under Amazon.com's Flag," *New York Times* (May 3, 1999): C4. On other early efforts to "save the web," see Spencer Reiss, "Internet in a Box," Wired (October 1996), http://www.wired.com/wired/4.10/scans.html; Bruce Sterling, "The Life and Death of Media" (speech delivered at the Sixth International Symposium on Electronic Art, Montreal, September 19, 1995), available at http://www.chriswaltrip.com/sterling/dedmed.html; John Markoff, "When Big Brother Is a Librarian," *New York Times* (March 9, 1997): IV: 3; James B. Gardner, comp., "Report on Documenting the Digital Age" (Washington, D.C., 1997); Nathan Myhrvold, "Capturing History Digitally: Why Archive the Internet?" (paper delivered at "Documenting the Digital Age," San Francisco, February 10–12, 1997), copies of Gardner and Myhrvold in possession of author.

[33.] Hamish Mackintosh, Interview with Brewster Kahle, "Webarian," *Guardian* (February 21, 2002): 4, http://www.guardian.co.uk/online/story/0,3605,653286,00.html; Molly Wood, "CNET's Web Know-It-All Goes Where You Won't," *CNET* (March 15, 2002), http://www.cnet.com/software/0-8888-8-9076625-1.html; "Seeing the Future in the Web's Past," *BBC News* (November 12, 2001), http://news.bbc.co.uk/hi/english/in_depth/sci_tech/2000/dot_life/newsid_1651000/1651557.stm. For a good explanation of the technical side of IA, see Richard Koman, "How the Wayback Machine Works," *O'Reilly Network* (January 21, 2002), http://www.oreillynet.com/pub/a/webservices/2002/01/18/brewster.html.

[34.] Google Employee, "Google Groups Archive Information Newsgroups," e-mail, December 21, 2001; Stellin, "Google's Revival of a Usenet Archive Opens Up a Wealth of Possibilities"; Danny Fortson, "Google Gobbles Up Deja.com's Babble," Daily Deal (February 12, 2001); Michael Liedtke, "Web Search Engine Google Buys Deja.com's Usenet Discussion Archives," *Associated Press* (February 12, 2001).

[35.] Lyman, "Archiving the World Wide Web."

[36.] Miller quoted in Gardner, "Report on Documenting the Digital Age." For an overview of OAIS, see Brian Lavoie, "Meeting the Challenges of Digital Preservation: The OAIS Reference Model," 2000, http://www.oclc.org/research/publications/newsletter/repubs/lavoie243/; on EAD, see Daniel V. Pitti, "Encoded Archival Description: An Introduction and Overview," *D-Lib Magazine* 5, no. 11 (November 1999), http://www.dlib.org/dlib/november99/11pitti.html. OAIS comes out of NASA and the space data community, not the librarians. But they have embraced it.

[37.] Raymie Stata, "The Internet Archive" (paper delivered at the conference "Preserving Web-Based Documents," Washington, D.C., April 23, 2002). On deep versus surface web, see Lyman, "Archiving the World Wide Web"; Roy Rosenzweig, "The Road to Xanadu: Public and Private Pathways on the History Web," *Journal of American History* 88, no. 2 (September 2001): 548–79, also available at http://chnm.gmu.edu/assets/historyessays/e1/roadtoxanadu1.html. Kahle himself indicates many of

the problems and limitations of the Internet Archive in Brewster Kahle, "Archiving the Internet: Bold Efforts to Record the Entire Internet Are Expected to Lead to New Services" (paper presented at "Documenting the Digital Age," San Francisco, February 10–12, 1997), copy in possession of author.

38. On robots exclusion, see http://www.robotstxt.org/wc/exclusion-admin.html. Apparently, the IA will retroactively block a site without direct request, if it simply posts the robots.txt file. This would seem to mean that if someone took over an expired domain name, they could then block access to the prior content. There is some evidence, however, that the IA does not actually "purge" the content, it simply makes it inaccessible. For an intense discussion of these issues, see the hundreds of online postings in "The Wayback Machine, Friend or Foe?" *Slashdot* (June 19–20, 2002), http://ask.slashdot.org/askslashdot/02/06/19/1744209.shtml. For a pessimistic assessment of the legality of the IA's practices (though not explicitly directed at it), see I. Trotter Hardy, "Internet Archives and Copyright" (paper delivered at "Documenting the Digital Age," San Francisco, February 10–12, 1997), copy in possession of author.

39. Insiders have commented to me that the IA would disappear if Kahle left the project. But there are very recent signs that the IA is broadening its base of financial support.

40. For a recent, brief overview of these trends, see Naomi Klein, "Don't Fence Us In," *Guardian* (October 5, 2002).

41. Thomas Brown, "What Is Past Is Analog: The National Archives Electronic Records Program since 1968" (paper delivered at the OAH Annual Meeting, Washington, D.C., 2002), copy in possession of author. In 1997, Kenneth Thibodeau estimated that the NARA invested only token amounts (2 percent of its budget) in electronic records. Gardner, "Report on Documenting the Digital Age."

42. Committee on an Information Technology Strategy for the Library of Congress, et al., LC21, http://books.nap.edu/html/lc21/index.html; Rosenzweig, "Road to Xanadu."

43. "Background Information about PANDORA: The National Collection of Australian Online Publications," PANDORA, http://pandora.nla.gov.au/background.html; Cathro, Webb, and Whiting, "Archiving the Web"; Colin Webb, "National Library of Australia" (paper delivered at the DAI Institute, "The State of Digital Preservation: An International Perspective," Washington, D.C., April 25, 2002, available at http://www.clir.org/pubs/reports/pub107/contents.html). For British efforts to cope with digital materials, see Jim McCue, "Can You Archive the Net?" *Times* (London) (April 29, 2002). On Sweden and Norway, see Warwick Cathro, "Archiving the Web," *National Library of Australia Gateways* 52 (August 2001), http://www.nla.gov.au/ntwkpubs/gw/52/p11a01.html/.

44. There is anecdotal evidence that this is being seriously considered.

45. National Archives and Records Administration, Proposal for a Redesign of Federal Records Management (July 2002), 10, http://www.archives.gov/records_management/initiatives/rm_redesign.html; Richard W. Walker, "For the Record, NARA Techie Aims to Preserve," Government Computer News 20, no. 21 (July 30, 2001), http://www.gcn.com/vol20_no21/news/4752-1.html/; GAO, Information Management, 50. So far, POP remains, as a NARA staff member explained in April 2001, "beyond the state of the art of information technology." Adrienne M. Woods, "Toward Building the Archives of the Future" (paper delivered at the Society of California Archivists' Annual Meeting, April 27, 2001), accessed online May 1, 2002, but not available as of June 20, 2002. See also Kenneth Thibodeau, "Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years" (presentation at the DAI Institute, "The State of Digital Preservation: An International Perspective," Washington, D.C., April 24–25, 2002, available at http://www.clir.org/pubs/reports/pub107/contents.html). In June 2002, the GAO reported that, in general, NARA's electronic records project "faces substantial risks" and "is already behind schedule." GAO, Information Management, 3.

46. Amy Friedlander, "The National Digital Information Infrastructure Preservation Program:

Expectations, Realities, Choices and Progress to Date," *D-Lib Magazine* 8, no. 4 (April 2002), http://www.dlib.org/dlib/april02/friedlander/04friedlander.html.

47. The quote is often incorrectly attributed to Carl von Clausewitz. It could be that it is simply a reworking of Voltaire's remark that "le mieux est l'enemi du bien" (the best is the enemy of the good) or of George S. Patton's dictum, "A good plan violently executed now is better than a perfect plan executed next week."

1. Kahle, Prelinger, and Jackson, "Public Access to Digital Materials." See, similarly, Michael Lesk, "How Much Information Is There in the World?" an online paper at http://www.lesk.com/mlesk/ksg97/ksg.html.

49. McCoy, "Struggle to Establish a National Archives in the United States," 1, 12. Indeed, one digital preservation program—LOCKSS (Lots of Copies Keep Stuff Safe)—relies on precisely this principle: http://lockss.stanford.edu/.

50. Lee Dembart, "Go Wayback," *International Herald Tribune* (March 4, 2002), http://www.iht.com/cgi-bin/generic.cgi?template=articleprint.tmplh&ArticleId=50002; "Seeing the Future in the Web's Past," BBC News (November 12, 2001). See also Joseph Menn, "Net Archive Turns Back 10 Billion Pages of Time," *Los Angeles Times* (October 25, 2001): A1; Heather Green, "A Library as Big as the World," *Business Week Online* (February 28, 2002), http://www.businessweek.com/technology/content/feb2002/tc20020228_1080.htm. The dream of a universal archive is also the nightmare of privacy advocates. In the paper era, the physical bulk of personnel files and bank, criminal, and medical records made them more likely to wind up in landfills than in archives. Even when preserved, the possibility of retrospective prying (was your neighbor's grandfather a deadbeat or a drunk?) was reduced by the sheer tedium of sorting through thousands of pages of records. But what if sophisticated data-mining tools ("tell me everything about my neighbors") made such searching easy? Even the "public" material on the web poses ethical challenges for historians. "The woman who is going to be elected president in 2024 is in high school now, and I bet she has a home page," exclaims Kahle. The Internet Archives has "the future president's home page!" Perhaps. But it also has the home pages of many other high school students, at least some of whom are going through serious emotional turmoil that they might later prefer to keep from public view. Kahle himself wrote a prescient 1992 article, the "Ethics of Digital Librarianship," which worries about "types of information that will be accessible" as "the system grows to include entertainment, employment, health and other servers." Menn, "Net Archive Turns Back 10 Billion Pages"; Wood, "CNET's Web Know-It-All"; Kahle quoted in John Markoff, "Bitter Debate on Privacy Divides Two Experts," *New York Times* (December 30, 1999): C1. See also Jean-François Blanchette and Deborah G. Johnson, "Data Retention and the Panoptic Society: The Social Benefits of Forgetfulness," *Information Society* 18 (2002): 33–45; Marc Rotenberg, "Privacy and the Digital Archive: Outlining Key Issues" (paper delivered at "Documenting the Digital Age," San Francisco, February 10–12, 1997), copy in possession of author; "Wayback Machine, Friend or Foe?"

51. Mike Featherstone, "Archiving Cultures," British Journal of Sociology 51, no. 1 (January 2000): 178, 166. For examples of enthusiastic prophecy about such changes, see Francis Cairncross, The Death of Distance: How the Communications Revolution Will Change Our Lives (Boston, 1997); Kevin Kelly, "New Rules for the New Economy," Wired 5, no. 9 (September 1997), http://www.wired.com/wired/archive/5.09/newrules_pr.html. For a sober and sensible critique, see John Seely Brown and Paul Duguid, The Social Life of Information (Boston, 2000), 11–33.

52. The phrase comes from my colleague Randy Bass; see Bass and Roy Rosenzweig, "Rewiring the History and Social Studies Classroom: Needs, Frameworks, Dangers, and Proposals," *Journal of Education* (2000), available at http://chnm.gmu.edu/assets/historyessays/e2/rewiring1.html.

53. See, for example, Geoffrey J. Giles, "Archives and Historians: An Introduction," in *Archives and Historians: The Crucial Partnership* (Washington, D.C., 1996), 5–13, who writes that "there is too much

archival material for the archivists and for the historian to deal with" and notes feelings of "envy" of "ancient and medieval historians, who have so little material with which to work."

54. A. R. Luria, *The Mind of a Mnemonist: A Little Book about a Vast Memory*, Lynn Solotaroff, trans. (New York, 1968), Jerome S. Bruner, foreword, viii. See the similar, but fictional, account in Jorge Luis Borges, "Funes the Memorious," in *Labyrinths: Selected Stories and Other Writings*, Donald A. Yates and James E. Irby, eds. (New York, 1964), 59–66.

55. Linton Weeks, "Power Biographer," *Washington Post* (April 25, 2002): C01. Carl Bridenbaugh's derisive view of sampling provides a good example of the traditional view that historians should look at everything. "The Great Mutation," AHR 68, no. 2 (January 1963): 315–31, also available with other Presidential Addresses at AHA.org/info/AHA_History/cbridenbaugh.htm">http://www.theAHA.org/info/AHA_History/cbridenbaugh.htm. Nevertheless, historians have always struggled with the problem of how to deal with large numbers of sources. Even medievalists worry about how to make sense of the huge numbers of documents that survive from twelfth-century Italy. Still, the digital era vastly increases the scale of the problem.

56. Stellin, "Google's Revival of a Usenet Archive Opens Up a Wealth of Possibilities"; Hedstrom, "How Do We Make Electronic Archives Usable and Accessible?" (paper delivered at "Documenting the Digital Age," San Francisco, February 10–12, 1997), copy in possession of author.

57. To be sure, a number of key figures in digital archives and library circles (for example, Daniel Greenstein, Margaret Hedstrom, Abby Smith, Kenneth Thibodeau, Bruce Ambacher) have doctoral degrees in history, but they do not currently work as academic historians. Still, it would be logical for academic historians to build alliances with these scholars who have a foot in both camps. Thus far, academic historians have been much more likely to build ties to historians working in museums and historical societies than to those in archives and libraries.

58. It is difficult to prove a negative, but one searches in vain through the participant lists at key digital archives conferences for the names of practicing historians. One exception was the Committee on the Records of Government, which had a historian, Ernest R. May, as its chair and another, Anna K. Nelson, as its project director. But perhaps significantly, that committee had a mandate that dealt as much with paper as electronic records: Committee on the Records of Government, Report (1985). Another partial exception was the February 1997 conference "Documenting the Digital Age" sponsored by NSF, MCI Communications Corporation, Microsoft Corporation, and History Associates Incorporated, which included a few public and museum-based historians but only one university-based historian. Similarly, history journals have provided almost no coverage of these issues. Archivists are not reading historians, either. Richard Cox analyzed the almost 1,200 citations in 61 articles on electronic records management published in the 1990s and found only a handful of references to work by historians. Cox, "Searching for Authority."

59. Cox, "Messrs. Washington, Jefferson, and Gates." Robert Townsend, Assistant Director of Research and Publications, AHA, kindly supplied membership information. One imperfect but telling indicator of the changing interests of professional historians: Between 1895 and 1999, the American Historical Review published thirty-one articles with one of the following words in the title: archive or archives, records, manuscripts, correspondence. Only four of those appeared after World War II, and they were in 1949, 1950, 1952, and 1965. Some representative titles include: Charles H. Haskins, "The Vatican Archives," AHR 2, no. 1 (October 1896): 40–58; Waldo Gifford Leland, "The National Archives: A Programme," AHR 18, no. 1 (October 1912): 1–28; Edward G. Campbell, "The National Archives Faces the Future," AHR 49, no. 3 (April 1944): 441–45. For a good, brief overview of the AHA's active, early archive and manuscript work, see Arthur S. Link, "The American Historical Association, 1884–1984: Retrospect and Prospect," AHR 90, no. 1 (February 1985): 1–17. NARA's "Timeline for the National Archives and Records Administration and the Development of the U.S. Archival Profession," NARA_timeline.html">http://www.archives.gov/research_room/alic/reference_desk/NARA_timeline.html, highlights the role of the AHA. It should be noted, however, that the AHA has made a notable contribution to archival issues through its central role in the National Coordinating Committee for the Promotion of History (NCC), which was crucial, for example, in winning the independence of the

National Archives in 1984. The new National Coalition for History, which has replaced the NCC, has also made archival concerns central to its work. Access to archives and primary sources was, of course, a central preoccupation—indeed, an obsession—of early "scientific" and professional historians. See Bonnie G. Smith, "Gender and the Practices of Scientific History: The Seminar and Archival Research in the Nineteenth Century," *AHR* 100, no. 4 (October 1995): 1150–76.

60. Deanna B. Marcum, "Scholars as Partners in Digital Preservation," *CLIR Issues*, no. 20 (March/April 2001), http://www.clir.org/pubs/issues/issues20.html. "Scholars," warns the CLIR Task Force on the Artifact in Library Collections, "may not see preservation of research collections as their responsibility, but until they do, there is a risk that many valuable research sources will not be preserved." CLIR, *Evidence in Hand*.

61. I am indebted to Jim Sparrow for a number of the ideas in this paragraph. For detailed coverage of "How the Bellesiles Story Developed," see *History News Network*, http://hnn.us/articles/691.html.

62. House Committee on Government Operations, *Taking a Byte out of History*, 4. For the assumption of selectivity among archivists, see, for instance, Richard J. Cox, "The Great Newspaper Caper: Backlash in the Digital Age," *First Monday* 5, no. 12 (December 2000), http://firstmonday.org/issues/issue5_12/cox/index.html.

63. Abby Smith, *The Future of the Past: Preservation in American Research Libraries* (Washington, D.C., 1999), http://www.clir.org/pubs/reports/pub82/pub82text.html; Marcum, "Scholars as Partners in Digital Preservation"; Nicholson Baker, *Double Fold: Libraries and the Assault on Paper* (New York, 2001). Compare, for example, Cox, "Great Newspaper Caper," and O'Toole, "Do Not Fold, Spindle, or Mutilate," with Robert Darnton, "The Great Book Massacre," *New York Review of Books* (April 26, 2001), http://www.nybooks.com/articles/14196. In 1996, the Modern Language Association (MLA) issued a statement arguing "that for practical purposes, all historical publications, even those produced by mass-production techniques designed to minimize deviations from a norm, have unique physical qualities that may have value as a carrier of (physical) evidence in a given research project." CLIR, *Evidence in Hand*.

64. GAO, Information Management, 16; Cox, "Messrs. Washington, Jefferson, and Gates." Cox's article responded, in part, to an earlier article by Raymond W. Smock that argues, "historians should not rely on archivists alone to make decisions about what history to save or to publish." Smock, "The Nation's Patrimony Should Not Be Sacrificed to Electronic Records," *Chronicle of Higher Education* (February 14, 1997): B4–5.

65. Robert Darnton, Sarah A. Mikel, and Shirley K. Baker, "The Great Book Massacre: An Exchange," New York Review of Books (March 14, 2002), http://www.nybooks.com/articles/15195.

66. See, for example, Vincent Kiernan, "'Open Archives' Project Promises Alternative to Costly Journals," Chronicle of Higher Education (December 3, 1999); Budapest Open Access Initiative, http://www.soros.org/openaccess. On questions of public domain and privatization, see Lawrence Lessig, *The Future of Ideas: The Fate of the Commons in a Connected World* (New York, 2001).

67. Justin Winsor, "Manuscript Sources of American History: The Conspicuous Collections Extant," *Papers of the American Historical Association* 3, no. 1 (1888): 9–27, AHA.org/info/AHA_History/jwinsor. htm">http://www.theAHA.org/info/AHA_History/jwinsor.htm. On the central concern with teaching in schools, see Link, "American Historical Association, 1884–1984," 12–15.