

March 10, 2007**History, Digitized (and Abridged)****By KATIE HAFNER**

SALINAS, Calif.

THE National Steinbeck Center, at the top of Main Street in this farming community, exhibits an array of artifacts from John Steinbeck's life and works: family memorabilia, a passport from the 1960s and movie stills from "The Grapes of Wrath." Downstairs, in a climate-controlled vault, is the original manuscript of "The Pearl," his novella published in 1947. There is also an exuberant letter that Steinbeck wrote to a distant relative when he was a teenager, as well as rare footage of him on 16-millimeter film, introducing a 1961 movie, "Flight."

Steinbeck aficionados wishing to examine the manuscript of "The Pearl," which he wrote in pencil in small, precise handwriting on a yellow legal pad, have to travel here — after making an appointment with a part-time archivist, who is in on Mondays, Wednesdays and Fridays.

The center takes great care to preserve these relics of Steinbeck, a Nobel laureate, yet it has no plans to take the collection a step further, to adapt to a digital age. As a result, the manuscript of "The Pearl" is no more likely to be digitized than is the camper with the canine-motif curtains that Steinbeck immortalized in his book "Travels With Charley," and that is parked in perpetuity in the center's main exhibition hall.

These Steinbeck artifacts are not the only important pieces of history that are at risk of disappearing or being ignored in the digital age. As more museums and archives become digital domains, and as electronic resources become the main tool for gathering information, items left behind in nondigital form, scholars and archivists say, are in danger of disappearing from the collective cultural memory, potentially leaving our historical fabric riddled with holes.

"There's an illusion being created that all the world's knowledge is on the Web, but we haven't begun to glimpse what is out there in local archives and libraries," said Edward L. Ayers, a historian and dean of the college and graduate school of arts and sciences at the University of Virginia. "Material that is not digitized risks being neglected as it would not have been in the past, virtually lost to the great majority of potential users."

To be sure, digitization efforts over the last 10 years have been ambitious and far-reaching. For many institutions, putting collections online, for both preservation and accessibility, is a priority. Yet for every letter from Abraham Lincoln to William Seward that can be found online, millions of documents bearing fine-grained witness to the Civil War will never be digitized. And for every CD re-release of Bessie Smith singing "Gimme a Pigfoot," the work of hundreds of lesser-known musicians from the early 20th century are unlikely to be converted to digital form. Money, technology and copyright complications are huge impediments.

It is not for a lack of trying.

At the Library of Congress, for example, despite continuing and ambitious digitization efforts, perhaps only 10 percent of the 132 million objects held will be digitized in the foreseeable future. For one thing, costs are prohibitive. Scanning alone on smaller items ranges from \$6 to \$9 for a 35-millimeter slide, to \$7 to \$11 a page for presidential papers, to \$12 to \$25 for poster-size pieces. (The cost of scanning an object can be a relatively minor part of the entire expense of digitizing and making an item accessible online.)

Similarly, at the National Archives, the repository for some nine billion documents, only a small fraction are likely to be digitized and put online. And at thousands of smaller, local collections around the country, the bulk of the material is languishing on yesterday's media: paper, LPs, magnetic tape and film.

Strapped for money, archivists around the country are looking to private partners for help. Google has donated \$3 million to help start an effort led by the Library of Congress that will digitize and share materials around the globe, and has also provided technical resources for digitizing various printed materials at the library. Google, on its own, is digitizing books at the Library of Congress, which has its hands full with other items. And a number of other companies and foundations, including Reuters, I.B.M. and the Andrew W. Mellon Foundation, have financed digitization projects around the world.

Even with outside help, experts say, entire swaths of political and cultural history are in danger of being forgotten by new generations of amateur researchers and serious scholars.

Consider the Library of Congress archive of one million photo prints from The New York World-Telegram & Sun; only 5,407 have been digitized. Of the 1.2 million images from U.S. News and World Report, the library has digitized only 366. Its collection of five million images from Look magazine, spanning the period from 1937 to 1971, creates what Jeremy E. Adamson, director of collections and services at the library, calls "a fascinating portrait of America through photo stories on social and political subjects, personalities, food, fashion and sports." Yet only 313 of those images have been digitized.

"It's a crying shame," Mr. Adamson said, "as today's public is acutely visually literate and comfortable with pictures as a means to understand the past and experience for themselves the direct look and feel of history."

The reason for not digitizing these collections? "Not enough money," Mr. Adamson said.

THE decision to put off digitizing a significant collection is seldom easy, archivists at the Library of Congress say. Plans to digitize The National Intelligencer, a newspaper published in Washington during much of the 19th century and filled with Colonial script not easily recognized by digitizing equipment, eventually had to be put on hold because of the high expense.

"If researchers conclude that the only valuable records they need are those that are online they will be missing major parts of the story," said James J. Hastings, director of access programs at the National Archives. "And in some cases they will miss the story altogether."

Maritime buffs, for example, hoping to use the Internet to piece together the story of the Silenus, one of the finest ships ever built in North America, will find a spotty narrative. The papers of its captain, Joseph King, who lived a brief but adventurous life, from 1782 to 1806, can be found courtesy of the Mellon Foundation, in a digitized archive from the Mystic Seaport's collection. Researchers will see how much Captain King paid for "1 potte lijn oli" in 1803, when the ship was in the Netherlands.

What they will not see is that two years after Captain King's death, at the Cape of Good Hope, the ship itself was advertised for sale on May 4, 1808, in Calcutta. This clue remains paperbound, on the front page of The Asiatic Mirror, an English-language newspaper published in Calcutta during that era, whose only known remaining copies now reside in large bound volumes in a remote storage room outside Washington. The relative obscurity of the newspaper, and its odd size, make it impractical to digitize.

A Google search will pick up the next chapter of the story at the Princeton University's special collection, which includes the papers of James and Dolley Madison. It reveals that in 1817, President Madison signed over the ship's papers to William Gallup.

"The story of what happened to the good ship Silenus between 1806 and 1817 will never be complete," said Mr. Adamson of the Library of Congress, "but what happened in 1808 in Calcutta is the kind of little crumb that can be picked up and become a significant research item."

The ultimate fate of information relating to potentially valuable but obscure people, places, events or things like the Silenus highlights one of the paradoxes of the digital era. While the Internet boom has made information more accessible and widespread than ever, that very ubiquity also threatens records and artifacts that do not easily lend themselves to digitization — because of cost, but also because Web surfers and more devoted data hounds simply find it easier to go online than to travel far and wide to see tangible artifacts.

"This is the great problem right now, and it's a scary thing," said the documentary filmmaker Ken Burns. "The dots are only connected by a few of us who are willing to go to the places to make those connections."

In its digitization efforts, the Library of Congress is focusing mainly on special collections, hewing to a philosophy that it should be digitizing objects that cannot be seen elsewhere. There are the obvious things, like the papers of Washington, Jefferson and Lincoln. And then there are the Farm Security Administration's collection of photographs from the Depression, and a set of mounted photographs of the America's Cup yacht race since the 1890s.

Elizabeth S. Dulabahn, a senior manager at the Library of Congress who oversees part of the library's digitizing effort, said the library was examining closely the behavior of those who use its Web site.

"We're trying to do a better job of understanding the kinds of information that people are looking for on the Web, and the kinds of searches that bring users to the library's site," she said. She cited Women's History Month and the centennial of the first Wright Brothers flight as "examples of events of interest to a broad constituency."

The Library of Congress and other archives are creating indexes that refer to the contents of a physical collection, in the hope that they will entice researchers away from their computers.

But the reality remains that a new generation of researchers prefers to seek information online, a trend made all too clear to Mr. Hastings of the National Archives last year, after Google, in an experiment of sorts, digitized 101 of the National Archives' films — including World War II newsreels and NASA footage — and put them up on its site, at video.google.com/nara.html.

"Before that happened, we had 200 requests total for the whole year in our research room," Mr. Hastings said. "The first month the films were available on Google, there were about 200,000 hits on them — a thousandfold increase."

In some cases, strange bedfellows have conspired to help solve the problem.

Over the years, the New Orleans Public Library has steadily been digitizing its photographs, but its documents have gone largely untouched. The collection, which rivals the holdings of many university special collections, contains millions of historical documents, going back to 1769 and the Spanish colonial era.

The records survived Hurricane Katrina unscathed, but are still at risk for damage and loss, said Irene Wainwright, an archivist at the library.

"I can't tell you how many people have suggested to us, 'Oh, you just need to digitize all that stuff down in the basement and you'll be all right,'" Ms. Wainwright said. "They have no idea how much effort that requires."

Enter the Genealogical Society of Utah, an organization financed by the Mormon Church, for whom the search for ancestors is a core mission. The society has embarked on a three-year, \$200,000 project to digitize all of the library's genealogically relevant records from 1805 to 1880 (www.familysearch.org).

"The records we gather document the lives of people," said Wayne J. Metcalfe, vice president of the society. "Births, christenings, land records and other documents that provide information about individuals who have lived on the earth."

To that end, genealogy experts affiliated with the Church of Jesus Christ of Latter-day Saints are fanning out, digital cameras in hand, making copies of genealogically relevant records in 200 cities around the world, including New Orleans. Over the next five years, the church expects to have hundreds of millions of digital images available.

Mr. Metcalfe said economies of scale helped his organization bring down the cost of capturing each image to roughly 20 cents — far less than what a commercial company might charge.

Similarly, I.B.M.'s digitization efforts — dating to the mid-1990s, when the company converted a healthy chunk of the Vatican Library's archives — are done in a way to benefit the company as well as the institution looking to digitize its holdings.

"We look for projects that will highlight I.B.M.'s most innovative technologies or help us develop those technologies with very specific partners who have a problem to solve," said Paula Baker, vice president for global community initiatives at I.B.M. The company looks for projects that require the newest technology.

Such is the case with its most recent multiyear, multimillion-dollar project: a virtual version of the vast Forbidden City in Beijing, which I.B.M. is building in partnership with China's Ministry of Culture. When it is finished, early next year, the site will include interactive, three-dimensional images of ancient thrones, artwork and military implements.

Ms. Baker added that each time I.B.M. embarks on a new venture, requests start coming in from other institutions in need. "When we do these projects everyone else comes out of the woodwork," she said. "But we have to be very selective."

Donald J. Waters, program officer for scholarly communication at the Mellon Foundation, said his foundation had also become increasingly selective over the years.

By way of example, Dr. Waters pointed to the papers of Matthew Parker, the archbishop of Canterbury in the 16th century who collected ancient manuscripts to prove the early existence of an independent English-speaking church that was responsible not to the pope but to the king of England. For centuries, those papers have been locked up at Corpus Christi College at Cambridge University. Mellon is financing a project to put them online.

"It takes a special skill to select stand-alone collections that have a durable appeal in the marketplace of scholars, which is the marketplace that Mellon cares most about," Dr. Waters said. "As interesting and as important as standout collections in individual libraries and archives might be, the mere fact of digitizing them does not mean that once they are online they will attract and sustain an audience."

The Parker collection, Dr. Waters said, meets all these criteria — it is a core collection for a variety of fields: linguistics, ecclesiastical and religious history, English history, art history, medieval studies. He added, however, that the materials have a long history of restricted access, largely to protect the materials because they are so important.

"Digitization would allow much broader access to the contents," he said, "which is sufficient for much research, without exposing the physical manuscripts to added handling."

WHILE copyright is not a concern for those digitizing documents that are hundreds of years old, copyright restrictions play a significant role when it comes to modern material. Even if the Steinbeck Center in Salinas were to find the money to digitize, say, the manuscript of "The Pearl," its copyright would limit its distribution.

"At this point, online materials are best for authors no longer under copyright," said Susan Shillinglaw, a professor of English at San Jose State University and scholar in residence at the Steinbeck Center.

When Leonard Bernstein's family donated the composer's papers to the Library of Congress in 1993, it was with the goal of digitizing portions of the collection and making them broadly accessible. Although more than a thousand items from the collection have been digitized and placed on the library's Web site, there is still an enormous quantity of material that, because of sheer volume and copyright concerns, is still accessible only to researchers who travel to the library.

For instance, the collection includes a seven-page letter that Jacqueline Kennedy wrote by hand to Bernstein at 4 a.m. on June 8, 1968, the day after the funeral for Robert F. Kennedy, thanking him for conducting Mahler's Requiem during the ceremony. The letter is an extraordinary window into her grief: "Your music was everything in my heart, of peace and pain and such drowning beauty," she wrote. But the library would

need permission from the estate of Mrs. Onassis to digitize it.

When it comes to sound recordings, copyright law can introduce additional complications. Recordings made before 1972 are protected under state rather than federal laws, and under a provision of the 1976 Copyright Act, may be entitled to protection under state law until 2067. Also, an additional copyright restriction often applies to the underlying musical composition.

A study published in 2005 by the Library of Congress and the Council on Library and Information Resources found that some 84 percent of historical sound recordings spanning jazz, blues, gospel, country and classical music in the United States, and made from 1890 to 1964, have become virtually inaccessible.

"Copyright is a very blunt instrument," said Tim Brooks, the author of "Lost Sounds: Blacks and the Birth of the Recording Industry, 1890 to 1919" (University of Illinois, 2004). "Once you have copyright, you have total control; there's very little room in the copyright law even for preservation, much less reissuing material."

Generally, rights owners like Sony BMG have reissued on CD only a small portion of the recordings they control.

For example, John Philip Sousa's own band made scores of recordings for Victor Records in the early 20th century. BMG bought Victor in 1986, and few if any of those recordings have since been reissued on CD. "There is probably an odd track out somewhere," Mr. Brooks said, "but they've certainly never done any kind of retrospective of him that I'm aware of." And of the hundreds of recordings made in the same period by Noble Sissle, an African-American tenor who recorded for several labels now owned by Sony BMG, few if any have made it onto CD.

THE result, Mr. Brooks said, is a series of gaps in the popular understanding of the nation's musical heritage. "It's as if before Bessie Smith, there was nothing," he said. "It has the effect of narrowing our own understanding of our own history."

Another factor that determines what is digitized is how straightforward it is to copy the material.

In some cases, said Theresa Salazar, curator of Western Americana at the Bancroft Library at the University of California, Berkeley, the two go hand in hand. "Agencies and organizations providing funding often want large volume for their money," Ms. Salazar said.

For example, she pointed out, objects like books can be handled in a straightforward way. It is easy to capture these materials because they are printed, and many of these titles are more or less the same size.

No one knows this better than Google, whose digitization efforts focus mainly on books.

In its quest to scan every one of the tens of millions of books ever published, Google has already digitized one million volumes. Google refuses to say how much it has spent on the venture so far, but outside experts estimate the figure at at least \$5 million. The company has also been scanning and indexing academic journals to make them searchable, and is working with the Patent Office to digitize thousands of patents dating back to 1790.

David Eun, Google's vice president for content partnerships, said that rather than dwell on what is being left behind, he preferred to take a more optimistic view.

"We're talking about a huge, huge universe of content," Mr. Eun said. "If you look at the glass as half-empty it becomes too overwhelming."

Copyright 2007 The New York Times Company

[Privacy Policy](#) | [Search](#) | [Corrections](#) | [RSS](#) | [First Look](#) | [Help](#) | [Contact Us](#) | [Work for Us](#) | [Site Map](#)