

ORPHAN WORKS REPLY COMMENTS

of THE INTERNET ARCHIVE

May 9, 2005

The Internet Archive thanks the Copyright Office for this opportunity to reply to the initial series of comments submitted in response to its Notice of Inquiry regarding orphan works.¹ The breadth and number of comments submitted make clear that the problem presented by orphan works spans a wide range of types of works (i.e., computer programs, books, pictures, movies, music, websites) and uses of works (i.e., commercial, non-commercial, personal, charitable, educational, preservational). Many commenters either proposed solutions that encompass all types and uses of orphan works, or acknowledged that an appropriate solution should be broad to address all types and uses.

The Internet Archive is in agreement with the general consensus that an appropriate solution should address all types and uses of orphan works, not just a solution that applies to libraries and archives. A solution to the problem, however, would not be complete without addressing the mission of libraries and archives to make culturally and socially important works and collections broadly available, particularly with respect to those libraries wanting to utilize digital technology in doing so. As many have suggested, digitization projects place a new and growing demand on those attempting to seek and obtain permission for copyrighted work.² The Internet Archive would like to take this opportunity to elaborate on its experience with digital archiving, and to discuss a proposed safe harbor³ solution that would address the important preservation and access goals central to the Internet Archive and many other organizations.⁴

*Put simply, we propose that libraries and archives should be permitted to **preserve** all works in their collections, including those determined to be orphans. We further propose that libraries and archives should be permitted to provide **access** to orphan works without fear of liability subject to a “Request and Removal” procedure that would respect the rights of copyright owners who come forward. Under this system, libraries and archives providing digital access to orphan works would be substantially protected from liability upon adhering to recommended best practices and guidelines for determining whether a work is an orphan, and*

¹ Notice of Inquiry, 70 Fed. Reg. 3739 (Jan. 26, 2005).

² See, e.g., Library Copyright Alliance OW0658, JSTOR et al. OW0680, DigiBarn Computer Museum OW0049, MIT Carol Fleishauer OW0515, North Carolina State University Libraries OW0606, Kernochan Center at Columbia Law School OW0666, Microsoft, Inc. OW0695, *Orphan Works Comments*, March 25, 2005, at <http://www.copyright.gov/orphan/comments/index.html>.

³ At least five other commenters have proposed solutions involving safe harbors or request and removal practices. See Stanford University Libraries OW0457, The J. Paul Getty Trust et al. OW0610, North Carolina State University Libraries OW0606, Library of Congress OW0630, JSTOR, et al. OW0680, *Orphan Works Comments* March 25, 2005, at <http://www.copyright.gov/orphan/comments/index.html>.

⁴ The Internet Archive is a nonprofit library that has collected billions of works (books, music, moving images, web pages, and software programs) and served millions of users since 1996. The Internet Archive works with the Library of Congress, the National Archives and Records Administration, the Bibliothèque Nationale de France and the British National Archives as well as policymakers and standards committees to find workable solutions to libraries' and archives' missions in the digital world. The Internet Archive's collections can be found at <http://www.archive.org>.

ORPHAN WORKS REPLY COMMENTS OF THE INTERNET ARCHIVE

upon compliance with appropriate procedures for protecting the interests of copyright holders, including implementing a “request and removal” procedure for the work in question.

Archival Preservation

Libraries and archives should be permitted to preserve all works in their collections, including those determined to be orphans.

Libraries and archives are often the only means of preservation and public access to socially, culturally, and historically important works. The Library of Congress was built around this fundamental premise, pronouncing its mission “to make its resources available and useful to the Congress and the American people and to sustain and preserve a universal collection of knowledge and creativity for future generations.”⁵ In light of the fundamental educational role archives and libraries play in society, copyright law recognizes the importance of preservation in several ways. Section 108 of the Copyright Act permits off-air recording of broadcast news programs by libraries under certain conditions for the purposes of archiving and public access, in recognition of the importance of archiving works.⁶ It further gives libraries and archives the opportunity to make copies of a work and distribute them for purposes of preservation and safekeeping provided that certain conditions are met.⁷ Recent amendments have allowed for digital reproduction,⁸ and the Family Entertainment and Copyright Act of 2005 has further expanded the list of allowable media to include musical works, pictorial, graphic or sculptural works, and motion pictures or other audiovisual works.⁹ Digital technology enables libraries and archives to fulfill their traditional function of preservation with unprecedented efficiency and integrity. Accordingly, any new rule regarding orphan works should permit, if not encourage, libraries and archives to use digital technology during the preservation process. In practice, this will often mean making a digital copy for preservation purposes.

A primary mission of the Internet Archive is to preserve and provide access to as many web pages, books, films, and other material of cultural, historical and literary significance as possible. Currently, the Internet Archive’s digital collections comprise 13,845 moving images,

⁵ Library of Congress, *The Mission of the Library of Congress*, at <http://www.loc.gov/about/mission/index.html>.

⁶ See 17 U.S.C. § 108(f)(3) (2000).

⁷ Before undertaking any copying under subsection (c), a library must make a “reasonable effort” to obtain “an unused replacement ... at a fair price.” See 17 U.S.C. § 108(c) (2000). Unfortunately, without additional clarity, the rights granted by this section are of limited utility. See also Library of Congress OW0630, *Comments of the Library of Congress In Response to the Copyright Office Notice of Inquiry “Orphan Works,”* at 2, March 25, 2005, at <http://www.copyright.gov/orphan/comments/index.html> (“The terms ‘reasonable investigation’ and ‘normal commercial exploitation’ are not defined and are subject to varying interpretation.”).

⁸ See Copyright Term Extension Act., Pub. L. No. 105-298, 112 Stat. 2827 (codified as amended at 17 U.S.C. § 108(h)-(i) (1998)).

⁹ Title IV of that Act is entitled Preservation of Orphan Works and provides that the limitation on rights of reproduction and distribution of copyrighted works does not apply to the authority of libraries or archives, during the last 20 years of any term of copyright of a published work, to reproduce, distribute, display, or perform in facsimile or digital form a copy or phonorecord of such work for purposes of preservation, scholarship, or research when certain conditions apply. Family and Entertainment Copyright Act of 2005, Pub. L. 109-9, 119 Stat. 218 (signed by the President on April, 27 2005) (to be codified at 17 U.S.C. §108(i)).

ORPHAN WORKS REPLY COMMENTS OF THE INTERNET ARCHIVE

20,611 texts and books, 34,825 audio recordings, 31,477 software titles, and over 40,000,000,000 web pages.¹⁰ In our experience, we have not encountered opposition to our practice of preserving works such as web pages that have been made publicly available unless the copyright owner objects either with automated means or by contacting us.¹¹ In fact, from time to time we are approached by large website operators who themselves have not retained copies of their websites. For example, USA Today discovered that they did not have a copy of their website from September 11, 2001 and asked us if we had made such a copy. And Yahoo!, Inc., upon its tenth anniversary, used the Internet Archive for early copies of its website. The Copyright Office also recognized the importance of the Internet Archive's preservation mission when it granted an exemption to the Digital Millennium Copyright Act that permits the creation of preservation copies of endangered software that happen to be copy-protected.¹² This exemption reflects the overwhelming societal interest in archiving works even in instances where the rights holder has used technology that legally and practically frustrates that societal interest.

A number of comments submitted to the Copyright Office address the need for archives and libraries to be able to preserve and provide public access to culturally valuable orphan works. As but one example among many, the Center for the Study of the Public Domain at Duke University Law School discusses in some detail the importance of preservation of films having long-term historical and cultural value.¹³ The Internet Archive shares the deep and well-founded concerns that those using digital technology for preservation and access have expressed regarding orphan works. If archives and libraries were explicitly permitted to make digital copies for purposes of preservation, these profound concerns regarding orphan works would be assuaged without undermining the rights of copyright owners who do come forward.

Orphan Designation

The Internet Archive agrees with the recommendation of many commenters that a categorical approach should be taken towards determining whether appropriate efforts have been made to locate the rightful owner of the copyright in a work. Under such an approach, a user would be required to follow generally applicable guidelines rather than apply to a board or regulatory body before using the work.¹⁴ Such an approach should take into account the distinct

¹⁰ See Appendix B, which provides additional details about the scope of the Internet Archive's collection.

¹¹ The DigiBarn Computer Museum, an online museum dedicated to the history of the personal computer, notes that with respect to over 100,000 artifacts over several years, not one copyright owner has invoked DigiBarn's "escape clause" permitting the owner of the copyright in any work to request that DigiBarn remove his or her work from its site. See generally DigiBarn Computer Museum *OW0049, Comment on Orphan Works Notice of Inquiry*, March 25, 2005, at <http://www.copyright.gov/orphan/comments/index.html>.

¹² See Exemption to Prohibition on Circumvention of Copyright Protection Systems for Access Control Technologies, 68 Fed. Reg. 62011 (Oct. 31, 2003), available at <http://www.copyright.gov/fedreg/2003/68fr2011.html>.

¹³ Among others were the UCLA Film and Television Archive *OW0638*, the National Film Preservation Foundation *OW0541*, the American Historical Association *OW0676*, JSTOR et al. *OW0680*, the Library Copyright Alliance *OW0658*, and John Ockerbloom *OW0674*.

¹⁴ For some types of media, the guidelines might be assembled by industry groups that work regularly with that media type. Variants on a "best practices" approach are recommended by a number of commenters, including the Center for the Study of the Public Domain at Duke University Law School *OW0597*, the UCLA Film and Television

ORPHAN WORKS REPLY COMMENTS OF THE INTERNET ARCHIVE

differences between the various types of works encountered by libraries and archives involved in digitization projects.

The Internet Archive digitally preserves publicly available traditionally published works such as books, music, movies, pictures, and other moving image media that are typically registered and catalogued in the Library of Congress. We believe that the public will be best served if archives and libraries involved in the preservation of such works can rely on relatively clear requirements for determining orphan status. A best practices approach, in which members of each community confer and develop guidelines for reasonable efforts to determine whether a work is orphaned, may be the optimal means by which to achieve this clarity.¹⁵

In addition, the Internet Archive also preserves more transient, often internet-specific works such as websites, shareware, software, and blogs that were publicly available at one time but are of such a temporary or ephemeral nature that they often are deleted, altered or become unavailable to the public almost immediately. The ephemeral nature of this transient content requires a unique approach when it comes to digital preservation and access. The Internet Archive stores over 500 terabytes of ephemeral web pages, book and moving images, adding an additional twenty-five terabytes each month. The short life span and immense quantity of these works prompts a solution that provides immediate and efficient preservation and access to orphaned ephemeral works. For instance, the average lifespan of a webpage is 100 days before it undergoes alteration or permanent deletion,¹⁶ and there are an average of fifteen links on a webpage.¹⁷ The high degree of interconnectivity, alteration, and malleability among internet-specific content alone is a cause for concern in the absence of clear measures for preservation of and access to such content.

The best practice guidelines for determining orphan status should take into account the ephemeral, transient nature of this type of work. *Since works that fall into the category of transient, ephemeral works are collected in mass quantity, last only temporarily, and provide owners an opportunity to initiate automated protections such as robot exclusions,*¹⁸ we

Archive OW0638, the Glushko Samuelson Intellectual Property Law Clinic OW0595, the J. Paul Getty Trust, et al. OW0610, the North Carolina State University Libraries OW0606, JSTOR et al. OW0680, the Recording Industry Association of America OW0687, Various Independent Film Interests OW0663, and Public Knowledge OW0629.

¹⁵ Generally speaking, we don't believe that the age of a work or its publication status should be determinative of its orphan status. The Library of Congress, among others, has noted that under the description of an orphan work provided by the Copyright Office in its Notice of Inquiry—a work whose copyright owner is difficult or impossible to locate—any operational definition should “focus on whether the owners of copyright-protected works can reasonably be located.” See Library of Congress OW0630, *Comments of the Library of Congress In Response to the Copyright Office Notice of Inquiry “Orphan Works,”* at 1, March 25, 2005, at <http://www.copyright.gov/orphan/comments/index.html>. Accordingly, “[a]ny definition should consider including unpublished works and works of any age.” *Id.*

¹⁶ This was determined by Alexa Internet in 2001 upon the launch of the Wayback Machine at the Bancroft Library, Berkeley, CA.

¹⁷ This was determined by Alexa Internet through its experience crawling the web during the 1990s.

¹⁸ A robots.txt file is a widely used method for controlling the behavior of automated programs that view and save copyrighted sites (all major such robots, including those of Google and Alta Vista, respect these exclusions). See School of Information Management and Systems, U.C. Berkeley, *The Oakland Archive Policy: Recommendations*

ORPHAN WORKS REPLY COMMENTS OF THE INTERNET ARCHIVE

recommend for purposes of efficiency and feasibility that the reasonable search requirements for these works be severely relaxed or eliminated completely. This would allow archives and libraries engaged in digital preservation of transient works to fully utilize digital technology to archive large amounts of such content systematically, and provide them greater flexibility in terms of providing access to these works.

As we discuss below, once a library or archive determines that a work is an orphan in compliance with accepted reasonable search guidelines, it should be permitted, without fear of liability, to provide immediate public access to the work subject to appropriate precautions designed to protect the rights of copyright holders who come forward.

Access to Archival Collections

An orphan works solution must promote public access to the vast amount of ephemeral information being created every day. Online archives and digital libraries collect and preserve publicly available works, both traditional and transient, for the future use of historians, researchers, scholars, and the general public. These archives and digital libraries strive to operate as trusted repositories for these materials, and work to make their collections as comprehensive as possible. Without the ability to provide public access to these collections, however, the laudable preservation efforts of archives and libraries nationwide are significantly less useful.

We recommend that once a digital work is determined to be an orphan, in accordance with clearly established criteria or generally accepted best practices, an archive or library should be permitted to provide public access to that work without incurring liability upon implementing a request and removal system for copyright owners that surface and object to online access to their work. The effect of such a regime would be not unlike the safe harbor and notice and takedown system already in place under the Digital Millennium Copyright Act. We are confident that libraries and archives can balance the interests of copyright holders and the public by implementing a request and removal system that allows rights holders to prevent their works from being archived and to remove their works from the archive.¹⁹

Similar to reasonable search “best practices,” request and removal procedures also could be developed by professional organizations in various relevant disciplines. All stakeholders, including content providers, copyright holders, and end-users, have a common interest in establishing a predictable and efficient framework. We have designed a set of general guidelines for proper request and removal of copyrighted works from archival websites as set forth in

for Managing Removal Requests and Preserving Archival Integrity (2002), at <http://www.sims.berkeley.edu/research/conferences/aps/removal-policy.html>.

¹⁹ This new regime would not, however, replace fair use defenses or Section 108 exceptions. Additionally, just as in non-digital libraries, by providing the public access to archival collections, this system is not meant to undermine other parts of copyright law that govern usage restrictions on a subsequent user’s use of a work contained in a library or archive’s collections. Furthermore, our proposal would not prevent users from making further use of any more broadly applicable orphan works regime that may be promulgated.

ORPHAN WORKS REPLY COMMENTS OF THE INTERNET ARCHIVE

Appendix A. These guidelines were adapted from the Oakland Archive Policy,²⁰ a uniform request and removal strategy developed in cooperation with policy experts, libraries and archives, and major search engine companies.

In the event that a copyright holder comes forward, a library or archive would be protected from liability under a safe harbor against all claims of copyright infringement so long as it adheres to established reasonable search guidelines, and implements request and removal procedures as outlined above. Where a library or archive fails either to conduct a reasonable search for the copyright holder in compliance with accepted guidelines, or to implement request and removal procedures as outlined above, it will not be able to avail itself of the safe harbor and the copyright holder can rely on his or her full panoply of rights under the Copyright Act.

Conclusion

The Internet Archive reiterates that its proposal,²¹ which is tailored to archives and libraries, is not meant to exclude other broader, more universally applicable plans based on a reasonable efforts search leading to a limitation of remedies. We support such approaches and present our proposal as a companion to a more comprehensive plan such as those proposed by the Copyright Clearance Initiative,²² the Center for the Study of the Public Domain at Duke University Law School,²³ and a coalition of Various Independent Film Interests.²⁴

In sum, we propose that at the very least, libraries and archives ought to be permitted:

- *to preserve the works in their collections, which often means making digital preservation copies;*
- *to designate a work as orphaned after a “reasonable efforts” search for the copyright owner – the requirements for determining which should be significantly relaxed or even eliminated for a certain class of transient, ephemeral, often internet-specific works; and*

²⁰ See School of Information Management and Systems, U.C. Berkeley, *The Oakland Archive Policy: Recommendations for Managing Removal Requests and Preserving Archival Integrity* (2002), at <http://www.sims.berkeley.edu/research/conferences/aps/removal-policy.html> (setting forth “best practices” developed by the Internet Archive and several other archiving organizations for eliminating access to content upon request by copyright owners and other interested parties).

²¹ The Center for the Study of the Public Domain at Duke University Law School and the DigiBarn Computer Museum have endorsed the Internet Archive’s proposal for a safe harbor for archives and libraries that engage in digital preservation and access subject to appropriate procedures designed to protect copyright holders’ interests.

²² Copyright Clearance Initiative OW0595, *Response to Notice of Inquiry on the Issue of “Orphan Works,”* March 25, 2005, at <http://www.copyright.gov/orphan/comments/index.html>.

²³ Center for the Study of the Public Domain at Duke University Law School OW0597, *Orphan Works Analysis and Proposal*, March 25, 2005, at <http://www.copyright.gov/orphan/comments/index.html>.

²⁴ Various Independent Film Interests OW0663, *Orphan Works Comment*, March 25, 2005, at <http://www.copyright.gov/orphan/comments/index.html>.

ORPHAN WORKS REPLY COMMENTS OF THE INTERNET ARCHIVE

- *to provide public access to the orphan works in their collections subject to implementation of a request and removal system in which the work is removed upon request by copyright owners who come forward.*

We have based this proposal on the simple idea that the public benefits when libraries and archives can freely preserve and provide public access to orphan works in their collections. We are confident that with the appropriate safeguards in place, this idea can be realized without burdening copyright owners, resulting in unprecedented public access to knowledge and culture.

We thank the Copyright Office for the opportunity to comment on an issue of great moment to all who are concerned with the preservation of, and public access to, so many of our nation's culturally, socially, and historically important works. We hope you found our discussion and suggestions insightful as to the unique problems that orphan works present for libraries and archives. We look forward to participating in any further proceedings on the issue of orphan works, and we encourage you to contact us if you require additional information or further clarification on any aspect of our proposal.

ORPHAN WORKS REPLY COMMENTS OF THE INTERNET ARCHIVE

About the Authors of this Comment

The Samuelson Law, Technology & Public Policy Clinic at the University of California—Berkeley, School of Law (Boalt Hall), submits this comment on behalf of the Internet Archive. Samuelson Clinic Student Interns Amanjit Arora and Dean Cheley participated in the preparation of this Comment.

Please refer any questions and comments to:

Internet Archive
The Presidio of San Francisco
116 Sheridan Avenue
San Francisco, CA 94129
(415) 561-6767
<http://www.archive.org>

Brewster Kahle, Digital Librarian
(brewster@archive.org)
Rick Prelinger, Board President
(rick@archive.org)
Jeff Ubois, <http://archival.tv>
(jeff@ubois.com)

Samuelson Law, Technology & Public
Policy Clinic
396 Simon Hall
School of Law (Boalt Hall)
University of California—Berkeley
Berkeley, CA 94720
(510) 643-4800
<http://samuelsonclinic.org>

Jack Lerner, Clinic Fellow
(jlerner@law.berkeley.edu)

ORPHAN WORKS REPLY COMMENTS OF THE INTERNET ARCHIVE

APPENDIX A

Recommended Request and Removal Procedures for Digital Libraries and Archives With Respect to Archived Web Pages and other “Transient” Content

Type of removal request	Response
Request by a copyright owner or webmaster of a private (non-governmental) web site.	<ol style="list-style-type: none">1. Libraries and archives should provide a “self-service” approach copyright owners can use to remove their materials based on the use of the robots.txt standard.2. Requesters may be asked to substantiate their claim of ownership (eg. changing or adding a robots.txt file on their site.)3. This allows libraries and archives to ensure that material will no longer be gathered or made available.4. These requests will not be made public; however, libraries and archives should retain copies of all removal requests.
Third party removal requests based on the Digital Millennium Copyright Act of 1998 (DMCA).	<ol style="list-style-type: none">1. Libraries and archives should attempt to verify the validity of the claim by checking whether the original pages have been taken down, and if appropriate, requesting the ruling(s) regarding the original site.2. If the claim appears valid, libraries and archives should comply.3. Libraries and archives will strive to make DMCA requests public via Chilling Effects, http://www.chillingeffects.org, and notify searchers when requested pages have been removed.4. Libraries and archives will notify the webmaster of the affected site, generally via email.
Third party removal requests based on non-DMCA claims.	<ol style="list-style-type: none">1. Libraries and archives will attempt to verify the validity of the claim by checking whether the original pages have been taken down, and if appropriate, requesting the ruling(s) regarding the original site.2. If the original pages have been removed and the Library or archive has determined that removal from public servers is appropriate, then the library or archive will remove the pages from their public servers.3. Libraries and archives will strive to make these requests public via Chilling Effects, and notify searchers when requested pages have been removed.4. Libraries and archives will notify the webmaster of the affected site, generally via email

APPENDIX B

What's in the Archive

What's in the Archive?

www.archive.org

The Internet Archive is a digital library that stores and provides access to human creative works in digital form. Universal access to our cultural heritage is within our grasp; with current digital technology we can build comprehensive collections, and with digital networks we can make these available to students and scholars all over the world.

Total number of MOVING IMAGES :: 13,845

PROGRAMS IN THESE SAMPLE COLLECTIONS:

Computer Chronicles: 617	Prelinger Archives: 2,002
SIGGRAPH: 119	Open Source Movies: 3,987
Net Café: 152	MSRI Math Lectures: 458
Independent News: 245	Feature Films: 450
Election 2004: 677	Open Mind: 81
Machinima: 332	Brick Films: 355
Shaping San Francisco: 88	Videogame Previews: 2,279
Mosaic Middle East News: 436	ArsDigita University: 44
Sabucat Movie Trailers: 103	AV Geeks: 67
Film Chest Vintage Cartoons: 34	Media Burn Archive: 100
Universal Newsreels: 273	Skill Replays: 71
Software EPK's: 60	Democracy Now: 41
Yourh Media: 357	

Total number of TEXTS :: 20,611

BOOKS IN THESE SAMPLE COLLECTIONS:

Project Gutenberg: 7,453	Million Book Project: 10611
US Patents: 2,846	Arpanet: 197
Children's Library: 1,477	Canadian Libraries: 854

Total number of AUDIO RECORDINGS :: 34,825

RECORDINGS IN THESE SAMPLE COLLECTIONS:

Live Music Archive: 22,626	Netlabels: 4,999
Democracy Now: 549	Open Source Audio: 5,141
Other Minds: 202	Naropa Audio Archive: 292
Presidential Recordings: 127	Tse Chen Ling Buddhist Lectures: 152

Total number of SOFTWARE TITLES :: 31,477

TITLES IN THESE SAMPLE COLLECTIONS:

Tucows Shareware Library: 31,470	Clasp: 7
----------------------------------	----------

Total number of WEB PAGES :: over 40,000,000,000