

# ARTICLE

## KNOWLEDGE ACCESSIBILITY AND PRESERVATION POLICY FOR THE DIGITAL AGE

*Peter S. Menell\**

### TABLE OF CONTENTS

I. INTRODUCTION ..... 1112

II. A HISTORICAL PERSPECTIVE ON PROMOTING, COLLECTING, PRESERVING, CATALOGUING, AND PROVIDING ACCESS TO KNOWLEDGE ..... 1117

    A. *Early History*..... 1117

    B. *The Renaissance and the Emergence of Copyright Protection* ..... 1118

    C. *Early U.S. Experience* ..... 1120

    D. *Modern U.S. Copyright Law* ..... 1128

        1. *The Deposit Requirement*..... 1129

        2. *The Catalog of Copyright Entries*..... 1130

        3. *The Library Provisions* ..... 1132

        4. *Television and Radio Archives* ..... 1136

III. PUBLIC VERSUS PRIVATE PRESERVATION OF AND ACCESS TO KNOWLEDGE IN THE DIGITAL AGE ..... 1138

    A. *Exploring the Values of Preserving and Facilitating Access to Knowledge in the Digital Age* ..... 1138

---

\* Professor of Law, University of California at Berkeley School of Law, and Director, Berkeley Center for Law and Technology. I would like to thank Craig Joyce, Tom Fletcher, Amy Kapczynski, Bobbi Kwall, Neil Netanel, David Nimmer, Claire Sylvia, and participants at the Intellectual Property Scholars Conference (Aug. 2006) and the University of Houston’s Copyright in Context Conference (June 2007) for comments on earlier drafts. I would like to thank Paul Finlayson, Tamar Gubins, and Will McGinty for research assistance.



2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1113

societies within reach and marks the beginning of a new era for scholars, authors, and other users of recorded knowledge. For public domain works, users would be able to retrieve and download the full documents. For works still under copyright protection, Google would provide a few sentences surrounding the search term as well as information about where the work could be procured legally (publisher sites, bookstores, and libraries). Just a few years ago, the cost and time required to digitize and render searchable ten percent of the vast stock of written human knowledge was thought to be prohibitive.<sup>2</sup> Yet Google has committed to making extensive collections of some of the world's leading libraries—approximately twenty-five million volumes—available within less than a decade and without any public expenditure.<sup>3</sup>

The inherent unpredictability of copyright law's fair use doctrine appears to be the principal impediment to this project. Shortly after Google's announcement, leading publishers and authors complained that Google's project infringed their copyrights and requested that Google delay scanning any copyright protected works until an agreement could be negotiated.<sup>4</sup> In an effort to avoid litigation, Google voluntarily suspended the scanning of such works for several months and implemented an opt-out mechanism by which copyright owners could remove their titles from the list of works to be scanned.<sup>5</sup> Nonetheless, the Authors Guild, followed shortly by five commercial publishers, brought suit alleging that Google's Book Search Project infringed their copyrights "by unlawfully reproducing and publicly distributing and displaying copies of

---

2. The University of Michigan estimates that, using the technology and resources at its disposal, it would have taken more than 1,000 years to fully digitize its seven million volume collection. Press Release, University of Michigan News Service, Google Library Partnership (Sept. 24, 2007), available at <http://www.umich.edu/news/index.html?BG/google/index>. Google plans to scan the entire collection in six years. *Id.*

3. Google projects that it can scan thirty-two million volumes in ten years time. Jeffrey Toobin, *Google's Moon Shot*, NEW YORKER, Feb. 5, 2007, at 30. While Google has not disclosed an estimated cost of scanning thirty-two million volumes, some estimate it will cost \$750 million. Jonathan Band, *The Google Library Project: Both Sides of the Story*, 1 PLAGIARY 6, 13 (2006), available at <http://quod.lib.umich.edu/p/plag/images/5240451.0001.002.pdf>. Other estimates put the cost at as little as \$10 per volume. Elisabeth Hanratty, *Google Library: Beyond Fair Use?*, 2005 DUKE L. & TECH. REV. 0010, ¶ 1 (2005), available at <http://www.law.duke.edu/journals/dltr/articles/2005dltr0010.html>.

4. See Chris Gaither, *Google Puts Book Copying on Hold*, L.A. TIMES, Aug. 13, 2005, at C1 (noting that publishers felt an opt out option would not protect their copyrights).

5. *Id.* (reporting that Google's sole stated purpose for suspending the project was to implement an opt-out procedure).

such works.”<sup>6</sup> Patricia Schroeder,<sup>7</sup> President of the American Association of Publishers, asserted that, while “Google Print Library could help many authors get more exposure and maybe even sell more books, authors and publishers should not be asked to waive their long-held rights so that Google can profit from this venture.”<sup>8</sup> The cases are currently awaiting trial in U.S. District Court for the Southern District of New York.

Unless a settlement can be reached or legislation passed, this litigation could drag on for several years.<sup>9</sup> Google’s main defense will be that the scanning, digitizing, and indexing of plaintiffs’ works, as well as providing snippets in response to user queries, fall within the fair use doctrine. The contours of this doctrine are notoriously vague,<sup>10</sup> and the Internet context further complicates the resolution.<sup>11</sup> This application of the fair use doctrine is a matter of first impression, and the outcome of the case is unclear.<sup>12</sup>

---

6. Complaint at ¶ 38, *McGraw-Hill Cos. v. Google, Inc.*, No. 05-CV-8881, 2005 WL 2778878 (S.D.N.Y. Oct. 19, 2005).

7. Since leaving Congress, Pat Schroeder, a former member of the House Judiciary Subcommittee on Courts and Intellectual Property, has continued to advocate for stronger intellectual property rights. See About AAP-Administration: Patricia Schroeder, <http://www.publishers.org/about/bio.cfm?StaffID=1> (last visited Nov. 7, 2007). From the helm of the Association of American Publishers (AAP), she opposes projects to digitize books or make them more widely available through libraries. See Linton Weeks, *Pat Schroeder's New Chapter*, WASH. POST, Feb. 7, 2001, at C1 (reporting the former Senator’s statement that the AAP has “a very serious issue with librarians”).

8. Press Release, Assoc. of Am. Publishers, Publishers Sue Google Over Plans To Digitize Books (Oct. 19, 2005), available at <http://www.publishers.org/press/releases.cfm?PressReleaseArticleID=292>.

9. *McGraw-Hill, et. al.* filed in the Southern District of New York on Oct. 19, 2005, and the pretrial conference is scheduled for November 27, 2007. Docket Entry of Judge Sprizzo, *McGraw-Hill Cos. v. Google Inc.*, No. 05-CV-8881, Oct. 19, 2005, available at <http://www.westlaw.com>.

10. See David Nimmer, “*Fairest of Them All*” and *Other Fairy Tales of Fair Use*, 66 LAW & CONTEMP. PROBS. 263, 287 (2003) (comparing sixty fair use decisions and concluding that “the problem with the four factors is they are malleable enough to be crafted to fit either point of view”).

11. See, e.g., *Perfect 10, Inc. v. Amazon.com, Inc.*, 487 F.3d 701, 725 (9th Cir. 2007) (vacating and remanding a preliminary injunction because the use of thumbnail images for search purposes likely fell within the scope of the fair use doctrine); *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 822 (9th Cir. 2003) (holding that images shrunk into thumbnails for search purposes fell within the fair use doctrine); *Sega Enters. Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1527 (9th Cir. 1992) (holding that the fair use doctrine protected “intermediate copying” of computer source code to find uncopyrightable elements).

12. See Gaither, *supra* note 4 (observing that lawyers not involved differed on whether Google’s plan would be found to be considered a fair use); see also Posting of James Boyle to Duke Law & Technology Review iBlawg, <http://www.law.duke.edu/journals/dltr/iblawg/?p=16> (Feb. 15, 2006) (predicting that the outcome of the Google litigation will turn on which judge hears the case).

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1115

Several scholars have usefully explored the application of modern fair use jurisprudence to Google's Book Search Project.<sup>13</sup> Rather than rehash the cases that have been suggested as relevant precedent, this Article explores the broader policy dimensions of digital archiving and the concomitant ability to search copyrighted works using Boolean search methodology. These dimensions have thus far been overshadowed by the focus on whether or not Google's Book Search Project passes muster under the fair use doctrine. But, given the profound and unanticipated ramifications of currently available technology, it is worth revisiting the underlying principles and purposes of copyright law in order to develop sound public policy relating to preservation, access, and searchability of human knowledge in the digital age.

The policy analysis described above, and any potential solution, should extend beyond Google's particular project for several reasons. First, by implementing the "opt-out" policy, Google has already made a significant compromise on its stated goal "to create a comprehensive, full-text searchable database of all the world's books"<sup>14</sup> so as to stand a better chance of prevailing in the pending litigation. While its motivations are certainly understandable for a publicly traded corporation with very deep pockets, such a compromise detracts from the larger social interest in promoting progress through a universally available, comprehensive search tool. Second, Google—notwithstanding its phenomenal success and noble aspirations—may well succumb to complicating influences as its business evolves. A legislative solution to the new opportunities and challenges presented by archiving and search technology could well produce a better system of incentives and safeguards. Such a solution could provide clear ground rules for competition and innovation in search technology and services.

The importance of these issues, and the fact that Congress did not foresee the possibilities for a universally accessible, comprehensive archive when it last seriously considered the role of libraries in the mid-1970s,<sup>15</sup> as well as the uncertainties and delay of adjudication, call for Congress to consider the larger

---

13. See Band, *supra* note 3, at 8–12; Emily Anne Proskine, *Google's Technicolor Dreamcoat: A Copyright Analysis of the Google Book Search Library Project*, 21 BERKELEY TECH. L.J. 213, 224–32 (2006); Kinan H. Romman, Comment, *The Google Book Search Library Project: A Market Analysis Approach to Fair Use*, 43 HOUS. L. REV. 807, 828–39 (2006).

14. Bob Thompson, *Search Me?*, WASH. POST, Aug. 13, 2006, at D1.

15. As discussed below, Congress has since made substantive adjustments to the library provisions of the 1976 Act. See *infra* notes 98–100 and accompanying text.

public policy ramifications of digital archiving. Furthermore, placing Google's Book Search Project within the broader context of copyright's history, values, and balances sheds valuable light on the fair use questions facing the courts.

The Article begins by doing just that—placing the Google controversy within copyright's historical landscape. The publishers and authors leading the offensive against Google characterize this case as rampant, garden variety copyright infringement.<sup>16</sup> Such a perspective overlooks copyright's larger purposes and liability structure. The goals of collecting, preserving, and cataloging human knowledge predate copyright laws. As copyright law developed, it incorporated these concerns within its objectives. Thus, a broader historical compass helps to frame the larger social interests relevant to public policy and fair use analysis. Drawing upon these interests, Part II examines the democratic, cultural, and economic dimensions of developing a comprehensive, searchable database of books and other library materials. It then analyzes the spectrum of institutional alternatives for promoting the goals of preserving and providing access to knowledge while safeguarding copyright law's incentives to create.

Based on this analysis, Part III contends that the Google Book Search Project generally (and greatly) promotes the policies and values of copyright. But the project is overly cautious from the standpoint of the public at large in that it allows copyright owners to "opt out" of its archive. At the same time, the project still poses the risks of widespread piracy, should its technical protection measures fail, and of loss of the knowledge archive through secrecy and changes in company practices. To address these concerns in the near term, this Article recommends that Congress better effectuate the preservation and access goals of the copyright system, without sacrificing protection or incentives to create, through a carefully crafted package of safe harbors, measured liability exposure for technology vulnerable to piracy, and public involvement in the development and management of a searchable digital repository of copyrighted works. In the longer term, this Article recommends that Congress update the deposit

---

16. See Complaint at ¶ 1, *Author's Guild v. Google, Inc.*, No. 05-CV-8136, 2005 WL 2463899 (S.D.N.Y. Sept. 20, 2005) ("By reproducing for itself a copy of those works that are not in the public domain . . . , Google is engaging in massive copyright infringement."); Complaint at ¶ 7, *McGraw-Hill Cos. v. Google, Inc.*, No. 05-CV-8881, 2005 WL 2778878 (S.D.N.Y. Oct. 19, 2005) ("Because Google's entirely commercial endeavor requires . . . massive, wholesale and systematic copying of entire books still protected by copyright for public distribution and public display, it infringes one or more of each Publisher's exclusive rights under the Copyright Act . . .").

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1117

requirement to provide for digital deposit of written works and plan for the development of a public comprehensive searchable archive. In the event that legislation is not forthcoming before the Google Book Search Project is litigated, the final section of this Article shows how the historical, statutory, and economic analysis provides a broader context for courts applying copyright law's fair use doctrine.

II. A HISTORICAL PERSPECTIVE ON PROMOTING, COLLECTING,  
PRESERVING, CATALOGUING, AND PROVIDING ACCESS TO  
KNOWLEDGE

Although the granting of exclusive rights to authors is commonly seen as the fundamental framework for public policy relating to the protection of works of authorship, such a perspective overlooks the larger societal context for collecting, preserving, and affording access to knowledge. Societies took an interest in the protection of knowledge long before copyright law came into existence. As copyright laws emerged, following the invention of the printing press during the Age of Enlightenment,<sup>17</sup> those larger societal concerns became part of the fabric of copyright. Such threads, however, are now obscured by a legal regime woven into an antipiracy garment. Yet, these core values remain a part of the public policy framework and emerge more clearly in the light of challenges and opportunities posed by advances in digital technology.

A. *Early History*

Dating back to the third century before the Common Era, King Ptolemy I established the Royal Library of Alexandria.<sup>18</sup> The library's collection was initially organized by Demetrius of Phaleron, a student of Aristotle, as part of a research center and repository of literature and knowledge of the Hellenic culture.<sup>19</sup> To hasten assembling of the collection, King Ptolemy III decreed that all visitors to Alexandria surrender books in their possession so that they could be copied for the library's archives.<sup>20</sup> More than a century later, Mark Antony donated 200,000 scrolls to the

---

17. Hanratty, *supra* note 3, at ¶ 36.

18. See Andrew Erskine, *Culture and Power in Ptolemaic Egypt: The Museum and Library of Alexandria*, 42 GREECE & ROME 38, 38–39 (1995) (opining that the library was likely created under King Ptolemy I and developed under King Ptolemy II).

19. See *id.* at 40 (noting a connection between Demetrius and King Ptolemy).

20. See *id.* at 39 (adding that the copies, not the originals, were returned to the owners).

library as a gift marking his wedding to Cleopatra.<sup>21</sup> At its peak, the library is thought to have amassed half a million scrolls,<sup>22</sup> estimated to have been between thirty and seventy percent of all books then in existence.<sup>23</sup> Archaeological discoveries indicate that the library housed vast lecture halls capable of accommodating thousands of students.<sup>24</sup> Scholars from throughout the Mediterranean region were attracted to the library.<sup>25</sup> Although ultimately destroyed by fire,<sup>26</sup> the library reflects early societal interest in collecting and preserving knowledge.

*B. The Renaissance and the Emergence of Copyright Protection*

The next major chapter in the development of laws and policies relating to collection and preservation of knowledge emerged in the wake of the invention of the printing press. As the technology historian Elizabeth Eisenstein has shown, the printing press transformed the scholar's perspective and role in society.<sup>27</sup> Prior to the availability of mass reproduction of texts, scholars focused their energies on searching for scattered manuscripts and copying them by hand. But with the widespread availability of important texts, scholars could shift their efforts to revising texts and discovering new knowledge. In addition, print shops and later libraries brought together scholars, artisans, and translators from various nations and religions. This coincides with the emergence of institutions of higher learning. In manifold ways, the printing press and the resulting preservation and concentration of knowledge played critical roles in both the Protestant Reformation and the Scientific Revolution.

During the Renaissance, public policies emerged that directly promoted the preservation and availability of knowledge.<sup>28</sup> In 1537, King Francis I, France's first Renaissance monarch, credited with promoting great cultural advances during

---

21. MATTHEW BATTLES, *LIBRARY: AN UNQUIET HISTORY* 24 (2003).

22. See Erskine, *supra* note 18, at 40.

23. Kevin Kelly, *Scan This Book!*, N.Y. TIMES MAG., May 14, 2006, at 42, 44.

24. David Whitehouse, *Library of Alexandria Discovered*, BBC NEWS, May 12, 2004, <http://news.bbc.co.uk/1/hi/sci/tech/3707641.stm>.

25. See Erskine, *supra* note 18, at 38.

26. Whitehouse, *supra* note 24.

27. See ELIZABETH L. EISENSTEIN, 1 *THE PRINTING PRESS AS AN AGENT OF CHANGE* 3 (1979).

28. It should be noted, however, that governments engaged in direct censorship during this time as well. The first copyright protections enforced religious and political conformity by regulating who and what could be published. See LYMAN RAY PATTERSON, *COPYRIGHT IN HISTORICAL PERSPECTIVE* 6 (1968) (attributing early development of copyright law in England to the need for censorship in a time of religious upheaval).

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1119

his reign, established the first law requiring the deposit of books and other cultural materials in a library for the purpose of preservation.<sup>29</sup> This concept, which came to be known as “legal deposit,” spread throughout the European states.<sup>30</sup> In 1611, Oxford University’s Bodleian Library entered into a perpetual covenant with the Company of Stationers—the guild of printers, bookbinders, and booksellers chartered by the Crown—for the deposit of published works.<sup>31</sup> England formally established a library deposit system through the Licensing of the Press Act of 1662.<sup>32</sup> It later incorporated this requirement into the first copyright law, the Statute of Anne, enacted in 1709.<sup>33</sup> The preamble characterized the statute as an “Act for the Encouragement of Learning by vesting the Copies of printed Books in the Authors or Purchasers of such Copies, during the Times therein mentioned.”<sup>34</sup> Section V provided

That Nine Copies of each Book or Books upon the best Paper . . . shall by the Printer and Printers thereof be delivered . . . for the Use of the Royal Libraries of the Universities of Oxford and Cambridge[,] the Libraries of the Four Universities in Scotland[,] the Library of Sion College in London[,] and the Library commonly called the Library belonging to the Faculty of Advocates at Edinburgh respectively.<sup>35</sup>

Thus, in addition to providing exclusive rights to authors, the earliest copyright law provided for the archiving of copyrighted works, which served both as evidence of what was protected and as a means of collecting knowledge in publicly accessible libraries throughout the realm.

Such deposit requirements emerged throughout Europe and the Americas during the 18th and 19th centuries.<sup>36</sup> Their role within copyright regimes came into question in 1908, when

---

29. Elizabeth K. Dunne, *Study No. 20: Deposit of Copyrighted Works*, in 1 *STUDIES ON COPYRIGHT* 409, 411 (Arthur Fisher Memorial ed. 1963).

30. *Id.* at 411–12.

31. R.C. BARRINGTON PARTRIDGE, *HISTORY OF THE LEGAL DEPOSIT OF BOOKS THROUGHOUT THE BRITISH EMPIRE* 17 (1938).

32. See Licensing of the Press Act, 1662, 14 Car. 2, c. 33 (Eng.) (noting that one copy of all printed books must be deposited with the Licensor). As the statute’s longer title—“An Act for preserving the frequent Abuses in printing seditious treasonable and unlicensed Books and Pamphlets and for regulating of Printing and Printing Presses”—suggests, the Act focused on censorship, but it also incorporated a deposit requirement. See *id.*

33. See Statute of Anne, 1709, 8 Ann., c. 21 (Eng.).

34. *Id.*

35. *Id.*

36. See Dunne, *supra* note 29, at 411–12 & n.5.

signatories to the Berne Convention for the Protection of Literary and Artistic Works (to which most European nations subscribed) prohibited formalities—such as deposit, registration, and notice—as conditions for protection of works of authorship.<sup>37</sup> Yet, most nations continued such requirements under a press law or a special enactment benefiting libraries.<sup>38</sup>

### C. Early U.S. Experience

Little in the way of formal knowledge preservation policy or copyright protection emerged in the Americas until the birth of the American Republic.<sup>39</sup> Following the Revolutionary War, the Articles of Confederation (effective March 1, 1781) governed until the ratification of the U.S. Constitution in 1789. Soon after independence, Connecticut, Massachusetts, and Maryland adopted copyright statutes.<sup>40</sup> The 1783 Massachusetts statute is particularly revealing of the broader social purposes of early American copyright protection. Its preamble stated:

Whereas the improvement of knowledge, the progress of civilization, the public weal of the community, and the advancement of human happiness, greatly depend on the efforts of learned and ingenious persons in the various arts and sciences: As the principal encouragement such persons can have to make great and beneficial exertions of this nature, must exist in the legal security of the fruits of their study and industry to themselves; and as such security is one of the natural rights of all men, there being no property

---

37. The initial Berne Convention, entered in 1886, did not bar formalities as a condition for copyright protection. H.R. REP. NO. 100-609, at 11–12 (1988). The abandonment of such requirements was added in the Berlin revision of 1908. *See id.* at 12 (“The principal achievement of the Berlin Revision Conference was the prohibition of formalities as a condition of enjoyment and exercise of rights under the Convention.”).

38. *See Dunne, supra* note 29, at 412.

39. There was relatively little literary activity in the colonies and, hence, little interest in protecting authors or publishers. *See PATTERSON, supra* note 28, at 183 (“Copyright was not secured by law in colonial America.”); Michael D. Birnhack, *The Idea of Progress in Copyright Law*, 1 BUFF. INTELLECTUAL PROP. L.J. 3, 28 (2001). The one exception was the Commonwealth of Massachusetts, which established a copyright law protecting the rights of publishers in 1672, thirty-eight years before England enacted the Statute of Anne. Howard B. Abrams, *The Historic Foundation of American Copyright Law: Exploding the Myth of Common Law Copyright*, 29 WAYNE L. REV. 1119, 1171–72 (1983). Legislatures occasionally extended copyright protection for particular works through special legislation. *See, e.g., id.* at 1172 n.216 (noting the Massachusetts House of Representatives once granted a copyright with a seven-year term to William Billings for his *New England Psalm-Singer*).

40. *See PATTERSON, supra* note 28, at 183.

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1121

more peculiarly a man's own than that which is produced by the labour of his mind . . . .<sup>41</sup>

This preamble begins with and emphasizes the goal of copyright law to advance knowledge and public welfare. This emphasis finds resonance in the Massachusetts Constitution, in the section entitled "The Encouragement of Literature, Etc.," which still reads:

Wisdom, and knowledge, as well as virtue, diffused generally among the body of the people, being necessary for the preservation of their rights and liberties; and as these depend on spreading the opportunities and advantages of education in the various parts of the country, and among the different orders of the people, it shall be the duty of legislatures and magistrates, in all future periods of this commonwealth, to cherish the interests of literature and the sciences, and all seminaries of them; especially the university at Cambridge, public schools and grammar schools in the towns; to encourage private societies and public institutions, rewards and immunities, for the promotion of agriculture, arts, sciences, commerce, trades, manufactures, and a natural history of the country; to countenance and inculcate the principles of humanity and general benevolence, public and private charity, industry and frugality, honesty and punctuality in their dealings; sincerity, good humor, and all social affections, and generous sentiments among the people.<sup>42</sup>

This grant of legislative power stresses the importance of education, access to knowledge, and encouragement of literature and science as cornerstones for a democratic society.

In 1783, James Madison and Ralph Izard persuaded the Congress of the Confederation to adopt a resolution recommending "to the several States to secure to the authors or publishers of any new books . . . the copyright of such books."<sup>43</sup> All

---

41. PATTERSON, *supra* note 28, at 187 (quoting U.S. Copyright Office, Copyright Laws of the United States of America, 1783–1862, at 1 (1962)). The Massachusetts statute served as the model for the copyright enactments of the New Hampshire and Rhode Island legislatures. Birnhack, *supra* note 39, at 29.

42. MASS. CONST. art. III, § 2.

43. Francine Crawford, *PreConstitutional Copyright Statutes*, 23 BULL. COPYRIGHT SOC'Y 11, 13 (1975). Madison's and Izard's resolution at the Congress of the Confederation reads:

Resolved, that it be recommended to the several States to secure to the authors or publishers of any new books not hitherto printed, being citizens of the United States, and to their executors, administrators, and assigns, the copy right of such books for a certain time not less than fourteen years from the first publication . . . such copy or exclusive rights of printing, publishing and vending the same, to be secured by the original authors, or publishers, their executors,

of the states except Delaware adopted copyright statutes by 1786.<sup>44</sup> These statutes differed in various respects—covered works, duration, remedies, reciprocity, and enforcement procedures.<sup>45</sup> James Madison complained at the Constitutional Convention that there was “want of uniformity in the laws concerning . . . literary property.”<sup>46</sup>

Against this backdrop, the newly formed republic adopted the U.S. Constitution as a framework for governing its affairs. The founders of the nation proclaimed “progress” as the goal for the enactment of laws creating patents and copyrights. As stated in Article I, Section Eight, Clause Eight:

The Congress shall have Power . . . To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors, the exclusive Right to their respective Writings and Discoveries.<sup>47</sup>

Although the legislative history relating to this clause is sparse,<sup>48</sup> there can be little question that U.S. copyright law is broader than just protecting the rights of authors.<sup>49</sup> Rather, as

---

administrators, and assigns, by such laws and under such restrictions as to the several States may seem proper.

*Id.* (omission in original).

44. *Id.*

45. *See generally id.* (detailing the differences in early state copyright statutes); *see also* Abrams, *supra* note 39, at 1173–74 (noting the diversity in early copyright statutes).

46. James Madison, *Vices of the Political System of the United States*, reprinted in 1 THE FOUNDERS' CONSTITUTION 166, 167 (Philip B. Kurland & Ralph Lerner eds., 1987).

47. U.S. CONST. art. I, § 8.

48. *See* EDWARD C. WALTERSCHEID, THE NATURE OF THE INTELLECTUAL PROPERTY CLAUSE: A STUDY IN HISTORICAL PERSPECTIVE 83 (2002) (commenting that “little is actually known about how [the Intellectual Property Clause] came about”); Thomas B. Nachbar, *Intellectual Property and Constitutional Norms*, 104 COLUM. L. REV. 272, 338 (2004) (noting that there was little “said about the Intellectual Property Clause during the ratification debates”). *But cf.* Dotan Oliar, *Making Sense of the Intellectual Property Clause: Promotion of Progress as a Limitation on Congress's Intellectual Property Power*, 94 GEO. L.J. 1771, 1791–1810 (2006) (developing a fuller account of the clause from the eight proposals by Pickney and Madison presented at the Constitutional Convention and dissecting the political alignments of key players in the debates). Note that the Constitution's formulation reflects historical usage. Copyright was intended to promote “science,” whereas patent protection spurred process in the “useful arts.” *See* MELVILLE B. NIMMER & DAVID NIMMER, 1 NIMMER ON COPYRIGHT § 1.03[B], at 1-90 n.11.2 (2007) (noting that recent Supreme Court decisions support the distinction of copyright as intended to promote science and patent as intended to promote inventions); Karl B. Lutz, *Patents and Science: A Clarification of the Patent Clause of the U.S. Constitution*, 32 J. PAT. OFF. SOC'Y 83, 84 (1950) (investigating the Intellectual Property clause and agreeing that the drafters intended to promote “science” by granting copyrights and the “useful arts” by granting patents); *see also* Act of April 10, 1790, ch. 7, 1 Stat. 109 (promulgating the act under the name of “An Act to promote the progress of useful arts”).

49. *See* L. Ray Patterson & Craig Joyce, *Copyright in 1791: An Essay Concerning the Founders' View of the Copyright Power Granted to Congress in Article I, Section 8, Clause 8 of the U.S. Constitution*, 52 EMORY L.J. 909, 946 (2003) (highlighting three

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1123

evidenced by the fact that all of the early copyright bills were called “Act[s] to promote learning,” the authors of this legislation believed that something greater was at stake.<sup>50</sup>

In an address to Congress on January 8, 1790, President George Washington emphasized,

[T]here is nothing which can better deserve your patronage than the promotion of science and literature. Knowledge is, in every country, the surest basis of public happiness. In one in which the measures of government receive their impression so immediately from the sense of the community as in ours, it is proportionably essential.<sup>51</sup>

In response, the newly formed House of Representatives resolved: “We concur with you in the sentiment that . . . the promotion of science and literature will contribute to the security of a free Government; in the progress of our deliberations we shall not lose sight of objects so worthy of our regard.”<sup>52</sup> In May of that year, Congress passed the first federal copyright law,<sup>53</sup> which sought to advance progress through the creation of limited rights for published maps, charts, and books, as well as through archiving, cataloguing, and advertising the knowledge furnished by those publications.

---

broader purposes of the founders’ conception of copyright law: (1) to promote learning; (2) to provide public access; and (3) to protect the public domain); *cf.* Oliar, *supra* note 48, at 1845 (arguing that the first part of the Intellectual Property Clause, which includes language about promoting progress, serves as a limitation on Congress’s power to regulate intellectual property). The U.S. Supreme Court would later emphasize the broader purposes of the Intellectual Property Clause:

“The copyright law, like the patent statutes, makes reward to the owner a secondary consideration.” However, it is “intended definitely to grant valuable, enforceable rights to authors, publishers, etc., without burdensome requirements; ‘to afford greater encouragement to the production of literary [or artistic] works of lasting benefit to the world.’”

The economic philosophy behind the clause empowering Congress to grant patents and copyrights is the conviction that encouragement of individual effort by personal gain is the best way to advance public welfare through the talents of authors and inventors in “Science and useful Arts.” Sacrificial days devoted to such creative activities deserve rewards commensurate with the services rendered.

*Mazer v. Stein*, 347 U.S. 201, 219 (1954) (alteration in original) (internal citations omitted).

50. See 8 NIMMER & NIMMER, *supra* note 48, Appx. 7, at 7-5, 7-25, 7-41 (collecting Colonial copyright laws). The 1709 English Statute of Anne, on which parts of the early colony copyright statutes were modeled, was titled “An Act for the Encouragement of Learning by vesting the Copies of printed Books in the Authors or Purchasors of such Copies during the Times therein mentioned.” Statute of Anne, 1709, 8 Ann., c. 21 (Eng.).

51. THORVALD SOLBERG, LIBRARY OF CONG., COPYRIGHT IN CONGRESS 1789–1904, at 115 (Greenwood Press 1976) (1905).

52. *Id.* at 118.

53. Act of May 31, 1790, ch. 15, 1 Stat. 124.

Of particular note for the purposes of this Article, Congress showed substantial interest in informing the public about works protected by copyright and preserving copies of copyrighted works. It accomplished that goal through a deposit requirement. Under the 1790 Act, Congress required that an author publish a record of the registered work in a local newspaper for a month.<sup>54</sup> Furthermore, the Act provided that “no person shall be entitled to the benefit of this act . . . unless he shall first deposit . . . a printed copy . . . in the clerk’s office of the district court where the author or proprietor shall reside.”<sup>55</sup> The Act further provided that “the author or proprietor of any such map, chart, book or books, shall, within six months after the publishing thereof, deliver, or cause to be delivered to the Secretary of State a copy of the same, to be preserved in his office.”<sup>56</sup>

The deposit requirement initially served primarily as record evidence of copyright protection.<sup>57</sup> But as the young nation developed and research institutions were formed, the preservation and access attributes of deposit came to be recognized as part of the copyright balance. In the 1846 Act establishing the Smithsonian Institution, Congress provided that one copy of each copyrighted work be delivered to the Librarian of the Smithsonian Institution and to the Librarian of Congress within three months of publication.<sup>58</sup> Charles Jewett, the first librarian of the Smithsonian Institution, extolled the preservation, access, and scholarly virtues of copyright deposit:

To the public, the importance, immediate and prospective, of having a central depot, where all the products of the American press may be gathered, year by year, and preserved for reference, is very great. The interest with which those who in 1950 may consult this library would view a complete collection of all the works printed in America in 1850, can only be fully and rightly estimated by the historian and bibliographer, who has sought in vain for the productions of the past. . . . Thus, in coming years, the collection would form a documentary history of American letters, science and art. It is greatly to be desired, however,

---

54. Act of May 31, 1790, ch. 15, § 3, 1 Stat. 124, 125. (“And such author or proprietor shall, within two months from the date thereof, cause a copy of the said record to be published in one or more of the newspapers printed in the United States, for the space of four weeks.”).

55. *Id.*

56. Act of May 31, 1790, ch. 15, § 4, 1 Stat. 124, 125.

57. See *Wheaton v. Peters*, 33 U.S. (8 Pet.) 591, 665 (1834) (“The deposite of the book in the department of state, may be important to identify it at any future period, should the copyright be contested, or an unfounded claim of authorship asserted.”).

58. Act of Aug. 10, 1846, ch. 178, § 10, 9 Stat. 102, 106 (1851).

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1125

that the collection should be *complete, without a single omission*. We wish for every book, every pamphlet, every printed or engraved production, however apparently insignificant. Who can tell what may not be important in future centuries?<sup>59</sup>

To facilitate building a national knowledge archive, Jewett advocated that publishers be able to transmit deposit copies free of postage, a proposal which Congress enacted in 1855.<sup>60</sup>

Practical problems ensued in managing the national knowledge archive. The Smithsonian Institution was inundated with materials considered of relatively low archival value (textbooks, music, and prints) and that were difficult to store, whereas publishers of substantial research works failed to comply with the deposit requirement.<sup>61</sup> This problem was exacerbated by a judicial decision holding that, although deposit of a copy with the district court was required to secure copyright protection, failure to deposit a work at the national repository had no effect on copyright protection.<sup>62</sup>

Congress ultimately reorganized the administration of the national library, repealing the deposit provision of the 1846 Act, establishing the Smithsonian Institution in 1859, and transferring the copyright deposits and records to the Department of Interior.<sup>63</sup> In 1864, President Abraham Lincoln appointed Ainsworth R. Spofford as Librarian of Congress, bringing new energy and vision to the role.<sup>64</sup> Spofford sought to invigorate the Library of Congress as a national repository and looked to copyright deposit as a means toward that end.<sup>65</sup> Congress supported this goal in its revision of the Copyright Act in 1865, requiring the Librarian of Congress to make a written demand of copyright registrants to comply with the deposit requirement with the Library of Congress if they had failed to do

---

59. SMITHSONIAN INSTITUTION, FOURTH ANNUAL REPORT OF THE BOARD OF REGENTS 35 (1850).

60. See Act of Mar. 3, 1855, ch. 201, § 5, 10 Stat. 683, 685 (pertaining to the Post Office Department).

61. See 11 SMITHSONIAN INSTITUTION ANNUAL REPORT 40 (1856); see also Dunne, *supra* note 29, at 423 (“Few except the leading publishers complied automatically, although most deposited on demand.”).

62. See *Jollie v. Jacques*, 13 F. Cas. 910, 910, 912 (S.D.N.Y. 1850) (basing interpretation on the fact that the national repository copy requirement was set forth in legislation establishing the Smithsonian Institution and not the Copyright Act, while expressing “some doubt” as to this construal).

63. Act of Feb. 5, 1859, ch. 22, §§ 5–6, 11 Stat. 379, 380.

64. MICHAEL H. HARRIS, HISTORY OF LIBRARIES IN THE WESTERN WORLD 250 (1984).

65. See Dunne, *supra* note 29, at 423 (describing the early policies of Spofford aimed at increasing the use of the Library of Congress as a depository).

so within a month of registration.<sup>66</sup> Failure to deposit the work within another month would result in forfeiture of the protections of copyright.<sup>67</sup> The burden of issuing a written demand was cumbersome, leading Spofford to call for a monetary penalty for failure to make a deposit,<sup>68</sup> which Congress enacted in 1867.<sup>69</sup>

In 1870, Congress streamlined copyright administration by centralizing all copyright business with the Librarian of Congress.<sup>70</sup> In a statement before the House of Representatives, Congressman Thomas Allen Jencks, chairman of the committee handling the legislation, explained that much of the national collection of deposited materials was effectively “beyond the reach of consultation” due to storage in an inaccessible location.<sup>71</sup> He noted that the Library of Congress had the requisite storage capacity, facilities, and skilled cataloguing staff to make the works available to the public.<sup>72</sup> The 1870 Act made clear that deposit of work within ten days of publication was required for copyright protection.<sup>73</sup> The Act continued the franking privilege (free postage for deposits) and imposed a monetary fine for failure to comply with the tight deposit deadline.<sup>74</sup> The Act also required the Librarian to make an annual report to Congress and to prepare a catalogue of entries.<sup>75</sup>

The resulting deluge of copyrighted works greatly expanded the Library of Congress’s collection.<sup>76</sup> This expansion coincided with the growth of other libraries throughout the nation.<sup>77</sup> After

---

66. Copyright Act of 1865, ch. 126, § 3, 13 Stat. 540, 540, amended by Copyright Act of 1909, ch. 320, § 63, 35 Stat. 1075, 1088.

67. *Id.*

68. Dunne, *supra* note 29, at 423.

69. See Act of Feb. 18, 1867, ch. 43, § 1, 14 Stat. 395, 395 (assessing a \$25 penalty to be collected by the Librarian of Congress for failure to deposit a work within one month of publication).

70. See Act of July 8, 1870, ch. 230, § 85, 16 Stat. 212.

71. CONG. GLOBE, 41st Cong., 2d Sess. 2683 (1870).

72. *Id.* at 2683–84.

73. Act of July 8, 1870, ch. 230, § 90, 16 Stat. 198, 213.

74. Act of July 8, 1870, §§ 94–95, 16 Stat. at 213.

75. CONG. GLOBE, 41st Cong., 2d Sess. 2684 (1870).

76. See Dunne, *supra* note 29, at 424 (“The practical result of the act of 1870 was to increase deposits so substantially that the Librarian asked for relief almost immediately.”).

77. See Jacob Rooksby, *Access and Storage of Knowledge in the New Millennium: The Google Book Search Library Project and the Future of Libraries*, TCHRS. C. REC., Jan. 11, 2007, <http://www.tcrecord.org/Content.asp?ContentID=12913> (noting that American universities began expanding after 1850 as a result of increases in money and interest in higher education).

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1127

1850, institutions of higher learning greatly expanded,<sup>78</sup> as did their libraries.<sup>79</sup> The libraries took on greater importance as more universities shifted toward the German model by adopting a seminar style of learning in which students conducted original research.<sup>80</sup> By the late 19th century, the library “ceased being a museum and became a more active part of the academic program” at colleges and universities.<sup>81</sup>

The proliferation of published works created a new problem for the nation: organizing the vast collections. At the university level, this led to the development of cataloguing systems such as the Dewey Decimal Classification, a proprietary system first developed by Melville Dewey in the late 19th century.<sup>82</sup> Congress also sought to address the information overload problem through a provision of the Copyright Act of 1891, which required the Library of Congress to publish “catalogues of . . . title-entries” at regular intervals.<sup>83</sup> This catalog would serve two purposes—it would function as a resource for customs officials seeking to prevent illegal importation of copyrighted works and as a tool for the public to conduct research.<sup>84</sup> The Library of Congress was not adequately equipped to handle search requests by the public, and the physical dimensions of the building were not conducive to simultaneous occupation by the public and staff.<sup>85</sup> The public took little notice of the catalog, despite efforts by the Copyright Office

---

78. Under the Morrill Land Grant Act of 1862, each eligible state received a total of 30,000 acres of federal land for each member of Congress the state had as of the census of 1860. 17 U.S.C. § 301 (2000). This land, or the proceeds from its sale, was to be used toward establishing and funding the educational institutions. 17 U.S.C. § 304 (2000). Many of the great universities of the United States trace their origins to this legislation. See HARRIS, *supra* note 64, at 232–33 (discussing the impact of the Morrill Land Grant Act on the growth of American universities).

79. See HARRIS, *supra* note 64, at 232–33 (noting that the libraries of these institutions “soon became some of the finest in the land”).

80. See *id.* at 233 (expressing that the German model “made library resources a high priority”).

81. *Id.* at 236.

82. BATTLES, *supra* note 21, at 138–39. In enacting the Digital Millennium Copyright Act (DCMA), Congress praised Yahoo! as “a ‘card catalogue’ to the World Wide Web” and safeguarded it from liability on that basis. See S. REP. NO. 105-190, at 49 (1998), *quoted in* 3 NIMMER & NIMMER, *supra* note 48, § 12B.05[A][1], at 12B-69.

83. See Act of Mar. 3, 1891, ch. 565, § 4, 26 Stat. 1106, 1108.

84. See Elizabeth K. Dunne & Joseph W. Rogers, *Study No. 21: The Catalog of Copyright Entries*, in 1 STUDIES ON COPYRIGHT, *supra* note 29, at 451, 453 (stating that the Catalog proved ineffective in preventing illegal importation).

85. See *id.* at 457 (citing the lack of staff to process requests from the public, ineffective methods for entering the official record into ledgers, poor quality material used in the physical entry, and increasingly crowded space as limitations to the public’s access to the Catalog).

to make it a more efficient research tool.<sup>86</sup> However, certain industries (namely music and motion picture) began to rely on the catalog, as it was the most comprehensive in the field.<sup>87</sup> Book publishers relied on alternative catalogs produced by third parties.<sup>88</sup>

Congress comprehensively overhauled the Copyright Act in 1909.<sup>89</sup> In the legislative history, Congress emphasized that the purpose of copyright is “[n]ot primarily for the benefit of the author, but primarily for the benefit of the public.”<sup>90</sup> While doubling the duration of protection, expanding the scope of works protected, and creating a compulsory license for musical compositions, Congress largely maintained the preservation and access provisions of prior legislation. The 1909 Act provides that works must be deposited promptly with the Library of Congress,<sup>91</sup> with failure to comply subject to a monetary penalty and potential loss of copyright protection.<sup>92</sup> Congress also expanded the Library of Congress’s cataloguing function.<sup>93</sup>

#### D. Modern U.S. Copyright Law

By the mid 1950s, advances in technology, growth of copyright industries, and the increasing importance of international commerce in copyrighted works generated pressure to update and systematically codify the U.S. copyright regime. To guide that process, the Copyright Office funded thirty-five studies of the history, functioning, and international differences

---

86. See *id.* at 457–58 (discussing attempts beginning in 1945 to improve the public service activities of the Copyright Office). In the late 19th century, the Library of Congress developed the Library of Congress Classification system which continues to be used by most research and academic libraries in the United States. Library of Congress, Library of Congress Classification, <http://www.loc.gov/catdir/cpsolcc.html> (last visited Oct. 16, 2007).

87. See Dunne & Rogers, *supra* note 84, at 458.

88. See *id.* at 461–62 (noting that these publications were “better organized for general bibliographic purposes”).

89. Copyright Act of 1909, ch. 320, 35 Stat. 1075 (current version in scattered sections of 17 U.S.C.).

90. H.R. REP. NO. 60-2222, at 7 (1909).

91. See Copyright Act of 1909, ch. 320, § 12, 35 Stat. 1075, 1078.

92. See *id.* Later decisions held that failure to deposit a work would result in forfeiture only if the copyright proprietor ignored an explicit demand by the Register of Copyrights. See *Washingtonian Pub. Co. v. Pearson*, 306 U.S. 30, 42 (1939); see also *King Features Syndicate, Inc. v. Bouve*, 48 U.S.P.Q. 237, 242 (D.D.C. 1940) (finding that the Register of Copyrights “has no power to refuse registration of a claim of copyright which is entitled to registration under the Copyright Act”).

93. See Copyright Act of 1909, ch. 320, §§ 56–60, 35 Stat. 1075, 1086–87 (providing that all copyright registrations shall be indexed, printed, and distributed to be used by customs officers, postmasters, and the general public).

of the major aspects of copyright law.<sup>94</sup> As referenced earlier, the most relevant studies for purposes of addressing the preservation and access functions of copyright law related to the deposit and cataloguing provisions. The revision process stalled because of wrangling among a variety of competing interests—ranging from juke box enterprises to cable television stations.<sup>95</sup> During this time, a new issue emerged: photocopying and the exposure of libraries for copyright infringement.<sup>96</sup> Furthermore, the sheer volume and growth of published material presented increasing challenges for libraries and archivists.

The revision process eventually reached fruition with the passage of the Copyright Act of 1976.<sup>97</sup> With regard to deposit and cataloguing functions of copyright policy, the 1976 Act largely retained the features of prior law. In the area of libraries, Congress reinforced the preservation and access goals of the copyright system by insulating public libraries from vicarious liability for the acts of patrons and authorizing limited photocopying to maintain the integrity and comprehensiveness of their collections.<sup>98</sup> Congress also created the American Television and Radio Archives within the Library of Congress.<sup>99</sup> Later amendments to the Copyright Act updated preservation and access rules in light of digital technology.<sup>100</sup>

1. *The Deposit Requirement.* The Deposit Study, prepared as background for the copyright revision process, endorsed the deposit system for its major contributions to enriching the Library of Congress's collection.<sup>101</sup> The study struggled to deal

---

94. See Harry G. Henn & Walter J. Derenberg, *Introduction* to STUDIES ON COPYRIGHT, *supra* note 29, at ix, ix (stating that these studies covered "practically all aspects of American copyright law").

95. See H.R. REP. NO. 89-2237, at 31–32 (1966) (discussing the inadequacy of then-current copyright law in light of developing technology).

96. See H.R. REP. NO. 89-2237, at 66 (1966) (observing the historical and scholarly importance of allowing archival institutions to make copies of unpublished works in order to deposit copies in manuscript collections).

97. Copyright Act of 1976, Pub. L. No. 94-553, 90 Stat. 2541 (codified as amended in scattered sections of 17 U.S.C.).

98. Copyright Act of 1976, Pub. L. No. 94-553, § 101, 90 Stat. 2541, 2546–47 (current version at 17 U.S.C. § 108 (2000)).

99. Copyright Act of 1976, Pub. L. No. 94-553, § 101, 90 Stat. 2541, 2546 (current version at 17 U.S.C. § 170(a) (2000)).

100. See Online Copyright Infringement Liability Limitation Act, Pub. L. No. 105-304, § 202(a), 112 Stat. 2877, 2877–86 (1998) (current version at 17 U.S.C. § 512 (2000)) (limiting liability related to material online).

101. See Dunne, *supra* note 29, at 440 (noting that the deposit system "has materially assisted the Library in building its collections on all aspects of American history, literature, law, music, and social culture").

with two issues: (1) what policies should govern retention and disposition of submitted works given the deluge of materials coming into the Copyright Office and the costs of managing and storing these items; and (2) how should deposit be most appropriately encouraged (whether through monetary fines, risk of forfeiture of copyright, or both).<sup>102</sup> With regard to retention and disposition policy, Congress afforded the Library of Congress greater flexibility in managing its collection.<sup>103</sup> With regard to deposit policy, Congress sought to ensure a comprehensive public collection of knowledge by requiring that the “owner of copyright or of the exclusive right of publication” deposit two copies or phonorecords of the work in the Copyright Office.<sup>104</sup> The 1976 Act imposed sanctions—\$250 per work and an additional \$2,500 penalty for willful violation—for failure to comply with a deposit demand from the Register of Copyrights within three months.<sup>105</sup> Congress did not, however, make deposit a condition of copyright protection. Therefore, violation of the deposit requirement would not result in forfeiture of copyright. Since Congress was well aware that deposit had only rarely served as a source of evidence in copyright litigation,<sup>106</sup> the deposit policy was now aimed principally at preservation and access.

2. *The Catalog of Copyright Entries.* During the first half of the 20th century, the Catalog of Copyright Entries served as a resource for researching registration records outside of the Copyright Office and as a national bibliography of current U.S. literary and artistic productions available in over 300 depository libraries throughout the United States.<sup>107</sup> An informal survey of depository libraries indicated that usage of the catalog varied geographically. In some regions, the catalog served to assist copyright record searching; in others, it assisted with bibliographic searching.<sup>108</sup> Among private subscribers, the catalog was particularly useful in the music and film industries.<sup>109</sup> The

---

102. See *id.* at 435–49 (discussing different types of deposit systems in an attempt to identify the best system to serve the basic purposes of copyright law).

103. See 17 U.S.C. § 704 (2000) (defining the means for transfer, retention, and disposal of deposited works maintained by the Library of Congress).

104. See Copyright Act of 1976, Pub. L. No. 94-553, § 101, 90 Stat. 2541, 2579 (current version at 17 U.S.C. § 407(a)–(b) (2000)).

105. See Copyright Act of 1976, § 101, 90 Stat. at 2579 (current version at 17 U.S.C. § 407(d) (2000)).

106. See Dunne, *supra* note 29, at 436 (recognizing that from 1929 through 1960 only seven known court cases have used deposit copies as evidence).

107. See Dunne & Rogers, *supra* note 84, at 453, 462.

108. *Id.* at 463.

109. *Id.* at 461–62.

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1131

catalog was of less value in the book industry, which developed better organized publications for its needs.<sup>110</sup> Usage of the catalog for customs purposes ended in 1953 due to a lack of efficacy.<sup>111</sup>

The Catalog of Copyright Entries Study offered only tepid endorsement for continuation of this feature of copyright law. The study fully endorsed the public record and access goals, as well as the catalog's role as a safeguard against destruction of the original records.<sup>112</sup> At the same time, though, it raised concerns about the catalog's limited and uneven usage by the public and its relatively high cost of preparation and production.<sup>113</sup> The study solicited input on several alternatives, such as the use of microform, making available duplicate sets of Copyright Office card entries for purchase, record searches by the Copyright Office, and more limited catalogs of particular areas (such as music and motion pictures).<sup>114</sup>

The comments offered by librarians divided on whether the catalog's cost was justified given its rather limited usage.<sup>115</sup> Most librarians supported a "more flexible" approach to the production of the catalog, suggesting that the Copyright Office should no longer be statutorily compelled to produce those parts of the Catalog that are not being used (e.g., for books). One librarian aptly characterized the growing information overload problem faced by library users, noting that the catalog covered fifty-six feet of shelf space in a discouraging, unbound format.<sup>116</sup> She went on to comment that:

For the use of the section on books and periodicals, we occasionally have an enthusiastic bibliophile who, having heard of this wonderful collection, arrives eagerly prepared to have all of his questions answered immediately. When faced with the expanse of separate indexes, his enthusiasm almost invariably dims and, after a short time or desultory poking about, he goes off to use the "National Union Catalog" of the CBI. The prospect of searching between fifty and one hundred separate indexes unless copyright date of the title sought is known discourages use . . . .<sup>117</sup>

---

110. *Id.* at 462.

111. *Id.* at 467.

112. *Id.*

113. *Id.* at 468–69.

114. *Id.* at 470.

115. S. COMM. ON THE JUDICIARY, 86TH CONG. 2D SESS., COPYRIGHT LAW REVISION: STUDIES PREPARED FOR THE SUBCOMMITTEE ON PATENTS, TRADEMARKS AND COPYRIGHTS OF THE COMMITTEE ON THE JUDICIARY (Comm. Print 1960).

116. *Id.* at 81 (comments of Ellen Jackson).

117. *Id.*

Congress ultimately retained the cataloguing function, although it afforded the Copyright Office greater leeway in how it would be compiled and published. Section 707(a) of the 1976 Act provides that

The Register of Copyrights shall compile and publish at periodic intervals catalogs of all copyright registrations. These catalogs shall be divided into parts in accordance with the various classes of works, and the Register has discretion to determine, on the basis of practicability and usefulness, the form and frequency of publication of each particular part.<sup>118</sup>

3. *The Library Provisions.* In the 1976 Act, Congress directly reinforced the preservation and access functions of the copyright system through a new set of provisions dealing with reproduction by and at libraries and archives.<sup>119</sup> These provisions augment the general fair use privilege and afford libraries greater leeway in copying and distributing copyrighted works.<sup>120</sup>

In section 108(a), Congress promoted access to copyrighted works through libraries and archives, while at the same time imposing standards to prevent commercial sales of copyrighted works from being supplanted by photocopying. The section exempts public libraries and archives from liability for the reproduction or distribution of a single copy of a work,<sup>121</sup> provided that the reproduction or distribution is not made for commercial advantage, the collections are open to the public, and the reproduction or distribution of the work includes a notice of

---

118. Copyright Act of 1976, Pub. L. No. 94-553, § 101, 90 Stat. 2541, 2592–93 (current version at 17 U.S.C. § 707(a) (2000)).

119. See Copyright Act of 1976, § 101, 90 Stat. at 2546–48 (current version at 17 U.S.C. § 108 (2000)) (excluding from the definition of copyright infringement reproductions made by libraries and archives); see also James M. Treece, *Library Photocopying*, 24 UCLA L. REV. 1025, 1031–38 (1977) (commenting on how the Copyright Act of 1976 dealt with the issue of photocopying by public libraries).

120. The House Report explains that

The doctrine of fair use applies to library photocopying, and nothing contained in section 108 “in any way affects the right of fair use.” No provision of section 108 is intended to take away any rights existing under the fair use doctrine. To the contrary, section 108 authorizes certain photocopying practices which may not qualify as a fair use.

H.R. REP. NO. 94-1476, at 74 (1976). The House Report also emphasizes that “the making of duplicate copies for purposes of archival preservation certainly falls within the scope of ‘fair use.’” *Id.* at 73 (noting the problem of disintegrating motion pictures).

121. Subsection (h) excludes musical, graphic, and audiovisual works from the specific exemptions of section 108, although the House Report emphasized that the fair use provision fully applies to such works. *Id.* at 78–79.

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1133

copyright.<sup>122</sup> “Commercial advantage” refers to the “immediate commercial motivation behind the reproduction or distribution itself,” and not to the profit-making status of the library or its parent organization.<sup>123</sup> Activities at a library in a for-profit institution can fall within this exemption, even if the reproduction ultimately helps a company to make a profit,<sup>124</sup> but “section 108 would not excuse reproduction or distribution if there were a commercial motive behind the actual making or distributing of the copies.”<sup>125</sup> In defining “commercial advantage,” the House and Senate reports explain that section 108 would not excuse “systematic” reproductions, made to avoid buying multiple subscriptions.<sup>126</sup> In addition, mere association with a library is not enough to qualify. Under 108(a)(1), a library may not contract out copying to a commercial organization, even if the library would be allowed to perform the same activities.<sup>127</sup>

Sections 108(b) and (c) afford libraries and archives particular flexibility to reproduce and distribute a copy of an unpublished work for purposes of preservation and security, replacement of damaged materials, and the provision to patrons of copies of out-of-print works.<sup>128</sup> In its original formulation, section 108 only allowed one copy, and that copy had to be made “in facsimile form,” which excluded digital copies.<sup>129</sup> Thus, as long as a library had an unpublished work in their collection, they could make one copy of it for preservation or replacement of a damaged copy.<sup>130</sup> The Digital Millennium Copyright Act (DMCA) added the option that a reproduction could be made in a digital format, but with the caveat that it could not be “made available to the public in that format outside the premises of the library,” a

---

122. Copyright Act of 1976, Pub. L. No. 94-553, § 101, 90 Stat. 2541, 2546–48 (current version at 17 U.S.C. § 108(a)(3) (2000)). This provision was easily followed when the original contained a notice of copyright, but the provision became cumbersome, if not impossible, when there was no such notice (e.g. unpublished works). See NIMMER & NIMMER, *supra* note 48, § 8.03[D], at 8-40. As a result, section 108(a)(3) was amended in 1998 so that, if the original did not contain any copyright notice, libraries could instead append a simple statement that the material might fall under copyright protection. S. REP. NO. 105-190, at 60–61 (1998).

123. H.R. REP. NO. 94-1476, at 75 (1976). Mere association with a library, however, does not bring an enterprise within this exemption. *Id.* at 74.

124. See NIMMER & NIMMER, *supra* note 48, § 8.03[a][1], at 8-36 to 8-37.

125. See H.R. REP. NO. 94-1476, at 75 (1976).

126. *Id.* at 74–75; S. REP. 94-473 at 70–71 (1975).

127. See H.R. REP. NO. 94-1476, at 74 (1976).

128. See 17 U.S.C. § 108(b)–(c) (2000) (detailing libraries rights to reproduce and distribute copyrighted works).

129. Copyright Act of 1976, Pub. L. No. 94-553, § 101, 90 Stat. 2541, 2547 (current version at 17 U.S.C. § 108(b)–(c) (2000)).

130. *Id.*

limitation designed to minimize unauthorized sharing and distribution.<sup>131</sup>

Sections 108(d) and (e) authorize libraries to make copies of copyrighted works under limited circumstances. Under section 108(d), a library may provide a patron, who so requests, with a small portion of a copyrighted work or a single article subject to the conditions that the copy becomes the property of the patron and is to be used only for research purposes.<sup>132</sup> The library may not honor the request if it suspects other motivations.<sup>133</sup> In addition, the library must highlight copyright protection status by “prominently” displaying notice on the order form and in the location where they accept the requests.<sup>134</sup> Section 108(e) authorizes libraries to provide reproductions of entire works within their collections if they are not available at a “fair price,” generally referring to out-of-print works.<sup>135</sup> To prevent this provision from impinging upon book sales, libraries must conduct a “reasonable investigation” to determine whether the work is available at a “fair price.”<sup>136</sup> As with section 108(d), the library is not permitted to provide a copy if it has notice that the work will be used for any purpose other than research and it must display notice of copyright protection.<sup>137</sup>

The main thrust of section 108(f) is to immunize libraries from indirect liability with regard to infringing actions of their patrons,<sup>138</sup> including the use of reproduction technology provided by the library.<sup>139</sup> Libraries are insulated from liability for infringing photocopying on their premises by patrons so long as the library posts a notice indicating that making a copy may be subject to copyright.<sup>140</sup> A user is still responsible for any infringing reproductions,<sup>141</sup> and a library cannot use section 108 to circumvent contractual obligations that it undertakes in

---

131. See Digital Millennium Copyright Act, Pub. L. 105-304, § 404, 112 Stat. 2860, 2890 (1998) (current version at 17 U.S.C. § 108(b)–(c) (2000)); see also S. REP. 105-190, at 61 (1998).

132. Copyright Act of 1976, Pub. L. No. 94-553, § 101, 90 Stat. 2541, 2547 (current version at 17 U.S.C. § 108(d) (2000)).

<sup>133</sup> See *id.* (stating that the library must have had “no notice that the copy . . . would be used for any purpose other than private study, scholarship, or research”).

134. *Id.*

135. See *id.*

136. *Id.*

137. See *id.*

138. See H.R. REP. NO. 94-1476, at 77 (1976).

139. Copyright Act of 1976, Pub. L. No. 94-553, § 101, 90 Stat. 2541, 2547 (current version at 17 U.S.C. § 108(f)(1) (2000)).

<sup>140</sup> See *id.*

141. *Id.* (current version at 17 U.S.C. § 108(f)(2) (2000)).

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1135

acquiring materials.<sup>142</sup> In addition, section 108(f)(3) specifically allows for limited reproduction of audiovisual news programs “intended to apply to the daily newscasts of the national television networks, which report the major events of the day.”<sup>143</sup> This section was added to 108 in order to protect organizations that preserve, archive, and catalogue television and radio news, because Congress saw “the importance of preserving television news” and wanted to encourage this endeavor.<sup>144</sup> To that end, Congress also created the American Television and Radio Archives at the Library of Congress.<sup>145</sup>

Congress also exempted libraries in profit-making organizations for isolated and spontaneous production of single photocopies so long as such activities do not constitute a systematic effort to substitute for subscriptions or purchases.<sup>146</sup> When the 1976 Act was considered, there was concern that this prohibition against systematic reproduction would inhibit interlibrary loan activities, so Congress added the stipulation that this requirement should not limit interlibrary loans, as long as libraries were not using the system to substitute for their own subscriptions.<sup>147</sup> Congress recognized that it might be difficult to distinguish “systematic reproduction” from acceptable copying.<sup>148</sup>

In crafting the Sonny Bono Copyright Term Extension Act of 1998—which added twenty years to the term of copyright protection—Congress limited the effect of the extension on access by carving out a limited exemption for libraries and archives to make copies of out-of-print and orphaned (works for which copyright ownership cannot be reasonably traced) works.<sup>149</sup> Section 108(h) exempts libraries from liability for reproducing, displaying, or performing works in their last twenty years of protection for purposes of preservation, scholarship, or research if, upon reasonable investigation, the library determines that the work is not subject to “normal commercial exploitation,” cannot be obtained at a reasonable price, and the copyright owner has

---

142. *Id.* (current version at 17 U.S.C. § 108(f)(4) (2000)).

143. H.R. REP. NO. 94-1476, at 77 (1976); *see also* Copyright Act of 1976, Pub. L. No. 94-553, § 101, 90 Stat. 2541, 2546 (current version at 17 U.S.C. §§ 108(f)(3) (2000)).

144. S. REP. NO. 94-473, at 69 (1975).

145. H.R. REP. NO. 94-1476, at 182–83 (1976) (“The need for such a repository has become more pressing as the importance of television’s role in American society has increased . . .”).

146. *See* Copyright Act of 1976, Pub. L. No. 94-553, § 101, 90 Stat. 2541, 2548 (current version at 17 U.S.C. § 108(g) (2000)).

147. *Id.*; H.R. REP. NO. 94-1476, at 77–78 (1976).

148. *See* S. REP. NO. 94-473, at 70 (1975).

149. *See* 2 NIMMER & NIMMER, *supra* note 48, § 8.03, at 8-55 to 8-57.

not properly notified the Register of Copyrights that the work is available.<sup>150</sup>

Congress also carved out limited exemptions for libraries in the anticircumvention provisions of the DMCA. To assist copyright owners in protecting their works against digital piracy, section 1201 of the DMCA imposes liability upon those who circumvent technological protection measures put in place by copyright owners to limit access to their works.<sup>151</sup> Section 1201(d) exempts nonprofit libraries, archives, and educational institutions from liability for circumventing technological protection measures to the extent necessary to determine whether to add copyrighted works to their collections.<sup>152</sup>

Taken together, these provisions reflect a clear national policy favoring flexibility and access to knowledge by the public, as well as the use of reproduction technologies to promote preservation.<sup>153</sup> Congress balanced the access and preservation goals with limitations and requirements aimed at avoiding disruption of content markets. Thus, libraries may not directly or knowingly assist in unauthorized commercial exploitation of copyrighted materials.

4. *Television and Radio Archives.* As part of the 1976 Copyright Act, Congress created the American Television and Radio Archives.<sup>154</sup> The archive is maintained by the Library of Congress “to preserve a permanent record of the television and radio programs which are a heritage of the people of the United States.”<sup>155</sup> Congress recognized the need to catalogue and maintain the large amount of audiovisual works created through television and radio, but still wanted to balance scholarly and

---

150. Sonny Bono Copyright Term Extension Act, Pub. L. No. 105-298, § 104, 112 Stat. 2827, 2829 (1998) (current version at 17 U.S.C. § 108(h) (2000)).

151. Digital Millennium Copyright Act, Pub. L. No. 105-304, § 103, 112 Stat. 2860, 2863–72 (1998) (current version at 17 U.S.C. § 1201 (2000)).

152. Digital Millennium Copyright Act, § 103, 112 Stat. at 2866 (current version at 17 U.S.C. § 1201(d) (2000)).

153. Other exceptions to copyright law that concern libraries include: (1) exemption for reproducing copies of published literary works for use by persons with disabilities (§ 121); (2) the § 504(c) remittance of statutory damages for a library or archive employee acting within the scope of their employment; and (3) the ability to import a limited number of copies acquired outside of the United States. *See* 17 U.S.C. §§ 121, 504(c), 602(a)(3) (2000).

154. American Television and Radio Archives Act, Pub. L. No. 94-553, § 113, 90 Stat. 2541, 2601 (1976) (current version at 2 U.S.C. § 170 (2000)); *see also* H.R. REP. NO. 94-1476, at 182–83 (1976) (noting that the newly-established archives “will provide a repository for the preservation of our nation’s television and radio heritage”).

155. 2 U.S.C. § 170 (2000).

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1137

archival needs against the rights of copyright holders.<sup>156</sup> In addition to maintaining and indexing the archive, the Library of Congress may make limited reproductions and distributions for scholarly purposes.<sup>157</sup>

Because of the American Television and Radio Archives Act, a subsection was added to the deposit requirement allowing the Library of Congress to record programs and to make one copy for the archive. The Library may even “demand” a deposit of a specific transmission (although it may not make a blanket demand for a series of transmissions).<sup>158</sup> According to the Library of Congress website, “because of the copyright law, the Library is very strong in American commercial television.”<sup>159</sup> Although this provision does not apply to all libraries, it reveals a clear public policy to preserve and catalog works for research and scholarship.

Subsequent laws have reinforced the national interest in film preservation. Congress passed the National Film Preservation Act of 1988 to promote faithful safeguarding of the integrity of films.<sup>160</sup> The Act directed the Library of Congress to establish a National Film Preservation Board (NFPB).<sup>161</sup> This Board would designate twenty-five films per year as “culturally, historically, or aesthetically significant” to be added to a National Film Registry.<sup>162</sup> In 1992, Congress reauthorized the NFPB, repealed the labeling requirement for altered films, and promoted film preservation by requiring the Librarian of Congress to prepare a comprehensive report on preservation activities and establish a comprehensive film preservation program in conjunction with other film archivists and copyright owners.<sup>163</sup> Subsequent acts have reauthorized the NFPB and created the National Film Preservation Foundation.<sup>164</sup>

---

156. See H.R. REP. NO. 94-1476, at 182, as reprinted in 8 NIMMER & NIMMER, *supra* note 48, Appx. 4, at 4-2, 4-200 to 4-200.

157. See 2 U.S.C. § 170(b) (2000).

158. Copyright Act of 1976, Pub. L. No. 94-553, § 101, 90 Stat. 2541, 2580 (current version at 17 U.S.C. § 407(e) (2000)); see also H. REP. NO. 94-1476, at 152 (1976).

159. Library of Congress, Collections Overviews—Special Materials/Formats, <http://www.loc.gov/acq/devpol/colloversviews/television.html> (last visited Aug. 26, 2007).

160. See National Film Preservation Act of 1988, Pub. L. No. 100-446, 102 Stat. 1782, repealed by Film Preservation Act of 1992, Pub. L. No. 102-307, § 214, 106 Stat. 267, 272.

161. *Id.*

162. National Film Preservation Act of 1988, Pub. L. No. 100-446, § 2, § 10(b), 102 Stat. 1782, 1782, 1787, repealed by Film Preservation Act of 1992, Pub. L. No. 102-307, § 214, 106 Stat. 267, 272.

163. Film Preservation Act of 1992, Pub. L. No. 102-307, § 203, § 214, 106 Stat. 267, 267, 272.

164. National Film Preservation Act of 2005, Pub. L. 109-009, 119 Stat. 224, 226;

\* \* \* \*

This historical progression reveals that the United States has increasingly promoted the societal interests in preservation and access to knowledge as part of its larger copyright system. Such policies play an important role in the overall balance of copyright protection and the pursuit of progress. We now turn to how these policies can best be effectuated in the digital age.

### III. PUBLIC VERSUS PRIVATE PRESERVATION OF AND ACCESS TO KNOWLEDGE IN THE DIGITAL AGE

Modern digital technology offers unprecedented opportunities to promote preservation and access to knowledge. As highlighted at the outset, Google seeks to offer anyone with an Internet connection the ability to conduct Boolean searches of much of the collective holdings of several of the world's leading libraries. It is certainly conceivable that much of recorded human knowledge could be at the fingertips of Internet users. Users might still have to "visit" an online bookstore, a pay-per-use or subscription archive, or a nearby library to view copyrighted works in their entirety. But, even accounting for these possibilities, users could more rapidly and confidently find the most relevant materials for their research. Yet, there is a delicate balance needed to ensure that universal search capability does not undermine incentives to produce and disseminate works of authorship. In order to assess public policy in this area, section A explores the rich mix of values implicated by the possibility of an Internet-accessible portal for conducting Boolean searches of much of recorded knowledge. Section B then compares a range of public and private institutions for enhancing accessibility and preservation of recorded knowledge.

#### A. *Exploring the Values of Preserving and Facilitating Access to Knowledge in the Digital Age*

Throughout much of U.S. history, the Smithsonian Institution, the Library of Congress, and university research libraries played the central role in preserving knowledge and making it available to the public. These institutions continue to play an important role in storing, preserving, and furnishing access to knowledge. They have also deployed digital technology to enhance accessibility of their collections, such as the

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1139

development of digitally searchable catalog systems.<sup>165</sup> But they now share the field with a host of new players. LexisNexis and Westlaw built successful subscription businesses around making a wide range of research material available through telecommunications networks and the Internet.<sup>166</sup> More recently, newspaper, periodical, and print publishers have made their content available on the Internet through various business models—advertising, subscription, and pay-per-view.<sup>167</sup> Carnegie Mellon University and other research institutions have begun digitizing and making available vast quantities of published works.<sup>168</sup> The Library of Congress, federal and state courts, and a host of other federal, state, and local governmental entities have uploaded considerable amounts of knowledge onto Internet-accessible servers. Nongovernmental organizations, such as the Internet Archive and Project Gutenberg, have instituted nonprofit initiatives to scan and digitize thousands of public domain and out-of-print books.<sup>169</sup> The latest entrants to this expanding bazaar of knowledge and knowledge service providers are search engine companies vying to expand their share of the Internet search marketplace, thereby increasing their ability to deliver relevant advertising. Yahoo!, Microsoft, and various other entities have formed the Open Content Alliance,<sup>170</sup> an ambitious project to provide public domain materials and copyrighted works for which publishers have consented. Google, the most popular search engine, has made the largest splash by its efforts to provide nearly universal searchability of recorded knowledge.

---

165. See James Gorma, *In Virtual Museums, an Archive of the World*, N.Y. TIMES, Jan. 12, 2003, at A1 (reporting that the Smithsonian National Museum of Natural History had catalogued nine million specimens online); Katie Hafner, *Books to Bytes: The Electronic Archive*, N.Y. TIMES, Apr. 8, 1999, at G1 (“Electronic materials are a growing part of library budgets.”); Irvin Molotsky, *Accessible via Modem, The Library of Congress*, N.Y. TIMES, Apr. 8, 1999, at G5 (reporting the digital cataloguing efforts of the Library of Congress).

166. Lynn Foster & Bruce Kennedy, *Technological Developments in Legal Research*, 2 J. APP. PRAC. & PROCESS 275, 281–82 (2000).

167. Katherine Q. Seelye, *Can Papers End the Free Ride?*, N.Y. TIMES, Mar. 14, 2005, at C1.

168. The Carnegie Mellon Million Book Project has scanned more than 1.4 million books. Carnegie Mellon University Libraries, FAQ about the Million Book Project, [http://www.library.cmu.edu/Libraries/MBP\\_FAQ.html](http://www.library.cmu.edu/Libraries/MBP_FAQ.html) (last visited Aug. 31, 2007). There are currently more than one hundred general digitization projects, in addition to a number of projects in specialized subjects. See GPO Access, Registry of U.S. Government Publication Digitization Projects, [http://www.gpo.gov/cgi-bin/dzr/dzr\\_browse.cgi?by=project\\_title](http://www.gpo.gov/cgi-bin/dzr/dzr_browse.cgi?by=project_title) (last visited Nov. 8, 2007) (listing U.S. government digitization projects).

169. Proskine, *supra* note 13, at 215–16.

170. See Posting of Brewster Kahle to Yahoo! Search Blog, [http://www.ysearchblog.com/archives/2005\\_10.html](http://www.ysearchblog.com/archives/2005_10.html) (Oct. 2, 2005) (announcing the Open Content Alliance).

What was once a sleepy pocket of the government and academic world has now attracted the attention of some of the most important information technology companies and captured the imagination of scholars at all levels—from middle school students writing school reports, to bloggers, to the most passionate archival researchers. One need no longer burn the midnight oil in dusty basements to study arcane historical questions. Given these dramatic changes, it is worthwhile to reexamine the principal values that underlay societal interests in preserving and promoting access to knowledge.

1. *Democratic Dimensions.* The nation's founders considered production of and access to knowledge essential elements of a democratic society. Such policies are prerequisites to the political awareness necessary in a participatory democracy.<sup>171</sup> As Professor Neil Netanel has explained, the structure of copyright law and its limiting doctrines play a critical role in promoting public education, political discourse, and equality.<sup>172</sup> Democratic values of freedom of expression and freedom of the press favor universal access to a comprehensive catalog of knowledge.

Democracy-enhancing spillovers are particularly important in the digital age. While the rise of the Internet has opened up communication channels to a much greater diversity of speakers, it has, at times, also produced a polarizing cacophony.<sup>173</sup> Expanding access to the most authoritative sources of information and enabling much more accurate and efficient search capability holds the potential to improve the quality of

---

171. See 1 ALEXIS DE TOCQUEVILLE, *DEMOCRACY IN AMERICA* 317–18 (Phillips Bradley ed., Henry Reeve trans., Alfred A. Knopf 1945) (1835) (“It cannot be doubted that in the United States the instruction of the people powerfully contributes to the support of the democratic republic . . .”).

172. Neil Weinstock Netanel, *Copyright and a Democratic Civil Society*, 106 *YALE L.J.* 283, 288–89 (1996); cf. YOCHAI BENKLER, *THE WEALTH OF NETWORKS: HOW SOCIAL PRODUCTION TRANSFORMS MARKETS AND FREEDOM* 273–300 (2006); Jack M. Balkin, *Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society*, 79 *N.Y.U. L. Rev.* 1, 33–38 (2004); Oren Bracha, *Standing Copyright Law on its Head? The Googlication of Everything and the Many Faces of Property*, 85 *TEX. L. REV.* 1799, 1842–55 (2007); Rosemary J. Coombe, *Objects of Property and Subjects of Politics: Intellectual Property Laws and Democratic Dialogue*, 69 *TEXAS L. REV.* 1853, 1877–80 (1991); Niva Elkin-Koren, *Cyberlaw and Social Change: A Democratic Approach to Copyright Law in Cyberspace*, 14 *CARDOZO ARTS & ENT. L.J.* 215, 224–32 (1996).

173. See CASS R. SUNSTEIN, *REPUBLIC.COM* 56–62 (2001) (observing that, “[i]n the face of dramatic recent increases in communication, there is an omnipresent risk of information overload”).

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1141

information available.<sup>174</sup> Better information has the power to sharpen and clarify discourse.

2. *Cultural Dimensions.* Societies have long recognized the cultural imperative of preserving and providing access to their history. The Smithsonian Institution, the Library of Congress, the National Archives, and numerous other museums, libraries, and monuments serve to celebrate, commemorate, and record the history and accomplishments of the American people. As Charles Jewett recognized, future historians will value the most complete and accessible record of knowledge.<sup>175</sup> Prior to the digital age, physical space, the costs of storage, and the impermanence of ink and paper (and later microfilm) limited the ability to preserve knowledge and make it publicly available. Since the late 19th century, librarians have struggled with how to balance the costs and benefits of knowledge retention. The digital age has greatly expanded the possibilities and reduced the cost of preserving the human record and making it available to limitless audiences.<sup>176</sup> Furthermore, the ability to effectively sift through the mountains of information adds value to the accretion of knowledge.

Digital technology has also expanded the boundaries of societies and cultures. The Internet enables peoples throughout the world to communicate and share knowledge nearly effortlessly. Whereas cultures in earlier eras tended to be insular and local, the Internet provides the means for greater connection among individuals and groups. The Internet transcends geographic and national boundaries. And the ability to access information serves a critical role in linking cultural networks.

The newly developed ability to preserve knowledge electronically has an important temporal dimension. Like endangered species, many forms of human knowledge are vulnerable to extinction. Therefore, societies run the risk of losing aspects of their cultural heritage by forestalling the process of digital archiving.<sup>177</sup> Of the 187,280 book titles

---

174. See CASS R. SUNSTEIN, *INFOTOPIA: HOW MANY MINDS PRODUCE KNOWLEDGE* 149–56 (2006) (discussing the popularity of the web-based encyclopedia Wikipedia as well as the quality of information available to its users).

175. See CHARLES C. JEWETT, *SMITHSONIAN REPORT ON THE CONSTRUCTION OF CATALOGUES OF LIBRARIES, AND THEIR PUBLICATION BY MEANS OF SEPARATE, STEREOTYPED TITLES* 8–9 (2d ed. 1853) (emphasizing “the benefits to be expected from a general printed catalogue of all books in the public libraries of America”).

176. See Diane Leenheer Zimmerman, *Can Our Culture Be Saved? The Future of Digital Archiving*, 91 *MINN. L. REV.* 989, 993–94 (2007) (hypothesizing an increase in access to culturally significant information as a result of mass digitalization).

177. Digital technology can also be vulnerable through obsolescence. As newer digital technologies develop, older platforms may cease to be used. This is particularly true of

published in the United States between 1927 and 1946, only about 2.2% were still in print in 2002.<sup>178</sup> Periodicals typically go out of print within days or weeks of publication. The Library of Congress has documented that fewer than 20% of U.S. feature films from the 1920s survive in complete form in American archives.<sup>179</sup> Only half of American feature films produced before 1950 survive.<sup>180</sup> Libraries have played the leading role in promoting preservation efforts.

Recognizing the importance of both preservation and accessibility, the European Commission has developed a framework (i2010) for creating a single European information space which promotes an open and competitive internal market for information and media services, strengthening investment in information and communication technology, and fostering inclusion, better public services, and quality of life through the use of information and communication technologies.<sup>181</sup> i2010 has funded the European Digital Library Project, among other initiatives.<sup>182</sup>

---

content storage technologies. See National Library of Australia, Technological Obsolescence, <http://www.nla.gov.au/padi/topics/13.html> (last visited Oct. 18, 2007) (discussing effects of changing technology). Some digital files from even a decade ago cannot be read by the latest systems, a problem sometimes referred to as lack of backward compatibility. See Joint Information Systems Committee, Digital Preservation Briefing Paper, [http://www.jisc.ac.uk/publications/publications/pub\\_digipreservationbp.aspx](http://www.jisc.ac.uk/publications/publications/pub_digipreservationbp.aspx) (last visited Oct. 18, 2007) (referring to disadvantages of “backwards compatibility”). In their desire to enhance performance and functionality (and displace older systems), technology companies develop new formats. As a result, information can become inaccessible not because the media deteriorates, but because of changes in the technology platform (and support for it). See *id.* (“The biggest risk to the accessibility of digital objects is the continued development of computing hardware and software.”). To avoid this threat to cultural preservation, public policy must be vigilant in ensuring that physical fragility of paper and ink is not replaced by technological change. Among the strategies that can be used to avoid this problem are promotion of backward compatibility and interoperability of systems, “migration of digital information to technologies from which they are accessible, the emulation of obsolete systems, and the preservation of obsolete technologies.” See *id.*; see also National Library of Australia, *supra*.

178. Deirdre K. Mulligan & Jason M. Schultz, *Neglecting the National Memory: How Copyright Term Extensions Compromise the Development of Digital Archives*, 4 J. APP. PRAC. & PROCESS 451, 472 (2002).

179. James H. Billington, *Preface* to ANNETTE MELVILLE & SCOTT SIMON, REDEFINING FILM PRESERVATION: A NATIONAL PLAN, RECOMMENDATIONS OF THE LIBRARIAN OF CONGRESS IN CONSULTATION WITH THE NATIONAL FILM PRESERVATION BOARD, at ix (Aug. 1994), available at <http://www.loc.gov/film/plan.html>.

180. Press Release, Nat'l Film Preservation Found., Treasures of American Film Archives Awarded \$500,000 in Arts Endowment Millennium Funds (Feb. 24, 1999), available at [http://www.filmpreservation.org/about/pr/pr\\_1999\\_02\\_24.html](http://www.filmpreservation.org/about/pr/pr_1999_02_24.html).

181. See Europe's Info. Soc'y, i2010—A European Information Society for Growth and Employment, [http://ec.europa.eu/information\\_society/europe/i2010/index\\_en.htm](http://ec.europa.eu/information_society/europe/i2010/index_en.htm) (last visited Sept. 10, 2007).

182. European Digital Library Project, <http://www.edlproject.eu> (last visited Aug. 24, 2007).

3. *Economic Dimensions.* The economic aspects of digital archiving and search technology are complex. They must be understood within the larger context of balancing incentives to create and disseminate creative works. The most direct economic effect of developing a comprehensive searchable archive of knowledge will be to increase the market for works that would otherwise be difficult to locate. As the Internet has proven with regard to many information goods—from books to music to movies—having a broad searchable catalog produces a long tail of demand that cumulatively can exceed the demand for even the most popular works. Whereas traditional retail stores stock only the more popular titles because of the costs of shelf space and restocking, Internet suppliers (such as Amazon.com and Netflix), with their relatively low marginal cost of warehouse storage and ability to provide online consumers with powerful search capabilities, can better satisfy diverse tastes.<sup>183</sup> Vendors of pure information goods can benefit even more significantly from this phenomenon because they have minimal marginal inventory and delivery costs.

Another important economic benefit of searchable digital archives comes from increased researcher productivity and quality of research. This derives from the ease with which online researchers can sift data precisely and efficiently, overcoming what Professor Frank Pasquale refers to as an “information overload” externality.<sup>184</sup> Traditional archival research that could take weeks or months of sifting through antiquated card catalogs, dusty records, and migraine-inducing microfiche reels can now be accomplished effortlessly and flawlessly in a fraction of the time. LexisNexis revealed this potential with its introduction of Boolean search technology for legal and news media.<sup>185</sup> Indexing technology greatly enhances social welfare by reducing the costs of finding the most valuable information. Search engines also offer algorithms

---

183. See CHRIS ANDERSON, *THE LONG TAIL: WHY THE FUTURE OF BUSINESS IS SELLING LESS OF MORE* 22–24 (2006) (discussing how the advantages of online retail, including increased product options and low costs, lead to increased margins); see also Erik Brynjolfsson, Yu Jeffrey Hu, & Michael D. Smith, *Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers*, 49 *MGMT. SCI.* 1580, 1580–82 (2003) (discussing advantage of online retailers, particularly Amazon.com and barnesandnoble.com, over “brick-and-mortar” retailers).

184. See Frank Pasquale, *Copyright in an Era of Information Overload: Toward the Privileging of Categorizers*, 60 *VAND. L. REV.* 135, 178 (2007) (explaining that the organization of material is necessary to combat the externality of “information overload”).

185. LexisNexis, *Introduction to Traditional Boolean Searching*, [http://support.lexisnexis.com/online/record.asp?ArticleID=GS\\_Boolean](http://support.lexisnexis.com/online/record.asp?ArticleID=GS_Boolean) (last visited Sept. 6, 2007) (explaining how to generate online LexisNexis searches).

that can use the collective behavior of millions of searchers to assist in pinpointing the most useful information.

A related benefit comes from the richness and rapidity of cumulative research—creative works that build upon prior works. By facilitating later generations of scholarship (and making it more accessible), digital search technology serves the greater goal of promoting innovation. Yet, this consideration cannot be taken too far. If pioneers (first generation researchers and their publishers) cannot derive an adequate return from their investments, there can be underinvestment in vital aspects of knowledge creation.<sup>186</sup> This is the conventional intellectual property trade-off, but it is important to recognize that lower costs of conducting research, controlling for all other factors, shifts the balance towards relatively lower protection for earlier and later creators. The works of both become potential inputs for later creators.

There are some other important economic dimensions to digital archiving and search technology. Google's ability to attract a large base of customers enabled it to revolutionize Internet marketing and advertising.<sup>187</sup> As it expands the domains of valuable searchable information, Google expands and strengthens its customer base, reinforcing the value of its advertising network. This advertising, in turn, drives investment in content expansion, much in the way that broadcast television is funded indirectly by product advertisers. The effect is particularly powerful on the Internet, a medium in which a single "network" can become dominant.<sup>188</sup> If Google has the largest share of customers, it will have the largest cachet among advertisers. As it better understands search behavior, Google is able to provide ever more accurate advertising placements, thereby increasing its value to advertisers.

Google and other search engines have made substantial headway in making the vast knowledge of the Internet accessible to the public at large. As Google has recognized, there are valuable first-mover advantages and natural monopoly dimensions to

---

186. See generally Peter S. Menell & Suzanne Scotchmer, *Intellectual Property Law*, in HANDBOOK OF LAW AND ECONOMICS 3 (A. Mitchell Polinsky & Steven Shavell eds., forthcoming 2007), available at <http://ssrn.com/abstract=741424> (providing a "comprehensive survey" of intellectual property law and economics).

187. Press Release, Google, Google Builds World's Largest Advertising and Search Monetization Program (Mar. 4, 2003), available at <http://www.google.com/press/pressrel/advertising.html> ("Google today announced its online advertising program . . . is the largest and fastest growing in the industry.").

188. See CARL SHAPIRO & HAL R. VARIAN, INFORMATION RULES: A STRATEGIC GUIDE TO THE NETWORK ECONOMY 173–75 (1999) (explaining that technology markets are often dominated by a single network).

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1145

opening up new search domains.<sup>189</sup> Scanning and digitizing books, and other tangible information goods, have relatively high fixed costs (principally labor) relative to the costs of delivering search results. By rapidly building an unparalleled book search service, Google can expand and reinforce its customer base, which translates into higher advertising revenue, higher user contact and loyalty, and new data for improving advertisement delivery. Potential entrants face a tremendous fixed entry cost if they wish to compete in the universal book search marketplace, with significant risk that they will not be able to divert loyal Google customers. From a larger societal standpoint, it is not at all clear that the benefits of competition would outweigh the costs of producing a second digital archive. Like the costs of laying electrical transmission lines, the socially optimal number of suppliers may be one.

As a result, basic Boolean book search technology will tend toward a single or just a few dominant suppliers. This does not mean, however, that Google will face no effective competition. Because of West Corporation's complementary assets in legal publishing, it was able to enter the legal search marketplace successfully with Westlaw after Lexis established a strong first-mover advantage. Similarly, book publishers and other content owners have distinctive advantages in offering fuller access to the materials being searched. Rather than merely providing information that a book is relevant to a user, the owner of the copyright in the book can provide full online access to the content through various business models. Content owners can offer premium products that Google cannot match. Westlaw, LexisNexis, JSTOR, HeinOnline, and other companies have developed successful subscription models for publishers of periodicals. Academic presses and other book publishers are beginning to develop enhanced services that go well beyond Google's source location capability.<sup>190</sup> Alternatively, publishers may wish to collaborate with Google through its Partner Program. Many publishers have partnered with Amazon.com, which has a large market share of the online book selling business, in offering greater access to the content of their books as a means of driving sales.

---

189. See *id.* at 168 (explaining that “[f]irst-mover advantages can be powerful . . . in information industries where scale economies are substantial”).

190. Nate Anderson, *Publishers Try to Stave Off Google, Amazon with Book Search*, ARS TECHNICA, Feb. 28, 2007, <http://arstechnica.com/news.ars/post/20070228-8949.html> (noting that Random House and HarperCollins launched web services to compete with the Google Book Search project).

A more complicated market structure question relates to competition in book search algorithms and complementary services. Unless competitors emerge in providing universal book search technology—such as currently exists in the marketplace for Internet search with Yahoo!, Microsoft, and Ask.com competing with Google—then Google will have a dominant position in book search innovation. The opportunity to expand its customer base (and resulting increased use of its advertising technology) will provide incentives for Google to advance this search technology, but the lack of direct competition, due to the high entry cost, could reduce the pressure to innovate.<sup>191</sup> Search innovators have the ability to license technology to Google, but the lack of a competitive marketplace for such technologies, as well as Google's ability to control the best laboratory for testing such technology, will lead towards a high concentration in this aspect of the industry.

The costs of archiving are another economic dimension of preserving and providing public access to copyrighted material. With the rise of digital publishing, the costs to preserve digital works are relatively low, and publishers now routinely prepare manuscripts in digital formats.<sup>192</sup> Similarly, film studios and record labels can now more easily store digital versions of their products.<sup>193</sup>

---

191. Monopolists have an incentive to maintain their monopoly power by developing new technologies before potential competitors can gain a foothold in the market. See Richard J. Gilbert & David M. G. Newbery, *Preemptive Patenting and the Persistence of Monopoly*, 72 AM. ECON. REV. 514, 514 (1982) (discussing monopolists' incentive to develop "sleeping patents" in order to prevent competition). In general, two factors operate: an efficiency effect and a replacement effect. See JEAN TIROLE, *THE THEORY OF INDUSTRIAL ORGANIZATION* (1988). Because monopoly profits will be higher in any given market than oligopoly profits, a monopolist's incentive to remain a monopolist (and hence pursue research and development intended to maintain such position) tends to be greater than a potential entrant's incentive to become a duopolist (the efficiency effect). On the other hand, a monopolist has less to gain from innovation since it is already earning high profits. Innovation by the monopolist may only displace all or part of the existing monopoly profits (the replacement effect). Hence, monopolists may not have as much to gain as potential entrants from innovating. In the case of drastic innovation, the replacement effect dominates because the entrant can become a monopolist. See Jennifer F. Reinganum, *Uncertain Innovation and the Persistence of Monopoly*, 73 AM. ECON. REV. 741, 743 (1983) ("It is found that, for drastic innovations, the incumbent always invests less than the challenger, so that the incumbency changes hands *more often than not*."). In the case of nondrastic innovation, the efficiency effect tends to dominate and hence monopoly power tends to persist. Cf. DREW FUDENBERG & JEAN TIROLE, *Dynamic Models of Oligopoly*, in *FUNDAMENTALS OF PURE AND APPLIED ECONOMICS* 1, 32 (A. Jacquemin ed., 1986) (arguing that in cases of drastic innovation, the "efficiency effect" is zero).

192. See Zimmerman, *supra* note 176, at 1005–08 (correlating the fall of digitalization costs with an increase in the amount of print works available today in digital form, including those published by Reid-Elsevier).

193. See, e.g., Press Release, Digital Cinema Implementation Partners, LLC, AMC Entertainment Inc., Cinemark USA and Regal Entertainment Group Form Company to Launch Theatre Conversion to Digital Cinema (Mar. 6, 2007), available at [http://www.ncm.com/CMS/articlefiles/690-AMC\\_Entertainment\\_Inc.\\_Cinemark\\_USA\\_and\\_Regal\\_Entertainment\\_](http://www.ncm.com/CMS/articlefiles/690-AMC_Entertainment_Inc._Cinemark_USA_and_Regal_Entertainment_)

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1147

Nonetheless, the real challenge of the next few decades lies in converting the large stock of knowledge stored in analog form to high quality digital formats. As a result of steep declines in preservation costs caused by advances in the technology for digitally encoding analog works, libraries working in conjunction with search providers are better situated than copyright owners to digitize analog works.

Digital archiving and search technology also promise significant long-term benefits and cost savings to libraries and archives. As libraries move more of their collections to digital form, patrons will be able to use those collections more efficiently. In addition, they will be able to store public domain works in digital archives, resulting in substantial savings in storage costs. Partnering with search companies will also defray some of the costs of preserving their collections. To the extent that libraries wish to preserve tangible versions, they will be able to relocate those materials to more secure and cheaper storage areas.

A final economic dimension relates to transaction costs. Affording copyright owners the right to block digital archiving, indexing, and Boolean search services dramatically raises the costs of preservation and access.<sup>194</sup> The challenge is greatest in the vast area of orphan works<sup>195</sup>—out-of-print books for which copyright owners are difficult to find. Yet, much of the “long tail”<sup>196</sup> of value, particularly for scholars, lies in this segment of the knowledge universe.

---

Group\_Form\_Company\_To\_Launch\_Theatre\_Conversion\_To.pdf (discussing the major film studios’ current “theatrical transition” from 35mm film to digital technology).

194. See generally Denise Troll Covey, Copyright and the Universal Digital Library (2005) (unpublished conference paper), available at [http://www.library.cmu.edu/People/troll/ICDUL\\_TrollCovey\\_FINALtmpREV.pdf](http://www.library.cmu.edu/People/troll/ICDUL_TrollCovey_FINALtmpREV.pdf) (2005) (comparing three Carnegie Mellon studies of reducing the cost of gaining permission to digitize works). As of 2005, the Million Book Project had permission to digitize 52,900 titles, at a projected cost of \$0.69 per title. *Id.* at 6. But, for many of these titles, blanket permission was given by publishers to publish all out-of-print titles or all titles published before a certain year. *Id.* Two other studies conducted by Carnegie Mellon, both seeking permissions for specific titles, had much higher transaction rates: \$200 and \$78 per title. *Id.* at 3–4.

195. The Register of Copyrights defines “orphan works” as those where the owner of the copyrighted work cannot be identified and located. U.S. COPYRIGHT OFFICE, REPORT ON ORPHAN WORKS 1 (2006), available at <http://www.copyright.gov/orphan/orphan-report-full.pdf>; Olive Huang, *U.S. Copyright Office Orphan Works Inquiry: Finding Homes for the Orphans*, 21 BERKELEY TECH. L. J. 265, 265 (2006). Fear of liability prompts many users of orphan works to simply avoid such works. For example, despite two years of research on copyright ownership, the Library of Congress decided not to post thousands of Hannah Arendt’s papers because there were questions surrounding ownership. See COMMENTS OF THE LIBRARY OF CONGRESS IN RESPONSE TO THE COPYRIGHT OFFICE NOTICE OF INQUIRY “ORPHAN WORKS” 6 (2005), available at <http://www.copyright.gov/orphan/comments/OW0630-LOC.pdf> (describing the project and its results).

196. See ANDERSON, *supra* note 183, at 10 (describing the birth of the term “the Long Tail”).

*B. Comparing Public and Private Approaches to Knowledge Preservation and Access*

Digital technology has brought about a new era in archiving and searching tangible works. The main questions revolve around the transition path to a widely accessible archive of recorded human knowledge that is both searchable and comprehensive—how will it be accomplished, how much will it cost and who will bear this burden, and how long will the process take? Companies like Lexis and Westlaw have led the way, in conjunction with and followed by the publishers of many leading periodicals. Libraries and governmental enterprises have begun the transition.<sup>197</sup> Several nongovernmental organizations have undertaken a range of initiatives—from Carnegie Mellon University’s Million Book Project (which is nearly at the 1.5 million book mark) to more modest approaches.<sup>198</sup> Publishers have also begun experimenting with this. And, significantly, the leading search engine companies have laid out the most ambitious plans. This section explores a spectrum of possible transition paths as a means of analyzing the advantages and disadvantages of different institutional options.

*Library of Congress—Public Funding and Management:* Going back to the first library in Alexandria, government enterprises have played the leading role in preserving knowledge. It is certainly conceivable that the Library of Congress would take on this task. It could develop the internal capacity to digitize its collection or outsource the work.

*Lending Libraries—Hand Indexing:* Public lending libraries consider it part of their core mission to provide the public with access to their collections. As a way of framing the copyright questions raised by the Google Book Search Project, consider as a thought-experiment the implications of libraries hand indexing each word in each volume on their shelves. Such a plan would, obviously, be extraordinarily expensive and prohibitively time-consuming. But would such activity be considered a violation of copyright law? It seems likely that this activity would pass muster under conventional fair use jurisprudence. It could be seen as a more versatile card catalog, allowing users to conduct Boolean searches so as to identify relevant books within the libraries collection.

---

197. See Carnegie Mellon University Libraries, *supra* note 168. (listing participants in library program); The Library of Congress, What the Library is Doing, <http://www.digitalpreservation.gov/library> (last visited Sept. 15, 2007) (discussing role of Congress in digital preservation).

198. See Carnegie Mellon University Libraries, *supra* note 168 (“To date the Project has scanned over 1.4 million books in China, India[,] and Egypt . . .”).

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1149

*Publishers (copyright proprietors):* As holders of copyrights, publishers and other copyright owners could provide enhanced search tools for their works. They could also join forces and develop their own search portals.

*Search Companies with Publisher Permission:* Search companies working in conjunction with publishers have taken significant initiatives aimed at expanding Internet users' access to knowledge. Project Gutenberg, the Million Book Project, and the Open Content Alliance fit this model, as does the Google Publisher Partner program.

*Google Book Search:* Google's initiative to scan and digitize collections of major libraries of the world without obtaining consent of copyright owners represents the most aggressive effort to produce a nearly universal book search capability. Its opt-out policy limits the comprehensiveness of the collection, although it seems unlikely that many publishers will exempt themselves from this project.<sup>199</sup>

Chart I compares these different institutional options along the principal social, legal, and economic performance criteria: the degree of comprehensiveness of the archive; legal uncertainty; the risk that this approach poses to copyright owners through enhanced exposure to unauthorized access; the extent to which each institutional approach would encourage innovation and ease of use; the likely effects of competition; the range of transaction costs; the extent to which each approach will result in rent-seeking behavior by interested parties; the time that each approach will take to achieve a comprehensive, searchable archive; and the economic feasibility of each approach.

---

199. Despite their vociferous dislike of the current Google Book Search Program and their desire to obtain licensing fees for the use of their materials, publishers would not want the negative economic consequences of being left out of a project that could result in additional sales and publicity. This principle is illustrated in the paradoxical case of freelance authors who wanted licensing fees for the use of their pieces in online databases. In *New York Times, Inc. v. Tasini*, 533 U.S. 483, 488 (2001), freelance authors appeared to win a victory when the Supreme Court held that making articles available in online databases did not constitute an allowable revision under 201(c). The authors could now enjoin the New York Times and others from republishing their pieces in online archives and on CD-ROMs. *Id.* at 492–93. The supposed victory turned hollow when, instead of paying additional fees to license the work, the organizations simply removed most of the freelance pieces from the online databases. See *Tasini v. N.Y. Times, Inc.*, 184 F. Supp. 2d 350, 352 (S.D.N.Y. 2002) (discussing action taken by the New York Times after the original ruling). Jonathan Tasini filed a new suit complaining that the New York Times should not be allowed to remove the articles under various theories (unconscionability, breach of the implied covenant of good faith), which the court dismissed for lack of standing. See *id.* at 353–55; see also Amy Terry, *Tasini Aftermath: The Consequences of the Freelancers' Victory*, 14 DEPAUL-LCA J. ART & ENT. L. 231, 232 (2004) (addressing the background of the *Tasini* case, its holding, and its consequences).

**Chart I**  
Institutions for Preserving and Providing Access to Knowledge in the Digital Age

Institution	Extent of Archive	Legal/Copyright Uncertainty	Piracy Risk	Ease of Use/Innovation	Competitive Effects	Transaction Costs	Public Cost	Political Costs	Time	Feasibility
<b>(1) Library of Congress: Public Funding and Management</b>	potentially universal	possible takings issue	modest	depends on procurement policies, likely to be stunted	depends	none	high	high	likely to be long	?
<b>(2) Lending Libraries: Hand Indexing</b>	variable - depends on library collection	possible takings issue; modest (c) risk	none	little	modest	none	high	low	glacial	no
<b>(3) Publishers (copyright proprietors)</b>	limited; typically only publisher catalog	none	modest	some-depth rather than breadth	significant risk of bottleneck	none	none	zero	gradual	yes
<b>(4) Search Companies with Publisher Permission</b>	no orphan works or in-(c) works lacking permission	none	modest	significant, but constrained by limited archive	some risk of bottleneck	significant	none	zero	modest	yes
<b>(5) Google Book Search</b>	nearly universal (less opt-out)	significant cloud of fair use uncertainty	modest	robust, complimentary assets	concentration (due to network effects), but potential spur to competition	none	none	low to medium	rapid	yes

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1151

The public provision approach has the potential to achieve a universal archive, but there are significant administrative, political, and cost drawbacks. Google's projected expense (as much as \$800 million)<sup>200</sup> exceeds the Library of Congress's annual budget.<sup>201</sup> And it seems unlikely that Google, or any other vendor, would be willing to provide these services at cost. Google has developed substantial proprietary technology for which it would want to be compensated. Adding this project to the Library of Congress's many other activities would require a substantial increase in its budget and would likely provoke strong opposition by the publishers. There could also be legal difficulties. Extrapolating from the legal battle with Google, authors and publishers might well assert that the scanning and digitizing of these images by the Library of Congress infringe their copyrights or constitute a taking without just compensation.<sup>202</sup> Although creating a truly public searchable archive would have great benefits, it seems unlikely that the government would have the administrative capability to manage this project effectively, and the price tag would represent a significant political battle. Both copyright owners and search engines would likely present strong opposition.

The library option, although hopelessly infeasible, makes the point that affording Boolean search capability to library patrons comports with norms about the proper role of libraries.

Relying on publishers to provide digital search technology would greatly diminish the reach of search technology. Publishers would be primarily interested in assisting the public in finding their own books, and they lack the technological experience to tackle this kind of project. Collaboration among publishers could well raise anticompetitive concerns.<sup>203</sup> Based on

---

200. See Toobin, *supra* note 3.

201. The 2006 congressional appropriation for the Library of Congress was \$609.7 million, of which \$396 million was discretionary. See Legislative Branch Appropriations Act, 2006, Pub. L. No. 109-55, 119 Stat. 565, 579-82 (2005). The rest was earmarked for special programs, including \$101 million for congressional research, \$59 million for the copyright office, and \$55 million for books for the handicapped. *Id.* For the 2008 fiscal year, the library is requesting a total budget of \$703.3 million. See James H. Billington, The Librarian of Congress, Statement Before the Subcomm. On the Legislative Branch of the H. Comm. On Appropriations, 110th Cong. (Mar. 22, 2007), available at <http://www.loc.gov/about/welcom/speeches/budget/request2008.html>.

202. The public, noncommercial nature of the project would weigh in the government's favor under a fair use analysis.

203. The U.S. Department of Justice has accused content companies of antitrust violations with regard to the creation of digital services. See Makan Delrahim, Dep. Asst. Atty. Gen., Antitrust Div., U.S. Dept. of Justice, Remarks to the Recording Artists' Coalition 5 (Dec. 18, 2003), available at <http://www.usdoj.gov/atr/public/speeches/204474.pdf> (listing recent antitrust settlements the Department achieved).

the record labels' failure to develop viable music distribution business models on their own,<sup>204</sup> it seems unlikely that book publishers will fare much better.<sup>205</sup>

Internet search engines are in a much better position to develop successful searchable archives and tools. The Open Content Alliance has made significant strides in developing the technology and working with publishers and libraries to build searchable archives.<sup>206</sup> The major problem with this approach is that it will leave vast portions of published works outside of the reach of Internet searchers because of the costs and impediment of copyright clearance. The problem is two-fold. Some copyright proprietors may refuse consent. More significantly, it is very difficult to clear rights in the many orphaned works. As a result, this approach will undoubtedly leave much knowledge outside of searchable archives.<sup>207</sup>

Google's more aggressive approach would produce a nearly universal archive of human knowledge within a relatively short time period and at zero public cost. There is little reason to believe that the project will produce a substantial risk of piracy. This is not to say that there is no risk. But much of Google's effort is devoted to indexing the long tail of works having relatively little commercial value (although their cumulative research and historical value is great). There is no doubt that popular works, such as the latest installment of the Harry Potter series, will be pirated—both digitally and in hard copy form. Pirates, however, need not go through the trouble of decrypting Google's technological protection measures to distribute the work illegally. They need only purchase a copy on the day it is released (or borrow one from a public library) and scan it into a computer system. The real value of Google's archive lies in its breadth. Anyone trying to profit from liberating an archive of that

---

204. In the *Napster* litigation, Judge Patel observed that the record labels' joint ventures in the digital distribution area "look bad, sound bad[,] and smell bad." *In re Napster, Inc. Copyright Litig.*, 191 F. Supp. 2d 1087, 1109 (N.D. Cal. 2002).

205. In addition, many works that would benefit from a comprehensive database, such as orphaned works, would fall through the cracks. Some publishers have not maintained comprehensive records of their works, and as rights have been transferred over the years or the works have gone out of print, the publishers themselves are not always aware of who holds the rights to a work. See Huang, *supra* note 195, at 265 (explaining orphan works' uncertain provenance creates liability concerns, prompting users to avoid them); see also Covey, *supra* note 194, at 3–4 (reporting that most publishers provided digitizers blanket permission to include out-of-print works).

206. Posting of Brewster Kahle to Yahoo! Search Blog, *supra* note 170 (describing major Internet providers' project to assemble both public domain and publisher-consented copyrighted material).

207. See Huang, *supra* note 195, at 276–77 (predicting orphan works will be avoided).

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1153

magnitude will undoubtedly show up on various enforcement radar screens. The greatest risk may well be from hackers who see breaching the Google archive as a technological achievement. Even under those circumstances, it seems unlikely that such a break-in can cause significant harm without attracting great attention or using substantial traceable server capacity. The reality of public libraries suggests that the digital piracy threat from Google's Book Search Project will be limited by the fact that just about any published book can be accessed for free from the many lending libraries throughout the United States. Google's advance in search technology and breadth does not change that reality.

Google's Book Search Project will undoubtedly stimulate tremendous innovation in both search technology and public access. It can be seen as a great donation to the heritage of humankind, an unprecedented resource for promoting progress, and a savvy business model for reinforcing its leadership role in the digital marketplace. The major impediment to this project moving smoothly forward appears to be the looming lawsuits. The outcome of that litigation is far from clear.<sup>208</sup>

#### IV. CALIBRATING COPYRIGHT LAW TO PROMOTE PRESERVATION, ACCESS, AND INCENTIVES IN THE DIGITAL AGE

The ultimate assessment of Google's ambitious Book Search Project should not be focused on a narrow analysis of the fair use factors. Rather, it should ask how society can best promote progress in the advancement of knowledge—through its production and access—in a digital age that brings universal search capability within the reach of everyone with a computer and an Internet connection. Another way to conceptualize this broader question is to imagine how the archival visionaries from the 19th century would have reacted to recent technological breakthroughs in information storage and access. What would Charles Jewett, first librarian of the Smithsonian Institution, or Ainsworth R. Spofford, the influential Librarian of Congress, think of the Google Book Print Project? It seems clear that they would have embraced this technological marvel and commended it to the nation. Recall that Jewett extolled the virtues of a

---

208. Compare Posting of James Boyle to Duke Law & Technology Review iBlawg, *supra* note 12 (opining that Google will lose at the trial level but prevail on its appeal in the Second Circuit), with Posting of Richard Epstein to Duke Law & Technology Review iBlawg, <http://www2.law.duke.edu/journals/dltr/iblawg/?p=23> (Feb. 22, 2006) (predicting Google will lose the case on both trial and appellate levels due in part to favorable venue in New York—home to the major publishers).

comprehensive “central depot” where knowledge would be preserved for reference and for future generations.<sup>209</sup> He emphasized the importance of a complete collection, “without a single omission.”<sup>210</sup> Spofford approached his responsibilities as a 19th century Ptolemy—seeking to build the universal knowledge resource.<sup>211</sup> It is in this spirit that Google’s project should be considered.

This Part approaches this inquiry from a fresh perspective. It explores how Congress should ideally craft copyright law amendments to promote the multi-faceted goals of the copyright system. It begins by framing the policy analysis for developing a comprehensive, searchable knowledge archive, drawing on the examples of the Human Genome Project and the Family Movie Act. It then analyzes the institutional and temporal issues involved in pursuing this objective.

A. *Framing the Policy Analysis: The Choice of Governance Institutions and Surgical Safe Harbors*

The opportunities and challenges of digital archiving and search technology bring copyright law into new areas of policy analysis. Two quite distinct policy areas shed valuable light on these issues. Following great advances in molecular biology in the late 1970s and early 1980s, the U.S. government, as well as other the governments and scientific establishments of other leading research nations, faced the question of who should have principal responsibility for building a valuable new database—a full map of the human genome.<sup>212</sup> This resource promised to be of immeasurable value to scientific research, but costly to assemble. Ultimately, the choice was made to create a public archive.<sup>213</sup> Although this project was not identical to the knowledge archive that Google is building, there are valuable lessons to be learned from this example. A second set of lessons can be derived from the experience leading up to the Family Movie Act of 2005. Protracted copyright litigation threatened to stifle, or at least

---

209. See SMITHSONIAN INSTITUTION, *supra* note 59, at 35.

210. *Id.*

211. See JOHN Y. COLE, JEFFERSON’S LEGACY: A BRIEF HISTORY OF THE LIBRARY OF CONGRESS 18 (1997) (remembering Spofford’s belief that, because no work existed which would not prove of use to the nation at some point, all should be cataloged).

212. See Roger Lewin, *Proposal to Sequence Human Genome Stirs Debate*, 232 SCI. 1598, 1599–1600 (1986) (commenting on the debate stirred by the proposal to sequence the human genome).

213. See Francis Collins et al., *New Goals for the U.S. Human Genome Project: 1998–2003*, 282 SCI. 682, 685 (1998) (relating the historical path of the project and future goals and timelines).

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1155

delay, the roll-out of technology that promised undeniable benefits to portions of the public without any significant risk to copyright owners. Due to a confluence of political constituencies, Congress was able to obviate the litigation by passing a carefully tailored legislative solution.<sup>214</sup>

1. *Lessons from the Human Genome Project.* Just as the printing press liberated monks, scribes, and medieval scholars from the laborious toil of hand copying manuscripts and hastened the age of enlightenment,<sup>215</sup> the ability to conduct Boolean searches of the immense holdings of the world's greatest libraries from any Internet-connected computer represents a dramatic advance in intellectual endeavor. Such technology will revolutionize everything from middle school research projects to the most arcane archival research. It also promises to advance scholarship, enrich political discourse, expand demand for obscure works of authorship, and spur search technology science. Public policy governing search technology, however, must address several complicating factors. First, such technology must not lead to piracy of copyrighted works—otherwise, copyright's engine for creativity will be undermined. Second, given the economics of book scanning, search commerce, and the Internet, public policy should pay attention to both competition and incentives to innovate.

The effort to map the human genome offers a useful point of comparison. By the early 1980s, molecular biologists came to realize that it would be possible and enormously valuable to develop a comprehensive map of the entire human genome<sup>216</sup>—several billion DNA base pairs coding several tens of thousands of genes. The map could serve as a foundational building block for all manner of biological research. Maintaining it as a public resource would spur downstream research. The cost of assembling this information resource, however, was a major obstacle.<sup>217</sup> There was also the question of encouragement of

---

214. Paul Farhl, *Now on DVD: The Sanitizer's Cut*, WASH. POST, Apr. 18, 2005, at A1 (describing movie studios ultimate capitulation to the technology in exchange for legislative quashing of stand-alone versions of permanently altered and sanitized films).

215. See EISENSTEIN, *supra* note 27, at 7 (listing various eras transformed by the printing press).

216. See John Abelson, *A Revolution in Biology*, 209 SCI. 1319, 1319 (1980).

217. See Lewin, *supra* note 212, at 1598 (explaining that "Big Science" initiatives were new to biology, and many biologists feared that the human genome project would leave little money available for continued funding of other research); Roger Lewin, *Shifting Sentiments over Sequencing the Human Genome*, 233 SCI. 620, 620 (1986) [hereinafter Lewin, *Shifting Sentiments*] (explaining that even with anticipated technological advances to reduce time and effort, fears remained that funds would be

innovation to accelerate the mapping process.<sup>218</sup> Providing intellectual property rewards for mapping the genome (through patents) could jeopardize rapid cumulative innovation through monopolization of basic knowledge and transaction costs of licensing use of protected information.

By the late 1980s, the U.S. government, as well as government research enterprises in China, France, Germany, and the United Kingdom, formally embarked on a collaborative, publicly-funded effort to map the genome and place it within the public domain for all researchers.<sup>219</sup> At the same time, some private entrepreneurs entered the race in the hope of obtaining valuable patents.<sup>220</sup> By the year 2000, President Bill Clinton and British Prime Minister Tony Blair announced that a “rough draft” of the genome had been achieved.<sup>221</sup> The genome was largely completed by April 2003.<sup>222</sup> Although the public project was criticized as costing much more than necessary,<sup>223</sup> this essential repository of knowledge is now in the public domain and being used widely by scientists throughout the world to conduct biological and medical research.

---

diverted from other research).

218. See Leslie Roberts, *New Sequencers to Take On the Genome*, 238 SCI. 271, 271 (1987) (“Not surprisingly, the proposal [to sequence the human genome] has given considerable impetus to the development of new automated DNA sequencing technologies to shave time and money off the task.”); Lewin, *Shifting Sentiments*, *supra* note 217, at 620 (summarizing consensus at the time as being against sequencing the human genome because, according to Leroy Hood of the California Institute of Technology, “it would be a serious mistake to jump into a full-scale sequencing effort with the cottage-industry techniques we have at the moment”).

219. See Major Events in the U.S. Human Genome Project and Related Projects, [http://www.ornl.gov/sci/techresources/Human\\_Genome/project/timeline.shtml](http://www.ornl.gov/sci/techresources/Human_Genome/project/timeline.shtml) (last visited Sept. 17, 2007) (presenting the Department of Energy’s log of major events related to the Genome project); see also Nicholas Wade, *Once Again, Scientists Say Human Genome Is Complete*, N.Y. TIMES, Apr. 15, 2003, at F1 (noting the involvement of the United States, United Kingdom, China, France, and Germany).

220. See Eliot Marshall, *A High-Stakes Gamble on Genome Sequencing*, 284 SCI. 1906, 1909 (1999) (noting the private pharmaceutical industry’s effort to patent the discovery of human genes); Eliot Marshall, *Companies Rush to Patent DNA*, 275 SCI. 780, 780 (1997) (same). One researcher, Walter Gilbert of Harvard University, even had plans to copyright the human genome. Leslie Roberts, *Who Owns the Human Genome?*, 237 SCI. 358, 358 (1987).

221. Press Release, White House, President Clinton: Announcing the Completion of the First Survey of the Entire Human Genome (June 26, 2000), available at <http://clinton3.nara.gov/WH/Work/062600.html>.

222. Nicholas Wade, *Week in Review, All of You*, N.Y. TIMES, Apr. 20, 2003, at WK2.

223. In 1998, Celera Genomics Corporation announced that it would sequence the human genome faster than the public effort for approximately \$200 million, about one-fifteenth of the cost of the publicly funded project. See Nicholas Wade, *New Company Joins Race to Sequence Human Genome*, N.Y. TIMES, Aug. 18, 1998, at F6.

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1157

Parallels can be drawn to the goal of creating a comprehensive, searchable archive of the contents of major libraries. The up-front costs will be relatively high and there are benefits—such as preserving disintegrating works and facilitating research—to accomplishing the project more rapidly. Private enterprise might well be able to assemble the archive more cost-effectively and expeditiously. But, as with the mapping of the human genome,<sup>224</sup> there may be long term disadvantages in terms of competition and search innovation from private ownership. Having the knowledge archive in the hands of one private enterprise might result in undue market power, as well as effective control over search-related innovation.

There are also significant differences between the Human Genome Project and the development of a comprehensive, searchable archive of books and other textual materials. Private enterprises that scan books will have significantly fewer intellectual property claims than might have developed if patents had been recognized for genome sequence data.<sup>225</sup> At best, Google or others will have the ability to protect their scans and Optical Character Recognition (OCR) files through contract and trade secrecy protection. Therefore, others will be free to do their own scans without running afoul of Google's rights.<sup>226</sup> Furthermore, whereas the Human Genome Project had a specific completion objective (a map of all base pairs), a comprehensive archive would be an ongoing project which would need to incorporate the

---

224. Cf. Michael A. Heller & Rebecca S. Eisenberg, *Can Patents Deter Innovation? The Anticommons in Biomedical Research*, 280 *SCI.* 698, 698 (1998) (anticipating research patents on DNA fragments that exclude use by other researchers as destructive to innovation, leaving no one with the right to use the whole or maximize the benefits of the knowledge).

225. The issue of intellectual property rights relating to sequencing data became a deeply controversial issue. The first head of the U.S. effort, James D. Watson, corecipient of the Nobel Prize for Medicine in 1962 for his path-breaking work deciphering the structure of the DNA molecule, was forced to resign from his post in 1992 as a result of his opposition to gene patenting. Carol Strickland, *Watson Relinquishes Major Role at Lab*, *N.Y. TIMES*, Mar. 21, 1993, § 13, at 1. The Patent Office ultimately developed guidelines precluding patent protection for sequence data unless specific, substantial, and credible utility was established, which the Court of Appeals for the Federal Circuit essentially sustained. See *In re Fisher*, 421 F.3d 1365, 1371 (Fed. Cir. 2005) (positing examples of “specific” or “substantial” utilities); U.S. Patent and Trademark Office, *Utility Examination Guidelines*, 66 *Fed. Reg.* 1092, 1092–93 (Jan. 5, 2001); see also Tashica Williams, *In re Fisher: Raising the Utility Hurdle for Express Sequence Tags*, 21 *BERKELEY TECH. L.J.* 123, 123 (2006) (discussing the utility standard and framing it as a timeline “tracking the invention’s ripeness”).

226. See ROBERT P. MERGES, PETER S. MENELL, & MARK A. LEMLEY, *INTELLECTUAL PROPERTY IN THE NEW TECHNOLOGICAL AGE* 35–37 (4th ed. 2006) (summarizing trade secret law and noting trade secrets are protected only against misappropriation, not independent discovery).

flow of written knowledge. There are also significant differences in degree between the two projects. For example, the benefits to biomedical research of unhindered access to the full genome database may well be greater than the benefits to search innovation of full access to the comprehensive knowledge database.

2. *Lessons from the Family Movie Act.* The copyright system has various doctrines, such as fair use and liability standards, that provide flexibility to adapt to technological change.<sup>227</sup> Yet, technological change can place undue stress on such doctrines and on the system as a whole, thus necessitating legislative adjustments. The digital revolution has produced an unprecedented torrent of legislative amendments—from the Audio Home Recording Act of 1992 to the Digital Performance Rights in Sound Recordings Act of 1995 and the Digital Millennium Copyright Act (DMCA) of 1998. Some of these amendments, such as the online service provider safe harbor provisions of the DMCA, have proven to be of great importance, helping to stimulate innovation and new services on the Internet.<sup>228</sup> Other adjustments have proven to be unnecessary (as in the case of the AHRA) or unwieldy (as in the case of the anticircumvention provisions of the DMCA). Nonetheless, digital technology continues to pose new challenges and opportunities.

In 2002, digital filtering technology posed an unprecedented challenge for copyright law. ClearPlay sought to offer consumers the opportunity to watch movies rated R and PG-13 without being exposed to nudity, violence, or profanity.<sup>229</sup> It developed digital scripts that would run in conjunction with legally procured DVDs to filter out objectionable content by skipping scenes or muting sound while the DVD played.<sup>230</sup> After motion picture studios and directors objected to such editing and threatened to sue, ClearPlay and other filtering companies brought a declaratory relief action seeking to establish the

---

227. See PAUL GOLDSTEIN, *COPYRIGHT'S HIGHWAY: FROM GUTENBERG TO THE CELESTIAL JUKEBOX* 15–16 (rev. ed. 2003) (citing fair use flexibility from sources as disparate as Howard Hughes's biography to 2 Live Crew's sampling); Peter S. Menell, *Envisioning Copyright Law's Digital Future*, 46 N.Y.L. SCH. L. REV. 63, 64 (2003) (“It would be a mistake . . . to view copyright as a static body of law.”).

228. By failing to anticipate the advent of peer-to-peer technology, the DMCA was unable to deal with the range of challenges. Nonetheless, the online service provider safe harbors provided reasonably clear and effective ground rules to encourage investment in Internet service companies and strong competition.

229. Alison R. Watkins, *Surgical Safe Harbors: The Family Movie Act and the Future of Fair Use Legislation*, 21 BERKELEY TECH. L.J. 241, 244 (2006).

230. *Id.*

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1159

legitimacy of their technology and business model.<sup>231</sup> The case centered on whether a digital script that instructed the playback device to skip scenes or mute the sound constituted a derivative work, and, if so, whether such a script was a fair use.<sup>232</sup> Given the uncertainties in both areas of copyright law,<sup>233</sup> the case could easily have dragged out through several court levels of litigation. Furthermore, courts in other circuits might well have gone in other directions. After several more years this issue might have worked its way up to the U.S. Supreme Court. From a larger public policy standpoint, however, the claims of the studios and directors were dubious. The studios themselves sanitized films for showing on airplanes. And there could be no doubt that purchasers or renters of DVDs would be fully entitled to turn down the volume, fast forward, or shut their eyes at designated intervals without being directly liable for copyright infringement. It seemed doubtful on policy grounds that companies that automated those actions should be barred from the market absent consent from copyright and trademark owners (as well as directors). Yet, there was certainly enough in the Copyright Act and case law to fight this out for several years.

Fortunately, Congress saw that copyright law should not stand in the way of protecting the rights of parents “to shield their children from violence, sex, and profanity” in a manner that was not likely to undermine the economic interests of copyright and trademark owners.<sup>234</sup> As with the Google Book Search Project, there was good reason to believe that the availability of digital filtering technology would expand the marketplace for

---

231. See *Huntsman v. Soderbergh*, No. 02-CV-1662, 2005 WL 1993421, at \*1 (D. Colo. Aug. 17, 2005) (noting the claim for a declaratory judgment but dismissing the case because the newly passed Family Movie Act rendered the issues moot).

232. See *Watkins*, *supra* note 229, at 246–48.

233. On the derivative work issue, compare *Lewis Galoob Toys, Inc. v. Nintendo of America, Inc.*, 964 F.2d 965, 968 (9th Cir. 1992) (enunciating a fixation limitation on the derivative work right), with *Micro Star v. FormGen, Inc.*, 154 F.3d 1107, 1110–11 (9th Cir. 1998) (interpreting the fixation limitation narrowly), and *Midway Manufacturing Co. v. Artic International, Inc.*, 704 F.2d 1009, 1014 (7th Cir. 1983) (finding game enhancement to be infringing derivative works even though it did not copy the underlying work). Note that the U.S. Copyright Office has rejected *Micro Star v. FormGen* as part of “the Ninth Circuit’s more lenient test for infringement of derivative works, which seemingly ignores the originality requirement, [and] appears to be in error as it runs contrary to all other Circuit Court precedent.” Mech. & Digital Phonorecord Delivery Rate Adjustment Proceeding, No. RF 2006-1, 71 Fed. Reg. 64303, 64311 & n.90, n.93 (U.S. Library of Congress Nov. 1, 2006) (final order). For an analysis of the fair use question, see generally Nimmer, *supra* note 10.

234. *Derivative Rights, Moral Rights, and Movie Filtering Technology: Hearing Before the Subcomm. on Courts, the Internet, and Intellectual Prop. of the House Comm. on the Judiciary*, 108th Cong. 2 (2004) (statement of Hon. Lamar S. Smith, Chair, House Subcomm. on Courts, the Internet, and Intellectual Prop.).

films. Thus, Congress wisely short-circuited years of contentious legal wrangling by passing the Family Movie Act of 2005.<sup>235</sup> Section 110 of the Copyright Act now exempts from liability:

the making imperceptible, by or at the direction of a member of a private household, of limited portions of audio or video content of a motion picture, during a performance in or transmitted to that household for private home viewing, from an authorized copy of the motion picture, or the creation or provision of a computer program or other technology that enables such making imperceptible and that is designed and marketed to be used, at the direction of a member of a private household, for such making imperceptible, if no fixed copy of the altered version of the motion picture is created by such computer program or other technology.<sup>236</sup>

The Family Movie Act expressly exempts from liability individuals who exclude limited portions of films that they are viewing in their homes and companies who create and provide software or technology to facilitate such activities.<sup>237</sup> Thus, the Act carved out a narrow safe harbor to insulate ClearPlay and similar businesses from copyright liability. In so doing, Congress provided a timely, definitive, and carefully tailored solution to a challenge posed by a specific new technology.

As was the case with digital filtering technology, Google's Book Search Project is on a course of protracted, uncertain fair use litigation. Based on the foregoing analysis, there is good reason to believe that Congress can and should step into the fray and preempt what will likely be years of litigation, distorted investments, and overly risk averse business models (such as Google's opt-out policy).<sup>238</sup> There is a strong public interest in

---

235. See Family Movie Act of 2005, Pub. L. No. 109-9, § 202, 119 Stat. 218, 223 (to be codified at 17 U.S.C. § 110(11) and 15 U.S.C. § 1114(3)(A)–(D)). The Family Movie Act also amended the Lanham Act, 15 U.S.C. §§ 1051-1141 (2000 & Supp. IV 2004), to exclude from trademark infringement liability conduct described in ¶ 11 of § 110 so long as the individual also provides “a clear and conspicuous notice at the beginning of each performance” notifying the viewer that the performance has been altered from the version originally intended by the director or copyright holder. Family Movie Act of 2005, § 202, 119 Stat. at 224.

236. Family Movie Act of 2005, § 202, 119 Stat. at 223 (to be codified at 17 U.S.C. § 110(11) (2005)).

237. Family Movie Act of 2005, Pub. L. No. 109-9, § 202(b), 119 Stat. 218, 223–24 (to be codified at 15 U.S.C. § 1114(3)(A)–(D)).

238. Professor Oren Bracha sees Google's opt-out as a good solution to the controversy. See Bracha, *supra* note 172, at 1802–03. While there is much to be said for his praise of Google's book search project, his legal and policy analysis overlooks the copyright system's larger core values of accessibility, preservation, and promoting the arts and the important real and symbolic value of a truly comprehensive publicly searchable archive.

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1161

hastening the availability of a universal book search engine. We know enough now about the ramifications of book search technology to make sound public policy decisions. In the spirit of the Family Movie Act and the online safe harbor provisions of the DMCA, the next section proposes a package of statutory changes that would better facilitate access to knowledge without jeopardizing the copyright system's incentives to create.

*B. Towards a Comprehensive Searchable Knowledge Archive*

The challenge of producing a comprehensive searchable knowledge archive can usefully be broken down into two separate, but related challenges: (1) archiving and making searchable the stock of existing written knowledge and (2) efficiently and safely capturing the flow of new written knowledge in a digital form and making it available for search requests. Both public and private entities can address these challenges. Policy analysis should focus on the comparative advantages and disadvantages of the institutional alternatives.

1. *Archiving the Stock of Recorded Knowledge: A Safe Harbor for Digital Archiving and Search Services.* In building an archive of the stock of recorded knowledge, the basic policy tradeoff is between more rapid and efficient assembling of the knowledge archive through private enterprise but with less innovation and competition in the longer term due to relatively high entry costs. Through foresight, investment, technological expertise, and complementary business assets (its advertising business model), Google is in a strong position to assemble an unprecedented searchable knowledge archive within a relatively short time period. Other enterprises—such as the Open Content Alliance—also have the capacity to preserve and make accessible the stock of human knowledge.<sup>239</sup>

Google's Book Search Project comes close to achieving the aspirations of the great knowledge visionaries—from Ptolemy III to Francis I, Charles Jewett, and Ainsworth R. Spofford. Google's powerful Boolean search engine also directly addresses the information overload problem that has plagued library science for the past century and a half. Had Congress even imagined these possibilities in 1976, the balances that it legislated in section 108 suggest that it would have established a safe harbor for digital archiving and search capability.

---

239. Posting of Brewster Kahle to Yahoo! Search Blog, *supra* note 170.

The principal impediment to private efforts to build a comprehensive and searchable knowledge archive is the looming threat of copyright infringement. Such litigation raises serious questions that will take years to resolve. Congress has the opportunity to update the Copyright Act to address what is now reality. It seems clear that the private sector should play a leadership role, as Google and other search companies and nongovernmental organizations already have, in bringing searchable digital archives to fruition. Such enterprises have shown tremendous innovative ability and are in the best position to make breakthroughs in technological innovation. Individual publishers lack the incentives and technological capacity to bring about a comprehensive searchable archive. Authors and publishers bringing the litigation are not seeking to further the broader social goals of promoting progress, but rather a narrow pecuniary objective.

With all due respect to the Library of Congress and the Smithsonian Institution, neither could have brought about a viable, comprehensive, searchable-knowledge archive nearly as quickly or effectively as Google and other search companies are poised to do. Given the great importance of digital archiving to both knowledge preservation and providing reliable access to the public at large, any legal safe harbor should demand safeguards for the public while still protecting copyright owners.

Congress can hasten the deployment of digital archiving and search technology by crafting a safe harbor specifically for companies that work in conjunction with established library organizations to scan and digitize their collections for preservation and search purposes. Like the section 108 provisions,<sup>240</sup> the safe harbor should be tailored so as not to substitute for acquisition of in-copyright published material. Thus, the search results would be limited to what was necessary to locate sources and to determine where they could be legally procured. Congress could specify general guidelines for providing a sensible balance and delegate to the Copyright Office authority to determine and update specific regulations governing these activities. Such guidelines would likely reflect the same types of judgments that Google has designed into its search technology—e.g., limited number of snippets per work, limitations on the size of snippets, registration by users to prevent multiples queries aimed at extracting significant sections of work, and exclusion of

---

240. See *supra* notes 119–151 and accompanying text (discussing the provisions of § 108).

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1163

specific types of works still under copyright protection (such as dictionary entries and works of poetry where a snippet could easily contain a significant portion of the work). To encourage improvements in search technology, the safe harbor would allow qualifying search engine providers to engage in internal research using the digital archive.

On the public side, search companies should have to make commitments to the public nature of digital archive information to fall within the scope of the safe harbor. First, they should have to provide the digital archive to both the library from which the materials were scanned and to the Library of Congress. Second, the software needed to access and utilize the archive should be publicly available or, if subject to trade secrecy, be held in software escrow<sup>241</sup> so that the government could gain access to it should the search engine/archiving enterprise enter bankruptcy.

To protect the interest of copyright owners, the digital archiving and search technology legislation should afford copyright owners a right of action against search companies and libraries that use insufficient technological protection measures to ensure the security of copyrighted materials. The liability for such breaches should not be grossly disproportionate to the harm caused. The current provisions for statutory damages were enacted before digital technology was appreciated. These provisions can result in unduly punitive remedies, as occurred in *UMG v. MP3.com*.<sup>242</sup> Given the huge number of copyrighted works that could be implicated, it is essential that Congress anchor the extent of liability to real-world impacts adjusted for detection and enforcement uncertainties so as not to unduly chill productive activities. Congress should provide the appropriate deterrent, but not crushing liability for activities that have socially redeeming value. The concern about unauthorized leakage of digital archives could also be addressed prospectively by empowering the Copyright Office to establish regulations for digital archiving and search technology providers, although such

---

241. Many software licensees require the licensor to store a current version of the source code with an independent software repository so that, in the event of bankruptcy, the licensee will have access to a human readable version of the program. In this way, the licensor can maintain the source code in secret while assuring the licensee that it will not lose access to critical technical information should the licensor's business fail. Walter D. Denson. *The Source Code Escrow: A Worthwhile or Worthless Investment?*, RUTGERS BANKR. L.J., Mar. 2003, at 2, [http://www.bankruptcy.rutgers.edu/source\\_code\\_escrow.pdf](http://www.bankruptcy.rutgers.edu/source_code_escrow.pdf).

242. *UMG Recordings, Inc. v. MP3.com, Inc.*, 56 U.S.P.Q.2d 1376, 1381 (S.D.N.Y. 2000) (holding that statutory damages in the amount of \$25,000 per compact disc copied were appropriate and, because the total number of copied discs was estimated to be 4,700, damages of \$118,000,000 were anticipated).

provisions could result in significant bureaucratic and political costs. The legislation could also mandate a certification process for companies wishing to take advantage of the safe harbor, as well as require that companies engaging in such activities post a bond and promptly disclose and remedy security breaches.

Although not strictly necessary to promote digital archiving and search, Congress should also consider addressing greater access to orphaned works in this legislation. A vast quantity of knowledge is not being actively protected. The Copyright Office has been studying policy alternatives for expanding use of orphan works.<sup>243</sup> In addition, a study group has been considering revisions to section 108 of the Copyright Act, the provision addressing reproduction and archiving by libraries.<sup>244</sup> Such reforms could well be integrated into legislation creating a safe harbor for digital search technology.

2. *Archiving the Flow of Recorded Knowledge: Digital Deposit.* With regard to archiving the flow of new written knowledge, the Library of Congress has an important institutional advantage as it already has a system for archiving tangible versions of books and other materials. Its main challenges relate to storage, archiving, and retention of published and registered works. It would make sense for Congress to require that publishers and authors deposit a digital version of their works with the Library of Congress.

At present, the overwhelming majority of authors and publishers now prepare their works in digital formats (such as Microsoft Word and Adobe's portable document format (PDF)). Therefore, the costs on the private sector of depositing copyright works in digital format would be modest. Congress should update the deposit system to require that publishers and authors deposit both tangible and digital copies of their works with the Library of Congress. The Library of Congress could establish, through ongoing rulemaking, the procedures and standards for digital deposit of copyrighted works. In this way, the Library of

---

243. See U.S. COPYRIGHT OFFICE, *supra* note 195, at 8–9 (recommending an orphan works amendment to the Copyright Act that limits remedies available to a copyright owner if a work is used, as long as the user performed an unsuccessful but reasonably diligent search to find the copyright holder); Huang, *supra* note 195, at 265 (discussing efforts by the copyright office to resolve the orphan works issue).

244. See Section 108 Study Group Mission Statement, <http://www.loc.gov/section108/mission.html> (last visited Aug. 30, 2007) (stating that the group's mission is "to conduct a reexamination of the exceptions and limitations applicable to libraries and archives under the Copyright Act, specifically in light of the changes wrought by digital media."). The Study Group will submit findings to the Librarian of Congress in mid-2007. *Id.*

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1165

Congress would have greater flexibility in its knowledge management and archiving policies. It could begin to experiment with new search technologies, avoid future costs of digital archiving, and possibly ease some of the storage burdens.

3. *Longer Term Knowledge Policy Prescriptions: A Public Comprehensive Searchable Archive.* Just as a public human genome database promotes downstream biological research, a public, searchable knowledge archive would promote competition and innovation in search technology. Due to the economics of amassing such a digital archive, a purely private solution will likely lead to a single provider or perhaps just a few competitors. This could well lead to a bottleneck at the search innovation level. Only the company with the dominant search archive would be in the position of conducting the best research and commercializing improvements in search algorithms; but without strong marketplace competition, that dominant company's impetus for innovation might lag. On the other hand, the alternative of public ownership and control of the digital archive could well produce bureaucratic and political forms of inefficiency. The government's ability to manage such a project would appear to be limited at present.

It makes sense, therefore, for public policy to aim toward the goal of building a comprehensive, publicly searchable archive. Implementing a digital deposit requirement would support this goal. But efforts would have to be undertaken to digitize and make searchable the much larger stock of knowledge. Various possibilities could be pursued. The government could develop internal capacity to digitize content or contract with Google or other entities to amass the archive. There would also remain questions about the design of the search technology, the restrictions on use by other entities, and safeguards to prevent piracy of materials within the archive.

#### V. THE RELEVANCE OF HISTORICAL, DEMOCRATIC, CULTURAL, AND ECONOMIC PERSPECTIVES TO THE APPLICATION OF THE FAIR USE DOCTRINE

If the legislative safe harbor proposed here cannot be enacted, then federal courts will face the unprecedented question of whether such activities qualify as fair use. Both sides will be able to find cases that arguably support their positions. Authors and publishers will emphasize the nature of the works, the

amount being copied, and Google's commercial nature.<sup>245</sup> Google will respond that its activities are highly transformative and do not harm the markets for books.<sup>246</sup> Authors and publishers will counter that Google does not transform the works so much as it reproduces them, and they will also contend that they are losing the potential market of licensing search information.

The mechanical quality of some fair use jurisprudence—marching through the four statutory factors—as well as the nebulous nature of “potential market harm” could well obscure the larger social benefits of Google's book search technology. This judicial tendency is exacerbated by the complexity and detail of much of the modern copyright code<sup>247</sup> and the relative infrequency with which most jurists encounter copyright cases. Whereas a technical approach is required for dealing with the tax code-like aspects of copyright law, fair use analysis entails a more flexible, open-ended approach. As noted in the legislative history to the 1976 Act, the list of fair use purposes set forth in the preamble to section 107—“criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research”—are intended to be illustrative, not exhaustive.<sup>248</sup> The House Report observes:

Although the courts have considered and ruled upon the fair use doctrine over and over again, no real definition of the concept has ever emerged. Indeed, since the doctrine is an equitable rule of reason, no generally applicable definition is possible, and each case raising the question must be decided on its own facts. On the other hand, the courts have evolved a set of criteria which, though in no

---

245. See *Basic Books, Inc. v. Kinko's Graphics Corp.*, 758 F. Supp. 1522, 1537–39 (S.D.N.Y. 1991) (holding unauthorized creation of anthologies of scholarly works for classroom use by a copy shop is not fair use); see also *Harper & Row Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 565, 569 (1985) (holding the fair use doctrine does not protect unauthorized copying of an unpublished manuscript where the portions copied constitute the “heart” of the unpublished work).

246. See *supra* note 11 (listing cases holding that a use fell within the fair use doctrine); see also *Sony Computer Entm't v. Connectix Corp.*, 203 F.3d 596, 602 (9th Cir. 2000) (holding intermediate copying of computer software to reverse-engineer the functional elements of the software is protected under the fair use doctrine), *cert. denied*, 531 U.S. 871 (2000); cf. *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 572 (1994) (holding unauthorized copying for use in a parody is protected by the fair use doctrine); *Bill Graham Archives v. Dorling Kindersley Ltd.*, 448 F.3d 605, 615 (2d Cir. 2006) (holding unauthorized copying of concert posters is protected by the fair use doctrine where the posters are not used for their expressive value but for their historical value).

247. See David Nimmer, *Codifying Copyright Comprehensively*, 51 UCLA L. REV. 1233, 1238 (2004).

248. See H.R. REP. NO. 94-1476, at 65–66 (1976) (“[T]he endless variety of situations and combinations of circumstances that can rise in particular cases precludes the formulation of exact rules in the statute.”).

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1167

case definitive or determinative, provide some gauge for balancing the equities. These criteria have been stated in various ways, but essentially they can all be reduced to the four standards which have been adopted in section 107: “(1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for or value of the copyrighted work.”<sup>249</sup>

The need for flexibility and attention to copyright law’s larger purposes is particularly relevant in dealing with the challenges posed by new technology. The House Report notes that “[t]he specific wording of section 107 as it now stands is the result of a process of accretion, resulting from the long controversy over the related problems of fair use and the reproduction (mostly by photocopying) of copyrighted material for educational and scholarly purposes.”<sup>250</sup> There is a risk that modern jurists, lacking a full appreciation of copyright law’s multi-faceted dimensions, will see the *Google* litigation simply as a case about unauthorized copying.

This paper has provided two antidotes to such reflexive reasoning. The historical review shows that both the constitutional framework supporting copyright law, as well as the arc of copyright legislation dating back to 1790, promote preservation and providing widespread access to knowledge. It is only where such access undermines the principal market for copyrighted works—as in systematic copying or copying for commercial advantage by libraries—that the rights of copyright owners trump the interests of research institutions and those searching for knowledge. Although the balances reflected in section 108 of the Copyright Act might be narrowly seen as technical rules, a broader understanding shows that they reflect a larger balance underlying the copyright system. Courts should look to copyright law’s overarching goal of promoting progress and its long-standing history of supporting preservation, cataloguing, and access to knowledge in framing the fair use inquiry.

The second antidote relates to the democratic, cultural, and economic considerations bearing on the fair use balance. As Part II of this Article developed, digital book search technology offers

---

249. *See id.* at 65.

250. *See id.* at 66.

dramatic democratic, cultural, and economic benefits to the larger populace without materially interfering with the exploitation of copyrighted works. To the contrary, such technology affords consumers the ability to find the works that best satisfy their needs. Given the dramatic benefits of digital archiving and search technology, courts should not become mired in overly mechanical application of the fair use factors. Rather, they should recognize that liberal construction of the fair use privilege serves to facilitate the creation of the greatest collection of knowledge in the history of humankind. Few innovations since the printing press hold as much promise for promoting progress in science and the useful arts.

## VI. CONCLUSION

Copyright and technology for the creation and distribution of works of authorship have evolved symbiotically throughout history.<sup>251</sup> Recent advances in digital technology have created the potential to make the vast stock of recorded knowledge searchable using sophisticated tools by anyone with an Internet connection. As Google is in the process of demonstrating, it is now feasible to scan the collections of the major libraries of the world within a matter of years, convert these works into an immense searchable digital archive, and enable Internet users to find the most relevant materials easily. Such technology appears to pose little if any economic threat to publishers and authors. To the contrary, it will undoubtedly expand demand for their works and spawn new markets for online delivery of copyrighted works. Unfortunately, this initiative is mired in protracted and uncertain litigation. Furthermore, Google has already compromised its grand vision of creating a universal archive by offering copyright owners the opportunity to remove their works from the archive.

Rather than endure years of legal wrangling over whether Google's Book Search Project qualifies as fair use, Congress should confront the potential opportunities and risks of digital technology preemptively and directly to strike the appropriate balance between protection of works of authorship on the one hand and accessibility and preservation on the other. By focusing on the economic, social, and cultural benefits of building a comprehensive publicly searchable database of literary and artistic works, Congress can effectuate the overarching purposes of "promoting progress" and preserving human knowledge

---

251. See generally Menell, *supra* note 227, at 63–64.

2007] *KNOWLEDGE POLICY FOR THE DIGITAL AGE* 1169

without sacrificing the beneficial economic incentives afforded by copyright law. A carefully crafted safe harbor, with appropriate safeguards to prevent piracy of in-copyright works, would fuel markets for copyrighted works while making accessible the vast stock of knowledge to current scholars and authors and preserving the largest possible record for future generations.

Congress should also bring its deposit requirements fully into the digital age by requiring that publishers and authors deposit a digital version of copyrighted works with the Library of Congress in addition to the tangible copy currently required. In this way, the Library of Congress could more efficiently manage its collection and accelerate the process of building a searchable knowledge archive. Over time, the government should go back in time and digitize the stock of human knowledge, thereby achieving the ultimate goal of providing a public comprehensive searchable archive. Such a resource could spur competition and innovation in search technology and services.

Beyond Congress prescribing the optimal knowledge access and preservation policy in the digital age, the analysis of this Article provides a broader context for courts applying copyright law's fair use doctrine to private initiatives, such as Google's Book Search Project, that provide a substantial public benefit with relatively modest risk of piracy. By appreciating the rich traditions and purposes of copyright law, courts can better assess the flexible, evolving balance reflected in copyright law's structure and defenses.