

New York Law Journal

Select '**Print**' in your browser menu to print this document.

Copyright 2009. Incisive Media US Properties, LLC. All rights reserved. New York Law Journal Online

Page printed from: <http://www.nylj.com>

[Back to Article](#)

Search Engine Finds Memories From Way Back

Kelly D. Talcott

08-22-2007

A recent decision by the U.S. Court of Appeals for the Eastern District of Pennsylvania, *Healthcare Advocates, Inc. v. Harding, Earley, Follmer & Frailey*, Civil Action No. 05-3524, tackles, in part, the inherently temporary nature of cache files.

Healthcare Advocates Inc. sued a client of the law firm Harding, Early, Follmer & Frailey for trademark infringement and misappropriation of trade secrets. Shortly after the complaint was served and before formal discovery had begun, employees of the Harding firm accessed, viewed and printed images of older versions of Healthcare's Web site that were stored on servers operated by the Internet Archive.

The [Internet Archive](#) is a not-for-profit group that attempts to search, record and store much of the publicly available information on the Internet. It employs a tool called the Wayback Machine that allows viewing of pages that the archive's robotic search tools automatically record and archive at various time intervals. [\[FOOTNOTE 1\]](#) The images are like snapshots of what a particular site looked like on the date it was recorded by the archive.

Not all Web sites are recorded by the archive. A site owner can "delist" a site by complying with a recognized standard for robot exclusion, which is a set of universal instructions that standard robotic search tools will follow and that are contained in a so-called robots.txt file that can be made part of the site's code. The archive's robots that read a "do not copy" instruction in a robots.txt file will (a) stop recording the site contents so long as the file remains, and (b) cause the archive to remove public access to the pages that have already been copied and recorded.

The Internet Archive can be a useful tool in litigation, since it may provide a record of past public statements made by or on behalf of a Web site's owner. It is with this in mind that the Harding firm used the Wayback Machine shortly after its client was served with Healthcare's complaint to review archived screenshots of Healthcare's Web site. Employees of the firm printed copies of each archived screenshot that they viewed.

Unbeknownst to the Harding firm, Healthcare Advocates had placed a robots.txt file on its site just days before Harding's employees accessed archived copies of the site via the

Wayback Machine. Under normal operating conditions, the archive should have checked Healthcare's site, detected the robots.txt file, and denied Harding access to the archived Healthcare pages.

However, unbeknownst to the Harding firm's employees, the Internet Archive servers were not operating properly on the dates they attempted to access the stored images of Healthcare's site, and they were able to view the archived pages despite the robots.txt instructions. The Harding firm used the old Web images in its defense of its client, and the lawsuit brought by Healthcare Advocates against its client was eventually dismissed.

The case was far from over for the Harding firm, however. Healthcare Advocates brought a second lawsuit, this one against the law firm, for copyright infringement as well as violations of the Digital Millennium Copyright Act, the Computer Fraud and Abuse Act, and common law torts of conversion and trespass to chattels. The focus of this second lawsuit was on the Harding firm's actions in accessing, viewing and copying the Healthcare content stored on the Internet Archive servers.

While the substantive claims raise moderately interesting issues of copyright law, the focus of this column is Healthcare's request that the court sanction the Harding firm for its failure to preserve the cache files of the computers that were used to access Healthcare's archived site content via the archive.

Cache files are created by Internet browsers to help speed up the process of viewing content from the Internet. Most of us have clicked back and forth between two pages of Internet content; perhaps, for example, from a list of novels responsive to our search for "legal thrillers," to a specific novel that we want to examine in more detail, then back to the list. When we first visit the page with the list of novels, our browser may store an image of that list page in our computer's cache memory. When we return to the list page after viewing the detail page about the novel that catches our interest, our computer probably pulls the image of the list of novels from its cache memory instead of going back to the Internet to download a new copy from the originating site. It is generally quicker and less of a burden on computer resources to retrieve the image from cache memory than it is to search and retrieve it again from the Internet.

Cache files are by their very nature temporary. Most browsers have tools that allow users to allocate more or less system memory to cache files, but eventually whatever amount of memory is allocated to cache files is filled, and older information is deleted to make room for the newer.

SPOILIATION INSTRUCTION

Healthcare Advocates argued that the Harding firm had infringed on Healthcare's right of reproduction when it saved copies of the archived pages onto the law firm's hard drives, not by any affirmative effort that the Harding firm had made to save the files, but in the form of the cached files that the firm's browsers saved automatically when the firm's employees viewed the archived pages. Because the Harding firm had not preserved those cache files, Healthcare argued, Healthcare was entitled to a spoliation inference at trial.

The court disagreed. As in *Columbia Pictures Industries v. Bunnell*, CV 06-1093 (U.S. District Court for the Central District of California), the analysis focused on the facts of the

particular case.

The court applied a Third Circuit balancing test that considered (1) the degree of fault of the party who altered or destroyed the evidence; (2) the degree of prejudice suffered by the opposing party; and (3) whether there is a lesser sanction that will avoid fundamental unfairness to the opposing party and, where the offending party is seriously at fault, will serve to deter such conduct by others in the future. [FOOTNOTE 2]

There were several facts that influenced the court's finding that "very little fault" could be attributed to the law firm for the loss of the cache files.

First, the court found that the firm "had no reason to anticipate that using a public Web site to view images of another public Web site would subject them to a civil lawsuit containing allegations of hacking." Indeed, many months had passed between the beginning of the Healthcare Advocates trademark infringement lawsuit against the Harding firm's client, and the dismissal of that lawsuit and the subsequent filing by Healthcare of a new copyright infringement lawsuit against Harding. Because the court saw no evidence to suggest that the Harding firm knew or reasonably should have known that a lawsuit against it was likely, the firm was under no obligation to stop using the computers that had been used to access the images so that the cache memory could be preserved.

Second, and most important to the court, there was no evidence that the cache files were affirmatively destroyed by the Harding firm. Instead, any cache files created as a result of the Harding firm's viewing of the archived Healthcare Web images were deleted automatically by virtue of the continued normal operation of the computers in question.

The court also found that Healthcare Advocates had not suffered significant prejudice as a result of the deletion of the cache files of the images that the Harding firm had accessed on the Internet Archive site. Healthcare's own computer forensics expert testified that he was able to piece together what had occurred from the data available on the forensic hard drive images that the Harding firm had provided in discovery.

As for the third consideration, the court did not find that any sanction against the Harding firm was warranted. "To impose a sanction on the Harding firm for not preserving temporary files that were not requested, and might have been lost the second another Web site was visited, does not seem to be a proper situation for an adverse spoliation inference," the court said.

ATTENTION TO DETAILS

What is clear from both *Columbia Pictures* and *Healthcare Advocates* is that even the most ephemeral of file storage media can be the target of one party's attempts to impose e-discovery preservation obligations on another.

In the case of the defendants in *Columbia Pictures*, it was information normally stored in RAM that included user IP addresses and the corresponding requests for dot-torrent file addresses of allegedly infringing content (NYLJ, June 19, 2007). There the court ordered the defendants to enable a log function of their server software that would capture in a more permanent form the user requests that were temporarily stored in RAM and otherwise would be overwritten as new requests came in.

No sanctions were assessed against the defendants for not doing this earlier, because the court found that their failure to retain the data was based on a good faith belief that the preservation of data stored only in RAM was not legally required.

In *Healthcare Advocates*, the cache files, though in a different format, were arguably similar to the RAM in the *Columbia Pictures* case. The information stored in the cache files would naturally be overwritten by the system as the computer was used in the ordinary course of business. There was nothing in Healthcare's trademark infringement lawsuit against the Harding firm's client to suggest that Harding would later be sued by Healthcare for copyright infringement. Here, too, the lack of notice was important to the court's ruling.

Also important to both courts was the intent of the defendants. The defendants in *Columbia Pictures* had never enabled the server log feature of their server software, nor were they asked to do so until the litigation was under way. Similarly, the Harding law firm was not on notice until after the lawsuit was filed against it that the plaintiff wanted to review the information in the cache files of the computers used to view the archived Web pages. In neither case was there any suggestion that the defendants had committed any affirmative acts designed to destroy potential evidence.

While neither of these cases seems to call for potential litigants to begin to treat ephemeral storage as automatically subject to litigation hold requirements, the cases do serve as reminders that discoverable information can exist in many different types of storage, some much more permanent than others. Parties need to be alert to what their e-discovery obligations might reasonably be under the particular situation at hand, and be prepared to tailor their actions accordingly.

Kelly D. Talcott, a partner in the New York office of K&L Gates, practices intellectual property and technology law.

::::FOOTNOTES::::

FN1 "Wayback Machine" pays homage to the "WABAC machine" from the 1960s cartoon segment "Peabody's Improbable History" that was part of "The Rocky and Bullwinkle Show." Mr. Peabody, a highly intelligent dog, used the WABAC machine to physically transport himself and his adopted boy Sherman to the past for visits to historically important events, where wry hilarity inevitably ensued.

FN2 *Schmid v. Milwaukee Elec. Tool Corp.*, 13 F.3d 76, 78 (3rd Cir. 1994).

This document was created with Win2PDF available at <http://www.win2pdf.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.
This page will not be added after purchasing Win2PDF.