

Article

Can Our Culture Be Saved? The Future of Digital Archiving

Diane Leenheer Zimmerman[†]

INTRODUCTION

Few things are more precarious than our hold on the physical embodiments of our cultural and intellectual history. And yet, our understanding of who and what we are as social beings and societies is largely informed by the continuity of our access to the books, correspondence, records, and other ephemera that capture the essence of earlier times and places. This is in large part why, after two thousand years, we still mourn the discontinuity that resulted from the destruction of the great library at Alexandria—and the hundreds of thousands of scrolls it contained—recording the literature, records, and learning of the great Hellenic culture of the Mediterranean basin.¹ The

[†] Samuel Tilden Professor of Law, New York University School of Law. Thanks are due to the colleagues who gave me much useful feedback on successive drafts of this Article, including Graeme Austin, June Besek, Stacy Dogan, Rochelle Dreyfuss, Harry First, Laura Gasaway, Dorothy Glancy, Joseph Liu, Michael Meurer, Anthony Reese, and Fred Yen. Special thanks to Peter Yu and participants in the conference entitled “W(h)ither the Middleman: The Role and Future of Intermediaries in the Information Age” at Michigan State University for encouraging me to turn my conference paper into an Article. The paper benefited greatly as well from the valuable input of participants in the Colloquium on Technology, Innovation, and Intellectual Property Policy at the University of Arizona, and from attendees at faculty workshops at Northeastern School of Law and the London School of Economics. Support for this research was provided by grants from the Filomen D’Agostino and Max E. Greenberg Research Fund at New York University School of Law and from the Law Department and the Suntory and Toyota International Centres for Economics and Related Disciplines at the London School of Economics. This Article is current through December 2006. Copyright © 2007 by Diane Leenheer Zimmerman.

1. The Ptolemies founded the library after the death of Alexander the Great. Andrew Erskine, *Culture and Power in Ptolemaic Egypt: The Museum and Library of Alexandria*, 42 GREECE & ROME 38, 40–41 (1995). The library,

looting and destruction of ancient sites in Mesopotamia and Afghanistan impoverish us because they irreparably disconnect the present from the past and thus impede access to the sources of both our shared and our particularized cultural roots. For example, we know other dramatists—possibly as great—worked alongside Aeschylus, Sophocles, Euripides, and Aristophanes, but all we will ever know of their work is what remains in the form of a random reference or fragment of text, not the vibrancy of their complete plays.²

The common tendency has been to shrug off this loss as one that is of serious concern only within the insular world of librarians and highly specialized scholars. But in truth it is the kind of loss that diminishes everyone, and it is a loss that could be avoided in the future, thanks to digitization. If so, the benefits would inure not just to a few, but to every man, woman, and child with access to a computer terminal. Ironically, however, we may never realize the chance to “save” culture because the very laws promoting the creation of communicative and artistic works turn out to be a barrier to preserving the works once they are created.³ The current dispute over the Google Book Search Library Project (Google Library) is just one example of how copyright laws complicate the preservation of new and not-so-new intellectual products.⁴ Unless lawmakers, copy-

scholars allege, was filled in part by seizing books from ships that docked in the port. Diana Delia, *From Romance to Rhetoric: The Alexandrian Library in Classical and Islamic Traditions*, 97 *AM. HIST. REV.* 1449, 1457 (1992); Erskine, *supra*, at 39. Historians report that, at its peak, the library held some half a million scrolls. Delia, *supra*, at 1458–59; Erskine, *supra*, at 40. It was intended to serve as a “centre of Greek culture and intellectual life.” Erskine, *supra*, at 42. The library was ultimately destroyed, but how and by whom continues to be a subject of controversy. Delia, *supra*, at 1460–65; Jon Thiem, *The Great Library of Alexandria Burnt: Towards the History of a Symbol*, 40 *J. HIST. IDEAS* 507, 508–09 (1979).

2. The playwright Agathon wrote apparently important tragedies in the fifth century B.C. but is known to us today largely through Plato’s Symposium. T.B.L. Webster, *Agathon*, 76 *J. HELLENIC STUD.* 115, 115 (1956) (reviewing P. LÉVÊQUE, *AGATHON* (1955)). Eupolis, a contemporary of Aristophanes, wrote at least fourteen comedies, but all that remains of them are fragments, the longest of which consists of “120 reasonably consecutive lines.” Ian C. Storey, *Dating and Re-Dating Eupolis*, 44 *PHOENIX* 1, 1 (1990). Even the surviving playwrights for whom we have at least some complete plays, their works are available to us only in abbreviated form. Sophocles, for example, is said to have written over 120 plays, of which only seven complete examples remain. *Sophocles*, in *THE COLUMBIA ENCYCLOPEDIA* (Paul Lagassé ed., 6th ed. 2000), available at <http://www.bartleby.com/65/so/Sophocle.html>.

3. See *infra* Part II.A.

4. See *infra* Part II.B. For a description of the Google Library, see Google

right owners, and the public can solve this conundrum, the opportunities created by this new technology will be lost and with them much of incalculable social and intellectual value. The purpose of this Article is to explore whether saving culture and saving copyright can be made compatible goals and to propose some possible ways to achieve both ends.

Until now, preservation of cultural and intellectual works largely meant saving physical objects, and responsibility for doing so was left mostly to institutions like museums and our great public, academic, and research libraries. The obstacles they have faced are formidable. Leaving aside the depredations of war and natural disaster,⁵ other issues such as limits on funds, the fragility of the various media, and lack of physical space have all conspired to keep even institutions that consider preservation of our cultural record a core mission from acquiring or preserving anything like the full range of authors' creations, much less evidence of the social contexts that gave birth to their works.⁶

Fiscal limitations inevitably circumscribe any given institution's efforts to collect and save the artifacts of culture. Even the wealthiest must choose what to acquire with their scarce funds, and gaps in collections are inevitable. Academic librarians, for example, confronting the so-called serials crisis, have in many cases been forced by increasing subscription costs to buy fewer journals or to shift their budget allotments away from collecting books and humanities publications in order to afford continued acquisition of expensive scientific publications.⁷

Book Search Library Project: An Enhanced Card Catalog of the World's Books, <http://books.google.com/googleprint/library.html> (last visited Mar. 5, 2007).

5. A short, but enlightening, account of the vulnerability of books to war, ignorance, and natural forces can be found in Bessie S. Rathbun, *Books, Bibliophiles, and Barbarians*, 43 CLASSICAL J. 293 (1948).

6. Libraries frequently need to prune out things they have previously acquired to conserve space and even to raise money to keep the rest of the collection available. A not-uncommon story appeared recently in an Indiana newspaper recounting the plans of a small town public library to sell off some six thousand books plus newspapers and magazines to save space and hopefully raise a few thousand dollars to pay bills and update technology. Kevin Cullen, *After 100 Years of Accumulation, Library Prunes Shelves for Book Sale*, J. & COURIER (Lafayette, Ind.), Aug. 27, 2006, at B7.

7. See Aaron S. Edlin & Daniel L. Rubinfeld, *Exclusion or Efficient Pricing? The "Big Deal" Bundling of Academic Journals*, 72 ANTITRUST L.J. 119, 125 (2004); Ann Okerson, *Midnight in the Garden of Good and Evil: Academic Publishing, Copyright, and Other Miasmas*, in THE TRANSITION FROM PAPER: WHERE ARE WE GOING AND HOW WILL WE GET THERE? (R. Stephen Berry &

Collecting is also constrained by space. Unless institutions are committed to continuous physical expansion, at a certain point they simply fill up. The Library of Congress, for example, can no longer make room for deposit copies of everything that is copyrighted.⁸ Even if institutional archives continue to collect once they run out of publicly accessible shelf or display space, they must keep parts of their collections in storage, with the result that these materials can be used only by special arrangement. This problem can sometimes be solved in the short run by transferring works into less space-demanding formats. Microfilm and microfiche, for instance, have long been used as media for keeping things like back issues of newspapers and old government records,⁹ although anyone who has used resources in this form can attest to the fact that the experience is cumbersome and unpleasant.¹⁰

Finally, originals are often made of materials that are easily destroyed by external forces or that simply self-destruct over time. For example, an estimated half of all motion pictures made before 1950 are now lost, largely because they were recorded on nitrate cellulose film that has either turned to dust or spontaneously burst into flame.¹¹ Books printed in the late nineteenth and early twentieth centuries were typically printed

Anne Simon Moffat eds., 2001), <http://www.amacad.org/publications/trans11.aspx>.

8. Section 407 of the Copyright Act of 1976 allows the Register of Copyrights to exempt categories of works from deposit, or to reduce the number of copies of registered works that are deposited to one. Copyright Act of 1976, Pub. L. No. 94-553, § 407(c), 90 Stat. 2541, 2579 (codified as amended at 17 U.S.C. § 407(c) (2000)). The Act permits the Library of Congress to use or dispose of the copy(ies) deposited with it. *Id.* § 407(b). For a discussion of these provisions and how they respond to the Library of Congress' space limitations, see 2 MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT § 7.17 (2006).

9. Converting printed material to microfiche or microfilm, if done by a third party like a library to save shelf space, can also raise copyright problems, as reflected in 17 U.S.C. § 108 (2000), *amended by* Pub. L. No. 109-9, § 402 (2005). See *infra* notes 101–25 and accompanying text. These problems are, however, except in a general way, beyond the scope of this Article.

10. Advocates of microfilm and microfiche, however, point out the stability of these materials (with a potential life expectancy today of five hundred years) and the comparative cheapness of producing and reproducing them. STEVE DALTON, NE. DOCUMENT CONSERVATION CTR., MICROFILM AND MICROFICHE, <http://www.nedcc.org/resources/leaflets/6Reformatting/01MicrofilmAndMicrofiche.php> (last visited Mar. 5, 2007). An added advantage is that documents stored in this form can be accessed without the use of complex equipment. *Id.*

11. See *infra* notes 57–60 and accompanying text.

on acidic paper that gradually became brittle and is now crumbling away.¹² Preservation techniques can sometimes save the original materials, but the process is often both painstaking and expensive. As a result, it is not uncommon for all or parts of physical collections to fall prey to heat, moisture, or deterioration caused by the materials from which they were made.¹³

But, with the advent of digitization, the survival and usability of culture and history need not rest wholly on the availability of space for, or the physical composition of, the original tangible embodiments of cultural production, much less on any single institution's budget. It might now become possible to keep copies of anything that is deemed actually or potentially of cultural significance for centuries and to do so in a medium that, hopefully, will be comparatively easy to conserve¹⁴ and in a format that can readily be searched. Instead of benefiting a fortunate few, these riches would be available to anyone with Internet access and a computer at her disposal, at least once the resources are in the public domain. A researcher at her desk in New Hampshire—or Ghana—would be able to find and read a document that exists in analog form only in one library in San Francisco. Because additional copies of such archived, digital materials would be reasonably inexpensive to make once the first copy is created, backups could be stored at multiple sites around the world to ensure that in the event of natural,

12. Laura N. Gasaway, *America's Cultural Record: A Thing of the Past?*, 40 HOUS. L. REV. 643, 647 (2003). Gasaway cites a 1987 study suggesting that a quarter of the books in U.S. library collections were at risk of crumbling away. *Id.*

13. A recent study shows that only a quarter of museums, archives, and libraries have the capacity to protect the physical integrity of their collections against such factors. See HERITAGE PRES., A PUBLIC TRUST AT RISK: THE HERITAGE HEALTH INDEX REPORT ON THE STATE OF AMERICA'S COLLECTIONS 2 (2005), available at <http://www.heritagepreservation.org/HHI/HHIfront.pdf>.

14. I do not mean to understate the technical difficulties that still need to be resolved for digital technology to serve as a permanent, stable storage medium. They are substantial. See, e.g., Mary Baker et al., *A Fresh Look at the Reliability of Long-Term Digital Storage*, 2006 PROC. EUROSYS 221, 222–23, available at <http://www.cs.kuleuven.ac.be/conference/EuroSys2006/papers/p221-baker.pdf>; Dorothy Warner, 'Why Do We Need to Keep This in Print? It's on the Web': A Review of Electronic Archiving Issues and Problems, 31 MICROFORM & IMAGING REV. 59, 61–67 (2002) (discussing various technical issues that need to be resolved to make persistent digital archives realistic). This Article, however, intends to solve the legal part of the problem—the lack of a coherent regime that would enable us to realize the enormous possibilities if and when the technological problems can be solved. At present, the technical difficulties, however challenging, are attracting considerable attention. Until recently, far less awareness of the legal difficulties has been apparent.

manmade, or technological disaster, at least one copy would survive and be available to future generations.

Admittedly, these digital copies will be only partial substitutes for originals. Examining a digital photograph of a Greek krater is not the same experience as examining the vessel itself. But photographs of the krater are nevertheless useful in many circumstances and invaluable should the original be stolen, lost, or damaged. As for works in text, the full value of a fine old manuscript cannot be captured in pixels. However, to the extent that what matters is the content—which can be perfectly reproduced in electronic form—the digital copy will be a quite adequate substitute. Thus, despite its limitations, a durable digital record of the human race's intellectual and artistic production, potentially accessible by anyone from anywhere in the world, would offer a public benefit of incalculable worth.

Lest there be doubt that this is so, consider the excitement and the sheer amount of press coverage generated by the Google Library.¹⁵ Although Google's primary purpose is to vastly increase the capability for online searches,¹⁶ librarians and other observers have been quick to recognize an added benefit: the copies made to create Google Library also form a digital backup for the collections of the participating libraries.¹⁷

Until Google announced its plan to digitize the full print collection of one major university¹⁸ and parts of the collections

15. See Google Book Search Library Project, *supra* note 4.

16. *Id.*

17. According to reports, the libraries participating in Google Library have planned various methods of ensuring that the electronic copies Google returns to them will be saved in durable and secure forms. Barbara Quint, *Google's Library Project: Questions, Questions, Questions*, NEWSBREAKS, Dec. 27, 2004, <http://newsbreaks.infotoday.com/nbReader.asp?ArticleId=16302> (noting Stanford's plan to keep at least three copies on magnetic tape cartridges and the University of Michigan's plan to store its copy on gold CD-ROMs with an estimated three hundred year life span). The University of California, however, will not receive its own digital copies but rather will only be allowed to access them from Google's own server. See Cooperative Agreement Between Google Inc. and The Regents of the University of California §§ 4.6–4.7.1 (Aug. 3, 2006), http://www.cdlib.org/news/ucgoogle_cooperative_agreement.pdf [hereinafter Cooperative Agreement]. As this Article indicates, however, there is more to preservation than conserving any particular digital copy. See *infra* Part III. For a comprehensive review of the requirements of adequate archiving of digital works, see Report from Donald J. Waters on Trusted Preservation Repositories for the Section 108 Study Group at the Library of Congress (Oct. 24, 2005) (on file with author) [hereinafter Waters].

18. The University of Michigan University Library, MBooks: Michigan Digitization Project, <http://www.lib.umich.edu/mdp/> (last visited Mar. 5, 2007).

of several others,¹⁹ the potential of digital archives did not seem to register on the mental radar screens of most people outside (and even inside) the academy. But a brief glance at the number and range of pre-Google projects that aimed to convert huge quantities of works into digital formats shows that at least some far-sighted individuals and entities have understood for years—in some instances for well over a decade—what could be at stake. For example, Project Gutenberg used volunteers to scan and proofread at least twenty thousand public domain books and has made them available online to be read and downloaded.²⁰ The Internet Archive, centered in San Francisco,

19. Harvard, Stanford, the University of California, and the University of Virginia will have portions of their collections, including both public domain and copyrighted works, digitized. Barbara Palmer, *Deals with Google to Accelerate Library Digitization Projects for Stanford, Others*, STANFORD REP., Jan. 12, 2005, <http://news-service.stanford.edu/news/2005/january12/google-0112.html>; Press Release, Univ. of Va., University of Virginia Joins Leading Research Libraries in Partnership with Google to Increase Discovery of Knowledge—And to Offer Library Books to Global Audience (Nov. 14, 2006), *available at* <http://www.lib.virginia.edu/press/uvagoogole> [hereinafter U. Va. Press Release]; Harvard University Library, FAQ: The University's Pilot Project with Google (Dec. 14, 2004), http://hul.harvard.edu/Site/news/2004_1214_faq.html; *UC System Will Join Google's Scan Plan*, LIBRARYJOURNAL.COM, Aug. 11, 2006, <http://www.libraryjournal.com/article/CA6361276.html>.

Other institutions, including the New York Public Library, will have parts of their public domain holdings scanned. Press Release, N.Y. Pub. Library, NYPL Partners with Google to Make Books Available Online (Dec. 14, 2004), *available at* <http://www.nypl.org/press/2004/google.cfm>. Oxford will contribute part of its collection of nineteenth-century works in the Bodleian. *Google to Scan Famous Libraries*, BBC NEWS, Dec. 14, 2004, <http://news.bbc.co.uk/2/hi/technology/4094271.stm>. The University of Wisconsin at Madison and the Wisconsin Historical Society will contribute public domain books as well as maps and documents. Press Release, Univ. of Wis.-Madison, UW-Madison Joins Google's Worldwide Book Digitization Project (Oct. 12, 2006), *available at* <http://www.news.wisc.edu/releases/13010.html>. Spanish language works in the public domain will be provided by the University Complutense of Madrid, which has the largest university library in Spain. Press Release, Google, University Complutense of Madrid and Google to Make Hundreds of Thousands of Books Available Online (Sept. 26, 2006), *available at* http://www.google.com/intl/en/press/annc/books_madrid.html.

20. See Project Gutenberg, <http://www.gutenberg.org> (last visited Mar. 5, 2007). Project Gutenberg began by focusing on books in English but is now adding works in other languages. Marie Lebert, Project Gutenberg, from 1971 to 2005 (Aug. 15, 2005), http://www.etudes-francaises.net/dossiers/gutenberg_eng.htm#languages. Its founder, Michael Hart, hopes to have one million books online by the year 2015. *Id.* In addition to allowing the public to read or download books without charge, Project Gutenberg makes the books available on CDs and DVDs (a donation is requested but not required). See Gutenberg: The CD and DVD Project, <http://www.gutenberg.org/cdproject/> (last visited Mar. 5, 2007).

is geared toward preserving materials that originated in digital form (including websites) and retaining them on the Internet.²¹ Its “way-back machine,” for example, can provide a snapshot of how the World Wide Web looked on any particular day in the past.²² Individual academic institutions are also putting various parts of their collections online,²³ as are other public repositories such as the Library of Congress.²⁴ The trend is international: European libraries are engaged in numerous digitization projects involving a wide range of subject matter.²⁵ Some initiatives deal with highly specialized material that, absent digitization and Internet-accessible databases, would be unavailable to most of the world. The New Zealand Electronic Text Centre, for example, digitized a vast array of “New Zealand and Pacific Islands texts and heritage materials” for searching, downloading, or browsing.²⁶ Most of these projects concentrate on works in the public domain, but a significant minority are attempting to include copyrighted works, an effort which raises questions about the need for permission of the copyright owner.²⁷ JSTOR, with funding from the Andrew W. Mellon Foundation, is constructing an archive of back issues of academic journals.²⁸ The

21. See Internet Archive, About the Internet Archive, <http://www.archive.org/about/about.php> (last visited Mar. 5, 2007).

22. See *id.* The Archive also has files of motion pictures and music. *Id.*

23. See, e.g., University of Virginia Library, Electronic Text Center, Goals and Missions, <http://etext.lib.virginia.edu/info/mission.html> (last visited Mar. 5, 2007). The University of Virginia’s Electronic Text Center includes over 2100 full texts of works ranging from children’s books to the Bible. University of Virginia Library, Electronic Text Center, Ebooks, <http://etext.lib.virginia.edu/ebooks/> (last visited Mar. 5, 2007).

24. See Library of Congress, Digital Collections and Programs, <http://www.loc.gov/library/libarch-digital.html> (last visited Mar. 5, 2007).

25. David Raitt, *Digital Library Initiatives Across Europe*, COMPUTERS LIBR., Nov.–Dec. 2000, at 26, 26, available at <http://www.infotoday.com/cilmag/nov00/raitt.htm>.

26. See New Zealand Electronic Text Centre, <http://www.nzetc.org> (last visited Mar. 5, 2007).

27. The Universal Library is attempting to get permission from copyright owners to include their books in its database, although it has encountered considerable difficulty in doing so. For a discussion of the issue, see The Universal Library: Content Selection, <http://www.ul.cs.cmu.edu/html/contentselection.html> (last visited Mar. 5, 2007).

28. See JSTOR: The Scholarly Journal Archive, <http://www.jstor.org> (last visited Mar. 5, 2007). JSTOR archives copyrighted academic journals. See JSTOR, Welcome to JSTOR, <http://www.jstor.org/about/desc.html> (last visited Mar. 5, 2007). A description of the agreements between JSTOR and journal publishers can be found in Sarah E. Sully, *JSTOR: An IP Practitioner’s Perspective*, D-LIB MAG., Jan. 1997, <http://www.dlib.org/dlib/january97/01sully>

goal of Carnegie Mellon University's Universal Library is digitizing one million books, both currently copyrighted and in the public domain, and making them searchable as well as accessible in full text.²⁹

The availability of this material opens windows in places where previously only walls existed. On a personal level, I recently completed a study of the background and influence of an important early twentieth-century Supreme Court case. The availability of databases of old newspapers and digitized copies of public domain art works permitted me to understand the personalities and enterprises that gave rise to this dispute at a level of detail I could not have achieved a decade earlier, except after an expenditure of time and money that would have been difficult to justify. I was lucky because I was writing about events that happened a century ago, and the surviving contemporaneous materials could be made available digitally because they were in the public domain. However, much of the material we need to understand the past disappears long before the point at which it enters the public domain. For that reason, copyright stands as a partial but important barrier to accomplishing much of what technological innovations now promise.

This Article will first discuss why copyright owners cannot be relied upon to serve the public's interest in preservation and then it will turn to why preservation of the past is an issue that copyright law must directly confront. Part II goes on to explore in depth the barriers to digital archiving in the current law, including those that may derail the Google Library project. Part III explores possible solutions that would allow digital archiving to develop its full potential as well as the possible international ramifications of the changes in domestic law that the Article suggests.

.html. Sully was general counsel to JSTOR at the time she wrote the article. *Id.* For a full history of the JSTOR project, see ROGER C. SCHONFELD, *JSTOR: A HISTORY* (2003).

29. Detailed information about the Universal Library project is available at Welcome to the Universal Library: Hosted by Carnegie Mellon University, <http://www.ul.cs.cmu.edu> (last visited Mar. 5, 2007).

I. THE PROBLEM OF PRESERVING COPYRIGHTED WORKS

A. THE PAST, UNRECOVERED

As projects like Google Library, the Universal Library, and Project Gutenberg demonstrate, a huge appetite currently exists for getting as much of our cultural product as possible into digital form.³⁰ The driving force behind most of these projects is the excitement of providing the public with access to works without regard to the location of their physical instantiations.³¹ Preservation seems likely to be thought of largely as an incidental but nice byproduct of providing access.³² At the outset, however, it is important to emphasize that, although intertwined, preservation and access are actually distinct and equally significant goals. After all, saving works without at least eventually making them accessible would seem pointless; and, without first ensuring that preservation is attended to, access cannot be assured. To deal with both of them adequately, however, may require disaggregating a bit.

The ease with which the dual objectives of preservation and access can currently be pursued is largely a function of whether a work is copyrighted or in the public domain. As noted above, copyright issues cause most of the newly created digital databases of books, journals, and art to focus on materials already in the public domain.³³ Getting permission to digitize copyrighted works is often technically complicated.³⁴ Even

30. See *supra* notes 15–29 and accompanying text.

31. See *supra* note 15 and accompanying text.

32. See *supra* note 17 and accompanying text.

33. Project Gutenberg is one example of a database that will contain only public domain works. See *supra* note 20. The Open Content Alliance, at its outset a product of cooperation between Yahoo! and the Internet Archive, appears likely to contain mostly (if not exclusively) public domain works because it wants to make the text of its works fully available online. See Open Content Alliance, FAQ, <http://www.opencontentalliance.org/faq.html> (last visited Mar. 5, 2007). In December 2006, Microsoft began testing its own book search service. Charlie Taylor, *Microsoft Creates Digital Library*, ELECTRICNEWS.NET, Dec. 7, 2006, <http://www.electricnews.net/news.html?code=9855866>. Initially, the service will include only public domain works, but it may eventually add copyrighted works. *Id.* Information on what Microsoft clearly hopes will be a rival to Google's service is available at Live Search's WebLog, <http://blogs.msdn.com/livesearch> (last visited Mar. 5, 2007).

34. As this Article will discuss more thoroughly, the transaction costs of obtaining permission are likely to be high. And, furthermore, it will be difficult if not impossible to identify the owner of many of the relevant copyrights—a

when archive projects can overcome the technical problems, they may find obtaining permission impossible because owners, fearing that online access will interfere with their ability to fully exploit the economic value of their copyrights, will refuse to give it.³⁵ Requiring digital preservation projects to wait until works enter the public domain might seem to be the reasonable response to this obstacle, were it not for the fact that today works must be old indeed to achieve public domain status, and many do not survive long enough to get there.³⁶ Those works that survive may remain only in the form of a few frail copies stashed away in the stacks and storerooms of scattered libraries or museums. As a result, it may be difficult to discover that these works exist, or the fragility of the works may restrict their usability.

A quick review of changes in copyright law over recent decades explains the problem. The 1909 Copyright Act,³⁷ in effect until the end of 1977,³⁸ gave protection to authors' works (assuming the copyright owner complied with the law's formal requirements)³⁹ for an initial term of twenty-eight years.⁴⁰ Au-

problem that gave rise to the Copyright Office's study and legislative recommendations on orphan works. U.S. COPYRIGHT OFFICE, REPORT ON ORPHAN WORKS (2006), available at <http://purl.access.gpo.gov/GPO/LPS67330> [hereinafter ORPHAN WORKS].

35. For a discussion of the risks of digitization to traditional publishers, see Gal Oestreicher-Singer & Arun Sundararajan, *Are Digital Rights Valuable? Theory and Evidence from the Ebook Industry*, 2004 TWENTY-FIFTH INT'L CONF. ON INFO. SYSTEMS 533. J.K. Rowling, the author of the Harry Potter series, for example, will not allow her books to be digitized out of concern over possible piracy. Robert Andrews, *Pirates of the Potter-ian*, WIRED NEWS, July 21, 2005, <http://www.wired.com/news/digiwood/1,68269-1.html>. Just recently, author John Updike received extensive publicity for railing against the electronic book, in part on the grounds that once books are available digitally, consumers will no longer pay for them, and authors will have to earn their living by lecturing or selling "personal access." John Updike, *The End of Authorship*, N.Y. TIMES BOOK REV., June 25, 2006, at 27; see also Sarah Glazer, *An Idea Whose Time Has Come Back*, N.Y. TIMES BOOK REV., Dec. 5, 2004, at 31.

36. One study found that over ninety percent of book sales occur in the first year after publication and that books are largely out of print by the third year. Gasaway, *supra* note 12, at 660.

37. Copyright Act of 1909, ch. 320, 35 Stat. 1075, *superseded by* Copyright Act of 1976, Pub. L. No. 94-553, 90 Stat. 2541 (codified as amended at 17 U.S.C. §§ 101-810 (2000)).

38. The 1976 Copyright Act was effective as of January 1, 1978. Copyright Act of 1976 § 301 (codified as amended at 17 U.S.C. § 301 (2000)).

39. The Copyright Act of 1909 stipulated that works would be protected only if they were published carrying a proper copyright notice. Copyright Act of 1909, ch. 320, § 10. A validly acquired copyright could be voided for failure to comply with deposit requirements as well. *Id.* § 12.

thors had the option to renew for an additional twenty-eight-year term,⁴¹ although they frequently did not exercise that option.⁴² As a result, copyrighted works usually fell into the public domain twenty-eight years from the date of publication, and the remainder became available after fifty-six years.⁴³ This limited time frame greatly increased the likelihood that physical copies of the works would survive long enough to enter the public domain.

The 1976 Copyright Act made two important alterations in the law. First, it provided that any expressive work would receive a copyright simply upon fixation in tangible form.⁴⁴ Previously, a work could not be protected by federal law until it was published, and even then, the work would not be eligible for copyright unless it carried a copyright notice as well.⁴⁵ This change greatly increased the number of works covered by statutory copyright. The second modification was to the term for new copyrights. Initially the term was changed to meet the standard set by the Berne Convention:⁴⁶ the author's life plus fifty years.⁴⁷ This change was accompanied by a lengthening of the second term of preexisting copyrights to make their total duration roughly equivalent to life-plus-fifty.⁴⁸ Then, in 1998,

40. *Id.* § 24.

41. *Id.*

42. See *infra* note 229 and accompanying text.

43. Starting in the 1960s, Congress began extending the second term of 1909 copyrights with an eye toward achieving some degree of parity between the owners of works covered by the 1909 Act and owners of works created under the statute then being drafted. Dennis S. Karjala, What Are the Issues in Copyright Term Extension—and What Happened?, <http://homepages.law.asu.edu/~dkarjala/OpposingCopyrightExtension/what.html> (last visited Mar. 5, 2007); see also Scott M. Martin, *The Mythology of the Public Domain: Exploring the Myths Behind Attacks on the Duration of Copyright Protection*, 36 LOY. L.A. L. REV. 253, 260–63 (2002). When Congress finally passed the 1976 Act, it added a total of nineteen extra years to any extant 1909 copyright during its renewal term. Tyler T. Ochoa, *Patent and Copyright Term Extension and the Constitution: A Historical Perspective*, 49 J. COPYRIGHT SOC'Y U.S.A. 19, 22–23 (2001).

44. Copyright Act of 1976, Pub. L. No. 94-553, § 102(a), 90 Stat. 2541, 2544–45 (codified as amended at 17 U.S.C. § 102(a) (2000)).

45. Copyright Act of 1909, ch. 320, § 9.

46. Berne Convention for the Protection of Artistic and Literary Works, art. 7, para. 1, Sept. 9, 1886, as last revised July 24, 1971, 25 U.S.T. 1341, 1161 U.N.T.S. 31, available at http://www.wipo.int/treaties/en/ip/berne/pdf/trtdocs_wo001.pdf [hereinafter Berne Convention].

47. The original version of 17 U.S.C. § 302 (2000) appears in the Copyright Act of 1976 § 302.

48. For the original version of 17 U.S.C. § 304, see Copyright Act of 1976

following the lead of the European Union,⁴⁹ Congress extended the copyright term again, now to life plus seventy years.⁵⁰ It also added twenty years to the second term of surviving 1909 Act copyrights.⁵¹ As a result, a work created and copyrighted in the mid-1920s had a significant chance of becoming nearly a centenarian before it fell at last into the public domain.⁵² A copyright acquired once the 1976 Act was in place could easily endure for longer than that.⁵³

In addition to the sheer length of time copyright now lasts, part of the reason that works disappear before their copyright term ends is that preservation of copyrightable works has traditionally been severed from their production. Preservation has largely been the business of third parties with rights only to the physical objects in which the intellectual property is embodied.⁵⁴ The parties expected to play the role of preservationists—libraries, museums, and the like—are therefore constrained in their preservation and conservation choices, not merely by space and resources, but also by the largely exclusive control the copyright regime confers on authors and their successors.⁵⁵ As long as preservation requires only manipulation of a copy, third parties can function well in their role, but if preservation requires making a reproduction, the existing law becomes a significant barrier.

Entities who actually have the legal rights to the intellectual property have often lacked resources to, or interest in, pre-

§ 304.

49. Council Directive 93/98/EEC, 1993 O.J. (L 290) 36 (EC). For a discussion of the importance of the European action, see Dennis S. Karjala, *Judicial Review of Copyright Term Extension Legislation*, 36 LOY. L.A. L. REV. 199, 207–08, 210–13 (2002).

50. Sonny Bono Copyright Term Extension Act, Pub. L. No. 105-298, § 102(b), 112 Stat. 2827, 2827 (1998) (codified as amended at 17 U.S.C. § 302 (2000)).

51. *Id.* § 102(d) (codified as amended at 17 U.S.C. § 304(a)(1)–(2) (2000)).

52. A work from the mid-1920s advantaged by the full term now permitted for a 1909 copyright is protected for ninety-five years from the date of first publication. *See* 17 U.S.C. § 304(b).

53. A novelist like Zadie Smith who published her first novel, *ZADIE SMITH, WHITE TEETH* (2000), at age twenty-five could easily live an additional sixty or seventy years, and her novel would then remain under copyright for seventy more years after that. *See* 17 U.S.C. § 302(a).

54. *See, e.g.*, 1 LIBRARY OF CONG., *FILM PRESERVATION 1993: A STUDY OF THE CURRENT STATE OF AMERICAN FILM PRESERVATION* 23–27 (1993), available at <http://www.loc.gov/film/study.html> (discussing how films were collected by large archives for preservation).

55. *See infra* Part II.

servicing their own works. Publishing houses, film studios, newspapers, and other actual producers of information products may not even keep back copies of the works they have produced,⁵⁶ or may store them haphazardly, with little effort to keep them from deteriorating over time. Several examples demonstrate the point.

Reports to the Librarian of Congress document huge losses suffered among older motion pictures. Although many are still technically under copyright, copies of only about twenty percent of feature films from the 1920s and fifty percent of feature films made prior to 1950 still survive.⁵⁷ Other forms of film, including documentaries and newsreels, are even less likely to remain in existence.⁵⁸ The problem is that these early movies were recorded on a cellulose nitrate film that both decomposes and is highly flammable.⁵⁹ Cellulose acetate film replaced the more fragile nitrate base film in the 1950s, but it turned out to be prone to its own form of decay known as “vinegar syndrome.”⁶⁰ To the extent that old films were preserved at all, the work has largely been done by private or institutional parties who managed to obtain prints of works that have entered the public domain; film studios until recently have not been major players in this effort.⁶¹

56. Old film was expensive and difficult to store, and producers often simply discarded prints or sold them to collectors. LIBRARY OF CONG., *supra* note 54, at 16–17. To avoid the expense and difficulty of dealing with nitrate base film, several major studios turned over their old titles to public film archives. *Id.*

57. National Film Preservation Foundation, Why Preserve Film?, http://www.filmpreservation.org/preservation/why_preserve.html (last visited Mar. 5, 2007).

58. LIBRARY OF CONG., *supra* note 54, at 5.

59. National Film Preservation Foundation, Nitrate Degradation, <http://www.filmpreservation.org/preservation/nitrate.html> (last visited Mar. 5, 2007).

60. National Film Preservation Foundation, Vinegar Syndrome, <http://www.filmpreservation.org/preservation/vinegar.html> (last visited Mar. 5, 2007).

61. The advent of the videocassette player and the DVD player has increased the willingness of studios to invest in preserving their products because of the newly perceived value of these older works. LIBRARY OF CONG., *supra* note 54, at 17. The television programming preserved on videotape has similarly been subject to decay as well as careless or deliberate destruction. One common practice, for example, was to re-record over prior programming, thereby destroying it. The problems in this field are chronicled in 1 LIBRARY OF CONG., TELEVISION AND VIDEO PRESERVATION 1997: A REPORT ON THE CURRENT STATE OF AMERICAN TELEVISION AND VIDEO PRESERVATION 13–44 (1997), available at <http://loc.gov/film/tvstudy.html>.

Book publishers, too, have been less than satisfactory custodians of their output. Of the more than ten thousand books published in the United States in 1930, only 174 were still in print as of 2002.⁶² In many cases, the only copies of older works still extant are either hidden away in private collections or lodged on library shelves, not in publishers' warehouses.⁶³ Many are undoubtedly lost for all time. One representative of an academic publisher was recently quoted in the press as saying that his organization did not even have a complete list of all the works it had published over its history.⁶⁴

Although sound recordings did not come under copyright in the United States until the 1970s,⁶⁵ and therefore present somewhat different issues, the losses of recorded musical materials are similarly stunning. For example, when the Radio Corporation of America (RCA) decided to demolish its warehouse in Camden, New Jersey, most of the master disks, test pressings, and other important materials were simply blown up with the building and bulldozed into the Delaware River.⁶⁶ Small record companies come and go, leaving behind their old recordings to be thrown out or abandoning them in storerooms.⁶⁷

62. Dierdre K. Mulligan & Jason M. Schultz, *Neglecting the National Memory: How Copyright Term Extensions Compromise the Development of Digital Archives*, 4 J. APP. PRAC. & PROCESS 453, 459 (2002). For further discussion of works that go out of print, see R. Anthony Reese, *The First Sale Doctrine in the Era of Digital Networks*, 44 B.C. L. REV. 577, 592–95 (2003).

63. See Reese, *supra* note 62, at 592–94 (discussing the quantity of books that go out of print each year and are no longer stored by their publishers).

64. Associated Press, *Google Moving Ahead with Digital Library*, MSNBC, Sept. 18, 2005, <http://www.msnbc.msn.com/id/9392186>; see also SCHONFELD, *supra* note 28, at 90.

65. See, e.g., *Capital Records, Inc. v. Naxos of Am.*, 830 N.E.2d 250, 252 (N.Y. 2005) (stating that in New York, common law copyright protects pre-1972 sound recordings).

66. Brief for the American Ass'n of Law Libraries et al. as Amici Curiae Supporting Petitioners at 26, *Eldred v. Ashcroft*, 537 U.S. 186 (2003) (No. 01-618), 2002 WL 1059710 [hereinafter Brief for the American Ass'n of Law Libraries]. This sort of loss could be even more extensive than it first seems because courts generally thought that the 1909 Copyright Act allowed composers and songwriters to distribute their works solely as recordings without being deemed to have "published" them. See *Rosette v. Rainbo Record Mfg. Corp.*, 354 F. Supp. 1183, 1190 (S.D.N.Y. 1973), *aff'd per curiam*, 546 F.2d 461 (2d Cir. 1976). Such works were therefore not subjected to the time-limited protection of statutory copyright. See *id.* at 1188–92. As a result, the compositions were not deposited with the Library of Congress and might therefore disappear along with the masters and copies of the recordings.

67. See Brief for the American Ass'n of Law Libraries, *supra* note 66, at 26.

B. WILL THE MARKET RECOGNIZE AND SATISFY THE MARKET FOR ARCHIVING?

Digitization makes the prospect of durable cultural archives that can easily be duplicated tantalizing. If realized, this effort would also help stem the demand for massive, continuous expansion of library stacks and storage facilities. Improvements in the technology that enables creation and protection of digitized resources are coming apace, and many more are likely in the near future. The equipment Google is using to create Google Library exemplifies the rate of technological progress. The equipment is reported to be both exponentially faster than the immediately preceding generation of scanners and far safer than earlier generations of the equipment for use on fragile books and documents.⁶⁸ The costs of digitizing, too, continue to fall.⁶⁹ Storage material is becoming increasingly durable,⁷⁰ although over time data will clearly need to be ported repeatedly to new formats or storage devices.⁷¹ One should not underestimate the major technical and financial challenges that will be involved in creating digital archives and keeping them current,⁷² but the technological and financial limits on the potential for digital preservation are only part of the problem. A third, more intractable barrier is the current state of the applicable legal regime.⁷³

At least in the United States, copyright is commonly understood as a mechanism through which to pursue a public benefit.⁷⁴ If the public interest is the compass that guides copyright policy, then it is not a stretch to argue that the boon that a stable, comprehensive electronic archive of cultural products

68. See *infra* note 185.

69. For example, the University of Toronto reported in 2005 that over a ten year period, the cost of digitizing a page of text dropped from one dollar to ten cents. *Canadian Libraries Join Race to Digitize Books*, CBC.CA ARTS, Dec. 29, 2005, <http://cbc.ca/arts/story/2005/12/29/canada-libraries.html>.

70. The University of Michigan plans to store digital copies of its library's contents, created through the Google Library Project, on gold CD-ROMs, which are expected to last three hundred years. See Quint, *supra* note 17.

71. See Warner, *supra* note 14, at 62–63.

72. Many of these problems are reviewed in Reese, *supra* note 62, at 635–44.

73. My guess is that the easier it becomes, legally, to create archives, the greater the incentive will be to invest in solving the technical problems.

74. See *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 429, 431–32 (1984); *Mazer v. Stein*, 347 U.S. 201, 219 (1954); *United States v. Paramount Pictures, Inc.*, 334 U.S. 131, 158 (1948).

would represent to scholars, students, journalists, and others, both here and around the world, is a powerful incentive for making the changes in the current copyright regime that would allow these benefits to be realized. The problem, of course, is that the benefits to posterity of archiving would not seem so attractive if, in the course of pursuing them, we undercut the very incentives to authors that ensure the creation of the cultural riches we want to preserve. The promise of saving culture can be realized only if we can reach some reasonable compromise between the interests of creators and users. The question that needs to be addressed is whether such a compromise is realistically achievable.

1. The Copyright Owner as Archivist

One could argue that this question need never be addressed and that the legal regime need never be modified. If the potential for digitization of content is as exciting as I have suggested, then either copyright owners themselves will figure out how to exploit it, or they will license some other commercial entity like Google to do it for them. Because this is a plausible argument, I would like to begin by explaining why I think it is unlikely that we can rely on the private sector—either copyright owners or parties cooperating with them—to satisfy the public interest in long-term conservation (and with it, the promise of current or eventual public access), and why sound public policy cannot rest on the assumption that the private sector will do so.

Today many works originate in digital form, even if they are then printed and released as hard copy.⁷⁵ Increasingly, they may be distributed digitally as well. At least some publishers of works that either start or end in digital format—notably Elsevier Science—have announced that they will, henceforth, take on the responsibility of keeping permanent digital copies of the works they issue.⁷⁶ In an increasing number of cases, the

75. A typical example is the magazine *Scientific American*, available in both digital and analog versions. Science & Technology at Scientific American.com, http://www.sciam.com/subscribe_combo.cfm (last visited Mar. 5, 2007).

76. *Elsevier Makes a Commitment to Electronic Archiving of Its Electronic Journals*, INFO. INTELLIGENCE ONLINE LIBR. & MICROCOMPUTERS, Dec. 1999, at 6, 6 [hereinafter *Elsevier*]. Responding to the concerns of libraries about the permanent availability of electronic journals, Elsevier Science publicly announced its commitment to archiving all the electronic journals it distributed through its ScienceDirect server. *Id.* Similarly, Ovid Technologies issued its

publisher is actually the only entity currently positioned to preserve the works consistent with the copyright laws because it may choose, as Elsevier Science has done, to give libraries only temporary or contingent possession of things it publishes digitally.⁷⁷ One common subscription arrangement is for all access to current and past issues of a digital journal to disappear if the library discontinues its subscription at any time in the future.⁷⁸ Back issues therefore will remain accessible to the library only as long as the publisher continues to find it worthwhile to preserve and maintain its digital files; for such works, libraries are incapacitated from playing the roles of preservationists and conservators that they have traditionally played in the world of hard copies.

own press release stating that it, too, would archive and provide permanent access to the journals it distributed through Journals@Ovid. *Ovid's Archiving Policy for Electronic Journals*, INFO. INTELLIGENCE ONLINE LIBR. & MICRO-COMPUTERS, Feb. 1999, at 5, 5–7. More recently, the effort to archive such journals has received support from a broad range of entities, including the Andrew W. Mellon Foundation and the Library of Congress. In 2005, JSTOR launched a new service called Portico. See Portico, About Portico, <http://www.portico.org/about/> (last visited Mar. 5, 2007). Public entities are also playing a role in archiving scientific and medical research papers and in making them accessible. The Wellcome Trust, a major funder of biomedical research, conditions grants on making the resulting papers freely available online. Wellcome Trust, Funding: Open Access Is Coming (June 30, 2006), http://www.wellcome.ac.uk/doc_WTX032117.html; Wellcome Trust, Wellcome Trust Position Statement in Support of Open and Unrestricted Access to Published Research (Sept. 20, 2006), http://www.wellcome.ac.uk/doc_wtd002766.html.

77. Elsevier Science offers a variety of licenses; only some permit former subscribers, upon payment of a fee, to enjoy continued access to back issues and even fewer carry the right to archive issues. The various licenses are described at ScienceDirect Info, Licensing and Policies: Primary License Options, <http://www.info.sciencedirect.com/licensing/primary/> (last visited Mar. 5, 2007).

78. Depending on the type of contract, a library subscribing to a digital journal or other resource may obtain perpetual access (comparable to an outright purchase of a book or magazine), or it may instead receive the right to access the work or works in question only so long as the library remains a subscriber. If the subscription is dropped, the library loses access not simply to future installments but to everything. For example, a site license agreement used by Videodiscovery, Inc. (VDY) specifies that if the terms of the license are violated or the subscription expires, the library's copy of the Videodiscovery Digital Library (VDL) "is to be returned to VDY and all on-site files derived from the VDL are to be expunged." Videodiscovery, Inc., Digital Library Site License Agreement 1 (Aug. 6, 2003) (on file with author). Such nonpermanent licenses are in common use. E-mail from Leslie Rich, Associate Director for Technology, New York University Law Library, to author (Mar. 22, 2005, 13:10 EST) (on file with author).

Having the copyright owner take on the role of digital preserver of its own works has some theoretically attractive features. Keeping the work in the form of an electronic file means that the copyright owner could produce and sell physical copies on demand without incurring the costs of making currently unneeded books and journals which then must be stored and protected. But the undertaking (to the extent that owners choose to participate) is likely in most cases to be forward-looking. If new books, journals, and photographs are produced in the first instance in digital form, saving them for the future might prove relatively easy as technology advances.⁷⁹ But few copyright owners are likely to take on the job of comprehensive digital preservation of old works that were originally produced in analog form. They may well lack the resources to scan their full backlists, and, even if resources are not an issue, they may no longer have copies of these older materials.⁸⁰ Also, many if not most of these owners are profit-making entities, and they would need to be convinced that spending their resources this way would generate enough return to be worthwhile—a doubtful proposition given the typically small market for older works.⁸¹

Even if copyright owners voluntarily undertake digital preservation of past, present, and future works, they are not likely to have the sort of firm commitment to preserving their own piece of our cultural history that good public policy requires. Sound archival practices require that multiple copies be kept and stored in a variety of locations,⁸² something owners may prefer not to do. Furthermore, even if owners decide to en-

79. Not all observers share the opinion that over time the cost and difficulty of digital preservation will decrease. Professor Anthony Reese in a recent article speculated that digital preservation might actually turn out to be more costly than preserving analog copies. Reese, *supra* note 62, at 641.

80. One news story quoted the marketing and sales director of a major university press as saying that his organization lacked a full record of what it had published over the years. He noted that “[b]lack in the 50s, 60s and 70s, there were no electronic files for those books.” Anick Jesdanudn, *Google Library Opponents Search Out Copyright Law*, CHI. TRIB., Sept. 19, 2005, at 4 (quoting Tony Sanfilippo of the Pennsylvania State University Press); see also SCHONFELD, *supra* note 28, at 90.

81. See Stephen Breyer, *The Uneasy Case for Copyright: A Study of Copyright in Books, Photocopies, and Computer Programs*, 84 HARV. L. REV. 281, 325 (1970) (estimating that the average tradebook earns whatever its return will be within two years); Reese, *supra* note 62, at 592–94 (noting the quantity of books that go out of print each year).

82. See *infra* notes 199–203 and accompanying text.

gage in digital archiving today, the vicissitudes of tomorrow's market or changes in corporate objectives could lead them to abandon the project—and the existing database.⁸³ The fact that works are actually fixed in digital form at a particular point in time also does not guarantee either eventual accessibility or that the media on which they are stored will survive and remain current.⁸⁴

Finally, archiving by copyright owners could easily result in a cultural record that is too fragmented to be readily usable.⁸⁵ If each entity works independently or even forms small groups, it may use technologies that are incompatible with those of other archives. And, assuming for the moment that the owner will eventually permit open access to its digital database, this approach may leave users in search of a particular kind of information floundering through the unlinked files of dozens of separate entities, each accessible on different terms. This situation is clearly inferior to an archive that combines the work product of many authors and owners and can be efficiently searched across lines of ownership or origination.

2. The Licensee Preservationist

An alternate model that could overcome some of the inefficiencies of the individual owners-as-preservationists model is a collective rights management scheme, which would allow copyright owners to license out the right to archive all of their works. But how such a third-party enterprise, which sounds more like a service to the community of owners rather than to the public, would sustain itself is unclear. Perhaps that role could be played by an entity like Google that could come up with a business plan for using the contents of archives in ways which generate income without harming copyright owners—shall we say by allowing searches of the archive that turn up snippets of books?⁸⁶ Google or some other company might conclude that, even after taking account of the transaction costs of

83. For a similar argument, see Waters, *supra* note 17, at 3.

84. For examples of digitized material that has either been lost or is considered insecure, see Warner, *supra* note 14, at 59–61. Warner notes that one archive disappeared entirely. The Committee on Institutional Cooperation spent six years archiving electronic journals on the Internet before running out of money, after which the archive simply vanished. *Id.* at 60–61.

85. I assume here that the material will not be kept in forms that would make it accessible to search engines, but that each database would have to be accessed separately and with some form of permission.

86. See *infra* notes 130–31 and accompanying text.

obtaining permission from copyright owners, it could make a profit on such a scheme. But what remains debatable is whether the business plan would succeed. Google clearly expects that it will turn a profit from scanning all those books by selling space on its site to advertisers.⁸⁷ Even without the added costs of seeking and negotiating licenses, it is unclear whether the service will, in the long run, prove sufficiently profitable to permit Google to continue providing it.⁸⁸

Nothing about this model, from a public policy perspective, makes it other than marginally more satisfactory as a way to capture the benefits of digitization for the public than the individual owners-as-preservationists model.⁸⁹ If the third-party archivist is a publicly traded company, as Google is, its business decisions, like those of each copyright owner, will necessarily be guided by its financial bottom line. If preservation is not profitable, third parties will act just like copyright owners: they will start cutting corners or may even abandon the project and cease to maintain or upgrade what they have already stored in the database.⁹⁰

In addition, archives operated as purely commercial enterprises can impose serious, permanent impediments to public access, even once the works they contain move into the public domain. If an enterprise has the sole physical copy of an object

87. See *infra* note 206 and accompanying text.

88. The three snippets that a search will produce in any given book (and related bibliographical information on the book itself) clearly is of some help to researchers, but it may not be all that effective a research tool in the end. See About Google Book Search, <http://books.google.com/googlebooks/about.html> (last visited Mar. 5, 2007). If a book contains your search term in ten places, but the three you get to see are not pertinent, you may well be left with the same uncertainty about the value of the particular work for your purposes as when you started the search. Yahoo!, Microsoft, and others are trying to compete with Google by creating their own searchable online databases of books. Anandashankar Mazumdar, *Yahoo and Partners Launch Rival to Google Print, but Seek to Avoid Copyright Issues*, 70 PAT., TRADEMARK & COPYRIGHT J. 630, 630 (2005); *Microsoft Intends to Create Service for Online Library Searches*, 70 PAT., TRADEMARK & COPYRIGHT J. 707, 707 (2005). They will, however, concentrate on public domain works and works where permission has been voluntarily granted, enabling them to provide the entire work to the user—and to avoid the potential cost and hassle associated with proactively seeking licenses from copyright owners. Mazumdar, *supra*, at 630; *Microsoft Intends to Create Service for Online Library Searches*, *supra*, at 707.

89. A centralized archive could achieve certain efficiencies that owner-by-owner archives cannot. A centralized archive would include a uniform set of technological choices to govern all the contents, and avoid artificial segregation of material by owner rather than subject.

90. See *supra* note 84.

or work, it can use its leverage as custodian and gatekeeper to limit access to the work forever in ways that serve its, rather than the public's, best interests.

One of the attractions of digital archiving is that the physical location of an object is no longer a barrier. At least in theory, digital copies could be transmitted anywhere and as often as there are persons who want to examine them. In reality, though, private digital archives are just as capable as physical ones of being used in ways that limit rather than facilitate access. Those who control digital archives can, if they choose, use a variety of techniques, including digital rights management tools, to determine who can see a work, how they can use it, and for how much money. A private digital archive could charge access fees well in excess of marginal cost, even once the work in question is past its copyright term.⁹¹ The monopolization problem is more acute in cases where all physical instantiations of the work have deteriorated and the digital copy is the only one that survives.⁹²

For all these reasons, simply relying on the market is unlikely to serve the interest of the public in the full range of potential benefits that could flow from digital archiving. If, as a society, we want these advantages, some reasonable modifications to the existing copyright bargain will be essential, but they should not solely be in the form of reduced rights for own-

91. This point is illustrated by looking at the debate over the fate of the Bettman Archives. The archives, which contain an irreplaceable photojournalistic record of the first three-quarters of the twentieth century (war images, photographs of Nazi Germany, pictures of celebrities), were bought by Corbis, a corporate entity owned by Bill Gates, in 1995, with the intent to digitize all of them. Stephen Smith, *Scanning the Century*, NEW STATESMAN (London), July 16, 2001, at 42, 42, available at <http://www.newstatesman.com/200107160031>. Because the originals are deteriorating, the digital copies will soon be the only ones remaining. *Id.* Corbis claims copyright in the digital file that holds the photographs once they have been scanned. E-mail from David Green, Corporate Counsel, Corbis Corporation, to Gerald Barnett, Senior Licensing Officer, University of Washington (Jan. 10, 2000, 16:30:56 PST), <http://legalminds.lp.findlaw.com/list/cni-copyright/msg09540.html>. Many archives simply, and inaccurately, also attempt to limit unconsented uses of the material they possess—even when in the public domain—by attaching “copyright” notices, or things that look very much like such notices, to their collections. Jason Mazzone, *Copyfraud*, 81 N.Y.U. L. REV. 1026, 1052–58 (2006).

92. See Smith, *supra* note 91, at 42–43. Some observers have expressed concern that digitization may in fact encourage custodians to allow physical copies to deteriorate or be disposed of. In the case of Corbis and its Bettman Archive photographs, the physical copies are reportedly stored in an underground limestone mineshaft in Pennsylvania, which may be, as a practical matter, the same as losing them. *Id.* at 42.

ers. To the extent that modified copyright law may give new opportunities to would-be archivists, those gains should be offset by attaching a set of responsibilities that archivists must shoulder. The point is not simply to free libraries and others to become serious archivists, but to be sure that the end product is the best possible servant of public knowledge. Enabling archiving, and then letting its practices grow haphazardly, would be unlikely in the long run to serve anyone's interests well.

II. THE COPYRIGHT QUANDARY

When conservation of copyrighted works meant taking care of the physical objects in which intellectual property resided, the frequency of conflicts of rights between third-party conservationists and copyright owners was circumscribed. Now that preservation by digital means is contemplated during the copyright term, however, the law as it currently exists ensures that conflict will occur. Whenever a book, journal, film, painting, or photograph in analog form is digitized or a digital work is backed up, a copy is made. The right to control copying is the core interest of copyright law,⁹³ and thus how copyright law deals with preservation by copying is the single most important factor in whether digital archives thrive or instead never realize their potential. As noted above, the very long term of modern copyright means that the law may still "protect" intellectual property against copying well beyond the point when the work is lost or destroyed for all time.⁹⁴ The copyright statute as currently drafted takes very little account of the problem of preservation during the term of the copyright.⁹⁵

As the current lawsuits against Google suggest, copyright owners are unlikely simply to acquiesce in preservation by digitization, however great the social benefits.⁹⁶ They would be

93. See, e.g., 17 U.S.C. § 106 (2000).

94. See *supra* notes 46–53 and accompanying text.

95. For a discussion of the provisions that currently exist, see *infra* Part II.A–B.

96. McGraw-Hill Co., Inc. v. Google Inc., No. 05-CV-8881, 2005 WL 2778878 (S.D.N.Y. filed Oct. 19, 2005); Author's Guild v. Google Inc., No. 05-CV-8136, 2005 WL 2463899 (S.D.N.Y. filed Sept. 20, 2005). Since the plaintiffs filed these lawsuits in the United States, other plaintiffs have filed suits abroad. Craig Morris, *French Publishing Group Charges Google with Violation of Copyright*, HEISE ONLINE, June 7, 2006, <http://www.heise.de/english/newsticker/news/73944>; John Oates, *French Publisher Sues Google*, REGISTER, June 7, 2006, http://www.theregister.co.uk/2006/06/07/france_sues_google/. In the fall of 2006, the original plaintiff in the French suit was joined by an asso-

concerned with unlicensed copying in any context, but their concern takes on a special edge when the copying is digital.⁹⁷ Copyright owners know that once a work has been converted to digital form, preservation is not the only possibility. Although photocopying and other analog technologies make the proliferation of unlicensed copies far easier than it would have been, say, in the 1920s, digitization means that massive numbers of perfect copies can be made and disseminated quickly and at virtually no cost—a point that the Napster-Aimster-Grokster dispute over peer-to-peer file sharing has amply demonstrated.⁹⁸ Furthermore, even if users do not download or print out their own copies of digitized works, once a copy is lodged on a server, multiple individuals may be able access it remotely. This feature raises the specter of a single central source satisfying the needs of countless users, driving down the number of copies the copyright owner can hope to sell or license.⁹⁹ Fear of

ciation representing four hundred publishing houses. Peter Sayer, *French Publishers Join Google Booksearch Lawsuit*, LINUX WORLD, Oct. 31, 2006, <http://www.linuxworld.com.au/pp.php?id=873767218&fp=2&fpid=1>. A publisher dropped a suit against Google in Germany after a copyright court judge informed the plaintiff that its chances of winning a preliminary injunction to stop Google's scanning project were slim. Posting of David Drummond to Google Blog, <http://googleblog.blogspot.com/2006/06/germany-and-google-books-library.html> (June 28, 2006, 10:45 am). The court, according to reports of the proceedings, found that German law did not apply to Google's actions in the United States, but it hinted that in any case the use of snippets of text would not violate German copyright. Posting of David Drummond, *supra*; *Partial Success for Google in a German Courtroom*, DEUTSCHE WELLE, June 29, 2006, <http://www.dw-world.de/dw/article/0,2144,2073046,00.html>; Richard Wray, *Legal Victory for Google in Library Project*, GUARDIAN UNLIMITED (London), June 29, 2006, <http://business.guardian.co.uk/story/0,,1808770,00.html>. More recently, however, a Belgian court objected to Google's display of snippets from Belgian, French, and German newspapers, and Google has reportedly ceased including them pending an appeal. Associated Press, *Google Posts Ruling of Court in Belgium*, INT'L HERALD TRIB., Sept. 24, 2006, <http://www.iht.com/articles/2006/09/24/business/google.php>; Paul Meller, *Court Halts Google's Belgian Newspaper Plan*, MACWORLD, Sept. 19, 2006, <http://www.macworld.co.uk/news/index.cfm?NewsID=15895>.

97. See Oestreicher-Singer & Sundararajan, *supra* note 35, at 534.

98. These three cases involved peer-to-peer file sharing of music without the permission of copyright owners, activities which the music industry claimed threatened to destroy its ability to sell phonorecords. In each case, the court found for the plaintiffs and against the entities facilitating the file sharing. *Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd.*, 545 U.S. 913, 941 (2005); *In re Aimster Copyright Litig.*, 334 F.3d 643, 656 (7th Cir. 2003); *A&M Records, Inc. v. Napster, Inc.*, 239 F.3d 1004, 1027 (9th Cir. 2001).

99. That actual effect is less clear. Empirical studies attempting to evaluate the economic impact of file sharing on sales of music have reached inconsistent conclusions. Compare Felix Oberholzer & Koleman Strumpf, *The Effect*

these possibilities is why publishers and authors expressed particular outrage after learning that Google agreed to give several of its participating libraries digital copies of their own books.¹⁰⁰

A. SECTION 108: THE LIBRARY EXCEPTION

It is hardly surprising, then, that current copyright law makes only the narrowest concessions to the need for conservation and preservation by someone other than the copyright owner¹⁰¹ and is far more generous toward analog than digital

of *File Sharing on Record Sales: An Empirical Analysis*, 115 J. POL. ECON. 1, 38–40 (2007) (concluding that downloading has no significant impact on album sales), with Stan Leibowitz, *File-Sharing: Creative Destruction or Just Plain Destruction?*, 49 J.L. & ECON. 1, 24 (2006) (finding a close link between changes in record sales and changes in file sharing).

100. In a highly publicized interview published in the *Washington Post*, Patricia Schroeder, President of the Association of American Publishers, told a reporter that publishers are “terrified” of librarians because once a library gets an electronic copy of a work, librarians then share it with other libraries and distribute it to their own users. Linton Weeks, *Pat Schroeder’s New Chapter: The Former Congresswoman Is Battling for America’s Publishers*, WASH. POST, Feb. 7, 2001, at C1, available at <http://www.washingtonpost.com/ac2/wp-dyn/A36584-2001Feb7?>. More recently, an editorial by University of Michigan President Mary Sue Coleman stirred up the publishing community with a defense of Google Library, arguing that copyright represents a compromise among competing values, the most important of which is access “to facilitate the sharing of knowledge, not to stifle such exchange.” Mary Sue Coleman, Editorial, *Riches We Must Share . . .*, WASH. POST, Oct. 22, 2005, at A21, available at <http://washingtonpost.com/wp-dyn/content/article/2005/10/21/AR2005102101451.html>. She added that students in her university, like students all over the country, have learned to depend on digital resources and ignore print ones. *Id.* This article added fuel to the belief that universities would make digitized copies of works available through their libraries once they had possession of them. In February 2006, President Coleman accepted an invitation to address the Association of American Publishers to reassure the members that Michigan intended to respect the copyrights in digitized materials. A video clip of this speech is available at Video: Digitizing the Library, CNET NEWS.COM, <http://news.com.com/1606-2-6036176.html> (last visited Mar. 5, 2007). As noted above, under Google’s recent contract with the University of California system, Google will not provide a separate set of copies to the university’s libraries. See Cooperative Agreement, *supra* note 17, §§ 4.6–4.7.1.

101. For critiques of the narrowness of the so-called library exceptions, see Ann Bartow, *Libraries in a Digital and Aggressively Copyrighted World: Retaining Patron Access Through Changing Technologies*, 62 OHIO ST. L.J. 821, 826–28 (2001); Gasaway, *supra* note 12, at 652–62; Laura N. Gasaway, *Copyright Ownership & the Impact on Academic Libraries*, 13 DEPAUL-LCA J. ART & ENT. L. & POL’Y 277, 292–99 (2003). Professor Gasaway is currently chairing the Section 108 Study Group, appointed by the Librarian of Congress to make recommendations for updating the library and archive exceptions to the Copyright Act. The Study Group is described at Library of Congress, Section

copying. The applicable provisions are contained in section 108 of the Copyright Act, entitled "Limitations on exclusive rights: Reproduction by libraries and archives."¹⁰² This exception to the copyright owner's exclusive right to make copies is limited to libraries and archives that are either public or open to unaffiliated researchers "doing research in a specialized field."¹⁰³ Other kinds of libraries, such as libraries operated by commercial entities for their staffs, do not benefit from these provisions. These provisions also do not seem to apply to a commercial enterprise that wants simply to play the role of preservationist, divorced from the services provided by more traditional libraries.¹⁰⁴ Furthermore, almost all of the provisions of section 108 that permit copying do so only for those works already in the library or archive's collection.¹⁰⁵ It seems fair to say, therefore, that the section as currently written was designed to facilitate use and conservation of library collections and is not generally oriented toward encouraging creation of the kind of comprehensive, permanent repository that this Article envisions.¹⁰⁶

The provisions of section 108 are a complex and highly nuanced set of rules that differ depending on whether the works to be copied are published or unpublished, and on the form in which they exist or are copied.¹⁰⁷ Specific limitations restrict

108 Study Group, <http://www.loc.gov/section108/> (last visited Mar. 5, 2007).

102. 17 U.S.C. § 108 (2000), amended by Pub. L. No. 109-9, § 402, 119 Stat. 218, 227 (2005). The only other provision that expressly allows copying is § 117, dealing with software. *Id.* § 117. This is not, however, a conservation provision but rather one that permits owners of software enough flexibility to use it on their computers and also to keep a single back-up in case of a malfunction. *Id.*

103. *Id.* § 108(a)(2).

104. The legislative history of the provision makes the limitation clear. A commercial entity like a copy shop would not even be privileged under § 108 to make a copy for a library that is itself entitled to make a § 108 copy. See H.R. REP. NO. 94-1476, at 74 (1976), reprinted in 1976 U.S.C.C.A.N. 5659, 5688. It is unlikely, therefore, that Google could use § 108 by claiming the status of a library or an archive.

105. For a discussion of the most important of these exceptions, see *infra* notes 107–25 and accompanying text. Another exception permits libraries to add to their collections by making copies of off-the-air news broadcasts. 17 U.S.C. § 108(f)(3).

106. Because the Section 108 Study Group is currently considering changes to the law, it is possible that the ability of libraries and archives to fulfill the preservation role contemplated by this Article will improve. See *supra* note 101.

107. Libraries cannot duplicate films, musical works, or pictorial, graphic or sculptural works, for example, for patrons. See 17 U.S.C. § 108(i).

how digital copies can be used. If an eligible institution owns unpublished works or phonorecords, it is entitled to make up to three copies of them in any format “solely for purposes of preservation and security.”¹⁰⁸ The institution may give one or more of the copies to other publicly accessible libraries or archives for research use.¹⁰⁹ However, if the copies are in digital form, they can only be used by the library’s patrons on its premises and cannot be accessed remotely.¹¹⁰

The restrictions on reproducing published works are considerably more stringent. The library or archive again is entitled to make three copies of a work, but only if the original is damaged, deteriorated, lost, stolen, or exists in an obsolete format.¹¹¹ A library or archive cannot do this, however, unless it first determines, “after a reasonable effort,” that a new copy cannot be obtained on the market at “a fair price.”¹¹² Copying purely for purposes of preservation is not permitted. Here, too, if the institution chooses to make copies in digital form, the copies can only be used on the library or archive’s physical premises.¹¹³

Congress understood that the addition of twenty years to the term of copyrights provided by the 1998 Sonny Bono Term Extension Act¹¹⁴ could exacerbate the problems libraries face in keeping their collections complete and usable.¹¹⁵ Therefore, the

108. *Id.* § 108(b).

109. *Id.*

110. *Id.* § 108(b)(2).

111. *Id.* § 108(c). A format is considered obsolete only if the device that would be needed to access the format is “no longer manufactured or is no longer reasonably available in the commercial marketplace.” *Id.* § 108(c)(2). A covered library can also make a copy of an entire work for its own patron or for one at another library if the copy is intended solely for personal use. *Id.* § 108(d)(1). But this exception applies only if the library, after a “reasonable” search, concludes that the work in question is not available on the market “at a fair price.” *Id.* § 108(e).

112. *Id.* § 108(c)(1). The statute does not define these terms, although the legislative history provides some indication of what a “reasonable” search might entail. *See, e.g.*, H.R. REP. NO. 94-1476, *supra* note 104, at 76.

113. 17 U.S.C. § 108(c)(2). Professor Gasaway points out an oddity in this provision. Increasingly, libraries obtain works in digital form and the accompanying license commonly permits remote access to them. The statute as written, however, suggests that if a copy of a digital work is made for replacement purposes, the copy could no longer be accessed remotely, even if the original could have been. Gasaway, *supra* note 12, at 656–57.

114. Pub. L. No. 105-298, § 103, 112 Stat. 2827, 2863–76 (1998) (codified as amended in scattered sections of 17 U.S.C. (2000)).

115. The Copyright Office notes that this part of § 108 is sometimes re-

Act gave qualifying libraries and archives the right to copy published (and unpublished) works, both for preservation purposes and for “scholarship, or research.”¹¹⁶ To use the exemption, however, the library or archive must first conduct a “reasonable investigation” that determines the work in question is neither available at a “reasonable price” nor is currently subject to commercial exploitation.¹¹⁷ Copyright owners can forestall exercise of the privilege by an opt-out in the form of a notice averring that the work is still available at a reasonable cost or that it is continuing to be exploited.¹¹⁸ If libraries can overcome these preliminary hurdles, they can make either digital or analog copies of published works, and there are few limits on how, where, or for what purpose they can be accessed once made.¹¹⁹ Libraries can apparently even use this provision to add entirely new works to their collections.¹²⁰

Section 108 is not only extremely complicated, but it also has some significant drawbacks as a means of facilitating preservation for the long term. The preliminary fact determinations on availability and cost that libraries must make prior to copying could prove to be cumbersome and, if a copyright owner questions any specific determination, expensive. Since section 108 does not spell out what is reasonable, the library or archive that takes advantage of the exemption provision faces the pos-

ferred to as an “orphan works” provision and adds that it was “[i]ntended to ameliorate the effects that the 20-year extension of term might have on libraries and archives in their use of older works” ORPHAN WORKS, *supra* note 34, at 45.

116. Apparently, Congress concluded that it could make the adjustment for libraries and archives covering that twenty-year period because doing so would raise no problems obligations under the Berne Convention. *See* Orphan Works, 70 Fed. Reg. 3739, 3742 (Jan. 26, 2005). Since Berne only demands protection for works for fifty years from the death of their authors, not seventy, Congress deemed that limiting the normal rights of owners to exploit works during the twenty year add-on was acceptable. *See id.* For a discussion of Berne, see *supra* notes 46–50 and accompanying text.

117. 17 U.S.C. § 108(h)(2)(A)–(B).

118. *Id.* § 108(h)(2)(C). According to the Copyright Office, however, as of January 2006, no copyright owner has ever invoked this provision. ORPHAN WORKS, *supra* note 34, at 46.

119. The statute permits reproduction, distribution, display, and performance of the works in question. 17 U.S.C. § 108(h)(1). It only excludes the right to make derivative works. *See id.*

120. The fact that libraries can make copies for purposes of scholarship and research and can distribute them without restriction suggests that libraries will be able to add works to their collections and not merely save the ones they already have. *See id.*; Gasaway, *supra* note 12, at 661.

sibility that a court will later decide either that the search for a commercially available copy was inadequate or that the price for which such a copy could be obtained on the open market was actually reasonable. Furthermore, the fact that a library's access to many of its digital journals disappears when the subscriptions end means that long-term preservation of such works is exceedingly problematic.

Also, section 108 does not address the specific requirements of sound archiving practices. An archive's value is greatly increased by its comprehensiveness. In order to preserve a certain category of journals, for example, the archive should contain copies of every issue of those journals ever published; similarly, collections of certain authors or kinds of books increase in usefulness the less patchy they are. Individual institutions, however, often have holes in their collections resulting not only from loss or destruction but from a failure to acquire something when it originally became available. Even with the benefit of the more lenient rules that apply during the last twenty years of copyright, an archiving library may find that works needed to fill out a collection have disappeared from circulation altogether by the time the provision kicks in. If we want an institution to embrace the role of preserving a cultural record, then the legal system must allow a more realistic way for the institution to fill such holes while it can still find the material at issue. Section 108 does not serve that role: during the life of the author plus the first fifty years after her death, it helps save only those works that are already a part of an institution's existing collection.¹²¹ From the language of the statute, it is not even clear how Congress expects a library that suffers the loss of a volume during that period actually to make the replacement copy—unless it already owns a second one from which it can make the needed reproduction.¹²²

A final problem is the existence of technological barriers that make copying some sources difficult. Although copyright owners themselves are increasingly making their new works available in digital form, they are also protecting them, not just

121. See *supra* notes 107–13 and accompanying text.

122. In reality, the libraries often rely on interlibrary loans to get the copies they need. But the text of the statute does not make explicit that this is permitted. The subsection dealing with interlibrary loans does not expressly deal with the possibility of borrowing to make a replacement copy; the provisions are geared toward using interlibrary loans to satisfy the requests of patrons. See 17 U.S.C. §§ 108(e), (g)(2).

with restrictive licensing terms, but with various forms of digital rights management (DRM) technologies that make them technically difficult to copy.¹²³ A library or archive that wants to make a copy of such a file for preservation purposes might first need to hack around the DRMs that protect the work, which would violate the Digital Millennium Copyright Act (DMCA).¹²⁴ Currently, one of the few exceptions to liability under the DMCA prescribed by the Copyright Office is to allow libraries to hack access controls on works in their collections that have become inaccessible because the devices protecting them are defective, damaged, or obsolete.¹²⁵ This limited right is scarcely comprehensive enough to allow an institution to make archival copies of all its digital works, even if copyright law were otherwise more favorable than it is now.

B. THE "GOOGLE PROBLEM"

Google, with its enormous actual and potential resources, is a new player in the digital preservation arena, but lacking any harbor under section 108,¹²⁶ it faces serious copyright obstacles to its current plans. Google decided to undertake the massive job of scanning into a searchable database the entire contents of one major academic library, some copyrighted and public domain works from four others, and public domain works from several more.¹²⁷

123. DRMs commonly protect movies and music downloaded from the Internet (or even purchased on compact disks or digital video disks) against unpermitted copying. See, e.g., Dierdre K. Mulligan et al., *How DRM-Based Content Delivery Systems Disrupt Expectations of "Personal Use,"* in PROCEEDINGS OF THE 3RD ACM WORKSHOP ON DIGITAL RIGHTS MANAGEMENT 77, 77 (2003) (discussing the use of DRMs to protect both film and music); Martin Peitz & Patrick Waelbroeck, *An Economist's Guide to Digital Music*, 51 CE-SIFO ECON. STUD. 359, 359 (2005) (discussing the use of DRMs in the music industry).

124. Pub. L. No. 105-304, 112 Stat. 2863 (1998) (codified as amended at 17 U.S.C. §§ 1201–1205 (2000)).

125. The DMCA authorizes the Copyright Office to promulgate regulations that exempt classes of works from the proscription on circumventing access controls where failure to do so would adversely affect the rights of users to make legitimate uses of the underlying work. 17 U.S.C. §1201(a). The rule referenced in the text is 37 C.F.R. § 201.40 (2006).

126. See *supra* notes 103–06 and accompanying text.

127. The company initially planned to scan the University of Michigan's entire book collection, parts of the collections at Harvard and Stanford (both copyrighted and public domain), and public domain holdings of the New York Public Library and the University of Oxford. Press Release, Google Press Ctr., Google Checks out Library Books (Dec. 14, 2004), <http://www.google.com/>

From the perspective of creating an archive, there is good reason to commend the plan. Because not every library contains the same works, or the same editions of works, drawing on collections across institutional lines allows Google to tap into riches that no single collection, however fine, could hope to offer. After all, minimization of the number of “blank spaces” is (and ought to be) a primary goal of such a project.

Once Google has made a digital copy of each work, what Google does with the copies will depend on the works’ copyright status. Books that are in the public domain will be put online in a form that makes them both searchable and accessible in full text.¹²⁸ Works still covered by copyright (actually or putatively—public domain status can sometimes be hard to be prove)¹²⁹ will be searchable, but not accessible except in the most limited way.¹³⁰ Users will be able to retrieve only three small snippets of text from these works—enough to provide the context within which search terms appear.¹³¹

If Google can carry out this plan, it will certainly have made the most credible start to date on creating a contemporary equivalent of the legendary Alexandrian library. The company has, however, the same potential problem that other middlemen archivists face—it neither owns the copyrights in the books it plans to scan nor has, in most cases, even a limited li-

press/pressrel/print_library.html. Recently, the University of California libraries and the University of Virginia also agreed to allow Google to digitize portions of their collections as well, including copyrighted works. UC System Will Join Google’s Scan Plan, *supra* note 19; U. Va. Press Release, *supra* note 19. More recently still, the University of Wisconsin-Madison and the University Complutense of Madrid have joined forces with Google, but will contribute public domain material only. See Press Release, Univ. of Wisc.-Madison, *supra* note 19; Press Release, Google, *supra* note 19.

128. Google Book Search Common Questions, <http://books.google.com/googlebooks/common.html> (last visited Mar. 5, 2007).

129. See Elisabeth Townsend Gard, Unpublished Work and the Public Domain: The Opening of a New Frontier 53–62 (unpublished manuscript, on file with author) (describing difficulties in determining the status of old works). Google itself apparently presumes that all works published after 1922 are copyrighted, even though many are not. See Google Book Search Common Questions, *supra* note 128.

130. Google Book Search Common Questions, *supra* note 128.

131. Google’s Search program allows a user to see no more than three snippets of any given copyrighted text, although the program also tells the user how many times the search term appears in the work. See Google Book Search Library Project, *supra* note 4. Searches will be monitored in some way so that canny and persistent users cannot piece together an entire text simply by running repeated searches. See Google Book Search Common Questions, *supra* note 128.

cense to copy them just once.¹³² Google, of course, intends to make more than one copy. In addition to retaining its own copies, Google has offered to supply several of its participating libraries with digital copies of any works they contribute to the database.¹³³ Whatever legal problems copyright poses for Google in creating searchable copies for its own use pale in comparison to the problems generated by distributing copies of protected works to other entities.

Section 108 aside, copyright law is not rich in exemptions that give anyone other than a copyright owner a right to copy or to copy and distribute entire works, however beneficent the purpose. Google seems to have two different approaches to justifying this massive amount of copying (although not necessarily its redistribution of copies to cooperating libraries). First, Google offered publishers an opportunity to “opt out” of having their works included,¹³⁴ suggesting it might make a novel—if not especially compelling—argument that those that do not opt out have granted Google an implied license to reproduce their works.¹³⁵

The second approach is fair use. Fair use, as codified in section 107 of the Copyright Act, is a defense available under certain circumstances to commercial, nonprofit, and individual

132. Indeed, if Google is ultimately found to be in violation of the copyright laws, it may be required to destroy that part of what it has already created that contains copyrighted works. *See* 17 U.S.C. § 503 (2000).

133. *See supra* notes 18–19. It appears, however, that only Google will have the full set of all scanned documents. *See* Quint, *supra* note 17. This makes the data less secure than it would be if the libraries at Michigan, Stanford, Harvard, Wisconsin, and elsewhere were instead made repositories of full copies of the entire Google database, with periodic updates and a formal responsibility to act as caretaker-agents for the public at large. Without a binding commitment of this sort, the database as a whole could be lost if Google chooses (or is forced by copyright liability) to abandon the project. The agreement with the University of California, as noted earlier, provides that the university will have access to copies of its works only through Google. *See* Cooperative Agreement, *supra* note 17, §§ 4.6–4.7.1.

134. The company declared a moratorium on scanning copyrighted works from August 2005 to November 2005, during which time copyright owners could submit lists of works that Google did not have permission to scan. Jeffrey R. Young, *Google Answers Complaints About Project to Scan Millions of Books, but Publishers Are Not Won over*, CHRON. HIGHER EDUC., Aug. 12, 2005, <http://chronicle.com/free/2005/08/2005081201t.htm>. For Google’s announcement to publishers on this subject, see Google Book Search Publisher Questions, http://print.google.com/googlebooks/publisher_library.html (last visited Mar. 5, 2007).

135. *See* Young, *supra* note 134.

actors alike.¹³⁶ It is to this provision that Google—which clearly does not qualify as a noncommercial archive or library under section 108—has turned for its primary defense of the legitimacy of the Google Library project.¹³⁷ The claim that showing merely three or fewer snippets of text in response to a user-initiated search is “fair” is a strong one. A few snippets clearly are not a substitute for the original. Furthermore, searchability is an invaluable benefit to the public that publishers currently fail to provide and are unlikely to provide in the future because of the vast degree of coordination and investment required to do so.¹³⁸

The bigger problem for Google Library is not its output but rather its input. Conducting searches and displaying snippets require Google not only to copy entire works but also to keep those works intact in its database.¹³⁹ This particular relationship between input and output is a function of the specific way Google plans to operate its Library.¹⁴⁰ But, in truth, any archive that includes copyrighted works without the copyright owner’s permission will have problems. Preservation archiving, by its nature, requires copying, and winning the argument that such copying is a fair use will be an uphill battle.

The most basic of the bundle of rights held by the owner of a copyright is the exclusive right to make copies of the entire protected work.¹⁴¹ Defendants, particularly commercial ones, who usurp that right rarely find refuge under section 107. However, courts sometimes recognize exceptions to this general principle. The most significant of these deviations appeared in

136. 17 U.S.C. § 107.

137. See Google Book Search: News and Views, Legal Analysis, <http://books.google.com/googlebooks/newsviews/legal.html> (last visited Mar. 5, 2007).

138. As Professor Wendy Gordon argued, the most unambiguous case for a finding of fair use exists when a bargain allowing a socially beneficial use to go forward cannot be struck, either because of intransigence on the part of a copyright owner who opposes the use (say, in a parody case) or because transaction costs are a barrier. Wendy J. Gordon, *Fair Use as Market Failure: A Structural and Economic Analysis of the Betamax Case and Its Predecessors*, 82 COLUM. L. REV. 1600, 1632–33, 1645 (1982). The kind of project Google contemplates may be just such an instance.

139. See Google Book Search Publisher Questions, *supra* note 134 (explaining that for copyrighted books, Google shows only short snippets of text but going on to admit that in order to index a book, Google needs to make a copy of the book).

140. See *id.* (implying that Google’s model is merely an extension of the company’s method of indexing web pages).

141. 17 U.S.C. § 106(1).

Sony Corp. of America v. Universal City Studios, Inc., in which a bare majority of the Court was willing to find noncommercial copying a fair use because it was only temporary.¹⁴² The preface to section 107 also recognizes the possibility of exceptions to the owner's exclusive right to make copies.¹⁴³ Clearly, both *Sony* and section 107 demonstrate that the duplication of entire copyrighted works is not per se excluded from fair use. Several courts have even extended the privilege to cases in which a commercial entity has made a full copy as a necessary¹⁴⁴—but temporary—step toward creating a permissible (that is, non-infringing) new work, such as interoperable software or hardware.¹⁴⁵ However, when the copy is persistent, as is the case with digital archives, courts generally deny fair use. The *Napster* court, for example, refused to find that downloading sound recordings onto a computer hard drive for personal use was fair use under section 107.¹⁴⁶ In another case, an innovative online music service copied thousands of CDs into a database and offered subscribers remote access to any CD the subscribers could prove they already owned.¹⁴⁷ The district court shut down the service on the grounds that the copying involved in creating the database was unexcused infringement.¹⁴⁸

Fearing that copyright owners will lose economic control over their works, courts have been hesitant in most instances to encourage the innovative application of new technologies to copyrighted works—especially by for-profit companies—unless the content is first licensed.¹⁴⁹ The most notable exception to

142. 464 U.S. 417, 421 (1984) (5-4 decision) (finding no infringement where the viewer “time-shift[ed]” television broadcasts by taping them for a later viewing).

143. 17 U.S.C. § 107 (providing the making of “multiple copies for classroom use” as an example of fair use).

144. See *Sony Computer Entm't, Inc. v. Connectix Corp.*, 203 F.3d 596, 602 (9th Cir. 2000) (explaining that computer code has protected and unprotected elements, and that because it is invisible to the naked eye, sometimes copyrighted material must be copied in order to access unprotected material).

145. See, e.g., *id.* at 599 (finding that copies made as an intermediate step toward achieving compatibility is fair use); *Sega Enter. v. Accolade, Inc.*, 977 F.2d 1510, 1514–15 (9th Cir. 1992) (finding that copies made as an intermediate step in creating a non-infringing product are fair use).

146. *A&M Records, Inc. v. Napster, Inc.*, 239 F.3d 1004, 1014–15 (9th Cir. 2001).

147. *UMG Recordings, Inc. v. MP3.com, Inc.*, 92 F. Supp. 2d 349, 349 (S.D.N.Y. 2001).

148. *Id.* at 350.

149. See, e.g., *Video Pipeline, Inc. v. Buena Vista Home Entm't, Inc.*, 342 F.3d 191, 194–95, 207 (3d Cir. 2003) (unofficial Internet movie trailers); *Bates-*

this practice, and the strongest support for Google, is *Kelly v. Arriba Soft Corp.*¹⁵⁰ Kelly challenged a search engine that looked for digitized images resident on the web.¹⁵¹ Arriba Soft made a database of thumbnail reproductions of these images to display in response to search queries.¹⁵² The user could then click on the tiny image and be linked to the website on which the original, full-sized picture appears.¹⁵³ Cognizant of the enormous value that was added to the Internet by search engines that allowed individuals to locate content, the Ninth Circuit refused to find that copying to create the thumbnails was necessarily infringing.¹⁵⁴ Although in *Kelly* users could see the entire image, the court said that the fact that the reproduction was at a lower resolution than the original made it unlikely that the copy would be a substitute for the original.¹⁵⁵ In one sense the leap between *Kelly* and Google is short: if showing the entire image is fair use, then the display of a mere three lines or so of text should be as well.¹⁵⁶ On the other hand, the copyright owner whose images were reproduced by Arriba Soft had already made its work freely accessible to users on the Internet, and the argument of implied consent in such a case seems stronger than in a case where the owner has not done so.¹⁵⁷

Although I personally believe that Google's treatment of copyrighted texts is a fair use, the matter is by no means clear.

ville Servs. v. Funeral Depot, Inc., No. 1:02-CV-01011-DFH-TA, 2004 WL 2750253, at *5–8 (S.D. Ind. Nov. 10, 2004) (online coffin sales). *But see* Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 431 (1984) (calling for “deference to Congress when major technological innovations alter the market for copyrighted materials”).

150. 336 F.3d 811 (9th Cir. 2003).

151. *Id.* at 815.

152. *Id.*

153. *Id.* at 815–16.

154. *Id.* at 820, 822.

155. *Id.* at 821–22. *But cf.* Perfect 10 v. Google, Inc., 416 F. Supp. 2d 828, 851 (C.D. Cal. 2006) (concluding that Google could not take advantage of the fair use exception because the thumbnails in question competed directly with reduced-size images marketed by the plaintiff for download onto cell phones).

156. *See Kelly*, 336 F.3d at 821 & n.37 (explaining that transformative, low quality works like those produced by Arriba Soft do not compete with, and may actually promote, the original).

157. *See Harper & Row, Publishers, Inc. v. Nation Enters.*, 471 U.S. 538, 551 (1984) (explaining that “factors such as implied consent through *de facto* publication on . . . dissemination of a work may tip the balance of equities . . .,” but going on to say that “[p]ublication of an author’s expression before he has authorized its dissemination seriously infringes the author’s right to decide when and whether it will be made public”).

First, the fair use argument in favor of Google would certainly not apply to archives that make the full text of works accessible to users.¹⁵⁸ And it is by no means clear whether a court would find some theory to impose liability on a defendant like Google if, for example, a hacker were to break into its database and “free” some or all of the copyrighted works contained therein.¹⁵⁹ The fact that Google intends to give additional copies of protected works to several of its participating libraries¹⁶⁰ both makes the data more susceptible to hackers¹⁶¹ and also dilutes Google’s control over the use of the database. If one of the additional custodians of the digitized copies should decide in the future to grant greater access to the works than Google now contemplates, fair use might no longer shield the company,¹⁶² and Google itself might bear some secondary liability.¹⁶³

Why, a reader might ask, must a would-be archivist rely on fair use? Would it not be simpler, not to mention legally preferable, to start by getting permission? The answer is no for several reasons. First and foremost, many copyright owners might simply be unwilling to have digitized copies of their works in the hands of anyone other than themselves or a particularly trusted agent.¹⁶⁴ Second, assuming that Google’s opt-out plan

158. See *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 591–92 (1994) (describing the difference between transformative and superseding works, and further saying that superseding works are unacceptable because they seek to fulfill the same needs as the original and to usurp its market demand). This distinction can explain Google’s loss in *Perfect 10*. See 416 F. Supp. 2d at 851 (“Google’s use of thumbnails likely *does* harm the market for [cell phone images]. . . . [U]sers will be less likely to purchase the . . . content . . .”).

159. JSTOR has experienced at least one such hijacking of its contents. Mark Cain, *Cybertheft, Network Security, and the Library Without Walls*, 29 J. ACAD. LIBRARIANSHIP 245, 245 (2003) (relating how cyberthieves broke into the JSTOR database and illegally downloaded fifty thousand articles, which represented about five percent of the database’s contents at the time).

160. See *supra* note 19.

161. See Cain, *supra* note 159, at 247 (explaining that the breach of JSTOR’s database was executed by exploiting security weaknesses of subscribing colleges’ networks).

162. See *supra* note 158 and accompanying text.

163. See *Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd.*, 545 U.S. 913, 929–30, 931 & n.9 (2005) (finding a strong argument for applying a theory of vicarious liability to the creators of a file sharing program that allowed users to violate plaintiffs’ copyrights, but ultimately reaching its decision on other grounds).

164. See Dan L. Burk, *Muddy Rules for Cyberspace*, 21 CARDOZO L. REV. 121, 124–25 (1999) (“[O]wners of copyrighted works, fearing that digital media will foster additional inroads on their entitlements have lobbied tirelessly for protection and expansion of their ownership rights.”).

is, as I suspect, inadequate, the transaction costs the company would incur in seeking affirmative approval would likely be enormous. Many owners would refuse to license their work, and many others would be impossible to track down.¹⁶⁵ Although no firm figures are available, the estimate of the number of orphan works—those for which permission would be virtually impossible to obtain because the current rights-holder cannot be located¹⁶⁶—is large enough that the Copyright Office has recommended changes in copyright law that would free them for use.¹⁶⁷ The more time that passes from the initial distribution of a work, the harder it becomes to trace the chain of title of any but the most unusually successful works.¹⁶⁸ Google says that if it had to seek permission, it would be unable to include an enormous percentage of copyrighted books in its database.¹⁶⁹

In sum, current copyright law does not make explicit provision for any but the most limited forms of digital archiving. Acquiring permission to do more is cumbersome and difficult. Copying full works for preservation purposes without obtaining permission is problematic, even if the resulting archive is kept dark (that is, inaccessible to the public). Supposing, however, that a dark archive might be deemed fair use, even limited intentional or accidental exposure to some portion of the public could strip its creators of protection. And adding any element of user access compounds the problems the archivist faces.¹⁷⁰ These uncertainties make going forward a risk that only a very

165. See *id.* This problem is accompanied by the fact that rights today are often divided, forcing the would-be archivist to parse through contracts to identify which party or parties need to grant permissions. See 17 U.S.C. § 201(d)(2) (2000) (allowing various exclusive rights of copyright owners to be transferred separately, and thus effectively making each such transferee a “copyright owner”).

166. ORPHAN WORKS, *supra* note 34, at 1.

167. *Id.* at 127 (recommending statutory language). The 109th Congress considered legislation to implement the recommendations. H.R. 5439, 109th Cong. § 2(a) (2006). The proposed bill would limit remedies available to owners of orphan works. *Id.* The bill will likely be introduced shortly in the 110th Congress. National Film Preservation Board, Legislative Initiatives of Interest to Moving Image Preservation Community, <http://www.loc.gov/film/legislation.html> (last visited Mar. 5, 2007).

168. See Olive Huang, Note, *U.S. Copyright Office Orphan Works Inquiry: Finding Homes for the Orphans*, 21 BERKLEY TECH. L.J. 265, 267–72 (2006).

169. Steve Seidenberg, *Copyright Clash*, INSIDE COUNSEL, Nov. 2006, http://www.insidecounsel.com/issues/insidecounsel/15_231/ip/724-1.html.

170. See Cain, *supra* note 159, at 246–47 (exploiting network security challenges, and discussing how the addition of remote user access can compromise or complicate them).

wealthy corporation can take. Moreover, present copyright law does nothing to ensure that archiving will proceed in a way that adequately protects the public's long-term interests. The questions posed, therefore, are: Can the law be changed fairly and effectively, and, if so, how?

III. CAN THE COPYRIGHT QUANDARY BE RESOLVED?

Any effort to change the law to enable the public to reap the enormous benefits of a sustainable archive would have to confront several questions. First, what kinds of entities should be empowered to engage in archiving? Should we extend the privilege to keep these cultural records only to the kinds of institutions—public and publicly accessible libraries and repositories—that are currently the beneficiaries of section 108? Or should the list of eligible candidates be expanded, and if so, how? Second, what duties should public law impose to ensure the fullest possible realization of the social benefits of preservation in return for the right to archive? Third, what is the fairest and most appropriate way to compensate copyright owners for the decrease in their control over their works that archiving rights would cause? Any response to this third question must take into account the most troublesome issue long-term preservation raises: whether some form of public access to works in archives should be permitted during the copyright term.

In attempting to deal with these questions, I am greatly aided by the experience of some model projects that are the brainchildren of the Andrew W. Mellon Foundation—in particular, one called JSTOR.¹⁷¹ The archive of scholarly academic journals in JSTOR—which is now an independent nonprofit entity¹⁷²—first became available to the public in 1997.¹⁷³ JSTOR's aim is to centralize the job of archiving complete, curated collections of journals in specific subject areas and then to make the digital versions of the journals available to individual libraries.¹⁷⁴ JSTOR attempts to create back files of journals with

171. This Article has benefited greatly from the existence of an extensive, detailed history of the JSTOR project, SCHONFELD, *supra* note 28, which was conducted in cooperation with the Mellon Foundation.

172. *Id.* at xvi.

173. *Id.*

174. JSTOR: The Need for JSTOR, <http://www.jstor.org/about/need.html> (last visited Mar. 5, 2007).

no missing articles or pages and no skipped years or months,¹⁷⁵ in a form that will not occupy ever-increasing amounts of shelf space in each subscribing library.¹⁷⁶ Although JSTOR's model cannot necessarily be generalized,¹⁷⁷ its experience is nonetheless both informative and illustrative of the benefits that accrue from creating centralized, curated digital archives.

A. WHO SHOULD BE PERMITTED TO SAVE THE PAST?

JSTOR's experience archiving scholarly journals suggests that even if Congress modified section 108 to allow libraries to engage in serious preservation, the change would be insufficient to allow modern technology to fully realize its potential for preserving culture. Certainly library professionals know a great deal about collection-building, preservation, and the technologies of modern communications. But for a variety of reasons, they can only operate within the confines of their own institutions, missions, and budgetary constraints. Although libraries are well-suited to act as digital archivists in some situations,¹⁷⁸ different kinds of entities (of the sort represented by JSTOR or even Google) may be equally or even better positioned to play the role of preservationist.

For one thing, some third parties might bring greater economic resources to the table than are available to individual libraries, or they might be able to spread the cost of archiving across numerous institutions.¹⁷⁹ Furthermore, a large independent entity might achieve economies of scale that would make feasible an otherwise daunting preservation project.

175. JSTOR: The Production Process, <http://www.jstor.org/about/process.html> (last visited Mar. 5, 2007).

176. JSTOR: The Need for JSTOR, *supra* note 174.

177. While JSTOR does obtain a nonexclusive license to copy works, JSTOR: The Production Process, *supra* note 175, it does so from a sector of the copyright community that may be more willing to grant such licenses. See Posting of Peter B. Boyce to liblicense-l@lists.yale.edu (Apr. 30, 2001, 16:10:02 EDT) (explaining that the reason why the American Astronomical Society has such liberal policies on sharing articles is "to further the distribution of knowledge").

178. SCHONFELD, *supra* note 28, at 365 (estimating that prior to JSTOR as many as two hundred libraries in the United States listed archiving as a major part of their missions).

179. See JSTOR: The Need for JSTOR, *supra* note 174 (explaining that while "it is not less expensive for a single library to convert paper to digital formats for the purpose of freeing up shelf space," in JSTOR's model "the costs can be shared [and] the savings captured").

One need look no further than Google for an illustration of what non-library archivists can contribute. First, Google is a major corporation with assets in excess of \$10 billion,¹⁸⁰ which allows it to devote resources well beyond the means of any single nonprofit library—or even any consortium of libraries—to such projects.¹⁸¹ While estimates of how much Google intends to spend vary,¹⁸² the company intends to scan the University of Michigan's entire seven-million volume print collection.¹⁸³ The University estimated that the cost of doing the job itself would be somewhere between \$600 million and \$1 billion.¹⁸⁴

The size of the project made it worthwhile for Google to acquire improved scanning technology that is far superior to the technology in ordinary use: it can copy books in just a few days without unbinding them and without causing wear and tear beyond what would be caused by the simple act of reading them.¹⁸⁵ The company expects to complete the Michigan scanning project in six years.¹⁸⁶ The university, by contrast, estimated that—using the equipment available to it—scanning seven million books would take more than a thousand years.¹⁸⁷

180. GOOGLE ANNUAL REPORT 42 (2005), http://investor.google.com/pdf/2005_Google_AnnualReport.pdf (listing Google's total assets as of December 31, 2005 at \$10,271,813,000).

181. In an interview, Stanford librarian Michael Keller said that financial and technical barriers have prevented the university from making headway in digitizing its own book collection. Barbara Quint, *Google and Research Librarians Launch Massive Digitization Project*, NEWSBREAKS, Dec. 20, 2004, <http://www.infotoday.com/newsbreaks/nb041220-2.shtml>.

182. Compare Arlette Grouner, *Books Scanned into Virtual Library*, INFO. ONLINE, Apr. 17, 2005, http://www.iol.co.za/index.php?art_id=qw113724981595B251 (estimating a cost of \$7 per book scanned), with *Google Partners with Libraries in Massive Digitization Project*, AM. LIB. ONLINE, Dec. 17, 2004, <http://www.ala.org/ala/online/currentnews/newsarchive/alnews2004/december2004ab/google.cfm> (estimating \$10 per book).

183. Quint, *supra* note 181.

184. Mike Wendland, *U-M's Entire Library to Be Put on Google*, DETROIT FREE PRESS, Dec. 14, 2004, at 1A.

185. Harvard University Library, *supra* note 19 (“[T]he Google scanning process is much gentler with books than other high-speed processes in use today.”). The development of this technology suggests the potential for progress in digital archiving, which might be fulfilled if legal barriers are reduced and investments in such technology are made worthwhile.

186. Google Book Search: News and Views, History of Google Book Search, <http://books.google.com/googlebooks/newsviews/history.html> (last visited Mar. 5, 2007).

187. UM Library/Google Digitization Partnership FAQ (Aug. 2005), <http://www.lib.umich.edu/staff/google/public/faq.pdf>. Similarly, Google estimated that it would take the university one thousand years to digitize all its

Unlike Google, JSTOR is a modest organization with an annual budget of about \$10 million.¹⁸⁸ But it has shown that, as a centralized archivist, it is well-positioned both to attract foundation grants for its work and to spread a significant portion of its costs across a large pool of beneficiaries; subscription fees from over three thousand libraries support JSTOR.¹⁸⁹ In this way, JSTOR has been able to stay afloat¹⁹⁰ and accomplish a task that no single library could easily duplicate.¹⁹¹ Unlike Google (at least at present), JSTOR has a clearly defined curatorial mission, focused on storing complete sets of defined classes of journals.¹⁹² Currently, JSTOR has more than six hundred journals from forty-seven disciplines.¹⁹³ If JSTOR finds gaps in a given journal's back issues, it often can fill them by asking subscribing educational institutions to search for the missing bits in their hard copy collections and to allow JSTOR to scan them.¹⁹⁴

The examples provided by Google and JSTOR suggest that any legislative provision for digital preservation of copyrighted works would ideally open the door to a variety of players in addition to libraries, each of which could bring different strengths

books. Google Book Search, *supra* note 186.

188. Scott Carlson, *JSTOR's Journal-Archiving Service Makes Fans of Libraries and Scholars*, CHRON. HIGHER EDUC., July 27, 2001, at A26, available at <http://chronicle.com/free/v47/i46/46a02601.htm>.

189. JSTOR: Facts and Figures, <http://jstor.org/about/facts.html> (last visited Mar. 5, 2007) (claiming 1696 American and 1507 foreign participating institutions as of December 22, 2006).

190. SCHONFELD, *supra* note 28, at 364 (detailing how increasing quality brought in more membership, leading to a stable budget).

191. JSTOR considers as one of its major achievements the elimination of non-archiving libraries free-riding on the efforts of the two hundred or so other libraries that did undertake the task of archiving. *Id.* at 169.

192. JSTOR: The Need for JSTOR, *supra* note 174.

193. JSTOR: Facts and Figures, *supra* note 189. Although JSTOR has included some biology and health journals, it does not include the extremely expensive hard-core science journals, and it is unclear whether the publishers of such journals would cooperate. However, access to a wide range of biomedical journals will be available without charge through a cooperative venture by the Wellcome Trust, the National Library of Medicine, the Joint Information Systems Committee, and several journal publishers. *Free Online Access to Nearly 200 Years of Medical Research*, PUBLICTECHNOLOGY.NET, May 15, 2006, <http://publictechnology.net/modules.php?op=modload&name=News&file=article&sid=4979>.

194. JSTOR also acquired missing issues from vendors that specialize in replacing lost or damaged issues. See JSTOR, Back Issues Needed, <http://www.jstor.org/about/issues/index.html> (last visited Mar. 5, 2007); see also Carlson, *supra* note 188.

to the table. And, as will be discussed more fully in the next section, no clear-cut theoretical reason exists for limiting the right to archive solely to nonprofit entities,¹⁹⁵ despite the tendency in copyright law to assume that permissive uses are nonprofit ones and that licenses should be negotiated for any profit-making use of copyrighted materials.¹⁹⁶ One clear advantage of a law that is flexible in designating who can act as an archivist is that the law may encourage institutions to concentrate their energies in different places. For example, instead of spending money creating and maintaining duplicative electronic archives, a library might instead choose to devote more of its scarce resources to the task of conserving those physical artifacts in their collections that are not superseded in value by electronic versions,¹⁹⁷ or to specialize in a particular aspect of digitization that other institutions are not prepared to undertake.¹⁹⁸

B. WHAT SHOULD THE LAW REQUIRE OF AN ENTITY THAT WANTS THE PRIVILEGE TO ARCHIVE COPYRIGHTED WORKS?

Simply creating a right in third parties to replicate copyrighted works for preservation would not in itself ensure that

195. The one consideration that might push toward an exemption solely for nonprofit entities is the Berne Convention and its implementation under the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS). *See infra* Part III.D. Also, archives may turn out to have the qualities of natural monopolies, a characteristic which might tempt a for-profit entity to engage in rent-seeking. However, there may be regulatory schemes that can counterbalance this problem. If experience still suggests otherwise, then a more limited privilege may be appropriate.

196. *See, e.g.*, *Am. Geophysical Union v. Texaco Inc.*, 60 F.3d 913, 922 (2d Cir. 1994) (stating that uses which bring about financial gain are less likely to qualify for the fair use exception).

197. Once books, journals, photographs, posters, and other copyrightable artifacts have been digitized, the digital version is unlikely to be a complete substitute for the physical embodiments of these works. A manuscript from the Middle Ages, for example, is not fungible with an electronic copy.

198. The physical preservation of rare books is one example. It is a costly enterprise and might be easier for libraries to undertake if they did not need to worry so much about preventing deterioration in their general circulation collections. *See* Maria Blackburn, *Preservation's Crumbling Future*, *JOHNS HOPKINS MAG.*, June 2006, at 32, 32, available at <http://www.jhu.edu/~jhumag/0606web/preserve.html>. A shortage of funds for physical conservation and preservation is a serious problem; less than a quarter of museums, archives, and libraries specifically allocate funding to protect the physical integrity of their collections. *HERITAGE PRES.*, *supra* note 13, at 2. On average, such cultural institutions set aside only two percent of their operating budgets for this type of work. *Id.*

the benefits of preservation would inure to society. To reach that goal, the law would have to impose several additional requirements. The point of digital archiving is to build a record that will be available to generations to come, and that objective requires several mandatory fail-safe provisions.¹⁹⁹ A single copy of an archive could easily be destroyed or damaged. Therefore, the law should permit and indeed require archives to make regularly updated duplicate copies and to store them in some number of geographically diverse sites so that human or natural disasters could not destroy this modern version of the Alexandrian library. The law should also require electronic archivists to make sound custodial arrangements for their databases should their creators at some later date abandon their projects. Finally, because technology moves rapidly in this field, legal commitments both to the use of durable storage material and to the performance of regular upgrades will be necessary to prevent the collected data from becoming useless.

Google has not publicly detailed how it plans to secure its database, the extent of its commitment to periodic upgrades, or the fate of its collection if the corporation ultimately shifts in other directions or is dissolved. It is a for-profit entity, and presumably its decisions—absent legal constraints—will be driven by commercial factors alone. JSTOR, by contrast, has spent a good deal of time thinking about the long-term security of its database and has publicly announced the details of its arrangements. It has taken two measures to ensure the long-term survival of the collections it has assembled. First, JSTOR promised that should it cease operation, the organization would provide copies of the existing digital page images to its participating libraries²⁰⁰—a form of data preservation. The law ought to make some variation on this theme a minimum requirement: the law should require archiving projects from the outset to designate a party who will take charge of the electronic files that have been generated, and that party should have the right

199. For a discussion of a range of technical considerations that should be taken into account, see Waters, *supra* note 17, at 2.

200. SCHONFELD, *supra* note 28, at 233. This assurance was important to the libraries. Because the JSTOR database is intended to be accessed by library users, libraries needed the assurance that they would not be left without the use of the back issues of the journals should JSTOR cease operating. *Cf. id.* (noting that several university librarians reviewed JSTOR's library license and offered feedback, leading to a number of changes that improved the license from the librarians' perspectives).

to maintain, update, and even take over and resume the project if the originator discontinues operations.²⁰¹

JSTOR, however, has gone beyond that simple form of planning. It is building an endowment that will act as an insurance policy for its database should demand drop to a point where JSTOR's current services to libraries can no longer be sustained.²⁰² The endowment will generate an annual income of \$1 million, which would support a small staff and provide the technology and infrastructure necessary to keep a dark version of the existing archive going indefinitely.²⁰³

As a quid pro quo for the privilege of including copyrighted work in an archive, the law should require the archive to do two other things. First, the archive should develop technological standards and protocols to ensure that the digitized sources remain in forms that can eventually be made widely accessible and easily searchable. This provision may entail vesting regulatory authority in the Copyright Office or some other body to set such standards when appropriate.

Second, the law should consider cost and terms of access. Any public domain material in such an archive should be available free of any charges that do not relate to the costs of creating and maintaining the archive. As such, the charge, if there is one, should reflect the cost of providing access to the material, rather than the value of the material itself. This objective will

201. Perhaps one solution would be for Congress to designate an entity, such as the Library of Congress, to manage these transitions. In an ideal world, Congress would also designate funds to pay for the maintenance of existing collections until another entity decides to take it up and operate it. The importance of redundancy in conservation is widely recognized. For example, a number of libraries cooperate in the LOCKSS (Lots of Copies Keep Stuff Safe) system, which backs up digital versions of scientific journals by creating a series of permanent web caches for the journals in question. See Vicky Reich & David S.H. Rosenthal, *LOCKSS: A Permanent Web Publishing and Access System*, D-LIB MAG., June 2001, <http://www.dlib.org/dlib/june01/reich/06reich.html>. Libraries can access one of these caches to repair or restore any of their own files that disappear or are damaged. *Id.* The Library of Congress, in conjunction with other digital preservation partners, including Emory University, is using the LOCKSS software now for other purposes, one of which is the preservation of digital material on "the culture and history of the American South." Emory University, MetaArchive Project Description, http://www.digitalpreservation.gov/partners/project_eu.pdf (last visited Mar. 5, 2007).

202. SCHONFELD, *supra* note 28, at 160.

203. *Id.* at 358–61. A dark archive, as I use the term, is one to which conservators, but not the public, have access. A deeply dark archive could be established that would allow access to no one, although maintaining its integrity under such circumstances would be, to put it mildly, challenging.

be somewhat complex to achieve because the success of the entire enterprise may turn on the availability of the option to charge something for providing the service. Scanning material, maintaining the necessary sites, and keeping up-to-date technologically is not cheap.²⁰⁴ Inputting material in a form that renders it searchable will also entail expense.²⁰⁵ Archiving entities need some way to recover these costs and, in the case of for-profit archiving entities, to make a profit.

Google, a for-profit company, currently plans to finance its operations indirectly by selling advertising.²⁰⁶ JSTOR, a non-profit entity, gets a sizeable portion of its funding from foundation grants.²⁰⁷ Other archives might find it difficult to attract enough revenue by either of these means—think, for example, of the prospect of selling enough advertising to support the archiving of humanities journals. Similarly, the amount of grant money that foundations will willingly devote to the cause is surely limited. Although some organizations will try to survive by depending largely on a volunteer (or to use the current jargon, peer-production) model, this resource, too, is likely to be limited. Thus, many archiving projects will either have to look to government grants or will have to charge users (including libraries) directly. The biggest problem in drafting enabling legislation could well be figuring out how to regulate the fee structure to account fairly for the cost of providing the good but to prevent rent-seeking based on the importance or popularity of the content.²⁰⁸

204. See Quint, *supra* note 17.

205. See *id.*

206. See Jonathan Band, *The Google Library Project: Both Sides of the Story*, PLAGIARY, Feb. 8, 2006, at 1, 8–9, available at www.plagiary.org/Google-Library-Project.pdf (stating that Google will not display advertisements on the page displaying the snippets from a particular work but it “hopes that by including a large number of books in its search index, it will differentiate itself from its competitors and attract more ‘eyeballs,’ which in turn will lead to more advertising revenue”).

207. See JSTOR Donor Acknowledgement, <http://www.jstor.org/about/dap.html> (last visited Mar. 5, 2007).

208. One reason fees may need some form of regulation—perhaps even in the form of a statutory rate—is that archives could turn out to be natural monopolies and therefore competition might impose discipline in fee-setting. The risk of rent-seeking may be greatest when for-profit entities are involved, but excluding them from participation may cut off the parties best able to command the resources needed to do the job.

C. MAKING THIS WORK FOR THE COPYRIGHT OWNER

The opportunities opened to us by the possibility of preserving our cultural heritage in searchable form are accompanied by the challenges of devising a fair compromise that encourages archives to be created without simultaneously discouraging the production of the cultural goods we want to save. Although copyright owners might not voluntarily embrace the necessary tradeoffs, legislatures can still establish potential forms of compensatory benefits.

1. What About Consent?

Let us start with the first of the two hardest nuts in the bowl to crack: consent. No one, whether a public library or an entity developed for the sole purpose of digital archiving on the terms and conditions set out in the last section, should be required to negotiate a license to copy a copyrighted work for purposes of preservation. This exception is a departure from the general norm of copyright, but there is precedent for it in U.S. law, particularly when an important public interest is at stake. Unconsented uses of copyrighted expression can be made if they are “fair” under section 107²⁰⁹ or if they fall within any of the numerous other express limitations on owners’ rights in the copyright law, such as the first sale doctrine in section 109.²¹⁰ Unless the law also treats preservation of a cultural record as an interest exempt from the owners’ ordinary rights, society will fail to realize the promise of archiving.

JSTOR managed to get permission for its use of scholarly journals even though it allows access to the full text, but its success in doing so may not readily translate into success with other types of materials.²¹¹ Some journals scanned by JSTOR

209. 17 U.S.C. § 107 (2000).

210. *Id.* § 109.

211. One factor that may have made journals particularly amenable to some form of compromise with JSTOR is the so-called serials crisis. The cost to libraries of specialized journals, particularly in the sciences (but to a lesser degree in other fields as well) spiraled upward during the 1980s and 1990s, leading libraries to cancel subscriptions or to shift money from purchasing books to acquiring journals. See Edlin & Rubinfeld, *supra* note 7, at 125. One of the reactions to the problem has been attempts by universities and other entities to create publicly accessible archives into which faculty and students deposit their work. See, e.g., *U-M Library Launches Deep Blue: More Access to U-M Scholarship*, MICH. NEWS SERVICE, May 25, 2006, <http://www.umich.edu/news/index.html?Releases/2006/May06/r052506> (announcing a publicly accessible database of research and papers produced by the University of Michigan).

are produced by learned societies or universities that are at least as interested in preservation and accessibility as they are in profit.²¹² Journals also typically own the copyrights to the articles they print, making the process of rights clearance easier.²¹³ Finally, many journal publishers did not see JSTOR's project as a serious threat to their primary market because JSTOR ultimately opted for a "moving wall" principle, which means that the five most recent years of any included publication are not accessible online.²¹⁴ At the time, few of the journals were making much money from permissions or their back lists, and JSTOR worked out a formula to compensate them for any losses they suffered as a consequence of JSTOR's activities.²¹⁵ The organization sweetened the deal by offering additional monetary compensation to journal publishers from any surplus income it generated.²¹⁶ JSTOR judged that the sums involved were small enough to be manageable.²¹⁷

Google claims that a large percentage of the copyrighted books would need to be excluded from Google Library if the company were first required to obtain permission from their owners.²¹⁸ Many sectors of the copyright community have embraced digital distribution only gingerly because of its perceived business risks.²¹⁹ For similar reasons, they are unlikely

faculty); see also *Elsevier*, *supra* note 76, at 6. Also, academics began to experiment with operating their own online journals, independent of traditional publishers. A version of this approach is represented by the Social Science Research Network, to which authors submit abstracts and working papers for immediate distribution. Social Science Research Network, <http://www.ssrn.com/> (last visited Mar. 5, 2007). An example of a scholar-run, free journal is the *Journal of Instrumentation*, which is run by scientists for the benefit of the community. Journal of Instrumentation, Author Benefits, <http://www.iop.org/EJ/journal/-page=benefit/1748-0221/1> (last visited Mar. 5, 2007). Lund University in Sweden is a clearinghouse for open access journals of many types. Directory of Open Access Journals, <http://www.doaj.org/> (last visited Mar. 5, 2007). This trend undoubtedly made many journal publishers more amenable than they otherwise might have been to cooperating with JSTOR.

212. See Boyce, *supra* note 177.

213. See Sully, *supra* note 28 ("[T]he publisher owns the copyright in the digitized images that [JSTOR] creates.")

214. SCHONFELD, *supra* note 28, at 134–38. A few participating journals would not agree to the moving wall principle, however, opting instead to limit permission to cover only journals published before a certain fixed date. *Id.* at 137–38.

215. *Id.* at 144.

216. *Id.* at 144–46.

217. *Id.* at 144.

218. Seidenberg, *supra* note 169.

219. For a discussion of these perceived risks, see Oestreicher-Singer &

to allow a third party to make these copies and then to distribute back-ups, too. In addition, even just a few years after publication, determining the actual ownership of many copyrighted works becomes quite difficult—and costly—if not impossible, as the orphan works problem demonstrates.²²⁰ Even if ownership determination were possible, the costs of multiple searches for thousands of rights-holders would often be too high for a prospective user to bear. And once the rights-holder is found, the licensing fee demands for use of the works in question, cumulatively, may be high enough to render a preservation project infeasible.

The solution need not be free use of works, although if the law were only to provide a right to create a dark, or publicly inaccessible, archive, it probably should be. It is difficult to see how the copyright owner's economic interest would be impinged upon in the case of a dark archive. If the law were to impose a compulsory license, however, it should set the rate at a low enough level that it will not inhibit the creation of the archive. The costs that will be incurred in building and securing an archive are comparatively high,²²¹ as is the public interest in having the work preserved. Any payment to the copyright owner for a dark archive should at most be token, not based on the claimed economic value of the underlying work.²²²

A different way to compensate copyright holders might be to require the archivist to provide them with digital renditions of any copied works they can show they own. Publishers who could otherwise not afford to do their own scanning may be able to use such copies to better exploit old works (through, for example, publishing on demand). Digital masters of publishers' backlists, without the costs of finding the books and making

Sundararajan, *supra* note 35, at 533.

220. For example, the Library of Congress reports that many old, deteriorating feature films were made as works for hire by production companies that are long defunct. LIBRARY OF CONG., *supra* note 54, at 5. A serious barrier to preserving them is the inability to figure out who now owns the rights in these films. *Id.* The problem is particularly acute for photographs. A library at Cornell University is reported to contain a collection of 350,000 photographs but to have no indication of the name of the photographer in the case of ninety-nine percent of the photographs. ORPHAN WORKS, *supra* note 34, at 25 n.32.

221. See, e.g., Band, *supra* note 206, at 8–9 (discussing the economics of creating the Google Library Project).

222. If the law permits public access to the copies, the value of the work might be reflected in a royalty component based on usage. See *infra* notes 231–34 and accompanying text.

copies, might well provide a welcome form of compensation for publishers.

2. What About Access?

The most controversial question undoubtedly is the proper timing of and conditions on public access to the archive once it is created. If Google (or JSTOR for that matter) came to copyright owners with an offer to create, for free, a stable dark archive of all their copyrighted works and to maintain it in perpetuity, and if the entity promised to allow public access to any work in the archive only once it entered the public domain, one could imagine at least indifferent acquiescence, if not outright enthusiasm, on the owners' part. Allowing access during the term of copyright, however, is the rub. Copyright owners want to charge for and control access to their copyrighted works.

Unfortunately, creation of durable and reasonably complete archives probably cannot occur unless their creators couple preservation with some form of public accessibility. Roger Schonfeld, in his extensive study of JSTOR, concluded that access to the back issues of the scanned journals was the added value that allowed JSTOR to attract the broad-based financial support from libraries that it needed to function as an archivist.²²³ He writes that participating libraries "bought" access, which in turn provided JSTOR the ability to archive.²²⁴ Schonfeld expressed that "[d]ark archives have generally been viewed with disdain, since, as public goods, it is almost impossible to develop revenue streams to support them."²²⁵

The JSTOR experience suggests that full text access is the deal-breaker. Google, however, clearly calculates that it can afford to create and maintain an archive containing copyrighted works if it does no more than make them fully text-searchable.²²⁶ Once the reader has identified what he wants, he

223. SCHONFELD, *supra* note 28, at 376.

224. *Id.*

225. *Id.* at 360. The experience of film preservationists is instructive in this regard. Film preservation is expensive. In 1992 alone, the Museum of Modern Art spent \$350,000 in laboratory expenses on film preservation and the University of California in Los Angeles spent almost \$300,000. LIBRARY OF CONG., *supra* note 54, at 24. Because film archives like these typically possess the film prints but not the intellectual property rights in them, they cannot raise funds for preservation by exploiting the works and are therefore limited in what they can do because they rely on donations and government grants for support. *Id.* at 23–25.

226. *Cf.* Band, *supra* note 206, at 8 (explaining that Google hopes, as it in-

then has to go to the library or bookseller who carries the work to see the entire text. It will be interesting to watch the experiment unfold—assuming that the courts do not intervene and close the project down. If Google is correct that operating an archiving project that offers no more than text searching is economically feasible, then many of the obstacles to finding a compromise in copyright will fall away. Providing searchability and conducting indexing are not the exclusive domain of copyright owners.

I am skeptical, however, that it will be enough, and from the perspective of the public interest, there are greater benefits from full text access than from a search that turns up only a few snippets of the work. A modification of copyright law that is aimed at the preservation of the cultural record should try hard to make access feasible. One possibility is to adopt some version of JSTOR's "moving wall."²²⁷ At least where literary works are involved, and probably to some extent film and music as well, the income curve for copyright owners peaks and then falls off in a fairly short period. As noted above, most books go out of print after the first three years, presumably because the market for them falls off sharply by that point.²²⁸ A study by the Register of Copyrights indicated that under the 1909 Copyright Act, renewal was sought after twenty-eight years for fewer than fifteen percent of all works initially copyrighted.²²⁹ This fact suggests that the large stakes for copyright owners are in the early years of a work's public availability and that allowing access to those works from a digital archive only once they have been out for a fixed number of years would work little hardship on the overwhelming majority of copyright owners. The size of the black-out period could vary depending on whether the work

creates the number of works in its Library Project, to increase the number of users, and in turn the interest of businesses in buying advertising space).

227. See *supra* note 214 and accompanying text.

228. See *supra* note 36 and accompanying text.

229. 2 STUDIES ON COPYRIGHT 1251 (Copyright Soc'y of the U.S. ed., 1963). In fact, in 1947 and 1957, only 10.6% and 12.86% of works were renewed. 1 *id.* at 618. These figures suggest that works exhaust most of their economic value quite quickly. As noted below, new technologies enable value to be extracted over longer periods of time. See *infra* note 230. For example, the advent of videocassettes and DVDs allowed motion picture makers to exploit markets for old films that might otherwise have disappeared (literally as well as figuratively) within a few years after release. Nevertheless, one has to suspect that only the rare work will be of interest to more than a few consumers after several years have elapsed. The fact that a motion picture is available on DVD does not tell us how often anyone rents or buys it.

is published or unpublished, and according to type (book versus film versus phonorecording, for example).²³⁰ An arrangement of this kind would greatly lessen the risk that valuable works would disappear before they make it into the public domain, would provide enough immediate value to attract financial support to the archive, and at the same time would pose minimal risk to the copyright owner during the period of her greatest economic vulnerability.

If access is added to the mix, the case for compensating the copyright owner with compulsory license royalties is stronger. JSTOR's experience working out terms that both it and publishers could live with suggests that a sliding scale fee based on the archivist's profits or economic surplus is one possible form of compensation.²³¹ Another possibility would be to set up a compulsory licensing or profit-sharing scheme with proceeds distributed to copyright owners based on frequency of usage, similar to the library lending rights schemes in European countries, which compensate authors for loans of hard copies.²³² This approach, which is not without difficulties,²³³ nonetheless does have the benefit of providing a metric for the value of the work, because the owner of a book or film that has become a classic would be paid a larger share of the applicable revenue stream than the owner of rights to a diary of interest to only two or three historians.²³⁴ Congress would need to decide

230. See *supra* note 229. Increasingly, digitization has allowed copyright owners to use their back lists more efficiently through rentals of movies and through on-demand delivery or publishing. New uses have stretched out the revenue stream for some works. It is doubtful, however, that the tail of the stream is very thick for any given work. If a licensing fee were imposed as a quid pro quo for access, copyright owners with works that remain at least modestly popular could continue to enjoy the benefits of that tail.

231. See SCHONFELD, *supra* note 28, at 144. Although the copyright owner's lost profits might seem to be a more appropriate measure of compensation, I have tentatively rejected that possibility, both because I believe these figures will often be difficult to establish with any certainty and also because I am doubtful that this measure will achieve the kind of compromise between conflicting interests necessary to achieve the goal of encouraging parties to undertake the expense entailed in permanent archiving.

232. See Peter Eckersley, *Virtual Markets for Virtual Goods: The Mirror Image of Digital Copyright?*, 18 HARV. J.L. & TECH. 85, 99 (2004); Jennifer M. Schneck, Note, *Closing the Book on Public Lending Right*, 63 N.Y.U. L. REV. 878, 887-97 (1998).

233. See Schneck, *supra* note 232, at 901-02.

234. See Eckersley, *supra* note 232, at 99. I do not mean to ignore or underestimate the difficulties that are incurred in managing the distribution of royalties that flow from compulsory licensing. One idea worth exploring is

whether the license fees would be generated by charging directly for each item accessed or instead through subscriptions paid by libraries or individuals.

Technologies that limit copying and retransmission may also turn out to have some value in this context. The law might, for example, permit individuals to make a single copy of a work for personal use but use DRMs to help control massive redistribution or commercial reuses.²³⁵ At the same time, many works will simply be unavailable for archiving under the proposed provision unless Congress amends the DMCA to allow archivers to disable access and copying controls on works published initially in digital form so that they can make the necessary copy.²³⁶ These suggested compromises are not intended to exhaust the possibilities but merely to demonstrate that a fair solution is imaginable.

D. CAN WE DO IT IF WE WANT?: THE SHADOW OF BERNE AND TRIPS

One final and obvious consideration in deciding if the United States would be able, if it chooses, to realize the potential of long-term cultural preservation is whether such a scheme would mesh with our obligations under the Berne Convention²³⁷ and the TRIPS Agreement,²³⁸ both of which impose minimum standards of copyright protection on their signatories. Article 9 of the Berne Convention states that the right to reproduce literary and artistic works is exclusively the right of the author.²³⁹ The provision permits only limited exceptions to that exclusive right “in certain special cases, provided that such reproduction does not conflict with a normal exploitation of the work and does not unreasonably prejudice the legitimate inter-

whether, if access to such archives were only available through library portals, data on usage could be automatically collected and used as the basis for distributing royalties. One important concern with such a record-keeping system would be user privacy; historically, libraries have shown themselves to be avid defenders of their users' interests and would seem to be the most trustworthy entities to act as privacy-sensitive intermediaries in this process.

235. See *supra* note 123 and accompanying text.

236. See *supra* text accompanying notes 124–125.

237. Berne Convention, *supra* note 46.

238. Agreement on Trade-Related Aspects of Intellectual Property Rights, Apr. 15, 1994, 108 Stat. 4809, 1869 U.N.T.S. 299 [hereinafter TRIPS Agreement]. Among other things, TRIPS makes most of the provisions of the Berne Convention enforceable by the World Trade Organization. *Id.* art. 9(1).

239. Berne Convention, *supra* note 46, art. 9(1).

ests of the author.”²⁴⁰ Furthermore, Article 8 of the World Intellectual Property Organization (WIPO) Copyright Treaty of 1996 also reserves to authors the “exclusive right of authorizing any communication to the public of their works, by wire or wireless means.”²⁴¹ These provisions seem to place real limits on the ability of the United States to experiment with cultural preservation.²⁴²

In the only enforcement action to date against the United States for violation of Berne or TRIPS,²⁴³ the United States was held in violation of its obligations by exempting small restaurants and businesses from paying royalties for playing broadcast music in public areas of those establishments.²⁴⁴ The panel that heard the case was disturbed by the extensiveness of the exemption²⁴⁵ and, more specifically, by the interference with “normal exploitation of the work” that creates “economic competition with the ways that rights holders normally extract economic value” from their work, thus “depriv[ing] them of significant or tangible commercial gains.”²⁴⁶

Unquestionably, publicly accessible archives of copyrighted works—even if the law provides compensation through a compulsory license mechanism and a “moving wall” provision—will in some instances lead to lower returns for at least a few copy-

240. *Id.* art. 9(2); *see also* TRIPS Agreement, *supra* note 238, art. 13 (limiting exceptions to copyright protections to special situations that do not interfere with the normal exploitation by the owner of the intellectual property interest).

241. WIPO Copyright Treaty art. 8, Dec. 20, 1996, 112 Stat. 2860, 2186 U.N.T.S. 152 [hereinafter WIPO Treaty]. Similar to the Berne Convention, *supra* note 46, and the TRIPS Agreement, *supra* note 238, the WIPO Copyright Treaty allows exceptions to copyright protections only in “special cases.” WIPO Treaty, *supra*, art. 10.

242. Legal scholars have expressed doubt that Google Library’s display of snippets of text would pass muster under domestic laws of countries traditionally more copyright-protective than the United States and with undeveloped or extremely narrow conceptions of fair use. Michael Warnecke, *Google’s Legal Stance on U.S. Book Search May Have Less Support Under European Law*, 71 BNA PAT., TRADEMARK & COPYRIGHT J. 206, 207 (2005). *But see, e.g.*, Wray, *supra* note 96 (noting the potential for legal victory in Germany on this issue).

243. Panel Report, *United States—Section 110(5) of the US Copyright Act*, WT/DS160/R (June 15, 2000) [hereinafter WTO Panel Report].

244. *Id.* The exemption at issue appears at 17 U.S.C. § 110(5)(B) (2000).

245. Data presented to the panel indicated that about seventy percent of restaurants, seventy-three percent of bars, and forty-five percent of retail stores qualified for the exception in § 110(5)(B). WTO Panel Report, *supra* note 243, ¶ 6.122.

246. *Id.* ¶ 6.183.

right owners. Publishers and authors are increasingly able through collective rights organizations like the Copyright Clearance Center to license copying for academic, personal, and other uses.²⁴⁷ The ability to digitize works makes print-on-demand an increasingly feasible response to any request for an out-of-print work. Facially, a major experiment in archiving cultural artifacts could land the United States in the midst of a major trade sanction debacle.²⁴⁸

One possible solution might be to make the archiving provisions apply only to those works claiming copyright under U.S. law, rather than under Berne. U.S. law, of course, has made this sort of distinction in other, albeit more limited, settings,²⁴⁹ but whether international regimes would accept a similar ploy that could be characterized as reducing the substantive rights of authors is unclear.²⁵⁰

There is, however, reason to hope that the rigidities of the international regime will not be an insurmountable barrier to the archiving project and that the size of the prize will lead to a cooperative search for a solution.²⁵¹ Google's announcement of

247. See Copyright.com, Corporate Overview, <http://www.copyright.com/ccd/do/viewPage?pageCode=au1> (last visited Mar. 5, 2007).

248. A second complication is that people can access online works from anywhere in the world, depending on the security measures the archivist uses. Presumably, however, if an archive included only U.S. works and complied with U.S. law, the international ramifications for an archive would be greatly reduced as well. Interestingly, the proposed legislation promulgated by the Copyright Office to deal with orphan works would cover all copyrights, whether U.S. or foreign. ORPHAN WORKS, *supra* note 34, at 121. The report explains that such a law would comply with international copyright obligations because it alters remedies, not rights. *Id.* Depending on the legislative approach taken to authorizing archiving, a similar argument—that remedies, and not rights, are what is at stake—could be made, although I think its success would be questionable.

249. The United States requires its own copyright claimants to register their claims as a prerequisite to filing a suit for infringement, but does not impose a similar requirement on those claiming under Berne. 17 U.S.C. § 411(a) (2000).

250. One difficulty will be that archivists will not readily be able to identify whether the work is of U.S. origin. I thank and credit June Besek for these helpful cautionary observations. See JUNE M. BESEK, NAT'L DIGITAL INFO. INFRASTRUCTURE AND PRES. PROGRAM, COPYRIGHT ISSUES RELEVANT TO THE CREATION OF A DIGITAL ARCHIVE 15 n.57 (2003). This approach might also encourage copyright owners to move first publication to other Berne countries to avoid the effect of the archiving rules. Cf. Graeme Austin, *Keynote Address*, 28 COLUM. J.L. & ARTS 397, 416–17 (2005) (discussing the issue of avoiding U.S. copyright constraints).

251. As noted above, Germany's recent decision on Google Library suggests that the laws of various European countries may contain greater flexibility

its plans, after all, did more than upset U.S. authors and publishers; it opened the eyes of the American public to the possibility that a fantasy might be realizable. It seems to have snapped the European Union to attention as well. The French government in particular was shaken by the vision of Google Library putting English literature online.²⁵² Its Minister of Culture said that the Google project “is confirmation of the risk of a crushing American domination in the definition of how future generations conceive the world.”²⁵³

On April 28, 2005, six European Community member states directed a letter to the Commission of the European Communities asking it to support the creation of a “virtual European library” that will preserve “cultural and scientific” content and increase accessibility.²⁵⁴ The resulting report from the Commission, issued to the European Parliament and Council in September of the same year, acknowledged several of the same problems noted in this Article, including the losses suffered in Europe each year of works recorded on fragile materials like videotape.²⁵⁵ The report also expressed enthusiasm about the public benefits of both better preservation and improved accessibility.²⁵⁶ But it, too, stumbled over the current state of copyright law.²⁵⁷ The report did not see insurmountable intellectual property obstacles to the creation of dark archives for copyright works, at least in theory,²⁵⁸ but it admitted that

than is at first apparent—or perhaps that interpretation of those laws may be influenced by recognition of the public benefits such online indexing might offer. See, e.g., Wray, *supra* note 96.

252. *European Libraries Fight Google-ization*, DEUTSCHE WELLE, Apr. 27, 2005, <http://www.dw-world.de/dw/article/0,1564,1566717,00.html>.

253. *Id.*

254. Europe’s Information Society, Digital Libraries Initiative Homepage: Timeline of Development, http://europa.eu.int/information_society/activities/digital_libraries/timeline/index_en.htm (last visited Mar. 5, 2007). The letter was signed by the diplomatic heads of France, Poland, Italy, Spain, the Chancellor of Germany, and the Prime Minister of Hungary. Letter from Jacques Chirac et al., President of Fr., to Jose Manuel Durdo Barroso, President, European Comm’n (Apr. 28, 2005), available at http://europa.eu.int/information_society/activities/digital_libraries/doc/letter_1/index_en.htm.

255. *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions—i2010: Digital Libraries*, at 5, COM (2005) 465 final (Sept. 30, 2005) [hereinafter EU Commission Report].

256. *Id.* at 3, 5.

257. *Id.* at 6.

258. European Community members may make limited exceptions to the exclusive right of the author to control copies in favor of libraries, museums,

online access did, even in some instances where the work itself was in the public domain but elements of the particular edition were not.²⁵⁹ The report added that getting permission to include copyrighted work, under the current legal regime, would in many instances be so difficult and costly as to be wholly impracticable.²⁶⁰ The report stated: “An online library offering works beyond public domain material is not possible without a substantial change in the copyright legislation, or agreements, on a case by case basis, with the rights-holders.”²⁶¹ Since the report was issued, the Commission has solicited input from interested parties including cultural and educational institutions across the European community²⁶² and has also appointed an Expert Group on Digital Libraries to assist it in “making Europe’s cultural heritage available online.”²⁶³ From both

and schools. European Parliament and Council Directive 2001/29/EC, 2001 O.J. (L 167) 13. However, states are not required to make exceptions, resulting in widely varying laws across member states. *Id.* at 10.

259. EU Commission Report, *supra* note 255, at 6.

260. *See id.* at 5–6.

261. *Id.* A staff report accompanying the commission’s missive lays out the problems in greater detail but without specifically addressing the changes in copyright law that might be considered. *Commission Staff Working Document Annex to the Communication from the Commission “i2010 Digital Libraries,”* COM (2005) 465 final (Sept. 30, 2005) [hereinafter *Commission Staff Working Document*]. It does, however, refer with seeming approval to licensing schemes in the Scandinavian countries that permit non-commercial exploitation of orphaned works based on collective licensing. *Id.* at 12.

262. EC Digital Libraries Initiative, Replies to Online Consultation, http://europa.eu.int/information_society/activities/digital_libraries/index_en.htm (select “Cultural Heritage,” “Actions on European Level,” “Consultations,” “Online Consultation”) (last visited Mar. 5, 2007). The replies are also summarized by the Commission staff. European Commission, Results Online Consultation “i2010: Digital Libraries” 2, http://ec.europa.eu/information_society/activities/digital_libraries/doc/results_online_consultation/en.pdf (last visited Mar. 5, 2007) [hereinafter *Online Consultation*] (“The *copyright issue is indeed the most contentious* part of the consultation. Whereas the rights-holders emphasize the adequateness of the current copyright rules, cultural institutions stress that change in the present copyright framework is needed for efficient digitization and digital preservation.”).

263. Press Release, European Comm’n, High Level Expert Group to Advise European Commission on How to Build the European Digital Library (Mar. 27, 2006), *available at* <http://europa.eu.int/rapid/> (search “IP-EC Press Release” for “Digital Library” on Mar. 27, 2006); *see also* Commission Decision 2006/178/EC, 2006 I.J. (L 63) 25 (setting up the High Level Expert Group on Digital Libraries); European Commission High Level Expert Group on Digital Libraries, Summary Minutes of the First Meeting (Mar. 27, 2006), *available at* http://europa.eu.int/information_society/activities/digital_libraries/doc/minutes_of_hleg_meet/summary_%20minutes_hleg_1st_meet_en_final.pdf [hereinafter *HLEG Minutes*] (providing the minutes of the Group’s first meeting).

sources have emerged concerns that, without appropriate modifications, modern copyright regimes will stifle the possibilities of effective preservation and access.²⁶⁴ The comments of these groups all suggest that the benefits of preservation and access are entitled to greater weight in the law than they currently receive.²⁶⁵

The fact that the European Commission has even raised the possibility that existing copyright law might be altered suggests a flexibility in Europe that may never before have existed.²⁶⁶ Changes in the international regimes depend on worldwide agreement. But the European example at least suggests that, as nations around the world come to realize the enormous potential benefits of digital preservation and access, a groundswell of support may rise up and make it possible for the international intellectual property community to make the changes in the balance between property interests and the public interest that are necessary to realize those benefits.

CONCLUSION

The technology that could enable us to preserve our cultural artifacts and make them accessible to any user with access to a computer terminal still has many wrinkles that need to be worked out. It now seems realistic, however, to believe that the more intractable barrier to success is the state of copyright law. Digital preservation requires copying, and copying works that are not in the public domain constitutes infringement. One price we pay for copyright protection that lasts for a century is the risk that because we must wait for works to exit copyright to effectively preserve them, these works may be lost before they can be saved. A modification of copyright law that

264. See Online Consultation, *supra* note 262, at 2; HLEG Minutes, *supra* note 263, at 3.

265. See Online Consultation, *supra* note 262, at 2. Most recently the British Library issued something it provocatively entitled a “manifesto,” urging reforms in copyright, including reforms that would enable effective archiving and preservation. BRITISH LIBRARY, INTELLECTUAL PROPERTY: A BALANCE 2 (2006), available at <http://www.bl.uk/news/pdf/ipmanifesto.pdf>. One of its complaints is that the law does not allow libraries to copy sound recordings and films to preserve them, leaving them with increasingly fragile and deteriorating collections. *Id.*

266. A further indication of flexibility may be found in a recent German case in which the court found a publisher’s application for a preliminary injunction against Google’s Library Project unlikely to succeed. See, e.g., Wray, *supra* note 96.

allows anyone to make digital copies is not the answer. But a carefully crafted compromise designed to respect the interests of copyright owners, impose responsibilities on archivists, and facilitate preservation in ways that best protect the public ought to be achievable. If it is not, then copyright will do an inadvertent disservice to the very creations it was designed to foster and will fail to serve the public's best interests in the process.