

PROCEED WITH CAUTION: HOW DIGITAL ARCHIVES HAVE BEEN LEFT IN THE DARK

By Alyssa N. Knutson

Digital preservation practice exists in a legal gray area. Following the proposed settlement of Google's Library Project litigation—a pair of class action copyright infringement lawsuits brought against Google in 2005 by the Authors Guild, the Association of American Publishers, and several publishing houses¹—the legality of digital archives and libraries remains unclear. Furthermore, cases such as the short-lived *Internet Archive v. Shell*² provide little guidance but highlight many of the dangers digital archives may encounter.

This Note lays out the principal liabilities to which digital archives are exposed, including their potential but imperfect defenses, using the recently settled but non-precedential *Internet Archive v. Shell* case as an example of the uncertainty surrounding digital archiving. Digital archives face many legal barriers, including practically perpetual copyright terms in the material they include, an uncertain fair use doctrine, a chaotic licensing scheme, and a proliferation of online contracting that threaten archivists' efforts to construct comprehensive digital libraries. While scholarly literature has already explored the application of the fair use defense to archives such as Google's Library Project³ and to some extent the Internet Archive,⁴ this Note further explores the fair use defense in light of the Google

© 2009 Alyssa N. Knutson. The author hereby permits the reproduction of this Note subject to the Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License, the full terms of which can be accessed at <http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode>, and provided that the following notice be preserved: "Originally published in the Berkeley Technology Law Journal 24:1 (2009)."

1. See generally Lawrence Jordan, *The Google Book Search Project Litigation—“Massive Copyright Infringement” or “Fair Use”?*, 86-SEP. MICH. B.J. 32 (2007) (providing a quick overview of Google Book Project and the New York litigation).

2. 505 F. Supp. 2d 755 (D. Colo. 2007).

3. See, e.g., ROBIN JEWELER, CONG. RESEARCH SERV., THE GOOGLE BOOK SEARCH PROJECT: IS ONLINE INDEXING A FAIR USE UNDER COPYRIGHT LAW? (2005), available at http://www.opencrs.com/rpts/RS22356_20051228.pdf; Jonathan Band, *Google and Fair Use*, 3 J. BUS. & TECH. L. 1 (2008); Douglas L. Rogers, *Increasing Access to Knowledge Through Fair Use—Analyzing the Google Litigation to Unleash Developing Countries*, 10 TUL. J. TECH. & INTELL. PROP. 1 (2007); Elizabeth Hanratty, *Google Library: Beyond Fair Use?*, 2005 DUKE L. & TECH. REV. 10.

4. See, e.g., Kinari Patel, *“Authors v. Internet Archives”: The Copyright Infringement Battle Over Web Pages*, 89 J. PAT. & TRADEMARK OFF. SOC'Y 410 (2007); Rebecca Bolin, *Locking Down the Library: How Copyright, Contract and Cybertrespass Block*

Library Project settlement and new threats to such projects including breach of contract actions and an unsympathetic preemption doctrine.

The legal uncertainty shrouding archives threatens current efforts to preserve the world's digital information, a socially and historically beneficial undertaking. Because of this unresolved legal area, archivists operate over cautiously. To avoid litigation of their status, or for a better chance to win at future litigation, archives have implemented and respected 'opt-out' policies, thus allowing copyright and webpage owners to refuse the inclusion of their materials in the archive. By doing so, the Internet Archive and similar archive projects have significantly compromised their overall goal of creating a comprehensive record of past knowledge. This compromise hurts the greater social interest in preservation and access to knowledge.

Furthermore, after Google's settlement⁵ it takes a rare company to enter this chilling environment. Without clear guidance on how the fair use doctrine may play out, individuals and institutions may be discouraged from creating or maintaining noncommercial preservation entities. Because the settlement did not go to the merits of the dispute, any entity that seeks to enter the digital archive field will continue to be subject to liability for copyright infringement, although it may not have Google's extensive resources to fight for fair use in court.

This Note concludes with a discussion of broad solutions to this legal limbo. These solutions highlight digital archives' socially beneficial nature but respect both copyright law's safeguards to incentivize creation and the personal right to enter into private contracts. Congress and the courts should encourage public digital preservation by implementing a safe harbor for digital archives under particular conditions and by rethinking traditional preemption analysis. We must create a new framework that enables us to endorse the prospects of nonprofit digital archiving while overcoming the present and emerging obstacles resulting from uncertain legal doctrines and Google's market-based approach to digital archiving.

Internet Archiving, 29 HASTINGS COMM. & ENT L.J. 1 (2006); Kelly Jines-Storey, Comment, *Does Rocky and Bullwinkle Hold the Key to Unlocking the Mystery of Fair Use in the Age of Internet Archiving?*, 35 CAP. U. L. REV. 1021 (2007).

5. See discussion *infra* Section II.A.4 (discussing how the proposed Google Library Project settlement effects digital archives' fair use defense); see also Jonathan Band, *The Google Library Project: Both Sides of the Story*, 1 PLAGIARY: CROSS-DISCIPLINARY STUDIES IN PLAGIARISM, FABRICATION, AND FALSIFICATION 6 (2006), <http://quod.lib.umich.edu/p/plag/images/5240451.0001.002.pdf> (providing background on the Google Library Project, the surrounding litigation, Google's fair use defense, and some copyright holders' responses to Google's arguments).

I. CURRENT STATUS OF ARCHIVES

A. Preservation Practice

Scholars widely acknowledge that preservation of and access to cultural artifacts is necessary for a robust cultural life. Historical preservation impacts a society's "collective memory"—the way in which a society sees itself, its values, and its traditions.⁶ Furthermore, since new ideas can build on past knowledge, access to collective knowledge leads to the creation of new creative expression.⁷ Access to information also allows citizens to exercise their First Amendment right to information and ideas, fundamental democratic values on which the United States was founded.⁸

As many scholars have documented, preservation of cultural property and knowledge is not new.⁹ From the ancient library of Alexandria to the Smithsonian Institution, a great deal of time and effort has been expended gathering, storing, preserving, and exhibiting cultural property.¹⁰ Early examples of preservation practice reflect societal interest in collecting and preserving knowledge. King Ptolemy III required that all books brought to Alexandria be copied for the library's archives.¹¹ The Renaissance saw the

6. See ARCHIVES, DOCUMENTS AND INSTITUTIONS OF SOCIAL MEMORY: ESSAYS FROM THE SAWYER SEMINAR 165 (Francis X. Blouin, Jr. & William G. Rosenberg eds., 2006), for a collection of essays discussing the role of memory institutions in "conditioning social memory and creating certain kinds of cultural understandings." See also JOSEPH L. SAX, PLAYING DARTS WITH A REMBRANDT: PUBLIC AND PRIVATE RIGHTS IN CULTURAL TREASURES 197 (1999) (arguing that cultural property forms the basis of our collective memory); Guy Pessach, [Networked] Memory Institutions: Social Remembering, Privatization and its Discontents, 26 CARDOZO ARTS & ENT. L.J. 71, 75 (2008) ("Landscapes of history and social remembering are major forces in the construction of ideologies and people's preferences.").

7. This is expressly the IP Clause's quid pro quo. See U.S. CONST. art. 1, § 8, cl. 8 ("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.").

8. Richard J. Peltz, *Use "The Filter You Were Born With": The Unconstitutionality of Mandatory Internet Filtering for the Adult Patrons of Public Libraries*, 77 WASH. L. REV. 397, 397 (2002) (arguing that libraries are "the quintessential venue for citizens to exercise their First Amendment right to receive information and ideas").

9. See Peter S. Menell, *Knowledge Accessibility and Preservation Policy for the Digital Age*, 44 HOUS. L. REV. 1013, 1014 (2007) ("[S]ocieties have long sought to preserve and catalog human knowledge and make it publicly accessible."); see generally *id.* (providing a detailed account of the history of archives and libraries).

10. J.H. Merryman, *The Public Interest in Cultural Property*, 77 CALIF. L. REV. 339, 344 (1989).

11. Andrew Erskine, *Culture and Power in Ptolemaic Egypt: The Museum and Library of Alexandria*, 42 GREECE & ROME 38, 39 (1995).

first law, the French “Ordonnance de Montpellier,” that required the deposit of books and other cultural property in a library for preservation purposes.¹²

Likewise, legal and social institutions in the United States have recognized the importance of the preservation of knowledge and cultural property. Congress recognized the public interest in preservation and access to knowledge when it established the Smithsonian Institution in 1846, requiring one copy of each copyrighted work be delivered to the Librarian at the Smithsonian Institution and to the Library of Congress within three months of publication.¹³ This requirement lives on today.¹⁴ In the Copyright Act of 1976, Congress further 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.¹⁵

B. Into the Digital Realm

Digital technology offers extraordinary opportunities for preservation of and access to knowledge. The problems of the past, such as lack of storage or geographical limitations, are disappearing as a variety of preservation institutions digitize their materials and make them more accessible to the public. Digital search tools allow instantaneous and global access to millions of users and effectively unlimited retrieval possibilities. Traditional institutions that have preserved knowledge and made it accessible to

12. JULES LARIVIÈRE, INT’L FED’N OF LIBRARY ASS’NS AND INSTS., GUIDELINES FOR LEGAL DEPOSIT LEGISLATION ch. 2 (2000) (a revised, enlarged, and updated edition of the 1981 publication by Dr. Jean Lunn IFLA Committee on Cataloguing).

13. Act of Aug. 10, 1846, ch. 178, §10, 9 Stat. 102, 106 (1851). Charles Jewett, the first librarian of the Smithsonian Institution, extolled the virtues of preservation without omission:

[I]n coming years, the collection would form a documentary history of American letters, science and art. It is greatly to be desired . . . 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?

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

14. The modern Copyright Act requires the owner or exclusive rights holder in a work published in the U.S. to deposit two complete copies of the work within three months after publication. 17 U.S.C. § 407(a) (2006).

15. Copyright Act of 1976, Pub. L. No. 94-553, §108, 90 Stat. 2541, 2546-47 (codified at 17 U.S.C. § 108 (2006)). *See also* Digital Millennium Copyright Act of 1998, Pub. L. 105-304, 112 Stat. 2860 (codified as amended at 17 U.S.C. §§ 512, 1201-1205, 1301-1332, 4001 (2006)) (updating digital preservation and access in the Copyright Act).

the public including the Smithsonian Institution, the Library of Congress, and university research libraries have now launched digital counterparts and projects.¹⁶ Similarly, traditional newspaper, periodical, and print publishers now make content available on the Internet, as do research institutions, courts, and government entities.

Digital archives, today's version of the traditional library, organize, maintain, and support collections of informational items in digital format, accessible to users through networks. Because traditional libraries are limited by storage space and the cost of maintaining that space, digital libraries have the potential to store and make available much more information at a fraction of the cost and space.¹⁷ The effects are clear; internet archiving has already preserved a great deal of culture. For example, the Internet Archive, a digital archive that collects, stores, and makes available old webpages, surpassed ten billion webpages by 2002, or 100 terabytes of information, an amount of material four times greater than all the books in the Library of Congress.¹⁸

16. For example, the Library of Congress's American Memory Project offers free access through the Internet to literature, speeches, sound recordings, movies, pictures, prints, maps, and sheet music that document the "American experience." Hannibal Travis, *Building Universal Digital Libraries: An Agenda for Copyright Reform*, 33 PEPP. L. REV. 761, 770 (2006); see generally Library of Congress, American Memory Home, <http://memory.loc.gov/ammem/index.html> (last visited Dec. 23, 2008). Similarly, JSTOR, Ithaka, the Library of Congress and the Andrew W. Mellon Foundation launched a project called Portico, an e-journal archive that preserves scholarly electronic journals to ensure their continued availability to scholars and researchers. June M. Besek & Philippa S. Loengard, *Maintaining the Integrity of Digital Archives*, 31 COLUM. J.L. & ARTS 267, 341 (2008); see generally About Portico, <http://www.portico.org/about/> (last modified Nov. 30, 2007).

17. While servers are not free strictly speaking, all-digital libraries require fewer labor resources such as hands-on maintenance of the materials, fewer space resources, and fewer material resources. See Lynn Silipigni Connaway & Stephen R. Lawrence, *Comparing Library Resource Allocations for the Paper and the Digital Library: An Exploratory Study*, 9 D-LIB Magazine, Dec. 2003, <http://www.dlib.org/dlib/december03/connaway/12connaway.html> (conducting a study on costs associated with paper libraries versus costs associated with all-digital libraries). The move to an all-digital library, however, requires a strong commitment to maintaining and backing up the digital copies.

18. Travis, *supra* note 16, at 769. How much information is currently contained on the Web and how fast it will grow is an interesting question which has generated a lot of debate. E.g., compare SCHOOL OF INFO. MGMT. SYS., UNIV. OF CALIFORNIA-BERKELEY, HOW MUCH INFORMATION? 8 (2003), available at http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/printable_report.pdf, with Edward T. O'Neill et al., *Trends in the Evolution of the Public Web* (1998-2002), 9 D-LIB MAGAZINE, Apr. 2003, <http://www.dlib.org/dlib/april03/lavoie/04lavoie.html>. As of June 2008, Cisco Systems, the leading supplier of networking equipment and network management for the Internet, estimated internet traffic at seven exabytes a month, which is about 160 tera-

A remarkable variety of institutions are developing or planning to develop digital archives. Some, such as Google's Library Project, are organized privately and are commercial entities open to the public. Other digital archives, like Burning Well, a public image depository, rely on submissions from a community of contributors and users.¹⁹ Project Gutenberg uses the services of numerous volunteers to proofread and prepare texts for online publication.²⁰ LexisNexis and Westlaw, which provide legal materials for a fee, are examples of private, commercial, and exclusionary archives.²¹

C. Internet Archive: The Digital Web Archive

The Internet Archive is a not-for-profit privately funded digital preservation institution with similar goals and spirit as the original public preservation institutions like the Library of Congress or Library of Alexandria. It has confronted the legal barriers and challenges typical of those digital archives face. Its methods of collecting, preserving, and displaying the information collected remain legally vulnerable under multiple areas of law, including copyright and state contract law. One of the Internet Archive's recent cases, *Internet Archive v. Shell*, sheds light on the uncertainty surrounding the legality of digital archives.

1. Brewster Kahle's Vision

Do you know what's carved above the Carnegie Library in Pittsburgh?—'FREE TO THE PEOPLE'—what a goal! . . . I can believe in this! At the Internet Archive, we think of our mission as 'universal access to all knowledge.'²²

The Internet Archive is a widely known and esteemed example of a grand vision of modern preservation practice. Based in San Francisco, California, the nonprofit Internet Archive seeks to preserve and maintain a

bytes per second. Bobby White, *Cisco Projects Growth to Swell for Online Video*, THE WALL ST. J., June 16, 2008, at B5.

19. BurningWell.org, BurningWell.org Frequently Asked Questions, http://www.burningwell.org/donors_faqs.html (last visited Dec. 23, 2008).

20. Project Gutenberg, Gutenberg: Volunteers' FAQ, http://www.gutenberg.org/wiki/Gutenberg:Volunteers%27_FAQ (last modified Mar. 24, 2008).

21. The distinction between nonprofit and for-profit digital archives makes a difference under copyright law's fair use analysis. See discussion *infra* Section II.A.2.

22. Heidi Benson, *A MAN'S VISION: WORLD LIBRARY ONLINE, Brewster Kahle Hopes to Realize His 25-Year Dream of an International Book Archive*, S.F. CHRON., Nov. 22, 2005, at A1.

comprehensive record of the Internet.²³ The Archive collaborates with institutions such as the Library of Congress and the Smithsonian to preserve a record of modern society's history and culture as a resource for future generations and present-day researchers, historians, and scholars.²⁴ In addition to copies of webpages taken at various points in time, the Archive also includes software, movies, books, and audio recordings.²⁵

Brewster Kahle, the founder and digital librarian of the Internet Archive, understood the importance of preserving what is stored and created on the Web. The Web has increasingly evolved into a storehouse of valuable scientific, cultural, and historical information. A wide variety of individuals use the Internet as a storehouse of knowledge that they can access when needed, often multiple times a day. However, webpages are ephemeral, sometimes only lasting weeks or days or even less.²⁶ For example, the ten most popular stories on CNN.com are updated every twenty minutes.²⁷ Therefore, the very content and layout of CNN.com's homepage changes seventy-two times a day. The Archive's goal is to preserve and store this intangible and impermanent content before it disappears.

Furthermore, Kahle envisioned an educational future for the Internet Archive as part of the eventual Universal Library of the Internet and digital culture.²⁸ By capturing different versions of webpages over time, the Archive documents the electronic past. The Archive can provide a snapshot of the "cultural state of society at any given point in time."²⁹ Further,

23. *Internet Archive v. Shell*, 505 F. Supp. 2d 755, 760 (D. Colo. 2007). A web archive is a "web server that caches snapshots of different websites at regular intervals and subsequently makes them available to the public." Boštjan Berčič, *Protection of Personal Data and Copyrighted Material on the Web: The Cases of Google and Internet Archive*, 14 INFO. & COMM. TECH. L. 17, 22 (2005).

24. *Internet Archive's Motion to Dismiss Counterclaims at 3*, *Internet Archive v. Shell*, No. 06-cv-01726-LTB-CBS (D. Colo. Nov. 16, 2006).

25. *Internet Archive, About the Internet Archive*, <http://www.archive.org/about/about.php> (last visited Dec. 23, 2008).

26. Of the roughly 50,000 software titles published over the years released, the vast majority are currently unavailable commercially. Moreover, in 2000, the average webpage was taken down after a mere seventy-five days, which meant that half of the websites were disappearing within a year's time or less. Travis, *supra* note 16, at 803.

27. *See CNN Homepage, CNN.com* (last visited Dec. 23, 2008).

28. Discussing his vision for the Internet Archive, Kahle stated:

I see this in a broader context than just making a time capsule. I'm not proposing that I know how we can build the ultimate digital library, but at least we can start the collection for those libraries that in a few years will become an integral part of our information ecology.

JOHN BROCKMAN, *DIGERATI: ENCOUNTERS WITH THE CYBER ELITE—WEB ADDITION* ch. 15 (1996), available at <http://www.edge.org/documents/digerati/Kahle.html>.

29. Patel, *supra* note 4, at 411.

Kahle believes that by documenting the past and archiving the whole Internet we can understand the phenomenon of the Internet and digital explosion in the future.³⁰ The Archive also serves as a record of our cultural past, documenting politics, entertainment, and more as content moves into the digital environment.³¹

The Internet Archive has also been used to preserve electronic information. For example, old webpages are often used as evidence in court.³² They can be useful in making novelty, statutory bar, or nonobviousness arguments in patent cases, and trademark attorneys can prove infringement or first use on particular trademarks.³³ The Archive can also be used as a backup for dead sites when websites disappear or webmasters move on.³⁴

2. *How the Internet Archive Works*

To collect data from the Web, the Archive employs the “Wayback Machine” to systematically browse the entire World Wide Web, copy the content from the browsed websites, and place it in a publicly searchable archive. The Wayback Machine uses a software program known as a “web crawler,” “spider,” or “robot” that moves across the Internet in search of

30. BROCKMAN, *supra* note 28.

31. For example, the Internet Archive had a project with the Smithsonian Institution to archive the 1996 presidential election to see how the Internet affected the political landscape. *Id.* In an early interview Kahle explained how one could use the Archive as a sociological research tool:

If you have a full copy of the Internet in one place, you can do clustering studies to understand the evolution of communities and their overlaps, as we move from being a global village where everybody is chanting the same theme song from a popular sitcom to having lots and lots of different communities out there. We can track demographic shifts and even experiment with new indexing technologies. A centralized resource may not be the correct long-term solution, but it's a way to get started quickly.

Id.

32. *See, e.g.,* *Telewizja Polska USA v. Echostar Satellite Corp.*, No. 02-C-3293, 2004 WL 2367740 (N.D. Ill. Oct. 15, 2004); *see also* Matthew Fagan, “*Can You Do a Wayback On That?*” *The Legal Community’s Use of Cached Web Pages In and Out of Trial*, 13 B.U. J. SCI. & TECH. L. 46, 67-70 (2007) (discussing use of cached webpages under the Federal Rules of Evidence).

33. The use of pages culled from Internet Archive’s Wayback Machine as evidence has been the subject of debate. *See, e.g.,* *Novak v. Tuscow, Inc.*, No. 06-CV-1909, 2007 WL 922306, at *5 (E.D.N.Y. Mar. 26, 2007) (holding contested webpages from Internet Archive not authenticated under Rules of Evidence because neither testimony nor sworn affidavits were proffered); *Telewizja*, 2004 WL 2367740, at *6 (allowing entry of data from Internet Archive service, however it was accompanied by an affidavit of a representative of Internet Archive attesting to its authenticity).

34. BROCKMAN, *supra* note 28.

new webpages. After the software downloads a page, it looks for cross references, or links, to other pages. To avoid downloading duplicates, the crawler evaluates uniform resource locators (URLs) against what is already archived in its database.³⁵ This technology generally operates with no human involvement or intervention. Google, Yahoo, and many other companies employ similar crawler technology for indexing websites for use in search engines.³⁶ After collection it generally takes six months or more for pages to appear in the Wayback Machine searchable archive because of delays of transferring materials to long-term storage.³⁷

The Wayback Machine does not seek the website owners' prior permission to reproduce the website content, but the Internet Archive website provides information advising website owners of how to prevent their material from being copied and how to remove the material from the archive.³⁸ First, authors who do not want their webpages archived can opt-out through the use of the robot exclusion standard. The robot exclusion standard, also known as the Robots Exclusion Protocol or robots.txt, is a method to prevent cooperating web crawlers from accessing all or part of a website that is otherwise publicly viewable.³⁹ Website owners who do not want their old site versions to be copied and ultimately accessible to the public include a denial text string into their robots.txt file on the computer that hosts the website.⁴⁰ Web crawlers will ignore particular websites that contain the denial string in their search. The Internet Archive's crawler program, which recognizes the robot exclusion standard, will refrain from making copies of any sites that include the robots.txt protocol. Therefore, sites that employ the robots.txt will not be accessible to the public through the Wayback Machine.⁴¹ While this is an efficient way for people to pre-

35. Brewster Kahle, *Preserving the Internet*, 276 SCI. AM. 82, 82 (1997).

36. For an analysis of the differences between search engines' caching of webpages and general archiving of webpages, see Berčić, *supra* note 23, at 20.

37. Internet Archive, Frequently Asked Questions, <http://www.archive.org/about/faqs.php> (last visited Dec. 23, 2008).

38. *Id.* Along with giving tips for how to exclude material from the Wayback Machine, the Archive asserts that it "is not interested in preserving or offering access to websites or other Internet documents of persons who do not want their materials in the collection." *Id.*

39. *Parker v. Yahoo!, Inc.*, 88 U.S.P.Q.2d 1779 (BNA), fn.9 (E.D. Pa. Sept. 25, 2008) (explaining how robots.txt works).

40. See *About /robots.txt*, <http://www.robotstxt.org/robotstxt.html> (last updated Apr. 1, 2008).

41. Internet Archive, Internet Archive Contacts: Removing Content from the Wayback Machine, <http://www.archive.org/about/exclude.php> (last visited Dec. 23, 2008) (stating that the robots.txt file will do two things: (1) remove all documents from a par-

liminarily opt-out of their pages being archived or cached, it is not legally required and up to the website publisher to implement.⁴²

The Internet Archive also removes material from the Wayback Machine archive on request from a website owner.⁴³ In its Terms of Use, the Internet Archive recognizes that authors and publishers of the “publicly available Internet documents” may not desire to have their documents included in the Archive.⁴⁴ Besides relying on website publishers to tag their web files for robot exclusion, the Archive allows the website owner to contact the Archive to remove that portion of the collection.⁴⁵ The Internet Archive also employs a Copyright Agent to handle standard DMCA § 512 notice and takedown requests from copyright owners who believe their material is being infringed.⁴⁶

3. *Litigation Against the Internet Archive*

It was only a matter of time before the Internet Archive would be brought into court. Brewster Kahle himself acknowledged the large legal implications of a project of this magnitude:

There are a bunch of legal and social issues as well. Most institutions cannot touch this because it hits every privacy, copyright, and export controversy. I feel like we've touched a raw nerve in attempting this project, since it can change the Net forever from an ephemeral medium to an enduring one.⁴⁷

The major impediment to Kahle's project moving forward smoothly appears to be the looming lawsuits, the outcome of which appears far from clear. Given the fact that the very activity of preservation requires wholesale copying, it was inevitable that the Archive would face legal challenges.

However, past cases have not focused on the legality of the Internet Archive's project, but instead have challenged the Archive's obligation to not archive webpages with exclusion mechanisms installed. The Internet Archive was sued in *Healthcare Advocates, Inc. v. Harding, Earley, Foll-*

particular domain from the Wayback Machine, and (2) tell Internet Archive not to crawl a particular site in the future).

42. See Bolin, *supra* note 4, at 30.

43. *Internet Archive v. Shell*, 505 F. Supp. 2d 755, 760 (D. Colo. 2007).

44. Internet Archive, Internet Archives' Terms of Use, Privacy Policy, and Copyright Policy (Mar. 10, 2001), <http://www.archive.org/about/terms.php>.

45. *Id.*

46. *Id.* See 17 U.S.C. § 512 (2006).

47. BROCKMAN, *supra* note 28.

*mer & Frailey*⁴⁸ as a co-defendant for breach of contract, breach of fiduciary duty, and negligent misrepresentation based on a failure to perform its duty of blocking access to the plaintiff's website via the Wayback Machine. The plaintiff had installed a robots.txt file on its website after the Internet Archive's web crawlers copied their website.⁴⁹ When it requests a website, the Archive's server automatically checks to see if the site has a robots.txt file; if it does, the Archive's servers are not to display the page.⁵⁰ However, in this case, the Archive's servers malfunctioned allowing the website to be viewed by the co-defendant law firm.⁵¹

Healthcare Advocates joined the Internet Archive in the suit for its failure to adequately secure and to remedy at least twelve separate acts of unauthorized access to Healthcare Advocate's website. Plaintiff alleged this failure constituted violations of the Digital Millennium Copyright Act, breach of contract, promissory estoppel, breach of fiduciary duty, negligent dispossession, and negligent misrepresentation.⁵² Plaintiff's breach of contract claim rested on the theory that the Internet Archive had an agreement with Healthcare Advocates to block archived historical content of the plaintiff's website through plaintiff's inclusion of a denial text string in the robots.txt file on the computer hosting the www.healthcareadvocates.com website.⁵³ The court dismissed Internet Archive early, after the parties' stipulation, leaving the issue of whether the Archive had breached a contract undecided.⁵⁴

More recently, in *Internet Archive v. Shell*, a pro se litigant sued the Internet Archive. She alleged that its copying of her website breached a contract contained therein.⁵⁵ The dispute arose when Suzanne Shell discovered the Wayback Machine had reproduced and archived the contents of her website approximately eighty-seven times between May 1999 and October 2004 and displayed the contents publicly.⁵⁶ Shell's website was devoted to providing information and services to individuals accused of child neglect and abuse and was registered with the U.S. Copyright Of-

48. 497 F. Supp. 2d 627 (E.D. Pa. 2007).

49. *Id.* at 632.

50. *Id.*

51. *Id.*

52. Second Amended Complaint for Violations of the Digital Millennium Copyright Act, Copyright Infringement, Violations of the Computer Fraud and Abuse Act, and Related Violations of Common Law at ¶¶ 118-53, *Healthcare Advocates, Inc. v. Harding, Earley, Follmer & Frailey*, No. 05-3524, 2006 WL 1784161 (E.D. Pa. May 26, 2006).

53. *Id.* at ¶¶ 117-25.

54. *Healthcare Advocates*, 497 F. Supp. 2d at 633.

55. *Internet Archive v. Shell*, 505 F. Supp. 2d 755, 764-66 (D. Colo. 2007).

56. *Id.* at 761.

file.⁵⁷ The site's terms of use deemed any copying of its contents as entering into a contract.⁵⁸ On December 12, 2005, Shell requested that her website contents be removed from the Wayback Machine, and the Archive complied.⁵⁹ Shell also demanded payment of \$100,000 and threatened to sue if the Archive failed to pay.⁶⁰

On January 20, 2006, the Internet Archive filed a declaratory judgment action in the U.S. District Court of the Northern District of California seeking a judicial determination that it did not violate Shell's copyrights.⁶¹ In Shell's answer she counterclaimed copyright infringement, conversion, civil theft, breach of contract, and violations of the Racketeer Influenced and Corrupt Organizations Act (RICO) and the Colorado Organized Crime Control Act (COCCA), and added members of the Archive's board of directors as third party defendants for the racketeering claims.⁶² The parties stipulated to transfer the case to the District of Colorado where the Archive moved to dismiss Shell's claims for conversion, civil theft, breach of contract, and RICO for failure to state a claim under Federal Rule of Civil Procedure 12(b)(6).⁶³ On February 13, 2007, Chief Judge Babcock dismissed all the counterclaims other than Shell's breach of contract and copyright claims.⁶⁴

On the motion to dismiss the court discussed Shell's counterclaim for breach of contract. Shell argued that copying her site constituted an acceptance of the site's terms of use and that the Internet Archive breached that contract when it failed to pay her the prescribed fees. The Archive responded that it never entered into a contract with Shell and that Shell failed to state a claim for breach of contract because the Archive only learned of Shell's terms of use after it copied the information.⁶⁵ The Archive further argued that while its automated web browser accessed Shell's website multiple times, it was not in fact aware of the terms of use, and Shell had not alleged that it or a human being at the Internet Archive was

57. *Id.* at 760.

58. The website contained the following notice: "If you copy or distribute anything on this site-you are entering into a contract. Read the contract before you copy or distribute. your act of copying and/or distributing objectively and expressly indicates your agreement to and acceptance of the following terms." *Id.* at 760 (capitalization omitted).

59. *Id.* at 762.

60. *Id.* at 761.

61. *Id.*

62. *Id.*

63. *Id.* at 761.

64. *Id.* at 770. The court's analysis of any of claims other than breach of contract or copyright is not the subject of this Note.

65. *Id.* at 764.

aware of the terms of use.⁶⁶ Absent such actual knowledge, the Archive argued, there cannot be a contract.⁶⁷

The court denied the Internet Archive's motion to dismiss on breach of contract grounds, holding that the counterclaim stated a viable cause of action for breach of contract and that Internet Archive's argument required a factual determination premature for this stage of the litigation. The court could not determine if the Archive knew at the time it copied defendant's site that doing so constituted acceptance of the site's terms of use.⁶⁸ The court noted that the record before it was unclear both as to the manner in which the terms of use were displayed and whether the Archive had received any notice of them given that it only accessed the site via an automated web crawler. Because Shell's complaint stated that notice of the contract terms "is published on every page of the website," the court noted that a factual investigation needed to be made on the location of the terms, "how a user reaches [them], and when a user becomes aware of [their] existence."⁶⁹ Moreover, the court stated that "while Internet Archive [might] be correct that the absence of human consent to this contract dooms Shell's claims, Shell ha[d] not had the opportunity to develop a factual record on this point."⁷⁰

The Internet Archive further argued that the U.S. Copyright Act preempted Shell's claim of breach of contract.⁷¹ The court, agreeing with the majority opinion among circuit courts, found that contracts requiring payment for use of copyrighted material protect rights beyond those protected by the Copyright Act; therefore, the Copyright Act does not preempt Shell's claim for breach of contract.⁷²

4. *Unanswered Questions and Potential Implications*

While this case did little to resolve the legally ambiguous environment in which digital archives operate, it does highlight several of their biggest legal risks. First, the Internet Archive did not challenge the sufficiency of Shell's copyright infringement claim, postponing instead for an uncertain fair use battle at a later stage of litigation. In addition, without a clear ruling on the breach of contract claim this case may have opened a Pandora's

66. *Id.* at 765.

67. *Id.*

68. *Id.*

69. *Id.*

70. *Id.* at 765-66.

71. *Id.* at 766-67; see 17 U.S.C. § 301 (2006).

72. *Id.* at 767. The Internet Archive also pled unconscionability, but the court noted that: "[A]bsent further factual findings, I cannot conclude that [the alleged] contract is unconscionable." *Id.*

box. After *Shell*, website owners can creatively draft contracts to stop any and all copying of their websites and then extract large fees by credibly threatening to take archivists through discovery, thereby increasing archivists' incentive to settle instead of fight. This possibility creates a new kind of copyright troll.⁷³ Without any clear ruling on the contract and copyright issues presented in *Shell*, more litigation will follow and will affect how these nonprofit digital archives operate in the future.

II. LEGAL RISKS OF ARCHIVES

"Kitchen sink" approaches to litigation, like *Shell*'s, where a complainant brings any claim or counterclaim against the respondent that arguably fits the facts, have a significant impact on the presence and permanence of archives. Although courts will throw out many of the claims at an early stage, archives still must expend considerable time and resources to defend themselves, exposing their vulnerabilities in the process. Even when the complainant's principal grievance is for breach of contract or copyright infringement, archives face and defend a multitude of claims such as RICO, wire fraud, trespass, conversion, and computer fraud and abuse. Until the uncertainty surrounding these major risks is dealt with more conclusively either by the courts or the legislature, archives face uncertain and burdensome litigation.

A. Copyright Liability

Copyright concerns are the most serious problem facing digital libraries.⁷⁴ Digital preservation resides in a murky legal gray area because of the need to copy digital information (one of the exclusive rights of the copyright owner) to preserve it. In addition, for an archive to be complete to the point of being socially beneficial there is a great interest in preserving in-copyright works without having to get the approval of each individual copyright holder, an expensive and time-consuming endeavor. Therefore, copyright law dominates the digital archive discussion.

Moreover, the proposed Google Library Project settlement and the lack of circuit level or Supreme Court rulings on similar digital archive

73. See *Field v. Google*, 412 F.Supp.2d 1106, 1123 (D. Nev. 2006) (finding bad faith on part of Plaintiff Field for taking affirmative steps to get his works included in Google's search results, where he knew they would be displayed with "cached" links to Google's archival copy, and deliberately ignoring protocols that would have prevented the caching).

74. See Travis, *supra* note 16, at 785 ("[A] government panel found that copyright was the 'single most significant barrier to preserving our cultural heritage' in digital libraries.").

matters has possibly made the question of whether a digital archive's copying constitutes fair use more uncertain. Whether digital archiving constitutes a fair use is an unsettled question, and Google's plan to make a Books Rights Registry and to make money from its endeavor may have shifted the analysis. The Google litigation could have settled the fair use issue, for better or worse. Instead, the settlement has left archives in the dark.

1. *Copyright Infringement*

The Copyright Act affords protection to "original works of authorship fixed in any tangible medium of expression."⁷⁵ Works published on the Internet are fully protected and subject to the same qualifications and limitations as non-digital works.⁷⁶ Digital works are "fixed" if they can be perceived, reproduced, or otherwise communicated for more than a transitory duration.⁷⁷ The copyright owner of a website or web content has the same exclusive rights under the Copyright Act as copyright holders of non-digital works.⁷⁸

Under a strict interpretation of copyright law, archiving meets the threshold of copyright infringement and therefore makes archivists liable without a defense. Archiving violates three exclusive rights of copyright owners: the right to reproduction,⁷⁹ the right to distribution,⁸⁰ and the right to display.⁸¹ First, archives necessarily make copies of each new webpage as their software crawls the Internet. This step is essential to the project of

75. 17 U.S.C. § 102 (2006).

76. They are still required to meet the subject matter requirements of § 102, must be fixed in a tangible medium, and be original works of authorship.

77. Whether fixation on a computer hard disk or random access memory (RAM) is enough is now controversial. *Compare* *Triad Sys. v. SoutheasterExpress Co.*, 64 F.3d 1330, 1333 (9th Cir. 1995) (granting preliminary injunction when defendant copied software into RAM of computer) *and* *MAI Sys. v. Peak Computers*, 991 F.2d 511, 518 (9th Cir. 1993) (finding that the "representation created in the RAM is 'sufficiently permanent or stable to permit it to be perceived, reproduced, or otherwise communicated for a period of more than transitory duration'"), *with* *The Cartoon Network L.P. v. CSC Holdings, Inc.*, 536 F.3d 121, 129-30 (2d Cir. 2008) (finding that the cable television company's embodiments of copyrighted television programs and movies in data buffers under 1.2 seconds did not last for a period of more than a transitory duration and therefore were not "fixed") *and* *CoStar Group, Inc. v. LoopNet, Inc.*, 373 F.3d 544, 551 (4th Cir. 2004) (holding information and data downloaded onto a user's RAM are not "fixed" because they are for no more than a transitory duration).

78. Those rights include: the right of reproduction (i.e. copying), the right to display, the right to prepare derivative works, and the right to distribute. 17 U.S.C. § 106 (2006).

79. § 106(1).

80. § 106(3).

81. § 106(5).

preservation. Second, archives distribute the copied web content when they make the material available via their website. Lastly, archives violate the exclusive right to display copyrighted material when they make the pages available on a website that is open to the public.⁸² As a result, digital archives like the Internet Archive will be found liable for copyright infringement without an exception or defense.

There are statutory exceptions to the exclusive rights of copyright holders like the library exception⁸³ or the DMCA safe harbor for ISPs.⁸⁴ Unfortunately, there are currently no similar exceptions for digital preservation by archives and their strongest defense, fair use,⁸⁵ is unpredictable, fact intensive, and uncertain.

2. *Library Exception*

The “library exception” of section 108 of the Copyright Act⁸⁶ is a narrow limitation on the exclusive rights of copyright holders. The primary purpose of the library provisions is to promote access to copyrighted works and reinforce preservation, while safeguarding against the commercial sale of works being supplanted by copying.⁸⁷

While the library exception allows for preservation, flexibility, and access to knowledge by the public, it also reflects concerns about the unauthorized commercial exploitation of copyrighted works and disruption of markets. Under section 108, a library may make a maximum of only three copies of a published work to replace a damaged, deteriorating, lost, or stolen copy, or if the existing format of the work becomes obsolete. Any copies must be only for the library’s own use.⁸⁸ Libraries are only allowed isolated and unrelated reproduction or distribution of a single copy of copyrighted materials for patrons requesting library materials.⁸⁹ Public libraries and archives are exempted from liability for the reproduction or distribution of a single copy of work, as long as the reproduction or distribution is not for commercial advantage, the collections are publicly

82. See H.R. REP. 94-1476 at 64 (1976) (“[D]isplay’ would include . . . the showing of an image on a cathode ray tube, or similar viewing apparatus connected with any sort of information storage and retrieval system.”).

83. 17 U.S.C. § 108 (2006).

84. 17 U.S.C. § 512 (2006).

85. 17 U.S.C. § 107 (2006).

86. § 108.

87. See Menell, *supra* note 9, at 1034-35 (explaining that the provisions “augment the general fair use privilege and afford libraries greater leeway in copying and distributing copyrighted works”).

88. § 108(c).

89. § 108(g).

available, and a notice of copyright is included in the reproduction or distribution of the work.⁹⁰ Furthermore, nonprofit libraries, archives, and educational institutions are exempted from liability for circumventing technological protection measures to the extent necessary to determine whether to add copyrighted works to their collections.⁹¹

Digital archives do not fall under the section 108 library exception. First, the exception does not apply to material the library or archive does not own. Second, under section 108 a library cannot distribute digital copies or make them available to patrons outside the library premises.⁹² Furthermore, although the DMCA allows for the digital preservation of copyrighted works, it states that pure digital libraries and archives that exist only on the Internet are not part of the library exception. The legislative history clearly shows congressional intent not to extend the library exception to libraries and archives existing wholly on the Internet. The Senate Judiciary Committee stated:

Although online interactive digital networks have since given birth to online digital 'libraries' and 'archives' that exist only in the visual (rather than physical) sense on websites, bulletin board and homepages across the Internet, it is not the Committee's intent that [17 U.S.C. § 108] as revised apply to such collections of information. The ease with which such sites are established online literally allows anyone to create his or her own digital 'library' or 'archives.' The extension of the application of section 108 to all such sites would be tantamount to creating an exception to the exclusive rights of copyright holders that would permit any person who has an online website, bulletin board or a homepage to freely reproduce and distribute copyrighted works. Such an exemption would swallow the general rule and severely impair the copyright owners' right and ability to commercially exploit their copyrighted works.⁹³

90. § 108(a).

91. Digital Millennium Copyright Act, § 103, 112 Stat. 2866 (codified at 17 U.S.C. § 1201(d) (2006)).

92. See § 108(b)-(c) (2006). However, the exception allows libraries and archives to reproduce, distribute, display, or perform in digital form a copy of a copyrighted work during the last twenty years of any term of copyright for purposes of preservation, scholarship, or research as long the work is not still being commercially exploited, the work can not be obtained at a reasonable price, or the copyright holder does not provide notice that one of the above conditions applies. § 108 (h)(1).

93. S. REP. NO. 105-190 at 62 (1998).

However, although the rigid library exception does not apply to digital archives' activities, an archive's actions in creating the archive could arguably fall under the fair use defense of the Copyright Act.⁹⁴

3. Fair Use

Digital archives' best chance to refute copyright infringement is the unpredictable fair use doctrine, which provides the best defense for digital archives, but is also the principal impediment to archiving projects.⁹⁵ Some scholars and commentators applying the fair use defense to digital archives believe that courts are likely to find fair use.⁹⁶ However, they fail to take into consideration the difficulty, fact intensity, and unpredictability of the doctrine that must be proven for each specific act of infringement in each lawsuit brought against a digital archive.⁹⁷

Section 107 of the Copyright Act "permits courts to avoid rigid application of the copyright statute when, on occasion, it would stifle the very creativity which that law is designed to foster."⁹⁸ The fair use defense creates a balance between exclusionary rights and free access, between social benefits and costs, and between the rights of authors to promote creative production and a democratic society's need for access to information and a free flow of ideas, information, and commerce.⁹⁹ The Copyright

94. § 108(f)(4) ("Nothing in this section . . . in any way affects the right of fair use as provided by section 107 . . ."). For a discussion of the fair use doctrine and its application to the activities of digital archives, see *infra*, Section II.A.3.

95. See Menell, *supra* note 9, at 1015-16 (discussing the instability of the fair use defense as an impediment to Google's Book Search Project).

96. See, e.g., Jines-Storey, *supra* note 4, at 1048-56 (finding fair use for the Internet Archive because there has been no demonstrable effect on the market by its copying of publicly accessible webpages).

97. See Bolin, *supra* note 4, at 25 (2006) ("Fair use is a doctrine too unstable to rely on when verbatim copying millions of webpages a day without consent."); see also Patel, *supra* note 4, at 424-26 (discussing the fact specific nature of the fair use defense); Travis, *supra* note 16, at 814, 816 (finding that the fair use doctrine has little to offer digital libraries because "courts have eviscerated it" post-*Sony*); Posting of Siva Vaidhyanathan to The Googlization of Everything Blog, My Initial Take on the Google-Publishers Settlement, http://www.googlizationofeverything.com/2008/10/my_initial_take_on_the_googlep.php (Oct. 28, 2008 7:21 EST) ("Fair use in the digital world is just as murky and unpredictable (not to mention unfair and useless) as it was yesterday.").

98. *Dr. Seuss Enters., L.P. v. Penguin Books USA, Inc.*, 109 F.3d 1394, 1399 (9th Cir. 1997) (citing *Iowa State Univ. Research Found., Inc. v. Am. Broadcasting Cos.*, 621 F.2d 57, 60 (2d Cir.1980)).

99. See *Harper & Row, Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 545-46 (1985) (balancing copyright's rewards to the individual against the interests of the public and arguing that "copyright is intended to increase and not to impede the harvest of

Act provides that the fair use of a copyrighted work for the purposes of criticism, comment, news reporting, teaching, scholarship, or research, is not an infringement¹⁰⁰ because copying of the works for those particular purposes is favored over protection of intellectual property.

Congress has identified four nonexclusive factors as especially relevant in determining fair use.¹⁰¹ In determining whether the use made of a work in any particular case is a fair use the factors to be considered include: the purpose and character of the use, including whether such use is for commercial purposes or is for nonprofit educational purpose and whether the use is transformative; the nature of the copyrighted work; the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and the effect of the use on the potential market or value of the original copyrighted work.¹⁰² There is no “rigid bright-line approach to fair use” and all four factors must be analyzed individually, in light of the purposes of copyright.¹⁰³

There remains a legal uncertainty as to whether courts would consider the copying and display of digital archives fair use.¹⁰⁴ This Note does not provide a detailed and thorough analysis of the fair use defense, but rather shows the unpredictable and fact-intensive nature of the imperfect fair use defense as applied to digital archives. The second and third factors do not seem to argue for or against a finding of fair use; however, the first and

knowledge”); *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 429 (1984) (“Because this task involves a difficult balance between the interests of authors and inventors in the control and exploitation of their writings and discoveries on the one hand, and society’s competing interest in the free flow of ideas, information, and commerce on the other hand, our patent and copyright statutes have been amended repeatedly.”); Orrin G. Hatch, *Toward a Principled Approach to Copyright Legislation at the Turn of the Millennium*, 59 U. PITT. L. REV. 719, 723 (1998) (“[C]opyright rights should be protected, unless it can be shown that the extent of protection is hampering creativity or the wide dissemination of works.”); Marybeth Peters, *Copyright Enters the Public Domain: The 33rd Donald C. Brace Memorial Lecture Delivered at New York University School of Law on Apr. 29, 2004*, 51 J. COPYRIGHT SOC’Y USA 701, 722 (2004) (“Striking the balance between meeting consumer expectations and limiting harmful copying and distribution is the key to preserving copyright’s standing in the eyes of the public.”).

100. 17 U.S.C. § 107 (2006).

101. *Id.*

102. *Id.*; see also *Harper*, 471 U.S. at 539-40.

103. *Campbell v. Acuff-Rose Music*, 510 U.S. 569, 584-85 (1994).

104. There does not seem to be a serious unauthorized distribution question for digital archives unless they are copying the digital content and then freely giving it to someone else. Whether they can be liable for contributory copyright infringement liability for their patrons’ use of the copyrighted works on the Wayback Machine is not within the scope of this Note.

fourth factors tend to be the most uncertain and raise the most interesting questions.

a) The Purpose and Character of the Infringing Work

The first factor of the fair use analysis is likely the most persuasive and most uncertain of the four factors when analyzing liability for digital archives. It considers a variety of elements including: whether the work is for commercial or nonprofit purposes; whether it fits in one of the categories laid out in statute (e.g., criticism, comment, news reporting, teaching, scholarship, or research or some other socially beneficial purpose); and whether the use is transformative.

The Internet Archive and other similar digital archives and libraries provide services available to the public and increase the availability of content published on the Internet, serving the policies of access and social benefit that the fair use doctrine was designed to promote. The Internet Archive in particular is a nonprofit organization with the goals of supporting research and scholarship. The funding for the archive comes from donations from foundations and the general public, and the website contains no commercial advertisements for personal profit.¹⁰⁵ Furthermore, the Internet Archive does not receive revenue from its users or by exploiting the information.

However, it is questionable whether a project such as a digital archive that employs web crawlers to systematically copy and display copyrighted content could be considered transformative. As the Supreme Court wrote in *Cambell v. Acuff-Rose*:

Although . . . transformative use is not absolutely necessary for a finding of fair use, the goal of copyright, to promote science and the arts, is generally furthered by the creation of transformative works. Such works lie at the heart of the fair use doctrine's guarantee of breathing space within the confines of copyright.¹⁰⁶

A transformative use "adds something new, with a further purpose or different character, altering the first with new expression, meaning or message."¹⁰⁷ The transformative nature of archives is closely analogous to cases in which search engines that indexed websites and included thumb-

105. See Internet Archive, Homepage, <http://www.archive.org/index.php> (last visited Dec. 23, 2008); Internet Archive, Donate, <http://www.archive.org/donate/index.php> (last visited Dec. 23, 2008).

106. *Campbell*, 510 U.S. at 579.

107. *Id.*

nails were found to be transformative fair use.¹⁰⁸ Indices with thumbnails were found to be fair use in part because use of the copyrighted works improved access to information on the Internet and provided a new electronic reference tool. Archives might also enjoy protection because they too put the web content to a new use by archiving and displaying the websites for the purpose of historic preservation. Technology such as that employed by the Wayback Machine allows visitors to see how the Web looked in the past and how it has evolved. Moreover, by serving as a record of a website on any particular day it also enables older websites to be used as evidence.

On the other hand, courts may be packing too much into the word “transformative.”¹⁰⁹ Digital archives, such as the Internet Archive, are simply copying the entire World Wide Web and the copyrighted works it contains and putting it somewhere else to be searchable in a different way at a different time. This arguably is not changing the nature of the work

108. In *Kelly v. Arriba Soft Corp.*, 336 F.3d 811 (9th Cir. 2003), the Ninth Circuit ruled on whether Arriba Soft had violated copyright law without a fair use defense in the use of thumbnail pictures and inline linking from Kelly’s website in Arriba’s image search engine. The court found fair use because the thumbnail images as previews were sufficiently transformative and the creation of the thumbnails did not substantially harm the market for the original photographs and may even have increased the public’s exposure to the original works. The court held that by improving access to information on the Internet, Arriba Soft created a new use for the works. *Arriba Soft*, 336 F.3d at 818-19, 821-22. See also *Worldwide Church of God v. Phila. Church of God, Inc.*, 227 F.3d 1110 (9th Cir. 2000) (holding that “where the use is for the same intrinsic purpose as [the copyright holder’s] . . . such use seriously weakens a claimed fair use”); *Núñez v. Caribbean Int’l News Corp.*, 235 F.3d 18 (1st Cir. 2000) (holding that copying a photograph intended to be used in a modeling portfolio and instead using it in a news article was transformative because it created a new meaning or purpose for the work).

Perfect 10 v. Google is the most recent case to apply the fair use defense to search engines. *Perfect 10 v. Google*, 416 F. Supp. 2d 828 (C.D.Cal. Feb 17, 2006), *aff’d in part, rev’d in part and remanded by Perfect 10, Inc. v. Amazon.com, Inc.*, 487 F.3d 701 (9th Cir. 2007), *opinion amended on rehearing by* 508 F.3d 1146 (9th Cir. 2007). Relying on *Arriba Soft* the court found fair use of photographs based on the significantly transformative nature of Google’s search engine and in light of its public benefit and only hypothetical harm to the market for the original work. The court found that the search engine transforms the copyrighted image into a “pointer” directing a user to a source of information. *Id.* at 721. Moreover, a search engine provides a social benefit by incorporating an original work into a new work, an electronic reference tool, to serve a completely different purpose. *Id.* *Perfect 10* was a huge win for Google and proponents of a wide application of the fair use doctrine because it held that even making an exact copy of a work may be transformative so long as the copy serves as different function than the original work.

109. For a discussion of the meaning of the term “transformative,” see Matt Williams, *Recent Second Circuit Opinions Indicate That Google’s Library Project Is Not Transformative*, 25 CARDOZO ARTS & ENT. L.J. 303, 305-07 (2007) (arguing that a transformative use involves “the creation of original expression that contains commentary”).

because the content remains the same. The copying of a website to be viewed in a different database at a different time is analogous to copying a song to include in a music collection to be heard at a later time such as in *UMG Recordings, Inc. v. MP3.Com, Inc.*¹¹⁰ Defendant MP3.com argued that its use was a transformative “space shift” because subscribers can enjoy their music on the computer without having to lug around the physical disks themselves.¹¹¹ The court rejected this argument finding that the unauthorized copies being retransmitted in another medium was an insufficient basis for any legitimate claim of transformation.¹¹² The song was not being transformed, but simply being moved and placed in a different collection. Furthermore, the archive may be archiving already archived webpages, which would clearly not be transformative because the Internet Archive would not be adding anything new. The Archive allows for public availability of websites long gone; however, the unavailability of a copyrighted work does not make it legally acceptable to copy. Further, providing all the webpages available in a unified source is not an acceptable transformative use or all not-for-profit public libraries could copy any book they wanted for their shelves without paying based on a beneficial unified source theory.¹¹³

Only a few courts have considered whether archiving should be considered transformative. In *Texaco*, the Second Circuit, found the copying of

110. 92 F. Supp. 2d 349 (S.D.N.Y. 2000).

111. *Id.* at 351.

112. *Id.*; see also *Am. Geophysical Union v. Texaco Inc.*, 60 F.3d 913, 923 (2d Cir. 1994), *cert. dismissed*, 516 U.S. 1005 (1995) (“[A]n untransformed copy is likely to be used simply for the same intrinsic purpose as the original, thereby providing limited justification for a finding of fair use.”); *Basic Books, Inc. v. Kinko’s Graphics Corp.*, 758 F. Supp. 1522, 1530 (S.D.N.Y. 1991) (quoting District Court Judge Leval that to be transformative “[t]he use . . . must employ the quoted matter in a different manner or for a different purpose from the original”).

113. *But see* Lloyd L. Weinreb, *Fair’s Fair: A Comment on the Fair Use Doctrine*, 103 HARV. L. REV. 1137, 1143 (1990) (arguing that the usefulness of “transformative” value as a criterion is overstated and posits that “[a] use may serve an important, socially useful purpose without being transformative, simply by making the copied material available”). In an example closely analogous to digital archives’ purposes of preservation and access to knowledge, he further states:

One may wonder whether the publication of material in a new “package” may not itself constitute a transformative use. For example, the publication of a volume of Salinger’s letters would have a purpose entirely different from that which prompted Salinger to write and send the letters contained in the volume. Preparing the collection involves effort and, perhaps, judgment of a kind that often is enough to sustain a copyright.

Id. at 1143, n.29.

scientific articles for archival purposes was not a fair use, despite the benefit of a more usable format.¹¹⁴ The court linked its analysis of the first factor to that of the fourth factor. Based on the fact that Texaco's scientists made copies of the articles and archived them, essentially creating personal libraries without paying for additional subscriptions or license fees, the court held that the first factor weighed against fair use. A court applying *Texaco* would be hard pressed to view archives more favorably because archiving essentially takes copyrighted content and places it in a more publicly accessible library without asking for permission or licensing the content. In the Ninth Circuit, the caching of thumbnail pictures in *Arriba* and *Perfect 10* was arguably like archiving because they were copying and storing large amounts of copyrighted material so that it could be publicly searchable. However, the use of the works went beyond merely storing the pictures and instead included new highly beneficial electronic reference tools. Therefore, it remains uncertain whether courts applying Ninth Circuit precedent will find that storing and displaying copyrighted content in a new database is transformative and whether the beneficial nature of a digital archive is enough to tip the balance in favor of fair use without a transformative finding.

b) Nature of Work and Amount and Substantiality Used

The second and third factors of the fair use determination, the nature of the work and the amount and substantiality of the copyrighted work used, are not dispositive in the fair use determination for digital archives.

In considering the nature of the work, courts look at whether the original work is published or unpublished, whether the right of first publication has been exercised, and whether the work is creative or factual in nature.¹¹⁵ In the case of the Internet Archive and similar digital archives, the material being copied has already been published and made publicly accessible at the time the information was made available on the Internet.¹¹⁶

114. 60 F.3d at 931.

115. *Harper & Row, Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 563-64 (1985).

116. There is also an argument that the copying and viewing of the webpages that were publicly accessible and free to view (and impliedly licensed to the user to copy and view) at a later date is analogous to broadcast television that was at issue in *Sony Corp. of America v. Universal City Studios, Inc.*, 464 U.S. 417 (1984), and different than other cases where the material content is closed. The public and the Internet Archive were impliedly invited to view the webpages as they existed on the date they were copied. Viewing the webpages at a later date, like recording and viewing movies at a later date (i.e., timeshifting), does not change the fact that the webpages were originally available for public viewing. See also *Field v. Google*, 412 F. Supp. 2d 1106, 1121 (D. Nev. 2006)

Furthermore, because the Internet Archive seeks to archive the entire Internet, the nature of the work is mixed between creative works and collection of facts and data. The various elements at play in this analysis, that the works are published but that some of the works may be creative and expressive in nature, make it difficult to determine whether courts would find for or against digital archives on this point. Notwithstanding a class action type-lawsuit, the outcome of this element would essentially depend on who is suing the archive and the nature of their work.

Next, courts consider the amount and substantiality of the portion used in relation to the copyrighted work as a whole.¹¹⁷ Digital archives like the Internet Archive create copies of the entire website, allowing users to browse the entire website as it would have appeared on that date. Qualitatively the heart of the work is copied, and quantitatively the whole of the work is copied. Moreover, all elements of expression are copied, such as the graphics, creative speech, word placement, etc. However, to serve its transformative purpose as a website archive, Internet Archive needs to copy as much of the original as possible. This analysis is analogous to the court's consideration of the copying for the use of parody; most if not all of the work is necessary for the transformative use to be successful and therefore this factor remains neutral.¹¹⁸

c) Effect on Potential Market for Original

The effect on the market for the original copyrighted work, generally considered the most important factor in the fair use analysis,¹¹⁹ is also uncertain. In determining the outcome of this factor, courts consider what effect the allegedly infringing use has on the current or potential markets for the original work.¹²⁰ The market is harmed if users will substitute the new work for the original¹²¹ or if there is "unrestricted and widespread conduct of the sort engaged in" that would result in a "substantially ad-

(finding this factor neutral because Field made his content available for free on his website and Google only used no more than necessary (even though it was all of it)).

117. *Harper*, 471 U.S. at 564.

118. See *Campbell v. Acuff-Rose*, 510 U.S. 569, 588-89 (1994). See also *Sony Corp.*, 464 U.S. at 449-50, for a case in which substantial copying—copying of entire programs for private viewing—was upheld as fair use. Likewise, see *Kelly v. Arriba Soft Corporation*, 336 F.3d 811, 820-21 (2003), where the Ninth Circuit held that copying an entire photo to use as a thumbnail in online search results did not weigh against fair use, "if the secondary user only copies as much as is necessary for his or her intended use."

119. *Sony*, 464 U.S. at 450-51.

120. *Harper*, 471 U.S. at 567.

121. *Arriba Soft*, 336 F.3d at 821.

verse impact” on the potential market for the original or its derivatives.¹²² A copyright holder need only establish with reasonable probability that the infringement resulted or will result in a loss of revenue.¹²³ The burden then shifts to the infringer to show that the damages would have occurred without the alleged infringement.¹²⁴

Whether the court finds that the copying has an effect on the market of the original work is closely linked to the purpose for which the copyrighted work has been copied. If the purpose is for research or scholarship, market effect may be difficult to prove.¹²⁵ As the Supreme Court stated in *Sony*, “[a] use that has no demonstrable effect upon the potential market for, or the value of, the copyrighted work need not be prohibited in order to protect the author’s incentive to create . . . The prohibition of such non-commercial uses would merely inhibit access to ideas without any countervailing benefit.”¹²⁶ Therefore, to rebut fair use, plaintiffs need to show proof either that the particular use is harmful or that if it should become widespread that it would adversely affect the potential market for the copyrighted work.¹²⁷

When applying the fourth factor to noncommercial digital archives like the Internet Archive courts could find there is no effect on the market. The Archive only copies publicly accessible webpages and displays them without charge to the public and without relying on advertisements or subscription fees to make revenue. Because the copyright holder made the work freely accessible in the first place, such as broadcast television in *Sony*,¹²⁸ and the Archive is not selling the works or making profit from the display, there seems to be no effect on the market.

On the other hand, courts could easily find a market effect when taking into account the archival nature of the service. Libraries also provide a public benefit but they cannot simply copy wholesale a book and place it on their shelves. Also, users might bypass paying for content from commercial sites by seeking older content through use of the Internet Archive’s services, which could result in the loss of advertising revenue and subscription fees for those commercial websites. For example, old arc-

122. *Campbell*, 510 U.S. at 590.

123. *Harper*, 471 U.S. at 567.

124. *Id.*

125. *But see Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 450 (1984) (“[E]ven copying for noncommercial purposes may impair the copyright holder’s ability to obtain the rewards that Congress intended him to have.”)

126. *Sony*, 464 U.S. at 450-51.

127. *Id.* at 451.

128. *Id.*

hives of many online newspapers cost money to view and databases such as LexisNexis charge fees to view their archives of court documents and laws. Users could circumvent paying these fees by finding the material in the Wayback Machine. Moreover, a complainant might demonstrate a loss of traffic and revenue from its site or that people are using the archives services as a replacement for original sites.

However, the Internet Archive could argue that it is only intended to be used when a past version, not a current version, of a site is sought, and can only be accessed through the Wayback Machine, where content less than six months old is not included. Furthermore, opt-out mechanisms fall in favor of the Internet Archive on the fourth fair use factor. Similar to the content owners in *Sony*,¹²⁹ many if not most copyright and website owners do not object to having their websites archived and therefore the best rule may be to require owners to opt-out of the archive project rather than requiring them to opt-in. Arguably a website owner cannot complain about the impact on the market when she could just opt-out of the project with minimal effort.¹³⁰ For example, the court in *Field v. Google*¹³¹ stated that because the complainant knew about the opt-out options of Google's web caching and failed to avail himself of those options, that knowledge formed the basis of an implied license.¹³² This could be a digital archive's best case against market effect in the fair use analysis but how it would fare in court when a web operator does not have actual knowledge remains to be seen.

129. *Id.* at 456 (finding that substantial numbers of copyright holders who licensed their works for broadcast on free television would not object to having their broadcasts time-shifted by private viewers).

130. There are a number of cases where immediate takedown precludes later relief. *See, e.g., Biosafe-One, Inc. v. Hawks*, No. 07 Civ. 6764(DC), 2007 WL 4212411 (S.D.N.Y. Nov. 29, 2007).

131. 412 F. Supp. 2d 1106, 1116-17 (D. Nev. 2006).

132. However, arguing that because people have the right to opt-out and therefore the archive project does not have an effect on the market of their original work begs the question of whether opt-out mechanisms are even legal, let alone good for archives. For a discussion on the legality of opt-out mechanisms, see Oren Bracha, *Standing Copyright Law on Its Head? The Googlization of Everything and the Many Faces of Property*, 85 TEX. L. REV. 1799 (2007). For an analysis on why digital archives should not have to rely on opt-out mechanisms to insulate themselves from liability, see Bolin, *supra* note 4 (discussing how right to withdraw under copyright "creat[es] a legal regime [that gives] authors the right to opt-out of history").

4. *Fair Use in Light of the Proposed Google Library Project Settlement*

The uncertainty over whether courts will find internet archiving a fair use is even more prominent after the proposed settlement of the Google Library Project litigation. The case, heralded as the one to resolve the fair use issue, has resulted in more uncertainty because the settlement was not on the merits.¹³³ Other individuals or entities who seek to do similar archival projects will face the same legal battles, without Google's power and wealth. Furthermore, Google's development of a market for book archiving might alter the market analysis because courts might be more likely to find potential markets where none exist or will likely exist, such as in the licensing of webpages to digital archives.

Authors and publishers brought suit against Google in 2005, claiming that Google violated their copyrights and those of other rightsholders of books and inserts, by scanning books from libraries, creating an electronic database of those books, and displaying short excerpts without the permission of the copyright holders.¹³⁴ Google denied all the claims and countered that digitizing these books and displaying only snippets of material was noninfringing or fair use of the material.¹³⁵

Under the settlement, which must be approved by a federal judge before it takes effect,¹³⁶ Google will pay authors and publishers \$125 million. Part of the settlement amount will be used to create a Book Rights Registry, which will allow copyright owners to register their works and receive a share of revenues of institutional subscriptions of books made available through the Google Book Search, as well as from sales of the books to online consumers. Another portion will go to resolving existing claims of rightsholders for books and inserts Google scanned without permission. Google will also allow users to purchase full books, which are saved to an "electronic bookshelf," offer institutional subscriptions, including free online portals for public and higher education libraries, and

133. *See generally* Settlement Agreement, *The Authors Guild, Inc., et al. v. Google, Inc.*, Case No. 05-cv-8136 (S.D.N.Y. Oct. 20, 2008).

134. Class Action Complaint, *The Authors Guild, Inc., et al. v. Google, Inc.*, Case No. 05-cv-8136 (S.D.N.Y. Sept. 20, 2005).

135. Answer, Jury Demand, and Affirmative Defenses of Defendant *Google Inc., The Authors Guild, Inc., et al. v. Google, Inc.*, Case No. 05-cv-8136 (S.D.N.Y. Nov. 30, 2005).

136. The fairness hearing will be held June 11, 2009 before Judge John E. Sprizzo. *Justia News & Commentary, The Author's Guild et al v. Google Inc., Docket #64*, <http://news.justia.com/cases/featured/new-york/nysdce/1:2005cv08136/273913/> (last updated Apr. 16, 2009).

continue to point users to buy or borrow the books found through their Book Search.¹³⁷

Some see promise in the Google Library Project settlement.¹³⁸ The proposed settlement will expand access to both in-print and out-of-print books through Google's Book Search project, preserving millions of books.¹³⁹ The settlement allows easy access for individuals anywhere in the United States to the collections of books included in the database.¹⁴⁰ Authors of out-of-print books win because they obtain otherwise unobtainable revenue from online purchase of older books which would not normally provide any royalties.¹⁴¹

Others do not share the same enthusiasm. For example, some fear the Google Library settlement raises a hornet's nest of concerns including privacy and antitrust issues and providing too many limits on access to culturally valuable copyrighted works.¹⁴² The settlement also leaves undecided

137. Settlement Agreement, *supra* note 133. For more information on the Google Library Project settlement, see Jonathan Band, *A Guide for the Perplexed: Libraries and the Google Library Project Settlement*, LLRX, Dec. 14, 2008, available at <http://www.llrx.com/features/googleprojectsettlement.htm>.

138. Posting by David Drummond, Senior Vice President, Corporate Development, and Chief Legal Officer of Google, to Google Blog, New Chapter for Google Book Search, <http://googleblog.blogspot.com/2008/10/new-chapter-for-google-book-search.html> (Oct. 28, 2008 7:14 EST).

139. *Id.*

140. *Universities See Promise in Google Book Search Settlement*, STANFORD REPORT, October 28, 2008, <http://news-service.stanford.edu/news/2008/october29/google-102908.html>.

141. *Google Revolution the End of the Publishing World?*, TODAY'S ZAMAN, Dec. 9, 2008, <http://www.todayszaman.com/tz-web/detaylar.do?load=detay&link=160437>.

142. See Fred von Lohmann, EFF Deeplinks Archive, Google Book Search Settlement: A Reader's Guide (Oct. 31, 2008), <http://www.eff.org/deeplinks/2008/10/google-books/settlement-readers-guide> (discussing user privacy questions opened by the settlement); Laura G. Mirviss, *Harvard-Google Online Book Deal at Risk*, THE HARVARD CRIMSON, Oct. 30, 2008, available at <http://www.thecrimson.com/article.aspx?ref=524989> (quoting University Library Director Robert C. Darnton that "the settlement contains too many potential limitations on access to and use of the books" and "[t]he settlement provides no assurance that the prices charged for access will be reasonable . . . especially since the subscription services have no real competitors [and] the scope of access to the digitized books is in various ways both limited and uncertain"); Posting by Neil Netanel to Balkinization Blog, Google Book Search Settlement, <http://balkin.blogspot.com/2008/10/google-book-search-settlement.html> (Oct. 28, 2008 9:02 EST) ("[T]he bottom line is that Google is left with a de facto monopoly over this 'universal library' service and . . . potential competitors face a higher barrier to entry than if Google had fought and prevailed on fair use (or if Congress enacts a statutory license for such uses)").

the issue of whether Google's scanning of entire books into a digital archive and its display of snippets is a fair use.

Many scholars believed that the courts would have ultimately held that it is a fair use, thus setting important precedent.¹⁴³ Now other library initiatives that do not have Google's wealth have to step up to fight or quit since most other preservation initiatives may not be able to bear the enormous legal costs of fighting so many copyright holders. Google has essentially carved out a special status for itself, with rights that others do not have and that libraries cannot afford to buy.

Moreover, because of Google's commercial plans, it may have made it more difficult to claim that such archival uses of copyrighted material do not harm a potential licensing market, an important aspect to a successful fair use argument. The most interesting part of the proposed settlement is the prospect of future revenues for both Google and the rightsholders. Rightsholders will receive a share of revenues of institutional subscriptions of books made available through the Google Book Search under the settlement, as well as from sales of the books by online consumers. They will also be paid for printouts at public libraries, as well as for other uses. Payments will flow through a Book Rights Registry, an ASCAP-like entity for writer's rights. The settlement therefore creates a new commercial opportunity for copyright holders to balance against fair use—if Google can create a licensing scheme, the archives have the potential too. Therefore, it is unclear whether the market effect factor will be bigger now after the settlement.

The uncertainty of the fair use doctrine surely had a hand in the bargaining of the settlement deal. Although the deal has yet to pass muster at the fairness hearing, in some ways its very existence suggests a change in fair use law for the worse. Google, as a private entity and not a library, has cast a large shadow over emerging not-for-profit institutions like the Internet Archive.

Surely, liability for copyright infringement without a defense or exception poses the biggest threat to digital archives and the Google settlement has left this legal area on shaky ground. But even if digital archives can get around the copyright hurdle, archives run straight into a conflict with website owners' increasing tendency to contract around the Copyright Act.

143. See generally Nari Na, *Testing the Boundaries of Copyright Protection: The Google Books Library Project and the Fair Use Doctrine*, 16 CORNELL J.L. & PUB. POL'Y 417 (2007); Rogers, *supra* note 3.

B. Liability Sounding in Contract

The uncertainty left after the *Shell* settlement¹⁴⁴ opened the doors to potential breach of contract liability for digital archives. The fact that even a beneficial nonprofit institution like the Internet Archive could not dispose of the *Shell* litigation at an early stage shows the uncertainty on the Web as to whether web crawlers, like the one employed by the Internet Archive, can be subject to breach of contract actions. Scholars and courts disagree on the legality of browsewrap agreements and whether an automated agent or an unsuspecting individual can fully and knowingly enter into a contract.¹⁴⁵ The cases to date seem to apply traditional contract rules of offer and acceptance to automated software programs, but even these cases are based on a very fact-specific inquiry.¹⁴⁶ Professor Mark Lemley offers some insight as to the way courts are ruling, but his article only highlights the unresolved issue of whether a court would hold a digital archive liable for a contract entered into by its universal web crawler.¹⁴⁷

1. *Electronic Contracting Cases*

In an era of digital content, copyright owners have increasingly turned away from the express statutory limits on their rights contained in the Copyright Act and instead have invoked the institution of contract law. These contracts proliferate on the web and govern most consumer transactions. More and more these non-negotiated contracts include restrictions on fair use, and most courts and commentators have rejected preemption as the appropriate tool for challenging these provisions.

Owners of websites often post website disclaimers and agreements called “user agreements,” “terms of use,” or “terms of service,” that establish rules for access to their websites and/or use of the content included on the website. Some agreements simply notify the users that the website publisher holds the copyright in the content on the site and inform the users that they may only use the material for personal, noncommercial purposes and cannot reproduce or distribute without the website owner’s permission. Essentially, these agreements simply restate their rights under the Copyright Act. More prevalent are contracts that have terms prohibiting certain uses or actions that include hefty fines for breach and often contain forum selection and dispute resolution clauses. These electronic

144. See *supra* Sections I.C.3-4 (discussing *Internet Archive v. Shell* and undecided breach of contract issues).

145. See *infra* Section II.B.1.

146. See *infra* Section II.B.2.

147. Mark Lemley, *Terms of Use*, 91 MINN. L. REV. 459 (2006).

contracts have often been held to be legally binding by the courts even when website users did not know about the contract terms.¹⁴⁸

Browsewrap contracts, such as the disputed contract in *Shell*, are formed by a user's action of simply visiting or viewing a website. Users enter into the contract without taking any affirmative action before the website performs its end of the contract and are generally entered into implicitly through continued use of the site or use of the site's contents. While this is still a developing area of law and the subject of much legal debate, courts have held browsewrap agreements to be valid as long as there is notice (i.e., the online agreement is conspicuous to website users).¹⁴⁹

2. *Can Electronic Agents Enter into Legally Binding Contracts?*

Today, possibly due to the acceptance of new technology, courts are more likely to find browsewrap agreements binding on automated users of websites, especially in the commercial context. In *Ticketmaster Corp. v. Tickets.com*, the U.S. District Court for the Central District of California upheld the validity of a browsewrap license entered into by a spider based solely on the evidence that the defendant knew of the license but neverthe-

148. See, e.g., *Feldman v. Google, Inc.*, 513 F. Supp. 2d 229 (E.D. Pa. 2007) (upholding forum selection clause in clickwrap agreement); *Hotmail Corp. v. Van\$ Money Pie, Inc.*, 47 U.S.P.Q.2d 1020 (N.D. Cal. 1998) (granting preliminary injunction for breach of contract for violating terms of service by using a Hotmail account for an improper use; contract upheld because clicking button after notice gave consent); *Caspi v. Microsoft Network*, 732 A.2d 528 (N.J. Super. Ct. App. Div. 1999) (finding forum selection clause contained in clickwrap membership agreement was enforceable because user had to click on an "I Agree" button). The history and legality of electronic contracting has been widely documented and analyzed. For background on electronic contracts, including shrinkwrap, clickwrap, and browsewrap licenses, see Ryan J. Casamiquela, *Contractual Assent and Enforceability in Cyberspace*, 17 BERKELEY TECH. L.J. 475 (2002); Rachel S. Conklin, *Be Careful What You Click For: An Analysis of Online Contracting*, 20 LOYOLA CONSUMER L. REV. 326 (2008); Christina L. Kunz et al., *Browse-Wrap Agreements: Validity of Implied Assent in Electronic Form Agreements*, 59 BUS. LAW. 279 (2003); Juliet M. Moringiello & William L. Reynolds, *Survey of the Law of Cyberspace: Electronic Contracting Cases 2006-2007*, 63 BUS. LAW. 219 (2007); Ian Rambarran & Robert Hunt, *Are Browse-Wrap Agreements All They Wrapped Up To Be*, 9 TUL. J. TECH. & INTELL. PROP. 173 (2007); Renée Zimurchyk, *Contractual Validity of End User License Agreements*, 11 APPEAL REV. CURRENT L. & L. REFORM 55 (2006).

149. See *Pollstar v. Gigmania Ltd.*, 170 F. Supp. 2d 974 (E.D. Cal. 2000) (refusing to enforce terms of online license agreement because the link to it was not sufficiently obvious; website provider placed a notice on its homepage that users would be bound by a license agreement reachable through a hyperlink if they continued on the site); *Specht v. Netscape Comm. Corp.*, 150 F. Supp. 2d 585 (S.D.N.Y. 2001) (finding arbitration provision contained in browsewrap license invalid because end users were able to download the free software without reviewing the license agreement).

less continued to send its spider to the plaintiff's interior webpages.¹⁵⁰ Likewise, in *Register.com, Inc. v. Verio, Inc.* the Second Circuit found that Verio's repeated access to and use of information from Register.com's WHOIS database evidenced consent to Register.com's terms of use, even though Verio used an automated robot to go to the website and did not see the terms of use until after it had completed using the database.¹⁵¹ In *Cairo, Inc. v. CrossMedia Servs., Inc.*, the Northern District of California similarly found that the terms of use found on CrossMedia's internet site were binding, given the continued use of the site.¹⁵² Even though defendant Cairo had no actual knowledge of the terms of use, its repeated access to plaintiff's site via a "robot" resulted in imputed knowledge to defendant even though the robot was not capable of, and did not, collect information on the contents of such terms.¹⁵³

While the majority of courts have yet to rule on whether an automated or electronic agent can enter into a legally binding contract and whether imputed knowledge under *Cairo* should be the best standard, the Uniform Electronic Transactions Act (UETA) and Electronic Signatures in Global and National Commerce Act (ESIGN) have attempted to deal with this issue. Under section 14 of the UETA, a contract may be formed by the interaction of electronic agents of the parties, even if no individual was aware of or reviewed the electronic agents' actions or the resulting terms and agreements.¹⁵⁴ Under section 101(h) of ESIGN, a contract may not be denied legal effect, validity, or enforceability solely because its formation, creation, or delivery involved the action of one or more electronic agents. However, this section allows the creation of contracts by electronic agents only so long as the action of the electronic agent is legally attributable to the person to be bound.¹⁵⁵ While courts following these statutes would likely find that an electronic agent entered into a legally binding contract, the statutes do not state that all contracts formed by bots are enforceable.

The *Shell* court did not discuss the issue of whether electronic agents have the authority to bind their principals to contracts and the electronic

150. *Ticketmaster Corp. v. Tickets.com, Inc.*, No. CV997654HLHVBKX, 2003 WL 21406289, at *2 (C.D. Cal. Mar. 7, 2003).

151. *Register.com v. Verio, Inc.*, 356 F.3d 393 (2d Cir. 2004).

152. *Cairo, Inc. v. Crossmedia Servs. Inc.*, No. C 04-04825 JW, 2005 WL 756610, at *5 (N.D. Cal. Apr. 1, 2005).

153. *Id.*

154. Uniform Electronic Transactions Act [UETA] § 14 (1999) (codified at 15 U.S.C. § 7002 *et seq.*).

155. Electronic Signatures in Global and National Commerce Act [ESIGN] § 7001(h), Pub.L. 106-229, 114 Stat. 464 (codified at 15 U.S.C. § 7001-31, Ch. 96 (2006)) (emphasis added).

agent issue has not received much judicial scrutiny. However, under the reasoning of *Verio* and *Cairo*, the Internet Archive most likely assented to the contract by visiting and copying Shell's website over eighty times over the course of four years. It appears in this case that there may have been adequate notice to put a reader on notice that an act of copying and distributing indicates agreement and acceptance of the terms.

However, unlike *Verio* and *Cairo*, where the defendants knew or might have known about the terms of use at some time while they were visiting the webpage, there is no evidence that the Internet Archive knew about the terms of use until after Shell contacted them. In fact, the Archive attempted to distinguish the facts from those in *Verio* by arguing that while its agents accessed Shell's website numerous times, no human being at the Internet Archive was aware of the terms.¹⁵⁶ The court recognized that the absence of human consent to this contract could doom Shell's claims but denied the argument on a motion to dismiss because Shell did not have the opportunity to conduct discovery on this point.¹⁵⁷

Professor Mark Lemley attempts to shed some light on this issue in his article "Terms of Use."¹⁵⁸ He distinguishes cases where courts found browsewrap agreements, including those entered into by automated agents, enforceable versus ones that were held not, finding that courts generally enforce agreements against sophisticated commercial entities who are repeat players and not against consumers.¹⁵⁹ It is an interesting question to determine where digital archives may fall in this continuum: as sophisticated companies that can enter into electronic contracts via an electronic agent, as consumers where browsewraps are generally not enforced, or as something else. If enforcement is limited to the context in which it has so far occurred, where do and should these preservation entities lie?

Accordingly, should there be a different standard for crawlers, such as the Internet Archive's or Google's that indiscriminately crawl the Web versus crawlers that target one or more particular sites? So far in the case law all browsing programs in question have been built to access one of a few particular websites, which were picked to some extent by their user.¹⁶⁰ For example, in *Ticketmaster*, defendant Tickets sent a spider to plaintiff

156. *Internet Archive v. Shell*, 505 F. Supp. 2d 755, 765-66 (2007).

157. *Id.*

158. Lemley, *supra* note 147.

159. *Id.* at 463, 472-77.

160. *See, e.g., Cairo, Inc. v. Crossmedia Servs. Inc.*, No. C 04-04825 JW, 2005 WL 756610, at *5 (N.D. Cal. Apr. 1, 2005); *Ticketmaster Corp. v. Tickets.com, Inc.*, No. CV997654HLHBKX, 2003 WL 21406289, at *2 (C.D. Cal. Mar. 7, 2003).

Ticketmaster's site to sweep it for particular information. Similarly, in *Cairo*, Cairo's robots crawled plaintiff's servers exclusively. Arguably, if the company set up the crawlers to crawl particular sites, it knew or should know the terms of use on the websites. But along that same line of reasoning, digital archives that have software that crawls the entire Internet should know that there will be terms of use on some of the websites and that they may be entering into them simply by visiting and viewing the sites.

3. *Intersection of Copyright and Contract: Copyright Preemption*

The liability of digital archives unknowingly entering into contracts while crawling the Web is exacerbated by the fact that the majority of courts find that federal copyright law does not preempt state contractual rights, even when the right is essentially the same right covered by the Copyright Act, as was the case in *Shell*.¹⁶¹ The fact specific and generally unwelcoming application of the federal preemption doctrine to state contract law places digital archives in a position where what they are essentially doing is copyright infringement with a potential fair use defense but are stuck litigating a contract that they cannot get around.

When it comes to the collision of exclusive rights under the Copyright Act and contractual rights, the majority of districts that have ruled on the issue have held that because contract rights depend on extra elements that distinguish them from exclusive rights under the Copyright Act¹⁶²—essentially the presence of a bargained-for exchange—contract claims are not preempted.¹⁶³ A minority of courts, however, reject this literal application of the extra-element test and will preempt a contract claim that alleges the extra element of a promise as long as the promise amounts only to a promise to refrain from doing one of the rights exclusively reserved for copyright owners in the Copyright Act (i.e., reproducing, performing, dis-

161. See discussion *infra* Section II.B.3.

162. State law claims are preempted by the Copyright Act when:

- (1) the particular work to which the claim is being applied falls within the type of works protected by the Copyright Act under 17 U.S.C. §§ 102 and 103, and
- (2) the claim seeks to vindicate legal or equitable rights that are equivalent to one of the bundle of exclusive rights already protected by copyright law under 17 U.S.C. § 106.

Briarpatch Ltd., L.P. v. Phoenix Pictures, Inc., 373 F.3d 296, 305 (2d Cir. 2004).

163. GOLDSTEIN ON COPYRIGHT 17.2.1.2 (3rd ed. 2007) (citations omitted); see also *ProCD, Inc. v. Zeidenberg*, 86 F.3d 1447, 1454 (7th Cir. 1996); *Nat'l Car Rental Sys., Inc. v. Computer Assocs. Int'l, Inc.*, 991 F.3d 426, 433 (8th Cir. 1993), *cert. denied*, 510 U.S. 861 (1993); *Taquino v. Teledyne Monarch Rubber*, 893 F.2d 1488, 1501 (5th Cir. 1990); *Acorn Structures, Inc. v. Swantz*, 846 F.2d 923, 926 (4th Cir. 1988).

tributing, or displaying the work).¹⁶⁴ Whether a browsewrap agreement that alleges no more than rights contained under the Copyright Act passes the extra-elements test depends on whether courts determine whether the browsewrap contains a bargained-for exchange.¹⁶⁵

Copyright scholars have argued that contractual licenses that contract away statutory rights should not be enforced.¹⁶⁶ Browsewrap agreements hurt innovation, fair use, and copyright policy by restricting use even in cases, such as here, where the challenged behavior is potentially legal under fair use. Moreover, they exceed and misuse the rights given to them under the Copyright Act by charging excessive fees and creating a monopoly unintended by our founding fathers.

Although courts and scholars disagree over whether copyright law should preempt contract law, they generally recognize that copyright law needs a remedy for contractual overreaching such as in the *Shell* litigation.¹⁶⁷ Currently, no test for preemption accommodates the interest in private contracting and the interests promoted by the Copyright Act. Scholars have proposed new models and methods for addressing the copyright preemption problem including using copyright misuse as a defense where the copyright owner uses a contract to expand their copyright “monopoly” by its scope under the Copyright Act¹⁶⁸ and refining the preemption analysis such as by drawing on law more related to contractual waivers of statutory rights.¹⁶⁹ However, most of the alternatives are after-the-fact solutions. Digital archives and search engines will still be subject to litigation

164. See *Wrench LLC v. Taco Bell Corp.*, 256 F.3d 446, 457 (6th Cir. 2001).

165. The extra-element test has been criticized heavily for its lack of guidance to the courts. Commentators believe that courts first decide independently whether they are going to apply preemption and then use the extra elements test to reach that conclusion. Schuyler Moore, *Straightening Out Copyright Preemption*, 9 UCLA ENT. L. REV. 201, 204 (2002).

166. See, e.g., Kathleen K. Olson, *Preserving the Copyright Balance: Statutory and Constitutional Preemption of Contract-Based Claims*, 11 COMM. L. & POL’Y 83, 84 (2006).

167. See, e.g., Christina Bohannon, *Copyright Preemption of Contracts*, 67 MD. L. REV. 616, 669 (2008) (“Copyright law has failed to develop a coherent account of contract preemption that harmonizes the individual interest in freedom of contract and the societal interest in federal copyright policy.”).

168. Mark A. Lemley, *Beyond Preemption: The Law and Policy of Intellectual Property Licensing*, 87 CALIF. L. REV. 111, 144-47 (1999) (arguing that we should use other doctrinal tools instead of preemption because using the preemption doctrine against contracts “is something like swinging a sledgehammer at a gnat”).

169. Bohannon, *supra* note 167, at 648-54 (proposing that courts consider, on a case-by-case basis, whether the copyright licensee is waiving rights that benefit him or rights that benefit others).

costs before they get to the preemption doctrine in court and it, like the fair use defense, is subject to unreliability.

III. CONCLUSION

As the Internet Archive case has illustrated, the opportunities for digital technology to freely preserve vast amounts of digital information and ensure easy access to all the world's knowledge are being impeded by exceptionally narrow statutory exceptions and fact-intensive and uncertain legal doctrines. The current law is not written for Kahle and the country's social entrepreneurs, and they cannot and should not make the economically beneficial concessions that Google, as a private commercial entity, has made.

In addition, time is working against the archivists. Internet content is ephemeral and even material in libraries is fragile and disappearing. Kahle's project is a noble effort to preserve digital culture from fading into the past. Similarly, Google's Library Project is making access to a broad range of out-of-print books that would have disappeared from our cultural memory as fast as yesterday's website. However, the restrictive laws and the unguided legal doctrines are creating a legally ambiguous environment for the individuals and institutions who are attempting the socially beneficial task of preserving knowledge. Unless they are very wealthy, like Google and Kahle, they will not be able to compete or have a chance at survival between the litigation costs and the uncertain outcomes.

Instead of allowing this legal uncertainty over whether digital preservation techniques constitute fair use continue unaddressed over the years as digital archives wait in the dark for the next lawsuit, Congress should preemptively confront the liabilities and create an adequate exemption for digital archiving.¹⁷⁰ There have been a variety of proposals to create such an exception. For example, Professor Peter Menell offers a framework for Congress to confront and deal with the uncertainty by crafting a safe harbor with safeguards that would recognize the appropriate balance between promoting progress and preserving human knowledge.¹⁷¹ Other scholars

170. By simply relying on the market to sort it out, as we are with the Google litigation, nonprofit digital preservation entities will be left in fear of costly and uncertain litigation and without the resources to compete against commercial entities that may be making compromises hurtful to their ultimate preservation goals.

171. See Menell, *supra* note 9, at 1064-66. Professor Menell would require companies to make commitments to the public nature of digital archive information to fall within the scope of the safe harbor and allow for a right of action for copyright owners against the archives for using insufficient technological protection measures to ensure the security of the copyrighted material. *Id.* Oren Bracha argues for a similar statutory safe haven

suggest a compulsory licensing scheme that would authorize reproduction of copyrighted works for the purposes of digital archiving and cultural preservation.¹⁷² Additionally, the Section 108 Study Group, a group selected by the U.S. Copyright office and the National Digital Information Infrastructure and Preservation program of the Library of Congress to update the Copyright Act for the needs of libraries and archives, has proposed a new exception to section 108 to permit libraries and archives to capture and reproduce publicly available online content for preservation purposes and to make those copies accessible to users for purposes of private study, scholarship, or research.¹⁷³ The Group also recommends that sites that use browserwrap agreements should be considered publicly available for purposes of the exception and subject to capture.¹⁷⁴

Furthermore, the future of copyright and digital preservation could be severely damaged by the advent of traditional contract principles being applied in the most unfamiliar terrain without exception. Until we develop intelligent spiders that can scan for the terms of use on websites, digital archives must be conscious that they may be liable for their scan first, opt-out later policy. We are in need of alternatives to the traditional legal doctrines of copyright and contract that allow for the promotion of preservation and access to knowledge while safeguarding copyright law's incentives to create and the personal right to create and enter into contracts. If the legislative safe harbors, new exceptions, or compulsory licensing schemes cannot be enacted, then the courts will have to face the question of fair use and breach of contract, and digital preservation practice will have to take the hit.

like § 512(c) of the DMCA's safe haven for ISPs. The statutory plan for digital archives would be to define the conditions under which digital libraries are exempt from copyright liability and employ opt-out options. Bracha, *supra* note 132, at 1861-65.

172. Guy Pessach, *Museums, Digitization and Copyright Law: Taking Stock and Looking Ahead*, 1 J. INT'L MEDIA & ENT. L. 253, 267-68 (2007) (discussing a "general, statutory compulsory license that authorizes reproduction for purposes of digital, cultural preservation," "together with a scheme that provides a just and fair compensation to copyright owners").

173. U.S. COPYRIGHT OFFICE & THE NATIONAL DIGITAL INFORMATION INFRASTRUCTURE AND PRESERVATION PROGRAM OF THE LIBRARY OF CONGRESS, SECTION 108 STUDY GROUP REPORT 80 (2008), available at www.section108.gov/docs/Sec108StudyGroupReport.pdf.

174. *Id.* at 84.

BERKELEY TECHNOLOGY LAW JOURNAL