Mass Digitization

Implications for Preserving the Scholarly Record

By Trudi Bellardo Hahn

Libraries and archives have a critical role in preserving the scholarly record; many players in the publication cycle depend on them for this. Preservation of scholarly books that are being digitized has lagged far behind preservation initiatives for electronic journals. The issue has become more critical, as large commercial companies such as Google, Yahoo, and Microsoft have begun mass digitization of millions of books in research libraries. Since December 2004, the pace of developments has been rapid, involving great risks on Google's part over the copyright issue. Google and certain participating libraries have not addressed the issue of whether or not all this effort to digitize huge numbers of books indiscriminately will serve students' and scholars' needs in the long run. Quality, secrecy, and long-term stability are all issues that suggest it may be foolish to expect that commercial companies will share librarians' values and commitment to digitized material preservation. The information profession must exert strong leadership in setting policies, standards, and best practices for long-term preservation of the scholarly record.

Libraries and archives that serve the scholarly community have a solemn responsibility to preserve the scholarly record. What these institutions do (or fail to do) will have an impact on all the players in the arena of what has been called the "publication cycle." The players in the cycle include publishers, editors, reviewers, librarians, archivists, readers, and, of course, scholars themselves. Converting and preserving scholarly materials are generally seen as the last steps in the cycle—if the cycle can be said to have an end. The experts in converting scholarly materials from paper or other tangible materials to digital formats, and in preserving those digitized documents, are only a small subset of this large community, and I am not one of those technical experts. Nonetheless, I am among the stakeholders in the community affected. My observations are from the perspective of an informed, objective, and concerned eyewitness to current developments.

Side-stepping the issues and developments surrounding digitized and born-digital journals, I will focus on the programs for mass digitizing of books and other, nonjournal scholarly materials by such companies as Google, Yahoo, Microsoft, and others. Interestingly, in the past, these companies were never considered part of the scholarly publication community—a fact that makes their abrupt and explosive entrance onto the scene not only unexpected, but also unsettling.

My issues and concerns are organized into five areas: pace of developments, foolish risk versus vision, justification for digitizing books, trust, and leadership. Each of these has implications for preservation and long-term access to digital documents. Preservation and access go hand-in-glove, but they are not the same. Most of my observations focus on Google; they are by far the biggest and most

Trudi Bellardo Hahn (thahn@umd.edu) is a visiting professor, College of Information Studies, University of Maryland, College Park.

This article is based on a presentation given at "Converting and Preserving the Scholarly Record," Eighth Annual Symposium on Scholarly Communication, State University of New York, Albany, October 24, 2006.

Submitted February 12, 2007; tentatively accepted pending revision March 31, 2007; revised and resubmitted April 7, 2007, and accepted for publication.

controversial player—the eight-hundred-pound gorilla. I will mention activities of Yahoo and Microsoft as well; the implications for preservation are similar.

Pace of Developments

Is this all happening too fast? Digitizing library books and making scholarly collections available on the Web have been around for more than a decade. Since the commercial world, in the form of such deep-pocket companies as Google, Yahoo, and Microsoft, has come into the academy, the pace has sped up enormously. Is the pace too fast to make good policy? Is it too fast to ponder and debate difficult issues and make decisions that will benefit all of us in the long term?

Some of us are still digesting Google's startling announcement in December 2004 that it will be working with five major research libraries to digitize more than fifteen million books from their collections in exchange for providing these libraries with digital copies of their books. Google will load the copies into their own digital library and make full-text versions available if they are in the public domain, or brief excerpts—snippets—if they are still under copyright protection. The project promises to cost millions of dollars—perhaps as much as a billion—and to take six to ten years to complete. Previously, the libraries involved had thought that such a digitization project would take far longer—when library staff at the University of Michigan were asked in 2004 how long it would take to digitize Michigan's seven million volumes, "the answer was more than a 1,000 years."[1] The project was—and still is—staggering in its speed and daring.

In the two years since that announcement, a rapid succession of announcements from Google and other organizations astonished us with the scope and potential for enormous impact:

- The *Seattle Times* reported on October 3, 2005, that Yahoo and Microsoft would team up with the nonprofit Internet Archive as well as several other large research libraries and archives in establishing the Open Content Alliance (OCA) to pool the collections of a large number of research libraries.[2] According to the article, OCA will only digitize those materials in the public domain, including handwritten manuscripts, unless copyright holders give explicit permission to digitize. The complete books will be freely available in a permanent archive. Funding and support will come from Yahoo and Microsoft as well as from participating libraries and other companies, such as Hewlett-Packard Labs, LibriVox, Octavo, Lulu.com, and Adobe. It appears that everybody wants to get in on the act! That Yahoo and Microsoft jumped on this bandwagon is not surprising. Yahoo is Google's archrival, and Microsoft's share of the searching market is growing all the time—it is, after all, "the default search engine built into the default Web browser available right out of the computer box."[3]
- The *Financial Times* (London) announced on November 4, 2005, that Microsoft is investing in a digitization project of 100,000 books from the British Library.[4]
- James H. Billington, Librarian of Congress, wrote an article for the *Washington Post* on November 22, 2005, about the Library of Congress (LC) receiving $3 million from Google to jump-start their digital archive of international cultural artifacts, the World Digital Library.[5]
- On March 7, 2006, the *Australian* reported that the European Commission (EC) plans to make at least six million books, documents, and other cultural works available by 2010. The EC will contribute $72 million to the digital library, and expects member states to make up the remaining $250 to $300 million to complete the project. Its goal is to combat other digitization projects that have an Anglo-American–centric view of history. The EC says it is not going after copyrighted works, but does not reveal details of the program, which has publishers worried that it might affect their own digital preservation programs.[6]
- The *Boston Herald* reported on June 15, 2006, that simultaneously with the Shakespeare in the Park festival in New York City, Google is launching a Web site that allows users to search all of Shakespeare's plays and poems—which, of course, are in the public domain.[7]
- A *New York Times* article on August 9, 2006, reported that the University of California would join the Google project, adding millions of books from the system's one hundred libraries. The digitization program will include copyrighted works. Google is talking to other libraries as well.[8]
- According to the *Milwaukee Journal Sentinel*, October 13, 2006, the University of Wisconsin has jumped on the Google bandwagon.[9]
- A few days later, the *Financial Times* (London) reported that Microsoft has a new partnership to scan books from Cornell University's library. Microsoft already has partnerships with the British Library and other library members of the OCA.[10]

As Van Orsdel and Born observed, perhaps the good news in all of these announcements is that book digitization projects have taken over the spotlight in the past two years and upstaged the serials crisis.[11]

Foolish Risk . . . or Vision?

We should be grateful to Google for sticking out its neck—for pushing the envelope on technological innovations, copyright, and other important aspects of digitization. At a symposium at the University of Michigan in March 2006, Google's Adam Smith said we need to "just do it" and "not let perfection be the enemy of the good," and that we need to "get it out there"—learn from mistakes, iterate the process, and make it better.[12]

On the other hand, Yahoo, Microsoft, the Library of Congress, and the OCA are staying in the background, which may be a good place to be. It not only is safer, but their smaller programs permit experimentation and policy setting to be done with older materials and those materials not under copyright. They are learning a lot, and introducing technological innovations with projects that do not risk lawsuits.

In regard to copyright, Google, publishers, and the participating universities all agree on the fundamental issue that intellectual property laws should be respected. Nonetheless, Google is being subjected to numerous lawsuits in the United States, France, Germany, and elsewhere because some publishers disagree on whether Google is infringing on copyright. Google is taking an extremely aggressive stance on copyrighted materials, insisting on an opt-out model that requires authors and publishers to contact Google and tell them they do not want their books included. Google says that opt-out is much easier, cheaper, and quicker than opt-in because Google would have to contact millions of copyright holders before even deciding which books they could digitize. Further, a large percentage of copyright holders would be virtually impossible to reach. This is one manifestation of the orphan works (copyrighted works whose owners may be impossible to identify and locate) problem.[13]

Authors and publishers say that Google is looking at this only from Google's perspective. What if a lot of companies and organizations—not just Google—get into large-scale digitization? That would put a big burden on the copyright holders who want to opt-out but might not even be aware that their books are being digitized.

Justification for Digitizing Books

Electronic journals and digitized versions of older print journals have become firmly established in research libraries' collections. Why digitize books as well? Twenty-first-century scholars are increasingly bypassing books—looking for background information in print library collections may slow down the scholar who wants to be productive. Even scholars in the humanities and social sciences are looking to their colleagues in the sciences, modeling their behavior after them because all scholars want to save time and be more productive. Initially, historians were hostile to JSTOR (a trusted archive of scholarly journals), but now most find it extremely helpful in their research.

College students use books and journals (at least if they have been trained to do so; otherwise, they simply use a search engine to find information on Web sites). For both students and scholars, however, the book is becoming increasingly irrelevant for learning and discovery.

Looking back a few centuries provides a perspective on how the pace of change is forcing us radically and rapidly to rethink our assumptions about scholarship. The transition from an oral to a written culture developed over many centuries. As Bengston said,

> During this slow evolution, our way of thinking fundamentally changed, from repetitive, oral, memory-based knowledge to visual and spatial memory, based on the physical object of the book. For centuries books were simply the most efficient and usable technology for the transmission of culture and ideas. We need only reflect on the past few years to sense how quickly and radically the ways that we write and communicate have been and will be altered.[14]

What do modern scholars and students really want or need? Have we factored their rapidly changing needs, preferences, and habits into our preservation programs? Predicting what, exactly, will happen to print books or even e-books in this century and beyond is impossible. Many people are confident that certain kinds of books "will cease to exist on paper: directories, reference works, textbooks, travel guides, to name a few."[15] No one can say, however, how much scholars and students will care about linear, narrative, book-length treatments. We do not know how much generations to come will care about preserving *words*, compared to visual and multimedia documents or even raw data. The only thing we may be certain of is there will be a tidal wave of interest in networked, digital media. Are our preservation programs responding to those trends?

For some time now, libraries have paid attention to cooperating in digitization projects that focus on unique collections as a cost-efficient way to give scholars all over the world access to rich resources and to preserve those valuable print materials that were deteriorating. Just a little more than two years ago, Brian Lavoie and Lorcan Dempsey at OCLC Research admonished libraries to be very careful and wise in allocating their insufficient budgets for preservation.[16] Accordingly, OCA members are carefully selecting which materials to contribute. Google's general approach, on the other hand, has been to throw a lot of money at the problem and grab as many books as they

can without selecting particular parts of collections. Daniel Greenstein, director of the California Digital Library, was quoted in January 2006 as saying that his discussions with Google officials disclosed that they are "more interested in grabbing a large quantity of materials than in carefully selecting certain collections of works."[17] Their attitude is simply, "the more of it, the better."[18] This approach is not true at all universities participating in the Google project, but it is a general strategy.

The question is whether a selective collection policy for digitization and preservation is better than a scattershot approach. It appears inevitable that the gems of our collections are not going to be exploited unless we digitize them. But should we be aiming for digitizing everything as fast as possible, so that we can provide what Mary Ann Coleman, president of the University of Michigan, referred to as "instant gratification of a one-in-a-million need," or should we be taking a more measured approach that addresses the most likely and important needs now and in the future?[19] I propose that we need to think more carefully about preservation priorities and match them to the norms of twenty-first-century scholarship. We need to spend our scarce resources for those digitization activities that not only will increase access, but will serve our long-term preservation goals as well.

Trust

At a symposium titled "Scholarship and Libraries in Transition: A Dialogue about the Impacts of Mass Digitization Projects," held at the University of Michigan on March 10–11, 2006, Clifford Lynch, executive director of the Coalition for Networked Information, was the wrap-up speaker.[20] He proposed that digitization is a form of insurance—in fact, one of the best forms of insurance we have. He said it is not a replacement for the physical object, but increasingly a good (albeit not perfect) surrogate. But is it really? If a foreign army came marching through your town, would you be preserving documents by tossing them into the hayloft? They would be out of the way of the marauders, yes, but they still would be subject to thieves stealing them, mice nibbling away at them, and rain leaking through the roof. Preservation is much more than finding a compact, convenient, and inexpensive place to stash materials.

The academy has enduring values and standards of preservation. Every academic library has in its mission statement something about archiving, conserving, or preserving the scholarly record for perpetual access. For example, on the University of Maryland Libraries' Web site is their mission statement: "Providing access to the use of the scholarly information resources required to meet the education, research and service missions of the University. The Libraries support this effort by building, organizing, maintaining and preserving these resources."[21]

Another way to express this statement of values comes from Coleman. In a speech to the Association of American Publishers in February 2006, she said, "General Motors does not need to maintain the tools for its 1957 Chevys, and would have a hard time manufacturing a car from that year. But a university is responsible for stewarding the knowledge of 1957, and for all the years before and after—the books and magazines; the widely known research findings and the narrow monographs; the arcane and the popular."[22]

Given that academic libraries accept a staggering responsibility with limited resources to meet that responsibility, they need to win people's trust that they will fulfill their mission to preserve evanescent digital materials.[23] They also need Google and other commercial enterprises as valued allies and partners. Karen Wittenberg, director of Columbia University's Electronic Publishing Initiative (EPIC), affirmed libraries' dependence on the for-profits—"We need to face the fact that commercial search engines are now the mechanism of choice for finding information, and we desperately need Google and other powerful players as valued partners with whom we will negotiate effective ways of collaborating that benefit our businesses and our users."[24]

Nevertheless, librarians and archivists need to be careful. They may get chummy with the staff of the for-profits—their employees are awfully friendly people, and how can you dislike a company such as Google, whose official motto is "Don't be evil?"[25] But we should never think of Google, Microsoft, or Yahoo as one of us. Google has a corporate mission "to organize the world's information," but it is for the goal of building and sustaining a massive and highly profitable media empire.[26] Google may sincerely believe that it operates according to higher principles, but some of its recent actions, such as its decision to abide by political restrictions placed on it by the Chinese government, prove that it is willing to compromise its principles in order to stay competitive.[27] A lot of money is to be made in digitizing books—when the content moves from physical to digital, its value jumps enormously. When there is money to be made, libraries and archives should be vigilant and alert. Too much chumminess with commercial enterprises raises three basic problems or issues related to trust: quality, secrecy, and long-term stability.

Quality

Quality is a serious issue in preservation that involves poor optical character recognition (OCR), poor originals that result in poor reproductions, missing pages, truncated text, and damage to the materials being digitized.

Is mass digitization preservation? Yes or no? Apparently no consensus exists, even among the representatives of the

Google 5 (as of January 2007, the Google 7 and expanding). Dale Flecker, associate director for Systems and Planning at the Harvard University Library, insists that the Google project is not planned as a preservation project; mass digitization is really only about providing access.[28] The attitude of the University of Michigan's administrators, however, is more complex. On the one hand, they acknowledge the seriousness of the preservation problem. For example, Coleman reported that Michigan was one of nearly 3,400 institutions that took part in the massive Heritage Health Index, which assessed how well our cultural institutions are tending to some 4.8 billion artifacts—the majority of which are books held at libraries.[29] Coleman said that the findings that came out in December 2005 were discouraging, and she warned, "As a country, we are at risk of losing millions and millions of items that constitute our heritage and our culture, because of a lack of conservation and planning. . . . So conservation efforts are paramount."[30] Michigan's response has been to create digital copies of works that are at-risk, out of print, or languishing in warehouses, an effort speeded up enormously because of the Google program. John Price Wilkin, University of Michigan associate university librarian, affirmed that Michigan thinks of it as a preservation project.[31] Even at Michigan, however, preservation and conservation staffs handle delicate materials that they feel are too fragile to scan. Michigan is not trusting the mass digitization program to protect their most vulnerable and valuable materials because they know that it does sacrifice quality.

Other Google partners have conceded quietly that the overall quality of the scans has not been great. Andrew Herkovic of Stanford University Library was quoted in an article by Helm in *Business Week Online* saying "Google has never pretended to knuckle under to quality demands that [preservationists] hope for."[32] In the same article, Sidney Verba, director of the Harvard University Library, said, "We at Harvard do a more careful and high-quality digitization when we do it for our own purposes, there's no question."[33] There is a question, however, whether Harvard is duplicating Google's digitization efforts. We also should ask why Google is not adhering to preservation standards when scanning.

Most of the institutions participating in the Google project concede that the main benefit of this project is not preservation, it is *access*—especially to students and scholars who would never otherwise be aware of the content of these books. If the quality is not good enough to read online, the hope is that the users will go to the library and find the original book. Given what we know about twenty-first-century behaviors among students and scholars, however, do we believe that very many of them will seek information beyond what they can find on their desktops? Deanna Marcum, associate librarian for library services at the LC, eloquently portrayed a scenario of the college student who, working from a "cozy, computer-equipped dorm room" can ignore the library completely and write a term paper—albeit with some questionable resources—entirely based on resources found online through a commercial search service such as Google.[34]

In any case, the price is right—it is pretty much free. Herkovic at Stanford was quoted in Helm's article, "If we were paying for this, if we were driving the [quality specifications], they would be different from what Google is offering."[35] Adam Smith, product manager of Google Books, responded in the same article by saying that "the primary goal right now is to put as much content online as possible, and address problems later."[36]

A news item in September 2006 reported that Google is turning to the greater engineering community for help improving the OCR technology it needs to index and archive books.[37] The technology Google is currently using is highly accurate at reading Latin characters, but still has trouble with other languages, handwriting, highly stylized fonts, smudged print, scientific treatises, and unique layouts. Google also has had problems with blurry or off-center scans that can confuse OCR engines and prevent the deciphering of a document's letters and words. Those pages would, therefore, not be indexed. At least Google is admitting that it has a problem, and one hopes that improvements in OCR and scanning will benefit all of us.

In the meantime, it appears that at least some materials being scanned will have to be scanned a second time—a waste of precious resources. Some researchers have suggested that preservation could end up costing much more than the original digitization of the books. If Google, Yahoo, or Microsoft have answers to the tough questions surrounding preservation of digital documents, they have not announced them or published them yet. It seems safe to assume that libraries and archives must accept that the responsibility for preservation is still theirs. It is, therefore, vital for all of us to know what the libraries participating in mass digitization intend to do.

Secrecy

Google in particular has been secretive—even cultivating an aura of mystery—about such things as their own high-speed book scanner. They refuse to divulge details of how it works or how fast it scans books. Google also does not say how many books it has scanned so far, or which books have been scanned.

Long-term Stability

Did anyone see a headline in a recent *Wall Street Journal* that read "Google Files for Bankruptcy?" The news item continued:

> Under the weight of too many lawsuits, rapid overextending of services (now more than twenty-nine different services), and mismanagement of its staggering empire, Google today filed for bankruptcy protection while it continues its operations. The chief executive of Google, who was recently appointed to the board of Apple Computer said, "This regrettable action became necessary only recently when good faith efforts to resolve outstanding debt with a creditor from the company's earliest days broke down." A spokesperson for Google declined to name the creditor.

Did anyone see that shocking announcement? No—of course not—I made it up, and it is nonsense. Google has been for some time the number one search engine in the United States and Europe, and probably everywhere else in the world. Its market share is well ahead of Microsoft, Yahoo, and Ask.com. It has amassed nearly $10 billion in cash.

However, a real news story appearing February 20, 2006, in the *Edge Singapore* revealed that Google shares had dropped nearly 25 percent as the company has grappled with growing competition from Microsoft and Yahoo, and "there could be a lot more tumbling ahead" because the stock prices do not reflect what the company is worth.[38] Google was facing increased pricing pressures on its online ad sales and mounting concern about what is known as click fraud as well as other challenges, such as lawsuits from newspaper and book publishers. It is not out of the realm of possibility that Google could shrink, redirect its mission, or even disappear altogether in the coming decades. These were the fates of other giants of American industry, such as Chrysler, IBM, and AT&T.

Another real headline appearing on October 6, 2006, in the *Washington Post* read, "Google Seeks Info from Book Scanners."[39] According to the news item, "Google Inc. has issued subpoenas for detailed information about its rivals' book-scanning projects as part of its defense against lawsuits attacking its own plans to put the contents of entire libraries online."[40] The article noted that the subpoenas were sent to Yahoo Inc, Microsoft Corporation, the Association of American Publishers, HarperCollins Publishers Inc., Bertelsmann AG's Random House Inc., and Holtzbrinck Publishers LLC. A similar request was also sent to Amazon.com Inc. The subpoenas included a request for "documents detailing every book the companies have made available online or plan to by the end of 2009"—the details are to include "lists of all authors, publishers, copyright holders and copyright status of each book scanned" as well as "all contracts or communications with publishers, copyright holders and libraries."[41] Does this tell us that Google is a little nervous? As of this writing, the targets of the subpoenas had refused to cooperate with Google; apparently, they feel the request for information is an attempt to capture trade secrets.[42] What would happen to mass digitization projects in research libraries if Google did collapse? Or if its stockholders decided that book search was a money loser and should be discontinued?

Some promising developments are appearing in the area of electronic journal preservation. Portico, developed by JSTOR and its partners, takes the "trusted third-party" approach. LOCKSS (Lots of Copies Keep Stuff Safe) distributes the task of preservation through local caching of subscriptions.[43] One step further is CLOCKSS (Controlled LOCKSS), a not-for-profit network of institutions, including OCLC, LC, other research libraries, as well as many publishers and learned societies.[44] The mission of CLOCKSS is to develop "a distributed, validated, comprehensive archive that preserves and ensures continuing access to electronic scholarly content."[45] Stemper and Barribeau provide a comprehensive review of all the efforts being made to preserve electronic journals.[46] These initiatives, however, while offering assurances that e-journals will be accessible far into the future, have not yet addressed the problem of preserving digital books. We know that libraries and archives have an avowed firm commitment to long-term preservation of and access to materials. As long as the major funders of our digitization efforts are commercial enterprises, however, can we count on sustainable access over the long term?

Leadership

A June 2004 report from the Association of Research Libraries endorsed digitization as an "accepted preservation reformatting option for a range of materials."[47] The report conceded that "ensuring high-quality image capture and providing for the long-term viability of digital objects is an admitted challenge."[48] The information professions, however, must take the leadership in developing standards and best practices, including developing "strategies to keep master files safe for the short-term, [which includes] the use of high-quality and reliable storage media, multiple back-up systems, periodic testing, and a schedule to refresh data."[49] These short-term strategies will at least keep the materials safe—safer than in the hayloft—while long-term solutions are being developed. This is the proper leadership role for librarians and archivists. Are we up to it, or will we let the eight-hundred-pound-gorilla companies drive the agenda and set the priorities? More specifically, are we even concerned that the gorillas are not dealing with preservation? A search in Lexis-Nexis for articles in the general and business news sections uncovered hundreds of hits on the topic of "(Google OR Yahoo) AND digitization AND books." However, as soon as the word "preservation" was introduced

into the search string, the count dropped to zero. Shouldn't that worry us?

We need to stop being reactive; we need to go after the preservation target in a strategic way. We *own* this problem of preserving books . . . at least for now. Ironically, many librarians were unhappy with a 2005 OCLC report because one of their key findings was that in the public's eye, the library brand is books.[50] That finding is troubling if people think of libraries as *only* about books. But we will be in much more trouble if our users stop thinking even that. Do you want a book? Go to Google or Yahoo or Amazon.com. Where will libraries be then? No brand recognition at all!

A news item on October 24, 2006, reported that Google, in partnership with the Frankfurt Book Fair literacy campaign and UNESCO's Institute for Lifelong Learning, is launching an online portal to connect literacy organizations.[51] In addition to allowing "organisations, teachers and others with an interest in literacy to search online for and share literacy information," the tool provides a zoomable, searchable map that enables users to locate literacy organizations around the world.[52] Searchers could find information in academic articles and digitized books, and share the information they find via groups, videos and blogs.

This is leadership on a scale that only a huge organization with extraordinarily deep pockets, a focused mission, and amazingly creative ideas can hope to mount. We have to applaud Google for this leadership, but are we also a little jealous, or worried? Perhaps we should be, if not the former, at least the latter.

Summary Thoughts

I have raised many questions without supplying answers. Why? Because this is all happening so fast. Research libraries with a mission to preserve collections and make them accessible to future generations will be affected, but we do not know exactly how yet. My cautions in each of the five areas are:

- *Pace.* We cannot slow it down; the pace car is Google, and other commercial drivers are nearly as pushy. But we must find the time to digest it all before making irreversible decisions about our precious collections.
- *Risks.* Let Google and others take risks if they wish, but we should not be taking risks with our collections, nor should we be risking our users' free access to our collections in the future.
- *Justification for digitizing books.* Given that we all have to get on this bandwagon, and given that we all have limited resources to do so, we should be thinking of ways to maximize value for scholars now and in the future. What sorts of materials will be of the most interest into the far-distant future? Can we afford to digitize older materials that are the "Long Tail" of our collections—items that will appeal only to a small number of researchers, or only one, or maybe none at all?[53]
- *Trust.* We must find partners—sister libraries, commercial entities, government agencies, and others. But we must keep a clear head about which of those organizations will be around for the long term, which of them share our values and our mission, and which truly understand preservation and conservation issues in regard to fragile, valuable, endangered, and irreplaceable artifacts. The Open Content Alliance model keeps a lot of control under the contributing libraries, where it should be.
- *Leadership.* The leadership needs to come from all parties in this endeavor. Because we are all mutually dependent, no one organization is in a position to dictate the discussions or the outcomes. Google and Yahoo need our content. We need stable, robust technology platforms for preservation and wider use of our collections. Scholars and students need more access and knowledge about how to use these collections. We all need to stay in close communication and collaboration . . . with our eyes wide open! In the end, research libraries alone will be held accountable for fulfilling that vital preservation mission.

References and Notes

1. Katie Hafner, "At Harvard, a Man, a Plan and a Scanner," *The New York Times*, Nov. 21, 2005, Section C; Column 2, Business/Financial Desk, www.nytimes.com/2005/11/21/business/21harvard.html?ex=1290229200&en=86f7d416af4055cd&ei=5090&partner=rssuserland&emc=rss (accessed Mar. 31, 2007).
2. Open Content Alliance. www.opencontentalliance.org (accessed Mar. 31, 2007).
3. Siva Vaidhyanathan, "A Risky Gamble with Google," *Chronicle of Higher Education* 52, no. 15 (Dec. 2, 2005): B7.
4. Jon Boone and Maija Palmer, "Microsoft in Deal with British Library to Add 100,000 Books to the Internet," *Financial Times*, Nov. 3, 2005, front page, first section, http://search.ft.com/ftArticle?queryText=microsoft+british+library&aje=true&id=051104001019 (accessed Jan. 29, 2007).
5. James H. Billington, "A Library for The New World," *The Washington Post*, Nov. 22, 2005, A29, www.washingtonpost.com/wp-dyn/content/article/2005/11/21/AR2005112101234_pf.html (accessed Jan. 29, 2007).
6. "E-Library for Europe," *The Australian*, Mar. 7, 2006, IT Broadsheet, 2.
7. Google Book Search, "The Complete Plays of Shakespeare Now at Your Fingertips," www.google.com/shakespeare (accessed Mar. 31, 2007); Jesse Noyes, "Shakespeare Has Google Web Site," *The Boston Herald*, June 15, 2006, Finance, 43.

8. Motoko Rich, "Arts, Briefly; Google Snags another Library," *The New York Times,* Aug. 9, 2006, Section E, Column 5, Page 2, The Arts/Cultural Desk.
9. Megan Twohey, "UW Deal Will Put Books Online," *Milwaukee Journal Sentinel,* Oct. 13, 2006, www.findarticles.com/p/articles/mi_qn4196/is_20061013/ai_n16786801 (accessed Mar. 31, 2007).
10. Rebecca Knight, "Microsoft in Digital Book Deal," *Financial Times,* Oct. 18, 2006, http://search.ft.com/ftArticle?queryText=microsoft+cornell+scan+books&aje=true&id=061018001258 (accessed Jan. 29, 2007).
11. Lee C. Van Orsdel and Kathleen Born, "Journals in the Time of Google," LibraryJournal.com (Apr. 15, 2006), www.libraryjournal.com/article/CA6321722.html (accessed Mar. 31, 2007); Adam Smith, "Google's Perspective" (informal presentation at "Scholarship and Libraries in Transition: A Dialogue about the Impacts of Mass Digitization Projects," University of Michigan, Ann Arbor, Mar. 10–11, 2006).
12. U.S. National Commission on Libraries and Information Science, *Mass Digitization: Implications for Information Policy: Report from "Scholarship and Libraries in Transition: A Dialogue about the Impacts of Mass Digitization Projects,"* Symposium held on March 10-11, 2006 at the University of Michigan, Ann Arbor, MI. May 9, 2006 (Washington D.C.: NCLIS, 2006), 11.
13. Ibid.
14. Jonathan B. Bengtson, "The Birth of the Universal Library," *Library Journal* Net Connect (Apr. 15, 2006), www.libraryjournal.com/article/CA6322017.html (accessed Mar. 31, 2007).
15. Andrew Richard Albanese, "The Social Life of Books," *Library Journal* (May 15, 2006): 28–30.
16. Brian Lavoie and Lorcan Dempsey, "Thirteen Ways of Looking at . . . Digital Preservation," *D-Lib Magazine* 10, no. 7/8 (2004), www.dlib.org/dlib/july04/lavoie/07lavoie.html (accessed Jan. 29, 2007).
17. Jeffrey R. Young, "Scribes of the Digital Era," *The Chronicle of Higher Education* 52, no. 21 (Jan. 27, 2006): 34.
18. Ibid.
19. Mary Sue Coleman, "Google, the Khmer Rouge, and the Public Good" (address to the Professional/Scholarly Publishing Division of the Association of American Publishers, Washington, D.C., Feb. 6, 2006), www.umich.edu/pres/speeches/060206google.html (accessed Jan. 29, 2007).
20. Clifford Lynch, "Web Cast: Closing Remarks," delivered at "Scholarship and Libraries in Transition: A Dialogue about the Impacts of Mass Digitization Projects," Mar. 11, 2006, www.lib.umich.edu/mdp/symposium/lynch.html (accessed Mar. 31, 2007).
21. University of Maryland Libraries. "Libraries Mission," www.lib.umd.edu/deans/index.html (accessed Mar. 31, 2007).
22. Coleman, "Google, The Khmer Rouge, and the Public Good."
23. Peter Hart and Ziming Liu, "Trust in the Preservation of Digital Information," *Communication of the ACM* 46, no. 6 (2003): 93–97.
24. Kate Wittenberg, "Beyond Google: What Next for Publishing?" *The Chronicle of Higher Education* 52, no. 41 (June 16, 2006): 20.
25. Google, "Google Code of Conduct," http://investor.google.com/conduct.html (accessed Mar. 31, 2007).
26. Google, "Google Corporate Information, Company Overview," www.google.com/corporate (accessed Mar. 31, 2007).
27. This issue is discussed in articles in *The Independent* (London), July 20, 2006; *Irish Times,* July 20, 2006; and *San Francisco Chronicle,* July 21, 2006.
28. "Google Print for Libraries: The Bold and the Cautious," LibraryJournal.com, www.libraryjournal.com/article/CA22558.html (accessed Oct. 22, 2006).
29. Coleman, "Google, the Khmer Rouge and the Public Good"; *A Public Trust at Risk: The Heritage Health Index Report on the State of America's Collections* (Washington, D.C.: Heritage Preservation—The National Institute for Conservation, 2005), www.heritagepreservation.org/hhi/full.html (accessed Mar. 31, 2007).
30. Coleman, "Google, the Khmer Rouge and the Public Good."
31. U.S. National Commission on Libraries and Information Science, *Mass Digitization.*
32. Burt Helm, "Google's Great Works in Progress," *Business Week Online,* Dec. 22, 2005, www.businessweek.com/technology/content/dec2005/tc20051222_636880.htm (accessed Oct. 21, 2006).
33. Ibid.
34. Deanna B. Marcum, "The Future of Cataloging," *Library Resources & Technical Services* 50, no. 1 (2006): 6.
35. Helm, "Google's Great Works in Progress."
36. Ibid.
37. Catherine Holahan, "Google Seeks Help with Recognition," *Business Week Online,* Sept. 7, 2006, www.businessweek.com/technology/content/sep2006/tc20060907_732714.htm?chan=top+news_top+news+index_technology (accessed Oct. 21, 2006).
38. Jacqueline Doherty, "Barron's: Google Trouble?" *The Edge Daily,* Feb. 21, 2006, www.theedgedaily.com/cms/content.jsp?id=com.tms.cms.article.Article_8b5a2df4-cb73c03a-23d27500-d95569df (accessed Jan. 29, 2007).
39. Jessica Mintz, "Google Seeks Info from Book Scanners," *Washington Post,* Oct. 6, 2006.
40. Ibid.
41. Ibid.
42. Keith Regan, "Yahoo Snubs Google in Digital Book Copyright Case," *E-Commerce Times,* Nov. 30, 2006, www.ecommercetimes.com/story/54494.html (accessed Oct. 21, 2006).
43. LOCKSS, www.lockss.org/lockss/Home (accessed Mar. 31, 2007).
44. CLOCKSS, www.lockss.org/clockss/Home (accessed Jan. 29, 2007).
45. Ibid.
46. Jim Stemper and Susan Barribeau, "Perpetual Access to Electronic Journals: A Survey of One Academic Research Library's Licenses," *Library Resources & Technical Services* 50, no. 2 (2006): 91–109.
47. Kathleen Arthur et al., *Recognizing Digitization As a Preservation Reformatting Method* (Washington D.C.: Association of Research Libraries, 2004), 2.
48. Ibid., 3.
49. Ibid., 4.

50. OCLC, *Perceptions of Libraries and Information Resources: A Report to the Membership* (Dublin, Ohio: OCLC, 2005), www.oclc.org/reports/2005perceptions.htm (accessed Jan. 29, 2007).
51. Ibid.
52. "Google Launches Online Literacy Project," *Business and Industry*, Oct. 4, 2006 (accessed in Lexis-Nexis [proprietary database] Oct. 21, 2006).
53. Chris Anderson, "The Long Tail," *Wired Magazine* 12, no. 10 (Oct. 2004), http://web.archive.org/web/20041127085645/http://www.wired.com/wired/archive/12.10/tail.html2004 (accessed Mar. 31, 2006).

Statement of Ownership, Management, and Circulation

Library Resources & Technical Services, Publication No. 311-960, is published quarterly by the Association for Library Collections & Technical Services, American Library Association, 50 E. Huron St., Chicago (Cook), Illinois 60611-2795. The editor is Peggy Johnson, Associate University Librarian, University of Minnesota, 499 Wilson Library, 309 19th Ave. South, Minneapolis, MN 55455. Annual subscription price, $75.00. Printed in U.S.A. with periodicals-class postage paid at Chicago, Illinois, and at additional mailing offices. As a nonprofit organization authorized to mail at special rates (DMM Section 424.12 only), the purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes have not changed during the preceding twelve months.

(*Average* figures denote the average number of copies printed each issue during the preceding twelve months; *actual* figures denote actual number of copies of single issue published nearest to filing date: July 2007 issue.) Total number of copies printed: *average*, 6,199; *actual*, 6,185. Sales through dealers, carriers, street vendors and counter sales: *average*, none; *actual*, 512. Mail subscription: *average*, 5,014; actual, 4,973. Free distribution: *average*, 196; *actual*, 189. Total distribution: *average*, 5,670; *actual*, 5,674. Office use, leftover, unaccounted, spoiled after printing: *average*, 529; *actual*, 511. Total: *average*, 6,199; *actual*, 6,185. Percentage paid: *average*, 96.54; *actual*, 96.67.

Statement of Ownership, Management and Circulation (PS Form 3526, September 2007) for 2006/2007 filed with the United States Post Office Postmaster in Chicago, October 1, 2007.