Perma: A Tool for Addressing Link Rot in Published Scholarship

Benjamin J. Keele
Research & Instructional Services Librarian
Ruth Lilly Law Library, IU Robert H. McKinney School of Law
Indiana University Indianapolis

I rely heavily on cited sources when researching a topic. Having found a relevant paper or book, I almost always look at some of the sources listed in the footnotes or bibliography for useful material. Authors, by citing the works on which they relied, save me significant research time and lead me to sources I might not have found otherwise. This is especially true in legal scholarship with its extensive footnoting. This practice in academic legal writing is not ideal for readability, but it is a boon for later researchers.

Consequently, when attempting to retrieve cited sources that are online, it is particularly frustrating to hit a 404--Not Found error or a page apologizing because the cited source is no longer there. This is called link rot: when URLs in citations no longer lead to the cited work. It happens when online sources move, are placed behind a paywall, or simply disappear. Regardless of the cause, link rot places obstacles in researchers' paths and weakens authors' arguments through lost support. An extensive literature has developed documenting the worrisome extent of link rot. A simple search of Google Scholar for link rot and the academic field of your choice will likely return several studies.

Computer scientists and biologists seem to have been the first to notice link rot's effect on their scholarship, but law libraries have launched an effort to mitigate link rot in legal scholarship. Perma (perma.cc) is a web site operated by law libraries to preserve online sources cited in legal scholarship. Developed at the Harvard Law School Library, Perma lets anyone enter a URL for a web page. Perma then archives the page, capturing the page's code and a screen capture in case the code
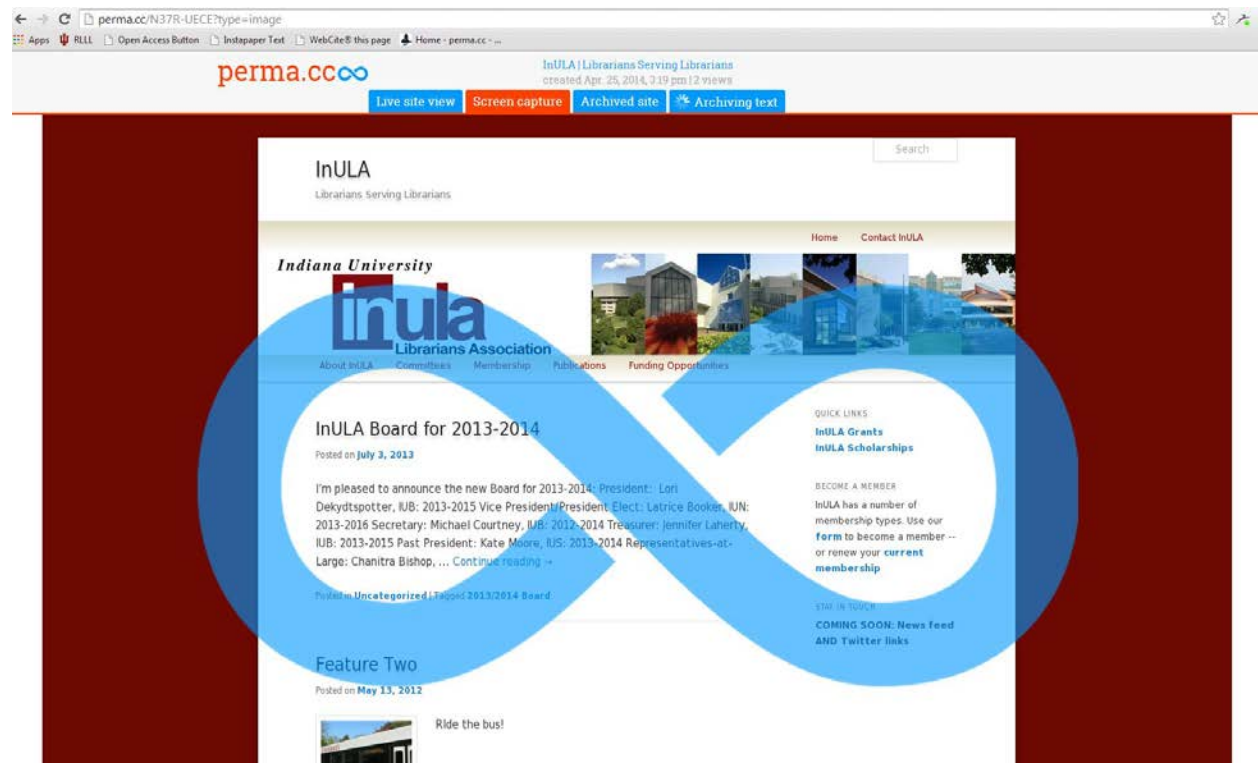
does not render well later. Perma creates a new URL for this archived copy that will remain stable for future reference. For example, I archived InULA's homepage http://www.indiana.edu/~inula/wp_2012/, and created this copy, http://perma.cc/N37R-UECE. Each copy is time stamped as well. This can prove handy in situations where precisely dated information is needed, like when a professor I was helping wanted to document the price of a product on Amazon on a particular day.

Anyone can submit online sources to Perma for archiving, but the copies will be retained for two years unless they are vested—designated for indefinite preservation—by a librarian or editor authorized by the law library. Law libraries can also provide servers that will mirror the Perma collection, thereby making the collection more reliable because it is diversified across geographic locations and institutions.

Authors and editors can archive sources and add Perma URLs at any time prior to publication. In my view, the source should be archived shortly after it is cited. I have seen instances in which cited sources have disappeared a couple months after the paper was accepted for publication. Perma makes the archiving process quite simple and requires no special expertise. At present, each source has to be manually archived, but perhaps in the future the process will be more automated. Both the original source URL and the Perma URL are included in the footnote, so later researchers can access the original, if possible, with the Perma copy as a backup.

To help avoid copyright problems, Perma complies with code in some web pages that prohibit archiving the page. In those cases, a copy is made, but it is "dark archived," i.e. not made publicly accessible. Librarians at Perma partner libraries can access the copy upon research request. This is similar to how libraries sometimes make copies of materials in library collections to fulfill research requests. Also, the copies in Perma that are publicly accessible are not indexed by Google or other search engines, so the only practical way a researcher can access a copy is by following the Perma URL in a citation.

Perma is not the first web archiving solution. The Internet Archive's Wayback Machine and Archive-It, WebCite, and archive.today are examples of other projects that address link rot in some way. What makes Perma distinctive is its scope (online sources cited in legal materials) and its shared operation by law libraries. Launched about a year ago, Perma has been adopted by a number of law journals, including IU McKinney Law's three journals. Courts are also considering Perma for archiving online sources cited by judges in their opinions. Hopefully with wide adoption, Perma will allow law libraries to preserve legal scholarship and digital sources for the long-term. Maybe it will even scale up to include other scholarly fields and libraries.