

ORPHAN WORKS COMMENTS OF THE INTERNET ARCHIVE

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

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>
for the Internet Archive

The Internet Archive submits these comments regarding orphan works to the Copyright Office in response to the Notice of Inquiry issued on January 26, 2005.¹

Orphan works—works for which the owner of the copyright in the work is difficult or even impossible to locate²—raise difficult problems for libraries and archives. In other contexts the Internet Archive and other libraries and archives have developed practical, workable solutions to related problems, and we suggest that the Copyright Office and Congress could craft a plan for orphan works that draws from, and builds on, these real-world solutions.

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.

In these comments, we identify various problems that archives and libraries now experience with respect to orphan works, and real-world solutions that we recommend. We also identify several solutions that we developed in response to similar problems in other contexts, which we hope can be used as a foundation for a solution to the orphan works problem.

Problem 1: For libraries and archives, preservation of orphan works often requires making copies of them, creating the risk of legal liability. Preservation is one of the most important activities in which archives engage, and archives often seek to preserve works for which no copyright owner can be found. In many cases, the simple act of preservation places libraries and archives on ambiguous legal ground. In the absence of clear legal rules that permit unimpeded preservation, many institutions are understandably uncomfortable engaging in preservation efforts, and as a result many works are not being preserved.

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

² Our definition of the term “orphan works” is based on the Copyright Office’s own usage in its Notice of Inquiry, *see id.* (defining orphan works as “copyrighted works whose owners are difficult or even impossible to locate”). In practice, archives such as ours may be different than many users since we collect materials in bulk. As a result, it is crucial that libraries and archives be permitted to use automated techniques to categorize works and clear the rights to those works.

Solution: Explicitly allow libraries and archives to make preservation copies. The Internet Archive makes millions of digital preservation copies of orphaned or other works each year. In practice, we have experienced very little resistance to such activity, and in fact in our experience copyright holders are generally pleased that their works are being preserved for posterity. We believe that part of the solution to the orphan works problem should involve permitting archives and libraries to preserve orphan works using digital technology. Successful examples point the way to future practice. For example, the Copyright Act permits off-air recording of broadcast news programs by libraries under certain conditions for the purposes of archiving and making available to the public,³ in recognition of the importance of archiving works. Relatedly, in 2003 the Internet Archive secured an exemption to the Digital Millennium Copyright Act that permits the creation of preservation copies of endangered software that happen to be copy-protected.⁴ 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 interest. These regulations and laws could be adapted to cover orphan works without inventing wholly new legal structures.

Problem 2: For libraries and archives, providing public access to digitized orphan works via computer networks creates the risk of legal liability, and an orphan works solution that requires rights clearance for each work would severely hamper archives' ability to provide such access. In physical libraries, access to printed orphan works is usually uncontroversial and largely free from legal uncertainty, since such works are customarily stored on shelves and physically provided to patrons upon request. In the digital world, providing access to scholars and the public is more problematic because it often involves access via computer networks that could expose archives and libraries to legal liability. Libraries and archives need clear guidance on how they can provide access to digital and digitized orphaned works without fear of liability. Moreover, many archives collect both published and unpublished materials in bulk and seek to provide digitized, network access to such materials. On a daily basis, for example, the Internet Archive receives thousands of orphan works for which there might be many different rights holders. To conduct a non-automated search for each work's copyright holders would be impracticable if not impossible; a requirement to conduct such a search would effectively prevent archives and libraries from offering networked access to all but a tiny fraction of available works.

Solution: Allow archives and libraries to provide access to orphan works subject to certain restrictions and an easy "opt-out" system for rights holders. In our experience, archiving millions of moving images, books, web pages, and musical works with millions of owners—works that are both non-orphan and orphan, both "born-digital" and analog—we have found that a system with use restrictions and an easy opt-out process works well. More specifically, we have found that our practice of (1) requiring through our Terms of Use that users

³ 17 U.S.C. § 108(f)(3) (2000).

⁴ 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>.

make only non-commercial uses of the works and refrain from non-infringing uses (effectively prohibiting the unauthorized creation of derivative works), and (2) implementing an easy “opt-out” system that allows rights holders to prevent their works from being archived and to remove their works from the archive, provides a balanced approach that facilitates archiving while addressing the concerns of a limited number of rights holders who object to non-profit archiving. These precedents could be leveraged to build protection more broadly into law or regulation for archives and libraries that seek to preserve and provide access to orphan works. Below, we discuss each of these precedents in detail in an effort to help the Copyright Office understand more thoroughly both the precedents and how they can be used in developing a solution to the orphan works problems.

- While difficult to implement with physical objects, the Internet Archive’s opt-out⁵ system, in which websites are archived via an automated process unless website owners notify the Internet Archive, works well in the digital world since pulling objects off of websites is relatively easy. The system allows libraries and owners to balance their interests and adjust this balance over time. The organizations that search and archive the World Wide Web and Usenet have begun to develop a uniform opt-out strategy, and have jointly developed the Oakland Archive Policy in cooperation with policy experts, libraries and archives, and major search engine companies.⁶ The Policy sets forth five categories of removal requests and suggested actions in response to each type of request. We recommend that the Copyright Office and Congress consider taking this existing practice, adapting it to cover orphan works, and building it into a law or regulation addressing orphan works.
- “Robot Exclusions,”⁷ an important part of the Internet Archive’s opt-out system, are a distributed and automated way to indicate an owner’s intent regarding how his or her web page may be archived. A similar approach might be implemented for certain classes of orphan works. This approach has helped the Internet Archive to respect website owners’ preferences as to whether their web site should be made publicly available. An owner can also modify his or her robot exclusion, thus “taking down” materials that had been collected in the past. This approach has been in use since 1995 with great success. Automated techniques such as these to categorize works could be implemented so that

⁵ By “opt-out,” we mean a type of permission system in which an archive takes certain actions unless it is notified of an objection. In contrast, in an “opt-in” system an explicit signal is required before an action is permitted. Both types of permission can be automated or based on human contact, and often both types of permission systems are implemented. The Internet Archive uses an opt-out system in which it follows signals from owners using standardized, automated approaches as well as direct contact with the Archive.

⁶ *The Oakland Archive Policy: Recommendations for Managing Removal Requests And Preserving Archival Integrity*, available at <http://www.sims.berkeley.edu/research/conferences/aps/removal-policy.html>.

⁷ A “robot exclusion” is a small, computer-readable file placed on a website that indicates what parts of the site can be archived. A commercial example is available at <http://www.nytimes.com/robots.txt>; this allows automated access to front pages but not archives, and some organizations are allowed special access. All of this is automated and requires no human intervention or central registration. The standard is specified at <http://www.robotstxt.org/wc/exclusion.html>.

works that may appear to be orphaned, but are not, are not archived without the owner's permission.

- The DMCA's "notice and takedown" provisions providing a safe harbor for internet service providers who promptly remove allegedly infringing content where rights holders file a standard objection⁸ in practice function much like an "opt-out" system in that service providers can assume that all content on their system is un infringing unless they are otherwise notified. This system has proven to be a workable approach for ISPs and archives having diverse collections of content created by others. In response to this provision, copyright holders have created automated systems for finding suspected infringement and contacting service providers. A similar procedural framework could be crafted to allow for archiving and web publication of orphan works, while providing a streamlined process for rights holders to object and have works removed.
- The Copyright Act provides for non-profit libraries to loan a limited number of copies of news videos for non-profit use.⁹ While this provision does not directly address orphan works and network distribution, it is an example of a successful law allowing restricted access that might serve as a model for upcoming legislation or regulation. An amendment to the Copyright Act allowing non-profit libraries to provide a limited number of copies of orphan works for non-commercial, non-derivative use would further the mission of archives and libraries in a manner consistent with this provision.
- The current interlibrary loan system, which evolved after the advent of photocopying technology, permits limited copying and exchange between libraries so long as such use does not displace sales of the work.¹⁰ Many of these agreements have been in place for quite some time and have successfully allowed libraries to copy and loan copyrighted works. This concept may serve useful in establishing a threshold level of use that must be met before the user of an orphan work must provide compensation to a later appearing owner.
- Creative Commons¹¹ has developed licenses that permit only non-commercial uses of a work and prohibit the creation of derivative works. The language in these licenses might serve as a starting point for expressing restrictions on uses of orphan works in a manner consistent with the notion of protecting against loss of sales discussed above. Today, millions of rights holders use them to express a wide range of intentions with respect to the use of their works, and the availability of such licenses has facilitated the Internet Archive's negotiations with owners of collections. The Copyright Office can build on

⁸ 17 U.S.C. §§ 512 (c), (g) (2000).

⁹ 17 U.S.C. § 108(f)(3) (2000).

¹⁰ See 17 U.S.C. § 108(g) (2000); *see also* *Reproduction of Copyrighted Works by Educators and Librarians*, available at <http://www.copyright.gov/circs/circ21.pdf>.

¹¹ See <http://creativecommons.org/about/licenses/>.

this experience by crafting restrictions akin to those found in Creative Commons licenses for copies of orphan works held by libraries and archives.

With these comments, we have attempted to describe concrete issues that this archive, and presumably other archives, have been confronting with respect to orphan works, and share some solutions that have been developed in analogous circumstances. It is encouraging that, in practice, pre-emptive archiving with access restricted to non-commercial, non-infringing uses, combined with an opt-out system, has worked very well for the Internet Archive. We respectfully suggest that the market-based and regulatory solutions discussed above provide the foundation for a solution to the orphan works problem for libraries and archives.

We have prepared these comments with the fundamental premise in mind that orphan works frustrate the intent of our copyright system, which is to incentivize creativity towards the ultimate goal of providing the public access to the fruits of that creativity. In the vast majority of cases, works for which the copyright owner is difficult or even impossible to locate are works that are languishing rather than thriving. They are inaccessible and therefore underutilized, frustrating the public access goal of copyright. Because their owners are not readily identified, these works remain unexploited and therefore provide no financial incentive to the author to secure future production.

In addition, it is important to emphasize that these solutions do not address other uses of orphan works by other types of institutions or by individuals. It is possible that some of these practices are more widely applicable, but other frameworks are likely necessary and are being suggested by others.

We hope that this comment has been useful to the Copyright Office and Congress in crafting a solution to the orphan works problem and creating rules that further assist libraries and archives in fulfilling their traditional public roles.

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:

- Jack Lerner, Samuelson Clinic Fellow, jlerner@law.berkeley.edu, (510) 643-4800;
- Brewster Kahle, Digital Librarian, Internet Archive, brewster@archive.org, (415) 561-6767;
- Jeff Ubois, jeff@ubois.com, <http://www.archival.tv/>; and
- Rick Prelinger, Board President, Internet Archive, (415) 561-6767.