

Citation-Based Web Archiving and Copyright

Scholars increasingly rely upon - and cite to - ephemeral content that exists only on the web. Unless preserved, this part of the scholarly record is likely to disappear. To help prevent this loss, the Harvard Law School Library created Perma and makes it available for free to all academic libraries and their communities.

Our colleagues sometimes ask whether citation-based web archiving is compatible with copyright law. In acknowledging this question, we thought it might be helpful to share some basic information about web archiving, copyright law and Perma.

In brief summary, copyright questions are important to us. We've taken great care to respect both the protections and limitations of copyright, while supporting efforts to protect the scholarly record. Under these circumstances, it is difficult for us to imagine a more straightforward application of fair use than library-administered, citation-based archiving of ephemeral web content for use in scholarly works.

What is the Historical Relationship of Web Archiving and Copyright?

Web archiving in general is a longstanding, commonly accepted practice that has not been a source of copyright liability or public allegations of infringement, and which provides an essential public benefit. For example, over its 25-year-plus history, the Internet Archive has archived more than 500 billion web pages, and its software actively, repeatedly preserves large swaths of the web.

What Does the Law Say That Might Relate to Archiving Cited Web Content for Purposes of Criticism, Commentary, News Reporting or Scholarship?

17 U.S.C. § 107 states that “the fair use of a copyrighted work, including such use by means of reproduction in copies ... , for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use, scholarship, or research, is not an infringement of copyright.”

Despite the pervasive nature of web archiving over the past many years, copyright holders have not pursued allegations of infringement based on web archiving, and thus courts have not had an opportunity to interpret the fair use limitation on copyright in the context of web archiving. The Supreme Court, however, recently held that Google's commercially motivated reproduction and large-scale redistribution of a portion of a copyrighted computer program was protected as fair use, as a matter of law. See *Google, LLC v. Oracle America, Inc.*, 593 U.S. ____ (2021).

What Care Does Perma.cc Take to Further Respect Copyright?

Perma has a much narrower scope than the Internet Archive or other general purpose web archiving services. This narrow scope is in keeping with the mission of the Harvard Law School Library, where Perma is built and maintained, and academic libraries more broadly.

Our focus is on working to preserve the small subset of publicly accessible, ephemeral web pages that are cited by URL in works of scholarship, commentary, criticism and news reporting. At present, after over seven years of operation, Perma's community of libraries, courts and others have prevented that loss by preserving more than 2 million web pages that have been cited in over 22,000 scholarly articles and over 2,000 court opinions.

We've been extremely attentive in designing Perma to ensure we respect copyright and remain true to our scholarship-enhancing mission. For example:

- Perma works only for web pages that are freely accessible on the web, and our terms reinforce that limitation.
- Perma (together with various citation authorities) encourage dual citation of both the original source URL and the archival Perma URL.
- Perma records include a prominent link to the live source URL and preserve only the single URL specified and cited by the user.
- Perma preserves only the cited content found at the specified URL.
- Perma does not crawl the web preserving entire websites or drill down into the preserved page and automatically archive embedded links.
- Perma does not preserve very large files.
- We provide simple instructions in our FAQ for site owners who do not want to allow archives of their content to be public.
- We do not make archives available to search engines, and viewing Perma records requires the given Perma URL, which means those who visit preserved material usually arrive via the scholarly citation.
- We have a DMCA notice-and-takedown process in place and would react responsibly to any requests should they materialize.

We designed these and other measures out of respect for the interests protected by copyright and to reinforce the clear application of fair use, so that the scholarly community can safeguard its work against the problem of lost sources.

How Do Libraries Contribute to Archiving Practices that Respect Copyright?

In addition to the measures listed above, we also encourage library partners to be thoughtful about the copyright implications of protecting web citations. We require academic users to be authorized by their academic library, which has visibility to the records created as well as the ability to terminate users. We also give libraries the ability to make archives private, either on a record-by-record basis or by default for all records, to upload modified or excerpted materials when appropriate and to download archived files and manage access locally.

Beyond our terms of service and the design choices built into the software, we do not dictate to libraries how they should administer Perma.cc for their scholarly community. We envision libraries adopting a range of practices and approaches and taking additional protective measures when they deem appropriate, although in our experience additional measures are rarely needed. We believe libraries are well-suited to this role.

* * *

Perma's ultimate aim is to take the most defensible aspect of an archiving practice that has not triggered liability across hundreds of billions of instances over decades of activity and, through software design and active partnerships with peer libraries, further enhance the protection of copyright interests to ensure that the risk is truly negligible for those who share our commitment to the long-term integrity of the scholarly record.

While it is always possible to imagine misuse or speculate about unusual hypotheticals and complex edge cases, we feel very confident about the compatibility of the core Perma use with copyright and the efforts of our friends and colleagues in the library community to help administer this beneficial service. As always, we encourage you to continue being thoughtful and responsible about these topics and to reach out to us with any questions.