

Archiving the Internet: The ‘Wayback Machine’ in the Courts

by Steve Kramarsky

Sometimes, the intersection of law and technology creates interesting legal or policy issues for lawyers and courts to explore; other times, it just creates headaches. Among the latter are the problems of proof that arise when legally significant events occur entirely on the Internet and not in the physical world. It is well established that web pages, blog entries, social media postings and other online activity can be evidence of liability, or even the sole basis for a claim—for example a claim of defamation or infringement, or breach of an online contract. But in the context of freely-editable, user-generated electronic content, it can be extremely difficult to establish, with legal certainty, what activity actually occurred and when.

Consider the following situation: a client comes to you to evaluate whether he has a claim for copyright infringement based on the use of certain artistic elements on website. But upon checking, you find that the site has been reworked and your client’s art is gone. In discovery, records of the old site will be available, but at that early stage, how are you to evaluate the claim? Or this: you represent a defendant accused of making defamatory statements about plaintiff on her blog. Your client has removed the posting, but plaintiff seeks to introduce a screenshot of the website into evidence. Your client insists the screenshot has been doctored and does not accurately reflect the posting. But how can you establish the truth for the court?

Scenarios like the above have become increasingly common with the rise of the Internet and the ease of access to content tools. We conduct more and more of our business, personal and legal affairs in electronic form, and increasingly those electronic interactions leave no tangible paper trail at all. For many substantial legal matters, the crucial evidence may be on a website or in an email. What representations did the company make on its website about the efficacy of its product? Was a user properly prompted to agree to certain terms and conditions (commonly referred to as a “clickwrap agreement”) before making a purchase? Did an international corporation tout its New York operations on its website at the time of a transaction? What public comments or statements were made about a person or event or transaction at a given time?

A case that turns on the who, what, and where of a representation can easily be derailed if that representation was made on the Internet and is subsequently removed. This doesn’t assume malfeasance—websites change every day for all kinds of perfectly valid reasons. But the effect is the same: if the website has been updated, the evidence may become unavailable for trial. Of course, it is always good practice to retain hard copy (such as a printed screenshot) of important website evidence, but even that may not be enough. Such

screenshots present authenticity issues and, if they are created by counsel, there is the risk that the lawyer will have to become a witness, describing how and when the screenshot was made. In addition, it can be difficult, months after the fact, to establish the timing of the screenshot and whether it presents an accurate picture of the website at the relevant time.

To get a more accurate picture requires a time machine capable of re-creating the web as it was on a given date. Luckily, at least for many web sites, such a machine exists. A recent U.S. Court of Appeals for the Second Circuit decision describes how to use it, and how to properly introduce records from it so that they can be accepted as evidence in court.

The ‘Wayback Machine’

Publicly available websites are, as a technical matter, nothing more than electronic computer files that can be downloaded, viewed and stored by anyone at any time—that’s what surfing the web is. There is, in theory, nothing stopping anyone from simply downloading the entire web—every single website from every computer connected to the Internet serving a web page—and keeping a record of it. The problem is one of scale. The amount of information involved is enormous and constantly changing, and we rarely know in advance what information we will need to recall later, so there is no way to target what is retained. The project of archiving the entire web as it changes each day would require massive amounts of computation power and storage. Happily, someone is doing it and has been for over two decades. That archive is called the “Wayback Machine.”

The Wayback Machine is a free archive of the entire publicly available web (and a great deal of other information) maintained by a non-profit organization called the Internet Archive. According to its website, the Wayback Machine has been archiving the Internet since 1996 and currently archives over 333 billion web pages. A single copy of this archive takes up over 30 petabytes of space—30 million gigabytes. Put simply, the Wayback Machine is an inconceivably large, entirely free archive that “captures and preserves evidence of the contents of the Internet at a given time.” *United States v. Gasperini*, 2018 WL 3213005, at *5 (2d Cir. Jul. 2, 2018).

The Wayback Machine is not perfect. It doesn’t capture every website on every date (some sites are not available to the public or ban the “crawler” programs used to archive websites) but it is surprisingly complete even for fairly obscure websites and blogs. Users of the Wayback Machine can search for the contents of a particular website at a particular time, and if the archive captured the website in question on that particular date (and chances are it did) they can recall an image of the website at that time. For example, the Wayback Machine has saved over 111,000 copies of the Wall Street Journal’s home page since 1996, often with numerous copies recorded on any particular day. The Wayback Machine gives counsel an opportunity to demonstrate exactly what a website looked like at a particular date and time—often crucial evidence in cases concerning Internet transactions and communications.

A Question of Admissibility

While this resource is obviously enormously useful, courts have struggled with the admissibility of website screenshots. Federal Rule of Civil Procedure 901 requires the proponent of evidence to “produce evidence sufficient to support a finding that the item is what the proponent claims it to be.” What is “sufficient” evidence that a website screenshot is what the proponent claims it to be—that is that the screenshot is representative of the appearance of a given website on a given date. Without sources such as the Wayback Machine, it is unclear and fact-specific what evidence would be sufficient to authenticate an Internet screenshot.

However, there is a clear trend towards admissibility of evidence regarding the historical appearance of a webpage by reference to the Wayback Machine or similar Internet archives. In *United States v. Gasperini*, the U.S. Court of Appeals for the Second Circuit considered whether the district court erred “by permitting the government to introduce screenshots of various websites taken by the Internet Archive, more commonly known as the ‘Wayback Machine.’” 2018 WL 3213005, at *5. The Second Circuit, in affirming the district court, explained that the evidence was sufficiently authenticated by “testimony from the office manager of the Internet Archive, who explained how the Archive captures and preserves evidence of the contents of the Internet at a given time” and because that same witness “compared the screenshots sought to be admitted with true and accurate copies of the same websites maintained in the Internet Archive, and testified that the screenshots were authentic and accurate copies of the Archive’s records.” *Id.*

Other circuit courts have reached the same conclusion. In *United States v. Bansal*, the Third Circuit considered the admissibility of records from the Wayback Machine. 663 F.3d 634, 667-68 (3d Cir. 2011). Following substantially similar logic, the Third Circuit held that such records were admissible to prove the contents of a certain website on a certain date.

Ensuring Admissibility

The Wayback Machine is not a silver bullet. While there is a trend towards admissibility, some courts have held that archived records may not be admitted on the grounds that they were not sufficiently authenticated—that is there was an insufficient showing that they were what the proponent said they were. E.g. *Novak v. Tucows*, 2007 WL 922306 (E.D.N.Y. March 26, 2007). In *Gasperini*, counsel addressed this problem by putting witnesses from the Internet archive on the stand to describe the collection process and satisfy the court’s authenticity and business record concerns.

Counsel seeking to introduce evidence from Internet archives would be well advised to follow that same plan. Attempting simply to introduce screenshots from a third-party archive may not meet with approval. Instead, that evidence should be supplemented with witness testimony describing the archive, how it works, and how the records to be introduced into evidence were produced and stored in the ordinary course of the archive’s business. This should address hearsay and authenticity issues, and go a long way towards ensuring that the evidence will be admitted.

This article first appeared in the *New York Law Journal* on July 16, 2018. John Millson, an associate at the firm, assisted with the preparation of this article.