1-1-2006

Locking down the Library: How Copyright, Contract, and Cybertrespass Block Internet Archiving

Rebecca Bolin

Follow this and additional works at: https://repository.uchastings.edu/hastings_comm_ent_law_journal

Part of the Communications Law Commons, Entertainment, Arts, and Sports Law Commons, and the Intellectual Property Law Commons

Recommended Citation
Available at: https://repository.uchastings.edu/hastings_comm_ent_law_journal/vol29/iss1/1

This Article is brought to you for free and open access by the Law Journals at UC Hastings Scholarship Repository. It has been accepted for inclusion in Hastings Communications and Entertainment Law Journal by an authorized editor of UC Hastings Scholarship Repository. For more information, please contact wangangela@uchastings.edu.
I. Introduction

Superbowl I, then just the 1967 Superbowl, Green Bay Packers vs. Kansas City Chiefs, was broadcast by both CBS and NBC. The

---

* J.D. Yale Law School, May 2006, Yale Information Society Project Student Fellow, registered patent agent. The author would like to thank Jack Balkin, Yochai Benkler, Eddan Katz, James Brimmelmann, Blair Kauffman, Lee Wilson, and Lila Bailey. This Article was made possible by Microsoft Corporation's generous support of the Yale Information Society Project.
The game was not very popular; the stadium had 30,000 empty seats, and neither CBS nor NBC kept a copy. It is gone forever. Without a dedicated custodian, culture itself faces the cold prospect of permanent deletion. Even NASA cannot locate the original tape of the moon landing. Today’s equivalent of the Superbowl and moon landing are on the Internet, the most fleeting media ever.

Archivists “fear the smell of burnt letters,” and today they fear the smell of a burning digital library, quietly smoldering forgotten bytes into nothing. The average life span of a Web page is under 100 days, and 44 percent of the Web sites found in 1998 could not be found at all 1999. Aging websites are not stored like precious leather-bound volumes in humidity-controlled library basements or in tall drawers of back-up microfiche. Our electronic cultural heritage would vanish from human history but for a dedicated set of archivists intent on preserving Internet publications. When websites are changed or deleted, they exist only in the archive’s racks of servers. The main archiving project, the non-profit Internet Archive, collects as much data every month as is present in the largest physical library in the world, the Library of Congress. Thus, the Internet Archive and other archiving projects have collected more human expression than

2. Id.
8. Scott Kirsner, Saving the World as We Know It, BOSTON GLOBE, Aug. 15, 2005, at E1.
9. I refer throughout this Note to Internet Archive itself, but the archive could be any public-minded archiver, from the Library of Congress to Google to universities. However, I believe that if archivers are granted exceptions in law, they should be supervised by a central authority and meet established standards. Unlike the clear physical library norms, there are currently no norms of good behavior for Internet archivists. See generally Ann Bartow, Electrifying Copyright Norms And Making Cyberspace More Like A Book, 48 VILL. L. REV 13, 93-94 (2003).
ever before, creating the largest “library” humankind has ever known.10

Unfortunately, the archive is both illegal and grossly incomplete because the information highway is now covered with legal asphalt.11 Though not all websites will qualify for copyright protection, any website meeting the baseline standard of creativity fixed in tangible media qualifies.12 Given the owners’ ironclad rights to publish, distribute, and copy their copyrighted works, the archives facially violate the Copyright Act.13 Internet archivists cannot depend on the traditional library rights to freely exchange, loan, or archive tangible media following the first sale doctrine.14 The Internet Archive’s wholesale copying would probably also run afoul of the fair use doctrine15 and other affirmative defenses were it to be dragged into court. Worse, the uncertainty surrounding copyright law has made archives wary of preserving too much, erring on the side of omitting items with hostile owners.

The superhighway is also covered with roadblocks such as passwords, payment, technical blocks, and restrictive contracts that would seem absurd for physical media publications. These conditions allow seemingly free, instant global publication while prohibiting unauthorized copies, even archival, historical copies. Most major publishers, including the Wall Street Journal and the New York Times, follow this lock-down strategy of highly restrictive contracts. This unprecedented ability both to globally publish and to refuse to relinquish any rights has created a de facto right on the Internet. As a legal matter, authors can withdraw and destroy all existing copies. Unlike traditional libraries, Internet archivists own no legal copy to preserve.

Part II of this Article will explain the unprecedented scope of the Internet Archive’s mission and the tools it uses to constantly collect a remarkable amount of data.

Part III will introduce cultural property as cultural artifacts as well as knowledge and production. Libraries once preserved, and the law protected, books, the ideas in them, and artifacts. However, in a world where ideas are no longer tied to physical cultural objects, a new kind of library must preserve a freeform intellectual property thought itself.

Part IV will explain copyright barriers to archiving Internet content and the traditional exceptions that applied to libraries. Because no copyright exceptions adequately protect Internet Archive’s mission, archives need a statutory protection they currently lack. Shaky affirmative defenses are chilling archives.

Part V will explain how contract and trespass prevent accessing or storing works. Sophisticated, hostile owners are exploiting nascent Internet property rights to opt out of history itself. Well-known authors have drafted tyrannical adhesion contracts to keep their content in controlled circulation, and out of the archive. This material is saved and commoditized by corporations, if saved at all.

Part VI explains how this hopelessly unprincipled legal landscape has created a new de facto right for copyright owners: the right to legally withdraw or destroy their work. In a world where no legal copies exist, the current legal regime – authors – can delete the only authorized copy from the Internet and thus from human history.

II. Background: What Archives Are Doing

Internet archives are using now standard Internet technology to collect as much data as possible. In the process they are creating the largest depository of information ever.

A. Collecting The Internet

Before the Internet archivist movement, deleted pages were wiped from history. No copy remained, and the Internet was a transitory publishing medium like a glorified bulletin board. Internet pages became 404ed into forgotten bits.

Internet archivists have independently created a repository of webpages so expansive it is comfortably called a modern-day Library of Alexandria. Several aggregating projects, including Google, MSN,}

---

17. Id. at 114.
Yahoo!, Library of Congress, university projects, Internet Archive, and several foreign national libraries have taken snapshots monthly, weekly, or even daily, of some subset of the Internet. Commercial search portals Google, MSN, and Yahoo! display search results that link to the pages as well as “cached” versions of the pages hosted by the service. For example, a Google search will return a list of results, and most results will also be available from Google’s servers in the form of the snapshot taken by Google when it last visited the site, generally within the last week. This archive is accessible via a link next to the search results titled “Cached.” The cached page can be minutes or months old, and Google always deletes pages that are defunct or whose host has requested that they be removed.

Internet Archive is a San Francisco non-profit organization dedicated to preserving the Internet at large. It was founded to save digital information and make it accessible to all. Its Way Back Machine service allows users to type a website address and access the monthly snapshots taken by Internet Archive. The Way Back Machine is now a standard research tool in lawsuits to show trademark usage at a given time. It reached ten times the size of the Library of Congress in 2001 and continues to grow.

All archives employ software called spiders, robots, or bots to quickly traverse the Internet. These programs scour the Internet, quickly retrieving the information stored on a website and then

18. Library of Congress has been doing some limited Internet Archiving, using a model that asks permission from authors. [reference redacted for anonymity—involves a letter from LOC asking for permission]
20. Some countries are trying to preserve national works, such as Sweden archiving the .sv top level C domain. Aedy, supra note 6.
visiting the linked pages.\textsuperscript{27} Robots can access hundreds of webpages a minute, and they can run constantly.\textsuperscript{28} Bots can be and are used for many purposes: to retrieve pricing information, to aggregate data for a search engine, to collect email addresses for spamming operations, to find plagiarized material, to archive the Internet, and more.\textsuperscript{29} New applications for bots are constantly being developed.\textsuperscript{30}

Over eighty percent of Internet users rely on an intermediate search page or aggregator; the most popular is Google.\textsuperscript{31} Most publishers want to be listed in Google’s results,\textsuperscript{32} and will even pay to have topical searches link to their website. These search engines also use robots, in Google’s case the same robots, to collect massive indexes of the Internet itself.

Because of their dubious legality, as I discuss in Part III, Internet archives do not want to litigate their status and are quick to remove materials at an author’s request. In 2003, the New York Times changed its previously free Internet-based archives to a paid service. They requested all past materials, including those in the Google cache be removed. Google complied immediately.\textsuperscript{33} However, the New York Times could have sued instead of politely submitting the form;

\footnotesize
\begin{itemize}
\item\textsuperscript{27} David Kramer and Jay Monahan, Panel Discussion: To Bot or Not To Bot: The Implications of Spidering, HASTINGS COMM. & ENT. L. J. 242, 242 (2000) ("[A bot is] a device which goes into a site and accesses information at superhuman speeds.")
\item\textsuperscript{28} V. Shkapenyuk & T. Suel, Design And Implementation Of A High-Performance Distributed Web Crawler, PROCEEDINGS OF THE 18TH INTERNATIONAL CONFERENCE ON DATA ENGINEERING 10 (2002).
\item\textsuperscript{29} Steve Fischer, When Animals Attack: Spiders and Internet Trespass, 2 MINN. INTELL. PROP. REV. 139, 146 (2001).
\item\textsuperscript{30} A computer science class at Johns Hopkins University developed a bot to scour public databases to collect and aggregate private information, such as addresses and social security numbers. Tom Zeller, Jr., Personal Data for the Taking: Students Surfing Public Records Learn It’s Easy to Find Out a Lot, N.Y. TIMES, May 18, 2004, at C1.
\item\textsuperscript{31} Lucas Introna & Helen Nissenbaum, Shaping the Web: Why the Politics of Search Engines Matters, THE INFO. SOC., 16(3):1-17, 2000.
\item\textsuperscript{32} Michael J. Schmelzer, Note, Protecting the Sweat of the Spider’s Brow: Current Vulnerabilities of Internet Search Engines, 3 B.U. J. SCI. & TECH. L. 12 (1997) (questioning whether an aggregator’s results can be copyrighted).
\end{itemize}
it could have won. These lawsuits are already brewing and being decided, and more are inevitable.

B. Republishing

An important goal of some archives is to act like a circulating library by making the material available, that is republishing it. Republication of copyrighted material is beyond the scope of this article, but a description of republishing in comparison to traditional libraries is important to determine the boundaries and goals of archives. In fact, the conflation of republication and archiving is probably what chills archiving the most today. Republication as distribution and archiving are conceptually related because they are both traditional roles of the library when dealing with physical media. However, I argue that a more nuanced view of information preservation is required when dealing with digital material, and archiving can be a separate project and concept.

In mid-1998, a Gallup poll showed that 67% of Americans visited a public library within the last year, 54% checked out a book, and 21% checked out other materials, like a CD or a video. Thus, traditional libraries can cost copyright owners revenue streams, but only in a limited way because of practical and legal access problems. The traditional library had little market impact because there were just too many barriers to universal access. There were only two copies of Oprah’s Book Club book. The library closes at seven. The library is across town. The fines are outrageous, and they only let you keep it two weeks. It will take a week to borrow that book from the next town.


The Internet has lowered barriers to entry and created an entirely new model of production of virtually costless expression.\textsuperscript{40} The Internet itself is like a library, but a fluid one that relies on archives to keep it intact. A traditional library expanded access to knowledge by distributing books more widely than otherwise possible. The Internet Archive overwhelms previous publication possibilities by republishing an unlimited number of copies worldwide.

Internet Archive’s mission as stated, like that of a traditional library, is to collect as much data as possible for all to use.\textsuperscript{41} However, the electronic format allows more content than a traditional library could ever dream of, as well as more delivery options and storage capacity. Internet-based delivery mechanisms ship information all over the world on demand, and universal access to all information is Archive’s stated goal.\textsuperscript{42}

When a world of information is delivered by Internet Archive, any computer becomes the Library of Congress. This new library represents a fundamental shift in the way media goods are consumed, mirroring new, cheaper means of distribution with very low start-up costs.\textsuperscript{43} Traditional libraries may have quite complete holdings, though still nowhere near the scale of Internet Archive’s holdings, even if you also count interlibrary loan (ILL). ILL is allowed by statute and governed by a set of library self-regulations that require documentation and limit the number of requests, for example, to five per periodical per year.\textsuperscript{44} These legal and norm-based restraints all intentionally dampen the library’s already limited impact on the market for the copyrighted goods.

A digital library lacking physical and artificial legal constraints predictably makes copyright owners very nervous. A remote library packed with free content not cleared with copyright owners starts to look like Napster.\textsuperscript{45} In a legal sense it is even worse because the

\begin{itemize}
\item \textsuperscript{40} Yochai Benkler, \textit{Freedom in the Commons: Towards a Political Economy of Information}, 52 DUKE L.J. 1245, 1250-1251 (2003).
\item \textsuperscript{41} \textsc{American Library Association}, \textsc{Libraries: An American Value}, http://www.ala.org/alaorg/oif/lib_val.html ("We must work to preserve this basic American right and ensure access to the broadest range of information.").
\item \textsuperscript{42} \textsc{Lessig}, supra note 10, at 114.
\item \textsuperscript{44} Laura N. Gasaway, \textit{Values Conflict in the Digital Environment: Libraries Versus Copyright Holder}, 24 COLUM.-VLA J.L. & ARTS 115, 146-48.
\item \textsuperscript{45} Bartow, supra note 38, at 38.
\end{itemize}
infringing materials are sitting in a central location, a set of servers in California.

Republishers have distinct obligations to copyright owners, difficult and important obligations that require more debate than this article can provide. The fundamental problem with new, Internet-based republication is that it threatens to be too efficient. The norms of the library, a traditional enemy of copyright law, are self-enforcing where a library circulates only one copy. The library facilitates borrowing and community ownership, but within limits imposed by physical transfer. It is harder to share in a limited way in a world of freely reproducible electronic publication.

A library that is too efficient, a digital circulating library, should not be confused with the concept of archiving without immediate distribution. Today, Internet archives can protect copyright while protecting history. New distribution possibilities should not overshadow new archiving potential. An uncirculated set of materials is no longer so burdensome, and having those servers around for the time when copyright expires may be the only way to preserve them at all. I suggest that a handful of uncirculated archival copies of copyrighted works, protected from unauthorized users, is in our interest as a nation and as a culture. The difficult issue of republication should not bear on an archive's acquisition of a historical copy. Even if the Wall Street Journal is right about its distributive copyright claims and owns the news, so to speak, it should never own history.

III. The Case For the Archives as Custodians of History

Libraries and archives are concerned about losing the data of culture itself. Internet Archive, like all archivists, believes we cannot prejudge the historical value of individual websites, so the archive must preserve everything as a potential historical artifact.

---

46. Id. at 77-78.
50. Id. at 521-22.
A library's democratic values include both education and preservation of culture.\textsuperscript{51} When the law blocks archiving, it becomes more than a legal problem. It becomes a cultural problem.\textsuperscript{52} Libraries once preserved cultural artifacts: physical things like scrolls, books, and artwork containing expression that revealed the ideas of a culture. Today those ideas float freeform with no book, no physical artifact, to preserve. The concept of "cultural property" should now include a duty to preserve intellectual property as published on the Internet.

The pages constituting the Internet are constantly changing, and without the archive, once removed they will simply be gone. The Internet is growing at the rate of seven million webpages a day, and the average lifespan of a webpage is forty-four days.\textsuperscript{53} Even among the websites that the reputable journals \textit{Science}, \textit{The New England of Medicine}, and \textit{The Journal of the American Bar Association} cite, thirteen percent were missing after twenty-seven months.\textsuperscript{54} The ubiquitous error message, "404 Site Not Found," is now used in slang as a noun and a verb.

\section{The Archive and Constitutional Values, Preserving the Commons}

Archives duplicate a value taken for granted that libraries perform: historical preservation. Libraries have long been important guardians of American culture. Thomas Jefferson's donation of his expansive library to the Library of Congress carried a commitment to preserve culture and to share his love for books of all kinds.\textsuperscript{55} The Library of Congress is more than a place to find an answer to a question; it is a "repository of American cultural tradition."\textsuperscript{56} Libraries have always been custodians of cultural artifacts, and their efforts to preserve works have ranged from protective leather cases for the earliest scrolls to microfiche to CD-ROM archiving to modern scanning operations.\textsuperscript{57}

\begin{thebibliography}{99}
\bibitem{51} Gasaway, \textit{supra} note 44, at 129.
\bibitem{52} Laura N. Gasaway, \textit{America's Cultural Record: A Thing of the Past?}, 40 \textit{Hous. L. Rev.} 643, 646 (2003).
\bibitem{53} \textsc{Peter Lyman}, \textit{Archiving the Worldwide Web} (2001), http://www.clir.org/pubs/reports/pub106/web.html
\bibitem{54} Robert P. Dellavalle et al., \textit{Going, Going, Gone: Lost Internet References}, 302 \textit{Science} 787, 788 (2003).
\bibitem{56} \textit{Id.}
\bibitem{57} Gasaway, \textit{supra} note 52, at 647-48.
\end{thebibliography}
In the Nineteenth Century, American libraries were founded to be public places of education and betterment, as democratic institutions. The library was a place of education that allowed the democratic governing populace to be sufficiently informed. As an economic matter, the library was an answer to the tension between market-based information production and intellectual property as a necessary public good. Libraries are an exception to the otherwise capitalist market allowing a limited monopoly on copyrighted works.

Libraries, or free public access, can be considered a solution to market failure. It seems predictable that a monopolist will not give his product away for free on demand or to hold onto a copy after his rents expire. Libraries also constitute a substantial demand market for copyrighted works. The adage that information should be free applies to individuals, not the library, which pays for its materials. Journals are expensive, as is any large collection of books, and the library acts as a community owner of information, a literary and cultural commons. A library’s democratic mission is to ensure access to knowledge and educational use.

Libraries are considered by many scholars to be a critical actual implementation of the First Amendment. Our democratic government’s reliance on education and information requires “the widest possible dissemination of information from diverse and antagonistic sources.” Libraries are vital to the First Amendment right to receive others’ free speech. Free speech’s value in a

58. See Gasaway, supra note 44, at 126-28.
60. Gasaway, supra note 44, at 132-33 (“Librarianship is essential in a capitalistic democracy because freedom of access to information is crucial in a democracy even though capitalism may not appreciate this necessity.”).
61. Bartow, supra note 38, at 90.
62. Id. at 96-97.
63. Id. at 94-95.
64. Id. at 89.
67. Peltz, supra, note 65, at 397.
Meiklejohnian view is that it contributes to the democratic public debate, and that debate is impossible without the works to debate. Access to works in the commons is also critical to making new works, a First Amendment right, and required to engage in their fair use, a statutory right.

Copyright shares these same goals of increasing information and creativity. As written in the Constitution, “Congress shall have Power . . . To Promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries.” A common moral justification of copyright views payment to authors as a secondary benefit; the real goal is to increase the amount of creative work produced in society. Eventually that work will no longer be copyrighted; it will be in the public domain, the commons, for all to use freely.

In 1970, copyright expert Melville Nimmer suggested that a photograph with extreme newsworthiness, or cultural value, became an idea and thus should not be able to be concealed from the public by the photographer who owns the copyright. Nimmer was certainly no copyright minimalist, but for cultural experience, which he calls news, he believed in public access. I suggest all copyrighted works have some value as cultural ideas in the broader scope of history, and may have value that no one can predict. This value is critical to First Amendment expression and the creativity encouraged by copyright.

68. ALEXANDER MEIKLEJOHN, FREE SPEECH AND ITS RELATION TO SELF-GOVERNMENT (1948).


70. U.S. CONST. art I, § 8, cl. 8.

71. United States v. Paramount Pictures, Inc., 334 U.S. 113, 158 (1938) (“The copyright law, like the patent statutes, makes reward to the owner a secondary consideration. In Fox Film Corp. v. Doyal, 286 U.S. 123, 127, Chief Justice Hughes spoke as follows respecting the copyright monopoly granted by Congress, ‘The sole interest of the United States and the primary object in conferring the monopoly lie in the general benefits derived by the public from the labors of authors.’”)

72. Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417,429 (1984) (“The monopoly privileges that Congress may authorize are neither unlimited nor primarily designed to provide a special private benefit. Rather, the limited grant is a means by which an important public purpose may be achieved. It is intended to motivate the creative activity of authors and inventors by the provision of a special reward, and to allow the public access to the products of their genius after the limited period of exclusive control has expired.”)

The First Amendment justification for the archive is not that the content is the archive’s own speech. The archive also resists classification as pure piracy. Instead, the archive as a historical resource is better viewed as the set of existing information from which others’ speech, ideas, and democratic debate are formed. The library is the commons itself.

Creativity builds on itself, and the commons is the foundation for creative expression. Copyright’s limits are intended to prevent past authors from exerting too much downstream control over future creativity based on their work. Upstream copyrights, if too rigid, can stifle individual expression and the progress of culture. Where the First Amendment protects authors from state control, copyright should protect private authors from private control over their inputs from culture.

Without preserving the basic inputs to creativity, the intent of copyright is shattered. Copyright was established to encourage creativity by allowing authors to gain financially from their works and then requires that the work be turned over to history. A copyrighted work may belong to an author for a limited time, but in the end, it belongs to history. Copyright and First Amendment rights rely on the preexisting speech of others. A copyright owner who can exploit his work during its limited monopoly and then destroy it has not only mocked the concept of the commons but has stolen the lifeblood of creativity from the future.

---

74. Eldred v. Ashcroft, 537 U.S. 186, 221 (2003) ("The First Amendment securely protects the freedom to make—or decline to make—one’s own speech; it bears less heavily when speakers assert the right to make other people’s speeches."); see generally Randall P. Bezanson, Speaking Through Others’ Voices: Authorship, Originality, And Free Speech, 38 WAKE FOREST L. REV. 983 (2003).


77. See generally Benkler, supra note 40.

78. See Netanel, supra note 39.

79. Benkler, supra note 40.

80. See generally LESSIG, supra note 10, at 9-11.

81. Id. at 19-20.
Cultural property belongs to everyone, and "[s]ociety is the ultimate loser when these works are modified or destroyed."82 Joseph Sax laments the destruction of great works of art in his book, Playing Darts With A Rembrandt: Public and Private Rights in Cultural Treasures.

If the proprietor of a great painting keeps it locked in his house and then destroys it, how have the rest of us been harmed? In one respect we haven't. Yet there would undoubtedly be a profound sense of loss. Perhaps the most obvious reason is that the community has a long view, and likely that the work would not have been locked away forever; so an opportunity has been lost. In addition, to destroy a work of art is an act of vandalism, a triumph of ignorance over genius; so there is the rending of a value that is important to the community, a symbolic loss that can occur to others even though the thing destroyed was not theirs.

Sax passionately fights for culture's need to preserve expression in singular works.84 The United States does protect singular works of visual art, as discussed in Part IV, from destruction; but, ephemeral Internet works have no such protection.

Works of art have been destroyed for political, "iconoclastic" motives,85 or because the public simply did not like a piece of art.86 A collector cut Toulouse-Lautrec's La Baraque de la Goulue in ten pieces to sell the piece more easily.87 Internet works are destroyed as pages are changed, servers crash, domain rights lapse, and websites are hacked. This destruction is, in a sense, more tragic than playing darts with a Rembrandt because it could have so easily been saved but for the legal restrictions placed on archives.

Our society has claimed some interest in protecting works of historical and cultural value. For example presidential papers are now protected by statute. The Grant, Harding, Pierce, Coolidge, and

83. SAX, supra note 82, at 2.
84. Sax believes in qualified notions of ownership when property, especially singular works of art, has significance to culture. Id. at 9 ("There are many owned objects in which a larger community has a legitimate stake because they embody ideas, or scientific and historical information, of importance. For the most part it is neither practical nor appropriate that these things be publicly owned . . . The conjunction of private and public interests, however, suggests that ordinary, unqualified notions of ownership are not satisfactory for such objects."); but see id. at 57-59 ("The owner-as-steward remains the law's awkward little secret.") Sax seems sympathetic to Rehnquist's dissent in Penn Station that a building was limited because it was too well done.
85. Id. at 13-20.
86. See id. at 29-30 (discussing Andrea Blum's unpopular public art).
87. Id. at 8.
Arthur administrations all destroyed large volumes of presidential papers. 

Lincoln's papers were sealed by request of his son until 1947; John Quincy Adams's papers were sealed until 1956. Though President Warren Harding's family could not get possession of a set of scandalous love letters soon enough to block their viewing, they did use copyright to block a biographer from quoting them. The biographer blanked out a dozen spaces in the book in protest. Shortly thereafter, the Presidential Records Act of 1978 eliminated private ownership of these papers and limited restrictions to only twelve years. The United Kingdom now allows official secrets to be released after thirty years; the United States declassifies most material after twenty-five years.

James Boyle analogizes cultural protection to environmentalism, and laments what he views as the propertization of information. His analogy is about internalizing the costs of pollution and about the interaction of ecosystems of information, but the analogy suggests another value, preservation. If we destroy our cultural environment by letting it simply vanish, the loss is irreparable. We must preserve our cultural landscape just as we must preserve our physical landscape.

Authors owe a duty to society described by one scholar as a "public trust." That is, authors owe to the future use of the current cultural landscape. Restrictions on future use of work can be analogized to servitudes on property, which "prevent full utilization of the land" against the public interest. This kind of servitude in some intellectual property is the technological rights management technology lamented by Lessig and others as alienating information and preventing fair use of the underlying material. On the Internet,

---

89. SAX, supra note 82, at 82.
90. Id. at 135-137.
91. Id. at 137.
92. Id. at 85.
93. Id. at 88-89.
94. Id. at 127-128.
97. Id. at 1137-38.
98. LESSIG, supra note 10; Benkler, supra note 66.
the servitude is more likely complete destruction, as though the owner of a parcel of land made it simply vanish so that no one could use it when his property rights expired. The servitude is easily self-enforced, as there is nothing remaining to enjoy. Selfish destruction runs counter to traditional allowances in copyright, as well as to society’s interest in maintaining cultural works.

**B. Cultural Property and Cultural Heritage**

The father of modern cultural preservation, Abbe Gregoire, considered all books, no matter how bad, to be part of the cultural legacy.\(^9\) These books were not just property; they were culture. Before 1954, “cultural property” was not a legal concept in English common law, though it was inadequately contemplated by the French “biens culturels” and the Italian “beni culturali.”\(^10\) These inspirations were drawn into the first explicit legal use of cultural property in the Hague Convention for the Protection of Cultural Property in the Event of an Armed Conflict, which protects, among other cultural treasures, libraries.\(^11\)

Cultural property once meant expatriated physical objects like the Greek Elgin Marbles in the British Museum, Nofrete in Berlin, and other plundered artifacts from times of conflict.\(^12\) Today, cultures are using cultural heritage law to protect music, dance, handicrafts, religious rites, and even methods of production.\(^13\) For these purposes, intellectual property law is often more effective than cultural property law.\(^14\) For some objects, group ownership is appropriate, and cultural property, as a legal concept, protects group property.\(^15\) Cultural property has been very successful at returning plundered

---


101. Id.; see also Hague Convention Respecting the Laws and Customs of War and Land, art. 56, 36 Stat. 2277, 2309 (1907).


103. Id. at 378.

104. Id.

artifacts, such as the notorious Elgin Marbles, and at limiting their disbursement, such as sacred Australian Aboriginal texts, artifacts, and films. Law has also compelled public ownership. For example, ancient Icelandic manuscripts have moved from quasi-private ownership to preservation and public access at the University of Copenhagen.

Cultural property law has adjusted to the fact that some chattels are more important than others, such as a Hindu family idol declared a legal entity entitled to a "next friend" in court. Law has also forced preservation of private property, such as historic buildings. Some historic buildings are protected by the 1966 National Historic Preservation Act (NHPA). The National Register of Historic Places lists federally-owned, historical landmarks. NHPA also gives funds to states to protect state-owned historic structures. Even some buildings too young to fall under historic purvey are subject to oversight boards. Historical building preservation regulations are so important to communities they are not viewed as eminent domain takings from individual.

Modern historians now recognize challenges to cultural preservation distanced from artifacts. In their landmark exposition, Lyndel Prott and Patrick O'Keefe foresaw a kind of culture that could not be preserved in objects like statues and books, as well as the limitations of the concept of "property" as a custodian for culture. They established the term "cultural heritage" as more apt:

[T]he existing legal concept of 'property' does not, and should not try to, cover all that evidence of human life we are trying to preserve: those things and traditions which express the way of life and thought of a particular society; which are evidence of its intellectual and spiritual achievements. On the other hand, they can be encompassed by the term 'heritage' which also embodies the notion of inheritance and handing on. This is central to our second

107. Prott & O'Keefe, supra note 100, at 314.
108. Id. at 316-317.
109. Id. at 310.
111. SAX, supra note 82, at 53-54 (discussing New York City's rejection of the Whitney Museum's and the Guggenheim Museum's planned changes).
objection to the existing legal concept of property; that 'property'
does not incorporate concepts of duty to preserve and protect. Cultural heritage evokes an obligation to hand on to future
generations our expression, ideas, and works, now jumbled on the
Internet in unprecedented teamwork and democracy. This work is
freeform intellectual property with no artifact in sight, and it is the
creative expression of our generation.

Intellectual property can be appropriate to protect cultures from
outside manipulation and theft, but it fails to protect our culture
from itself because intellectual property has no per se sense of duty of
preservation. Worse, as I will discuss in Parts IV and V, intellectual
property and contract are sabotaging those within the culture who
wish to preserve it.

To Susan Scafidi, champion of cultural property, misappropriation is the worst case scenario for cultural
products. The motivation behind Scafidi's recommendation for cultural
authorship is a fear of inauthentic commercial exploitation and
blurred cultural messages. Her framework is concerned with
distortion, not preservation. I believe Scafidi's concerns pale in a
world in which collective cultural heritage can be destroyed by the
whim of individual authors after publication. Right now, a future
senator's teenage Myspace page, the front (web)page of a news
organization, or today's op-eds may be lost. Complete or significant
destruction is now the worst case scenario. Perhaps even a distorted
history is preferable to a forgotten one.

When preservation of things is no longer the duty of private
owners, such as art collectors, it is important to assess where that duty
lies. Prott and O'Keefe impose a duty of preservation for cultural
property, and someone must shoulder that duty for the Internet. A
great deal of the Internet is a very old kind of cultural heritage,
written expression, manifested in a new media that lacks the cultural
property of the past. Libraries, as the guardians of printed culture,
have the duty to preserve books. But when there are no books to buy,
culture must be preserved some other way.

113. Prott & O'Keefe, supra note 100, at 307.
114. See Benkler, supra note 59.
115. See Fechner, supra note 102.
116. SUSAN SCAFIDI, WHO OWNS CULTURE?: APPROPRIATION AND AUTHENTICITY
IN AMERICAN LAW 124 (Rutgers University Press, 2005).
117. Id. at 148-151.
IV. Copyright Problems

Copyright gives authors a set of rights that translate poorly into modern activity on the Internet. Most people probably do not go an hour on the Internet without violating some outdated copyright technicality. Copyright was established to regulate a technology, the printing press, and has been updated along the way. Now a new technology, the Internet, has made traditional copyright protections seem irrelevant. The rights to reproduce, display, perform, and publish copyright works call into question the legality of basic activities on the Internet and in the archives.

The conflict between a library's free public access and the copyright holder's monopoly is a fundamental conflict in copyright law. Librarians want more access for everyone while content owners want to control access to extract the legally allowed rents. Ironically, the library, an historic balancer of copyright law, has no explicit librarying right, though some believe a library right should be codified. Internet archives cannot depend on the delicate statutory copyright balances that protect traditional libraries, and copyright must adapt to new technology to allow archiving.

A. Facial Copyright Infringement

Copyright grants four rights in 17 U.S.C. § 106 which are all facially violated by the Internet archives: the right of reproduction, the right of public performance, the right of public display, and the right of public transmission.

119. Litman, supra note 11, at 34-35.
121. See Netanel, supra note 39 at 283 (“If copyright is cast too narrowly, authors may have inadequate incentives to produce and disseminate creative works or may be unduly dependent on the support of state or elite patrons. If copyright extends too broadly, copyright owners will be able to exert censorial control over critical uses of existing works or may extract monopoly rents for access, thereby chilling discourse and cultural development.”).
122. See Gasaway, supra note 44, at 115-116 (“Librarians tend to view information as a necessary public good, such as food, shelter, and warmth; that should be made available at a reasonable cost. Commercial producers and publishers of copyrighted works, however, tend to view their works as private property that can be commercialized.”)
124. For a more detailed description of these rights as applied more generally to Internet use, see David L. Hayes, Advanced Copyright Issues in the Internet, 7 TEX. INTELL. PROP. L. J. 1 (1998).
The most fundamental copyright is the right of reproduction, or the right to make copies and publish work as an author sees fit. This right applies to both digital copies as well as to partial digital copies.\textsuperscript{125} Even in a simple Internet transmission of data, it is hard to know when or where or how many copies have been made.\textsuperscript{126} Search engine robots make their own copy of each website at every visit and use that information to search the data as well as to create an archive. The archivers must store that copy, preserving the evidence of unauthorized copying.

The Copyright Act's right of public performance includes the right "to transmit or otherwise communicate a performance."\textsuperscript{127} This section was written for isochronous transmissions, such as broadcasting, so its application to asynchronous behavior on the Internet is uncertain.\textsuperscript{128}

The right to "to display a work publicly" uses the same definitions as the right to public performance, and a transmission, synchronous or asynchronous, is a public display.\textsuperscript{129} Congress explicitly noted that display includes remote access to works in a library.\textsuperscript{130}

A copyright owner has the exclusive right to "distribute copies or phonorecords of the copyrighted work to the public by sale or other transfer of ownership, or by rental, lease, or lending."\textsuperscript{131} Thus, distribution of copies of copyrighted works to the public is infringing.\textsuperscript{132}

For copyright purposes, using a website is quite different from using a record or a book. Accessing a website requires a series of packet exchanges, transmitting pieces of data, between several computers. Unlike previous media, accessing a copy of the work may require making a large number of copies before it can be viewed by

\textsuperscript{125} See generally, MAI Systems v. Peak Computer Inc., 991 F.2d 511 (9th Cir. 1993).

\textsuperscript{126} David L. Hayes, Application of Copyright Rights to Specific Acts on the Internet, 15 No. 8 THE COMPUTER LAWYER 1, 3-4 (1998).

\textsuperscript{127} 17 U.S.C.A. § 106(4) (West 2005).

\textsuperscript{128} Hayes, \textit{supra} note 124, at 30.


\textsuperscript{130} H.R. REP. No. 94-1476, at 64 (1976) ("[Display includes] the transmission of an image by electronic or other means, and the showing of an image on a cathode ray tube, or similar viewing apparatus connected with any sort of information storage and retrieval system.")

\textsuperscript{131} 17 U.S.C. § 106(3) (West 2005).

\textsuperscript{132} These rights are similar to the WIPO rights of communication and distribution. \textit{WIPO Copyright Treaty}, Dec. 20, 1996, \textit{S. Treaty} Doc. No. 105-17 (1997).


the user, who also makes a temporary local copy in RAM.\textsuperscript{133} Distribution, transmission, and access of the work may be required as well, all facial copyright violations.\textsuperscript{134}

A new exception in copyright law responds to a court ruling in \textit{MAI Systems} that a temporary copy of a computer program was infringing.\textsuperscript{135} Now temporary copies in the RAM of a computer are allowed.\textsuperscript{136} However, this exception is limited to computer programs and not to copyrighted website content. Thus, using the logic of \textit{MAI} that a statutory exception is required, given the narrow exception in the Copyright Act it appears that accessing any website requires infringing copyright.

Indeed, access to material without a license is strictly infringing. To find contributory infringement for infringing webpages, for example those hosting illegal movies or pictures, the court must first find direct infringement from browsing by making copies in RAM by website visitors.\textsuperscript{137} Archiving bots commit millions of facial copyright violations daily. The archive has no way to preserve works except to make an unauthorized copy, potentially infringing the right to copy, the most important copyright.

\section*{B. Why Traditional Library Copyright Exceptions Do Not Apply: No Sale, No Owner}

A monopolist’s rights and public rights to access are traditionally balanced in carefully drafted copyright exceptions that allow libraries to buy, own, and circulate vast quantities of information. There is no codified libraying right, instead there are enough exceptions for a traditional library to just slip by. A traditional library relies heavily on the section 109 first sale doctrine, which allows the unrestricted resale and loan of legally purchased tangible media, as well as some leasing of material.\textsuperscript{138}

Because Internet works are not sold as tangible works, first sale does not apply: no sale, no owner.\textsuperscript{139} In order to archive webpages,

\begin{itemize}
\item \textsuperscript{133} Hayes, \textit{supra} note 124, at 5.
\item \textsuperscript{134} See Hayes, \textit{supra} note 126, at 1-2.
\item \textsuperscript{135} MAI Systems v. Peak Computer Inc., 991 F. 2d. 511 (9th Cir. 1993).
\item \textsuperscript{136} See Albritton, \textit{supra} note 34, at 137, 141.
\item \textsuperscript{137} See Michael Dockins, Comment, \textit{Internet Links: The Good, the Bad, the Tortious, and a Two-Part Test}, 36 U. TOL. L. REV. 367, 383-85 (2005).
\item \textsuperscript{138} 17 U.S.C. § 109 (2004).
\item \textsuperscript{139} See Bartow, \textit{supra} note 9, at 113 (“So, if there is no sale, there is no owner; and if there is no owner, there are no first sale rights.”). 
\end{itemize}
archives must make unauthorized copies. Internet archives have no ability to legally purchase and distribute tangible media, so authors are not imposing illegal restraints on physical property.

Libraries do have limited duplication rights for their holdings, but none so broad as to constitute copying the entire Internet. These exceptions do fairly well at covering works previously owned in hard copy, for example microfiche archiving, but they do not address archival copies of electronic media, especially media that could never be owned.

A handful of exceptions in section 108 allow libraries to reproduce orphan or out-of-print works in special situations. These exceptions are riddled with gaps and inconsistencies. The Registrar of Copyrights agrees that legislation is badly needed on orphan works. These exceptions do fairly well at handling orphan and out-of-print works in physical media that libraries already own, but nothing at all for digital media. For example, the exception allowing copies of out-of-print works in their last twenty years of copyright does not help Internet archivists because they are dealing with materials with a lifespan of far less than a year.

Out-of-print works have narrow exceptions to licensing requirements. Some argue that fair use rights, compensated importation, or compulsory licenses are the solution to access to out-of-print works. In the context of the Internet, these discussions

140. id. at 93 ("[A] library cannot "share" material over the Internet without reproducing and transmitting it. It is therefore necessary as well as expedient to make multiple copies of copyrighted digital works to facilitate any sort of sharing."). But see Ryan, supra note 123, at 162-63 (arguing cached copies can be kept forever by some set of users).
142. See Gasaway, supra note 44, at 139-142.
143. Marybeth Peters, Copyright Enters the Public Domain, The 33rd Donald C. Brace Memorial Lecture Delivered At New York University School of Law (Apr. 29, 2004) at 713 ("Unfortunately, the terms of Section 108(i) make it inapplicable to motion pictures, musical works and pictoral, graphic and sculptural works, even though the intent was that the provision would apply to all types of works.").
145. See Gasaway, supra note 44, at 129.
about the market concerns for publication seem wholly inappropriate. For example, Wendy Gordon argues that fair use for out-of-print work is inappropriate if the copying would affect the market for an authorized reprint.149 This view of out-of-print works assumes authors rely on fees and could later want to embark on the economic journey of republication.150 For websites, an expensive reprint that requires demand is unthinkable. An author does not abandon publication because it was too expensive; a web author needs no demand at all to make publication feasible. Publishing works on the Internet is now the cheapest publication has ever been: free or nearly free.151 Out-of-print works on the Internet are an historical resource that will be lost after their abandonment.

More works are orphan works than ever before. Today, only 18% of older tangible works are considered “surviving works;” the remainder are orphaned.152 On the Internet this figure is probably even more dismal as many works simply vanish and others are entirely anonymous. Consider an anonymous blog with a defunct, anonymous email contact. The Copyright Term Extension Act, coupled with the Berne revocation of registration, has made orphan works a very broad, pressing area of copyright law requiring reform desperately.153 Traditionally, first sale would protect at least some limited access when a work is withdrawn.154 Some library would still have a copy, and the library could theoretically ILL the work, even if there are very few copies remaining. Without a first sale doctrine, this is no legal copy, and no protected access to those works. Libraries cannot buy any Internet work they want,155 so they lose the library’s bedrock copyright exception.

150. Basic Books, Inc. v. Kinko’s Graphics Corp., 758 F. Supp. 1522, 1534 (1991) (“Kinko’s copying unfavorably impacts upon plaintiffs’ sales of their books and collections of permissions fees. This impact is more powerfully felt by authors and copyright owners of the out-of-print books, for whom permissions fees constitute a significant source of income. This factor weighs heavily against defendant.”).
C. Copyright Defenses

There are defenses to facial copyright infringement, but none seem quite right for Internet archiving. Implied licensing and fair use are uncertain doctrines too shaky to rest culture on. No one really knows what would happen in court for an archive, so the uncertainty of the doctrines chills archivists' behavior for good reason. It is not doctrinally sound to force the concept of preserving history despite an author's resistance onto a set of narrow exceptions aimed at very different behavior.

In *Netcom*, the Church of Scientology sued a provider of bulletin board services for hosting copyrighted material of the church. The court acknowledged that Netcom was indeed storing the material on its servers, but declined to find Netcom liable because "there must be some element of volition or causation which is lacking when a defendant's system is merely used to create a copy by a third party." Historically, and statutorily, copyright infringement is a strict liability offense; no volition is needed. However, when presented with basic functionality of the Internet, the court created an exception out of whole cloth rather than destroy the Internet.

In a more recent case, *Field v. Google*, a Nevada district court relied on *Netcom* to find the Google Cache lacked volition. However, *Field* was not litigating the legality of Google's copy, but the copy downloaded by a theoretical user when he accessed the cache. The court found that Google's servers' automatic response lacked volition, as they automatically responded to user inquiries. Thus, though Google copied, stored, and transmitted the copy, its code did so automatically upon the response of a third party who actually made the copy, so Google lacked volition. This volition exception is even more painful than *Netcom's*, and extended even slightly beyond the facts at issue leads to the absurd result that your robot or other code can be your non-volitional agent. Thus, the

158. *Id.* at 1368 ("Although copyright is a strict liability statute, there should still be some element of volition or causation which is lacking where a defendant's system is merely used to create a copy by a third party.")
160. *Id.* at 9.
161. *Id.* at 10.
exceptions are stretched beyond reason and are unreliable for Google or Internet Archive.

Internet archives, if sued, would probably rely on fair use. This affirmative defense is not the solution to archives' copyright problems. Fair use is a doctrine too unstable to rely on when verbatim copying millions of webpages a day without consent. Traditional libraries only rely on fair use for relatively small-scale efforts to archive, copy, and excerpt works, for example copying course packets or copying a work in reserve for an academic course.\textsuperscript{162}

Fair use has four statutory factors: (1) the purpose and character of the use, including whether such use is of a commercial nature or is for non-profit educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for and value of the copyrighted work.\textsuperscript{163}

The doctrine of fair use is "the most troublesome in the whole law of copyright."\textsuperscript{164} Fair use is a difficult, fact-intensive, and unpredictable doctrine. In this case, the first factor, purpose, is the archive's most sympathetic factor. Though Google Cache is commercial, and Internet Archive's users could be using the archive for commercial purposes, a library has a special kind of public interest that might be considered the highest kind of education purpose. Thus fair use doctrine could use this first factor to manufacture the elusive library right.\textsuperscript{165} The second factor values more creative or unpublished work,\textsuperscript{166} but the archives pick up all kinds of work, so all types of work would be infringed. The third factor, amount copied, prohibits excessive copying. Thus, verbatim reproduction in the archive is the worst kind of copying.\textsuperscript{167} The most important factor, market impact, cuts against archives that distribute. Noncirculating archives have no

\textsuperscript{162} See Joshua H. Foley, Comment, Enter the Library: Creating a Digital Lending Right, 16 CONN. J. INT'L L. 369, 373 (2001).
\textsuperscript{164} Sony, 464 U.S. at 474 (Blackmun, J., dissenting).
\textsuperscript{165} See Bartow, supra note 38, at 89-95.
\textsuperscript{167} See Marcus v. Rowley, 695 F.2d 1171 (9th Cir. 1983) (finding "wholesale copying" to preclude fair use); Diane Leenheer Zimmerman, The More Things Change, the Less They Seem “Transformed”: Some Reflections on Fair Use, 46 J. COPYRIGHT SOC'Y U.S.A. 251, 257 (1998) ("[P]resence or absence of transformation has become the linchpin on which post-Campbell fair use cases tend to turn.").
market impact by definition. The case does not look promising for the Internet archives, but no one really can say because of the fickle nature of fair use.

Search engines that use short-lived copies of websites to speed up searches and make access to the original copyrighted material more efficient may have a fair use claim. Even limited republishing for indexing purposes has a colorable a fair use claim. However, this claim is more questionable when indexes plan to keep the work forever, even for historical purpose. Usually, republishing because a work is out of print or not sufficiently available is no legal defense at all.

Field dedicated much of the opinion to the fair use defense for the cache. It is critical to note that Field was not litigating the legality of Google's copy, but the copy a theoretical user would download from the cache. Since the fair use doctrine is fact intensive, this user's standing is very different than a theoretical inquiry into the legality of a historical archive, or even an archive used for something else, such as to generate Google's search results. The Field court granted summary judgment for Google. It was impressed by the function of the archive, even with unauthorized display of changes in the work, and despite Google's for-profit status. The Field court found it relevant that the copied content was available elsewhere for free, mentioning this fact ten times. The court analogized the free content in the cache to timeshifting. The court pointed out that the cache did not impact Field's desire to create or the market for his work because his work was available for

168. With circulation, wholesale copying can be presumed to have market impact, Dr. Seuss Enters., L.P. v. Penguin Books USA, 109 F.3d 1394, 1403 (9th Cir. 1997), even where the copiers are non-commercial, Princeton Univ. Press v. Mich. Document Servs., 99 F.3d 1381 (6th Cir. 1996).

169. See, e.g., Sega v. Accolade, 977 F.2d 1510 (9th Cir. 1992) (emphasizing functionality).


173. Id. at 1115.

174. Id. at 1120.

175. Id. at 1106.

176. Id. at 1118. The timeshifting justification in Sony did not directly involve republication to third parties, but the Court ignores this fact.
free.\textsuperscript{177} The court then added another fair use factor used in that circuit, good faith.\textsuperscript{178}

Though this seems like a victory at first blush for archives, it is not. This kind of fair use inquiry is problematic because it is significantly colored by the fact that the identical work is published for free. Google Cache deleted old, inaccessible versions, which are much more sympathetic in a fair use inquiry. If the work were destroyed by the author, requested to be removed by the author, or if the work were password protected or for pay, the Field inquiry would topple quickly. Internet Archive has also been sued for unauthorized republication for revealing past versions of a webpage, and Field does not reliably translate to these facts.\textsuperscript{179}

The fair use doctrine is not stable enough to predict what kind of archiving is allowed. Internet Archive is wary of this doctrine, as it should be. We should certainly not vest our cultural rights in this fickle doctrine. Further, Internet Archive and Google Cache's republication missions are intertwined in this analysis, so the hazy outlook of Internet Archive's republishing carries over into its collecting policies. Field and other lawsuits are about the right to publication, but they chill archiving and ownership for historical reasons as well. This problem will remain until there is a statutory right to archive.

Since copyright translates so poorly to what actually happens on the Internet, there must be some implied license or no one could legally access copyrighted work.\textsuperscript{180} A person accessing a copyrighted webpage must have an implied license to copy, at least temporarily in her RAM.\textsuperscript{181} However, it is unclear where this license ends, and a court could easily forbid caching's harmful results,\textsuperscript{182} as well as more permanent copying.

When dealing with the implied license theory, the Field court did not use an objective standard for the author, but relied on the fact that Field was "aware of these industry standard mechanisms,"\textsuperscript{183}

\begin{itemize}
  \item \textsuperscript{177} Id. at 1121-22. The court does not distinguish between Field's volition publication for free and the free, universal access of a traditional library.
  \item \textsuperscript{178} Id. at 1122-23.
  \item \textsuperscript{179} Healthcare Advocates, supra note 35.
  \item \textsuperscript{180} Christine D. Galbraith, Access Denied: Improper Use of the Computer Fraud and Abuse Act to Control Information on Publicly Accessible Internet Websites, 63 MD. L. REV. 320, 323 (2004).
  \item \textsuperscript{181} Kramer & Monahan, supra note 27, at 252-53.
  \item \textsuperscript{182} Hayes, Advanced Copyright Issues, supra note 126, at 6.
  \item \textsuperscript{183} Field, supra note 36, 412 F. Supp. 2d at 1116.
\end{itemize}
meaning robot exclusion protocols. The fact that he omitted the robot commands despite that knowledge created an implied license that Google could cache his work. The court ruled Field's knowledge made his publication without appropriate notice a "conscious decision" to waive his copyright.

This analysis is a limited victory for Google, at best. Unsophisticated authors lack the capacity to make a conscious decision to waive their rights, and the Googlebot has no way of knowing whether a user is unsophisticated or consciously omitted the exclusion notice. Worse, this implied license can be and is being explicitly revoked by sophisticated parties who do have the ability to make restrictive explicit demands.

V. Trespass and Contract Problems

Well-established technical flags and contracting can tell users that an author plans to protect his content by forbidding caching or archiving. Using these tools, major web authors have expressly revoked whatever copyright entitlements or implied license allow archiving. This regime uses nascent cyberspace property rights to block not only vicious competitors' robots but also archival robots. I will call this an "Efficacious Promulgated Superseding Entitlement Regime," or EPSER, because it uses contracting and cybertrespass to override intellectual property law.

Even very restrictive contracts, ones which waive traditional intellectual property allowances, are enforced. In the words of Judge Easterbrook, "a simple two-party contract is not 'equivalent to any of the exclusive rights within the general scope of copyright' and therefore may be enforced." These contracts, with people and with robots, now lock down content from archivers.

184. Id.
185. Id.
186. Kramer & Monahan, supra note 27, at 253 ("OK. You now have an implied license to make that copy. But an implied license can be expressly revoked.").
187. Margaret Radin coined this term to apply to widespread contract schemes to override traditional property structures, including intellectual property. Margaret Jane Radin, Boilerplate Today: The Rising of Modularity and Waning of Consent, 104 MICH L. REV. 1223, 1233, n.38 (2006).
A. Technological Blocks

Content owners can regulate the behavior of visitors through code by simply blocking them. Website owners can use this technology to prevent access even to non-copyrighted data. For some authors, these preventions may be more cost- and time-effective than notice and subsequent legal action.

Servers and networks are private property. This means they can reject any traffic or any visitor. The simplest way to do this is to block a human or robot visitor’s IP address, or the location of his computer. This method is used often in chat rooms or message boards to block offensive members or to block unsavory robot traffic, such as robots interested in harvesting email addresses. Blocking IP addresses is often not very effective since IP addresses can change frequently, are often dynamically assigned by ISPs, and can be forged or spoofed.

Authors can also lock down their content by controlling access using passwords, even for free content, like the New York Times, or running entirely private networks, like the AOL network or a business intranet. These passwords can require payment, membership, or identity verification. This technique restricts access dramatically.

Website authors can also make dynamically scripted CGI pages, or pages which change content based on time, characteristics of the visitor, and more. Dynamic pages do not limit access per se, but they can change what robots have access to versus regular users.

---

189. James Grimmelmann, Note, Regulation By Software, 114 YALE L. J. 1719, 1730 (2004) ("Think of the calculator and the password-protected Internet site. There is no need to punish someone who uses a calculator program to write a letter—it is simply not possible to use the program in that way. Similarly, as long as the password-checking software on the site works, it constrains your access immediately.").
190. O'Rourke, supra note 26, at 1984-6.
191. Dockins, supra note 137, at 403.
195. Fischer, supra note 29, at 142-43.
197. Google actually removed BMW's German website for presenting a different version to the Googlebot than regular users would see. See Nancy Gohrig, BMW Back in With Google, IDG NEWS, Feb. 8, 2006, at http://reviews.infoworld.com/article/06/02/08/75182_HNbmwgoogle_1.html?SEARCH%20ENGINES.
Some websites employ special tests called CAPTCHAs which robots cannot pass, usually including distorted text.\(^{198}\)

No technology can stop all robot activity,\(^{199}\) but technology can make a large practical and even legal difference for an author hostile to the activities of robots. In the context of trespass to chattels on the Internet, lack of notice with technical protections can signal consent to access.\(^{200}\) Usually, however, restrictions are used to show intent to exclude. For example, one robot operator was held to have exceeded authority by visiting a website protected by passwords, which were viewed as implied notice to exclude robots.\(^{201}\)

B. Robot Readable Notice: Robot Exclusion Standards, Metatags and http Protocols

Guardians of the Internet responded to robot behavior with the Robot Exclusion Standard, an informal agreement among website publishers and robot operators. Though the robot exclusion header does not carry the force of law, it is a useful tool for authors to notify robots about copyrights.\(^{202}\) Terms of access for visiting robots are located in a file called “robots.txt” at the root or top directory of a group of websites. This file contains many notices to robots, including areas that are forbidden for robots to visit and maps of the site. There are also two specialized commands—"noarchive" and "nocache"—that publishers can make in this file that relate to the archives. These commands are simple and ban archiving or caching in all or a portion of the sites. Robot exclusion headers are generally followed.\(^{203}\)


\(^{199}\) Jeffrey M. Rosenfeld, Spiders and Crawlers and Bots, Oh My: The Economic Efficiency and Public Policy of Online Contracts That Restrict Data Collection, 2002 STAN. TECH. L. REV. 3, 3.


\(^{201}\) See EF Cultural Travel BV v. Zefer Corp., 318 F.3d 58, 63 (1st Cir. 2003) (“We agree with the district court that lack of authorization may be implicit, rather than explicit. After all, password protection itself normally limits authorization by implication (and technology), even without express terms.”).


Metatags are HTML commands invisible to a user of a website, though the user can see them by viewing the page’s source code. Metatags contain search terms and descriptive information about a website as well as directions to visiting robots. The two relevant metatags are also “noarchive” and “nocache” which can be narrowed to a specific website and a specific agent’s robot. For example, the New York Times uses both tags on its front page. All of the major archives obey the noarchive tag. Caching restrictions are also embedded in HTTP itself. Authors can control what level of caching allowed to the second. A robot passing through a website can read its robot protocol restrictions, and thus its actions become assent just like clicking a click-wrap contract.

Robot protocol files are a kind of click-wrap, probably fully enforceable in contract law. Common law and the Uniform Computer Information Transactions Act (UCITA) support contracting with robots. UCITA is a model contract law governing “information in electronic form which is obtained from or through the use of a computer.” UCITA has been in development for years and proposed as federal law. So far, two states have passed versions of UCITA, Maryland and Virginia. Even unconscionability, copyright preemption, and lack of meaningful consent do not bar robots from contracting with website owners under UCITA or standard contract law.

Because signaling costs are low, network owners can avoid the cost of closing the network with the tags. Most robots, such as Google’s indexing bot, have no hold-out problem when sites opt

204. NYTTimes.com, <meta http-equiv="Pragma" content="no-cache"/> <meta name="robots" content="noarchive"/>. The Pragma command changes the implementation, but not the spirit of the command. See http://support.microsoft.com/ kb/222064/.


206. Rosenfeld, supra note 199, at 44-45.

207. Id. at 37-38.

208. Id. at 35.


211. Rosenfeld, supra note 199, at 35-41.

212. Bellia, supra note 200, at 2251.
out,\textsuperscript{213} they simply move on to another website. However, all robots have a chilling effect problem because noarchive scares them away—Internet archivists flee when they encounter these restrictions. Webpages marked with “noarchive” are not being archived for the future.

Besides being a binding contract, robot protocol is also somewhere between a notice-based restriction and a code-based restriction to access.\textsuperscript{214} It is notice through code. Though notice about copyright is no longer required, it can be relevant in court. Actual notice of copyright (©, year, and owner) was once required but repealed in the codifications of the Berne Convention, effective March 1, 1989.\textsuperscript{215} Failure to give notice of copyright does not dedicate a work to the public. Even for criminal violations of “willful” infringement under 17 U.S.C. 506(a), notice can be probative but is not required.\textsuperscript{216}

A recent lawsuit\textsuperscript{217} claims that robots.txt is an effective technological protection for copyrights, as defined in the DMCA, 17 U.S.C. § 1201(a). This would make circumventing a robots.txt file a crime. Further, this plaintiff claims a robots.txt file retroactively gives notice about allowances for a published website. Neither of these legal claims is very convincing. However, the point remains that a robots.txt file is a chilling notice to archivers. Presence of “noarchive” could be used to show notice and even willfulness of the violation; archivers would at least be violating a contract. But with this simple word, the author opts out of the library, and of history.

Because the robots generally obey the protocols, there has been no litigation, and because there has been no litigation, the robots cautiously follow the protocols. It is difficult to know just how much web content is labeled with these filters because robots do not visit it.\textsuperscript{218} We do know that many major content providers, such as New

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{213} McGowen, \textit{supra} note 75, at 379.
\item \textsuperscript{214} Bellia, \textit{supra} note 200, at 2212-14.
\item \textsuperscript{215} 18 U.S.C.A. § 102 (West 2005).
\item \textsuperscript{216} \textit{See} United States v. Cross 816 F.2d 297 (7th Cir. 1987) (affirming conviction where FBI warned defendant); United States v. Heilman, 614 F.2d 1133, 1138 (7th Cir. 1980) (affirming conviction where defendant was aware of similar prosecutions); \textit{but see} United States v. Whetzel, 589 F.2d 707, 712 (D.C. Cir. 1978) (affirming conviction based in part on past actions).
\item \textsuperscript{217} \textit{Healthcare Advocates, supra} note 35.
\item \textsuperscript{218} \textit{See generally}, Herbert Snyder, \textit{How Public is the Web?: Robots, Access and Scholarly Communication}, ASIS (1997), at http://www.slis.indiana.edu/faculty/hrosenba/\textit{www/Papers/asis981.html#4} (concluding many “Most Wired” universities use robot exclusion protocols).
\end{itemize}
\end{footnotesize}
York Times and Wall Street Journal, have removed their content from the archive. In the short term, these major news sources are limited to private hands, but in the long term, history itself is preserved by a corporation, if preserved at all.

Robot protocol restrictions are blocking off areas of the Internet from archiving. For example, whitehouse.gov’s robots.txt file lists the entire directory structure, effectively confining all robots to the front page alone. The Library of Congress’s Thomas at thomas.loc.gov also blocks robots. Thomas prohibits all robots except Googlebot and Ultraseek. This means that Yahoo! and MSN do not index Thomas’s homepage and that archiving bots stop at the robots.txt file. Even the Library of Congress has locked its content to the archivers.

C. Terms of Service and Fine Print

Authors can revoke implied rights and impose restrictive terms through terms of service written for users but unintelligible to robots. Virtually all commercial websites have some terms of use or conditions of service, but all may differ. The legal terms are usually found by clicking on a link at the bottom on the page stating like “terms of service,” “conditions of use,” “legal,” or sometimes on the copyright notice. This form of contract can be called “browse-wrap” because continuing to browse may show assent.

Robots are probably liable for extreme contract restrictions in terms of service, even beyond robot exclusion protocols. Courts usually enforce terms of service, even quite restrictive ones robots cannot read. Robot operators do not necessarily require knowledge, and the assent needed to consent to browse-wrap

---

219. Womack, supra note 33.
221. Thomas.loc.gov, Robots.txt, at http://thomas.loc.gov/robots.txt (last visited Apr. 12, 2006) (“User-agent: Googlebot I Disallow: | User-agent: Ultraseek I Disallow: /cgi-bin | User-agent: * I Disallow: /”). This robots.txt file gives Googlebot access to all files and directories (blank disallow), Ultraseek access to all except cgi-bin (cgi-bin), which contains scripts like Thomas’s search function, and denies all access to any director to any other bot.
222. Hayes, Advanced Copyright Issues, supra note 124, at 6.
contracting can be as effective as a legal file, even if a robot cannot read it. 226

Terms of service lock down content in ways an author selling the content in print simply could not accomplish. Disney’s terms of service are intended to be a contract with children. 227 Microsoft’s terms of service might reach so far as to forbid users to criticize Microsoft. 228 In the real world, these restrictions on, say, a newspaper would be wholly unenforceable. However, modern electronic contracting has lost sight of concepts like unconscionable or expectations of the parties, viewing the contract as the product itself. 229 Where the contract and product start to collapse, such as locked down intellectual property, this automated contracting becomes the worst kind of adhesion contracting. 230

Terms of service can unambiguously block archiving. The New York Times, surely an important U.S. author, allows access with a free passworded account, but with restrictive terms of service. The Member Agreement forbids any copying not for a member’s private use without explicit approval of the New York Times. 231 Some electronic works on the site never enter a library, such as some content in the New York Times’s paid Internet features, Times Select, which never exist in physical print. 232 Though physical copies of the New York Times are preserved by libraries around the country,

226 See Register.com v. Verio, 356 F.3d. 393 (2d Cir. 2004) (holding bot owner liable for violating terms of service). However, even the Second Circuit has been unclear about what kind of assent is really needed. See Juliet Moringelo, Signals, Assent And Internet Contracting, 57 RUTGERS L. REV. 1307,1326-30 (2005) (“With only Specht and Register.com providing real guidance, the state of the law governing browse-wrap terms can almost be described as follows: bad-guys (screen-scrapers collecting information for a competitor’s web site) lose, good guys (consumers like the plaintiffs in Netscape) win. While the policy behind these results might be appealing, the cases provide no framework at all for ascertaining whether or not browse-wrap terms generally should be enforced.”)

227 Radin, supra note 223, at 1133.

228 MSN.com, Terms of Use, § 4, http://privacy2.msn.com (last visited Oct. 7, 2006). (“You will not use the MSN Web Sites in any way that is unlawful, or harms Microsoft, its affiliates, resellers, distributors, service providers and/or suppliers (each, a “Microsoft Party” and collectively, the “Microsoft Parties”) or any customer of a Microsoft Party, as determined in Microsoft’s sole discretion.”)

229 Radin, supra note 223, at 1128, 1155-57. (“Perhaps we could make the problem go away simply by replacing the ordinary discourse contract-as-consent model with the economists’ contract-as-product model.”)


232 Some libraries, including Yale, do have academic contracts to the online material from New York Times, but Times Select is a contract only users can buy. The content may or may not overlap.
Times Select material is archived, if at all, by the New York Times only. It could be destroyed or forgotten. Times Select is using restrictive contract to avoid publishing to libraries and to prevent anyone from owning an archival copy. In the long term, Times Select has contracted to erase itself from the historical repository.

The Wall Street Journal contract does not even explicitly allow personal printing, and it allows no electronic copies whatsoever. The Church of Scientology terms of service forbid all copying including printing and probably caching and temporary storage in RAM. Looking through major news publications, I could not find a single terms of use agreement that would allow for historical archiving. Sophisticated authors are contracting away that feature, which generates no revenue, or looking to commoditize the archive, like the New York Times has already done.

Contracts with institutions like universities also block archiving. Expensive database subscriptions, such as Lexis-Nexis or Medline, are severely restricted to paying users by contracts, passwords, and other protective measures. UCITA and common law validate clickwrap licenses on libraries, which are sometimes not clear until the work is already bought. For subscribing institutions, these contract terms are very important and must be well-documented to

233. Wall Street Journal Online, Subscriber Agreement, http://online.wsj.com/public/page/subscriber_agreement.html (last visited April 14, 2006) ((i) You may occasionally distribute a copy of an article, or a portion of an article, from a Service in non-electronic form to a few individuals without charge, provided you include all copyright and other proprietary rights notices in the same form in which the notices appear in the Service, original source attribution, and the phrase “Used with permission from The Wall Street Journal Online” or “Used with permission from Barron’s Online.” Please consult the Dow Jones Reprints web site if you need to distribute an article from a Service to a larger number of individuals, on a regular basis or in any other manner not expressly permitted by this Agreement. (ii) You may occasionally use our “E-mail This” service to e-mail an article from a Service to a few individuals, without charge. You are not permitted to use this service for the purpose of regularly providing other users with access to content from a Service.”)

234. Scientology.org, Notice for Materials Copyrighted to Church of Scientology International, http://www.scientology.org/csi.htm, (“Users are not authorized to download or transmit any of these materials electronically, or to otherwise reproduce any of the materials in any form or by any means, electronic or mechanical, including data storage and retrieval systems, recording, printing or photocopying.”) Despite these harsh terms of service, scientology.org has a minimally restrictive robots.txt file blocking one directory and no other restrictions given to robots. Scientology.org, Robots.txt, at http://www.scientology.org/robots.txt (last visited April 12, 2006), (“User-agent: * I Disallow: /AdBooking/ I Disallow: “”).

235. LESSIG, supra note 10, at 281-82.

avoid breach. To keep track of all these contracts, Yale has a very large website with charts of acceptable terms for each database, as well as several library licensing specialists.237

For example, Yale’s Ovid (Medline) subscription allows ILL, and personal printing, but no archiving.238 Lexis’ Academic Universe forbids ILL, and Title Source II allows only staff to print and no walk-in usage.239 These databases use contract to supersede the traditional and statutory balance on library functions, such as ILL.240 A hardcopy book held by Yale’s library could never contain these restrictions. Archival copying, called “e-reserves” in the chart, is allowed on a small minority of Yale’s database subscriptions.

Additionally, the DMCA prohibitions of circumventing copyright protections can limit factors like printing access restrictions to passwords.241 Archives qualify for two DMCA exceptions: circumvention to determine whether they should acquire a work,242 and circumvention of obsolete computer media for the purpose of archiving it.243 Neither can apply to work that is never sold. The DMCA keeps content out of libraries by giving protections like New York Times’s passwords the force of law.

### D. Cybertrespass

The bots that have garnered the most legal attention have been a different breed from archival bots; they are malicious, economically driven competitor’s bots. In 2000, a California court applied a trespass to chattels claim and issued an injunction against a competitor’s use of a bot to collect pricing information from the popular eBay auction site in *Ebay, Inc. v. Bidder’s Edge*,244 creating a new movement in Internet enclosure. Cybertrespass relies loosely on the real property

---

237. Yale University, Permitted Uses of Online Resources, http://resources.library.yale.edu/online/licensing.asp (last visited Apr. 10, 2006).
239. Yale University, Permitted Uses of Online Resources, http://resources.library.yale.edu/online/licensing.asp (last visited Apr. 10, 2006).
240. Gasaway, *supra* note 44.
concept of trespass to chattels, that is interference with personal property.245

The doctrine of cybertrespass suggests that robots can be liable for just visiting areas they are not wanted.246 Cybertrespass has created a new exclusionary property right in websites.247 This exclusionary power could be even more restrictive than a copyright. Chattels could be used to enclose copyright material in an entirely different set of property rights, eliminating fair use and other copyright exceptions,248 just like restrictive contracting. Thus both contract and property can lock down content.

The case for cybertrespass is an economic one based on past bad behavior of robots, like the robots using eBay’s own site against it.249 Advocates for a bright-line right to exclude without real damages laud the clarity and efficiency of the trespass doctrine.250 When the policies of the Internet and the bots patrolling it as commons are too harsh, they argue, content owners should be able to shut themselves off from the commons.251

The problem with the economic basis for trespass is that it assumes all robots are equal, that is that robots are agents of private parties doing some economic action that affects the network dynamics of the system.252 Archiving robots, even if dramatically increased in scale, would be unable to burden a server the way Bidder’s Edge’s bots did.253 Archival copying, by definition, has no impact on the market until the copyright has expired, enough lag to be no competition at all. Bots that cause some colorable harm to the

246. Id.
248. Bellia, supra note 200, at 2197.
250. Id. at 337-43.
251. Id. at 345.
252. See, e.g., id. at 327-29.
253. Even the most benign bots do have the ability to destroy poorly designed websites. For example, Googlebot has accidentally destroyed sites while accidentally taking administrator power over a site. Web authors are well aware, and plan for, this problem. See, e.g., Rael Dornfest, Google Web Accelerator Considered Overzealous, O’REILLY RADAR, May 6, 2005, http://radar.oreilly.com/archives/2005/05/google_web_acce_1.html (last visited Sept. 16, 2006).
systems of an author, as Bidder’s Edge did to eBay, remain distinguishable.\textsuperscript{254} Also, archival robots have a special public interest and unique market failure at heart, cultural preservation.

Cybertrespass, as implemented in \textit{Bidder’s Edge}, has blurred the boundaries of denying access to copyright materials. It seems that the right of enclosure has the possibility to create a new intellectual property right to limit access to undesirable readers.\textsuperscript{255} Cybertrespass can allow an author the unmatched ability both to globally broadcast his work and to opt out of the library by denying library robot access or archiving. Under the first sale doctrine, it would be copyright misuse to forbid sale (or resale) of a book to a critic or a library; but eBay’s website, because it is not published in a physical book, can do just that.\textsuperscript{256} eBay may have the right to refuse certain customers, or even competitors, from its private servers, but eBay should not have the right to deny historians the ability to preserve its historically significant catalog.

This Article draws no conclusion about the doctrine of cybertrespass in general, but in the context of the archival preservation, cybertrespass should yield. Cybertrespass is chilling the archives because it is a murky doctrine at best, and there is no historical preservation exception.\textsuperscript{257} Objects of cultural value are often in private hands, where access is limited.\textsuperscript{258} However, on the Internet, limited access correlates to a limited archive, and, accordingly, destruction of cultural property, as discussed in Part II. Real property analogues and contracts in websites create a world in which sophisticated authors can display appropriate robot commands and effectively remove their content from the vaults of libraries, even though the content is effectively in global circulation.

**VI. A New Kind of Moral Right**

The combination of unsympathetic copyright law and an intimidating contract and cybertrespass ESPER has created a world in which an author has the right to own the only legal copy. When an author removes that copy from access, he has destroyed it altogether. This right to withdraw has never previously existed in United States

\begin{footnotes}
\textsuperscript{254} See Bidder’s Edge, 100 F. Supp. 2d at 1058.
\textsuperscript{255} See Burk, \textit{supra} note 247 (tort creates new IP right); O’Rourke, \textit{supra} note 26, (policy & preemption concerns)
\textsuperscript{256} See Fischer, \textit{supra} note 29, at 170.
\textsuperscript{257} Id.
\textsuperscript{258} Sax, \textit{supra} note 118, at 1544.
\end{footnotes}
copyright, but other law has cobbled a de facto right to destroy work, creating a legal regime giving authors the right to opt out of history.

The right to withdraw, or droit de retrait ou de repentir, is a very rare European moral right, or droit moral, and theoretically gives an artist a limited right to reclaim and destroy published work. The French concept of droit moral is a restriction on property rights, but it is not intended to preserve, at least not in the way we think of preservation of cultural artifacts like historic buildings.\textsuperscript{259} Droit moral belongs to the artist, and it is intended to protect his reputation and personality as embodied by his art.\textsuperscript{260} Thus, some embodiments of droit moral do not restrict destruction at all, or they prevent some other reputation harming action, such as display of mutilated works.\textsuperscript{261}

The international codification of this moral right, the Berne Convention’s “integrity right,” is an affirmative right against defacement or “derogatory” treatment which would have previously required libel law in some countries.\textsuperscript{262} Destruction of an entire work, ironically, does not trigger Berne inquiry because it is not in itself prejudicial to an author’s reputation, thus some other act must accompany the destruction, such as a public destruction.\textsuperscript{263} Moral rights preserving some artists’ rights have also been popular in non-Western jurisdictions.\textsuperscript{264} For example, Indian courts have been sympathetic to pleas using Berne Convention codifications of the integrity right.

The United States’ implementation of the Berne Convention is also based on the dignity of the artist.\textsuperscript{265} The Visual Artists Rights Act of 1990 (VARA) amends the 1979 Copyright Act, particularly in 17

\begin{footnotes}
\item[259.] Sax, supra note 118, at 21-22
\item[260.] Sax, supra note 118, at 22.
\item[261.] Id.
\item[262.] COPINGER AND SKONE JAMES ON COPYRIGHT 625 (Kevin Garnett et. Al. eds., 14th ed. 1999).
\item[263.] Id. at 630.
\item[264.] Mira T. Sundara Rajan, Moral Rights and the Protection of Cultural Heritage: Amar Nath Sehgal v. Union of India, 10 INT’L J. CULTURAL PROP. 79, 80 (“In countries as diverse as India, Russia, and Mali, moral rights have been adopted to serve objectives of cultural policy.”).
\item[265.] Id. at 86 (“The cry is ‘Ils ne passeront pas!’ and in such a situation, Indian courts will always be found dynamic and responsive.”).
\item[266.] H.R. Rep. No. 101-514, at 15 (1990), reprinted in U.S.C.C.A.N. 6915, 6925 (“An artist’s professional and personal identity is embodied in each work created by that artist. Each work is a part of his or her reputation. Each work is a form of personal expression (oftentimes painstakingly and earnestly recorded). It is a rebuke to the dignity of the visual artist that our copyright law allows distortion, modification, and even outright permanent destruction of such efforts.”).
\end{footnotes}
U.S.C. §§ 101, 106A, and 113. Berne requires moral rights to last as long as economic rights, but Congress codified the integrity right to last for a significantly shorter span, the life of the author.\textsuperscript{267} VARA protects "intentional distortion, mutilation, or other modification of the work which would be prejudicial to [the artist's] honor or reputation" or the "destruction of a work of recognized stature."\textsuperscript{268} However, VARA's strict definitions, including exclusion of works for hire, have ensured that extraordinarily few artists have qualified for protection by VARA in court.\textsuperscript{269} Digital artwork never qualifies for VARA protection.\textsuperscript{270}

The French droit moral allowing withdrawal, droit de retrait, theoretically gives an artist ultimate control over the work, allowing an artist to withdraw it entirely from culture, even after selling.\textsuperscript{271} However, withdrawal is the most restricted moral right and requires special indemnifications to purchasers.\textsuperscript{272} Sometimes, French courts have allowed withdrawal for unpublished manuscripts where publishing contracts with indemnification clauses would allow it.\textsuperscript{273} In extreme cases, some artists think they can reclaim or rework already sold paintings and generally fail.\textsuperscript{274} Even where the right to withdraw exists in France, it is often ineffective when it is practically impossible to reclaim sold works.\textsuperscript{275} Droit de retrait ends when the artist loses a property right in the work; French courts deny requests to recall sold books.\textsuperscript{276} Thus, even in France, publication of books revokes the author's moral right to destroy his own work because it is another individual's property as well as cultural property.

\begin{thebibliography}{99}
\bibitem{271} Susan P. Liemer, Understanding Artists' Moral Rights: A Primer, 7 B.U. PUB. INT. L.J. 41, 54-55 (1998) ("Even after deciding she has completed that process, she may change her mind.").
\bibitem{274} Sax, supra note 118, at 42-43 (referencing Soutine and Whistler).
\bibitem{275} Paul Goldstein, Goldstein on Copyright § 15.24.4 (1989).
\end{thebibliography}
Droit de retrait stops at the border and French authors have no right to reclaim works in other jurisdictions.\textsuperscript{277} Italy, Germany, Spain, and Belgium have codified a limited droit de retrait.\textsuperscript{278} In these jurisdictions, an author must withdraw the work only because he no longer supports it, and he must compensate those with pre-existing rights.\textsuperscript{279} Application outside France has been even more limited.\textsuperscript{280} Canadian courts have upheld droit de retrait only when the works at issue are unpublished manuscripts and authors and publishers disagree.\textsuperscript{281} No droit de retrait regime has ever allowed a mass recall of globally published books.

The United States never recognizes the moral right to revoke any kind of work, especially a published work. President Clinton’s Taskforce on the Internet produced a whitepaper that does not explicitly reject moral rights for Internet works, but suggests that the United States resist expanding the scope of the Berne Convention moral rights.\textsuperscript{282} The United States has only codified the right to integrity in VARA and the right to attribution.\textsuperscript{283}

Unpublished works are perhaps the most sympathetic application of droit de retrait, and the only circumstance in which a U.S. author can withdraw in limited circumstances. An artist presents a published work which he considers worthy of his reputation, whereas unpublished work, such as an undelivered speech by Churchill, could be destroyed because it presents an unfavorable image.\textsuperscript{284} Indeed, artists such as Rouault or Brahms, who famously discarded unfinished, unpleasing art, deprive the world of no value

\textsuperscript{279} Dietz, Legal Principles, supra note 278.
\textsuperscript{280} Id. at 61.
\textsuperscript{284} Sax, supra note 118, at 43.
because, by their own definitions, this was never their true work. Sax argues the same is true for unpublished manuscripts, even when the request is posthumous destruction. Strahilevitz agrees with Sax, arguing that an author should be able to destroy unpublished works in a will to balance avoiding waste with compelled speech. Practically speaking, executors destroy unpublished papers constantly.

Libraries constantly struggle with public access to unpublished materials, such as intimate or embarrassing personal letters, medical records, and attorney-client communications. At Yale, Langston Hughes's unpublished papers were denied to an unofficial biographer. These exclusionary actions can be based on spite, economic competition for a biography, or the economic interests of heirs as owners of lucrative property. Unpublished works have a privacy right that publications on the Internet simply do not. Even a personal blog is effectively published to the entire world. Additionally, copyright duration dilutes the privacy interest in even the most sensitive material in the archive. Only in extreme cases could that privacy right outweigh the public interest of the archive in the long term.

Despite the United States rejection of droit de retrait, we have manufactured one using bits from other law. Contract and trespass have become an ill-planned ESPER allowing authors to withdraw from section 109 allowances, eliminating sale, limiting transfer, and

285. Id. at 43.
286. Id. at 44-47.
288. Sax, supra note 118, at 118-120.
289. Id. at 143.
290. Id. at 123 (“The Jung family in Switzerland, which was unhappy with critical statements in Noll’s previous writings (he had called Jung “the most influential liar of the 20th century”), insisted that the library deny Noll access to the papers. The library complied.”)
291. Id. at 117. Sax suggests letters should only be restricted from public access between twenty-five and fifty years after the death of all contributors. Id. at 127-128
292. Obviously, to codify any kind of right, it is important to define what is published. In the context of online publishing, I believe anything available without restrictions to the public or anything available for a fee is published. It becomes important to distinguish no-fee protected websites, such a family's photoalbum passworded for privacy, from publishers using passwords to invoke ESPERs, such as New York Times. Even though the New York Times restricts access, it is clear the work is published. An easy way to differentiate these actions might be by the number of visitors, an easy proxy for cultural impact. A more difficult way would be by intent. Another solution, drawing on the FBI investigation standards, might allow archiving anything an archive could get access to, by whatever means, even fraudulent. I leave out email from this inquiry.
limiting archiving. In a world with no first sale doctrine, authors own the only copy. Each website becomes a singular work of art, owned by a capricious author, unchecked by law. Contract and cybertrespass have overridden American copyright principles and created a de facto copyright the United States has never embraced and created a world with an unprecedented right to withdraw work from human memory.

VII. Conclusion

The Internet is disappearing fast, and the cultural expression of our generation is going with it. Intellectual property is thought and creativity, culture itself, and the bedrock of future expression and production. In a world with no books to be preserved by libraries, archives have a special duty to make sure that freeform ideas are saved somewhere so future generations can use, understand, and build on them.

The law affecting Internet archives has lost touch with the constitutional value of speech and of the commons. An unsympathetic copyright regime offers no protection for the archives' historical mission, and that mission is too important to settle with a judicially manufactured copyright exception. Copyright should be unambiguously on the side of those seeking to save culture.

A regime of contract and cybertrespass has further blocked archival efforts, eliminating any doubt that copyright might be the solution for archiving. Authors are contracting out of libraries, and out of history. Restrictive contracting and complicit copyright policies have allowed authors the ability to globally broadcast their work and simultaneously lock it away from the inquiry of history. Internet archives are losing this legal battle, but without real change society will be the ultimate loser. We must, as a culture valuing speech as much as our forefathers, lock down a copy of culture before it is locked away forever.
***