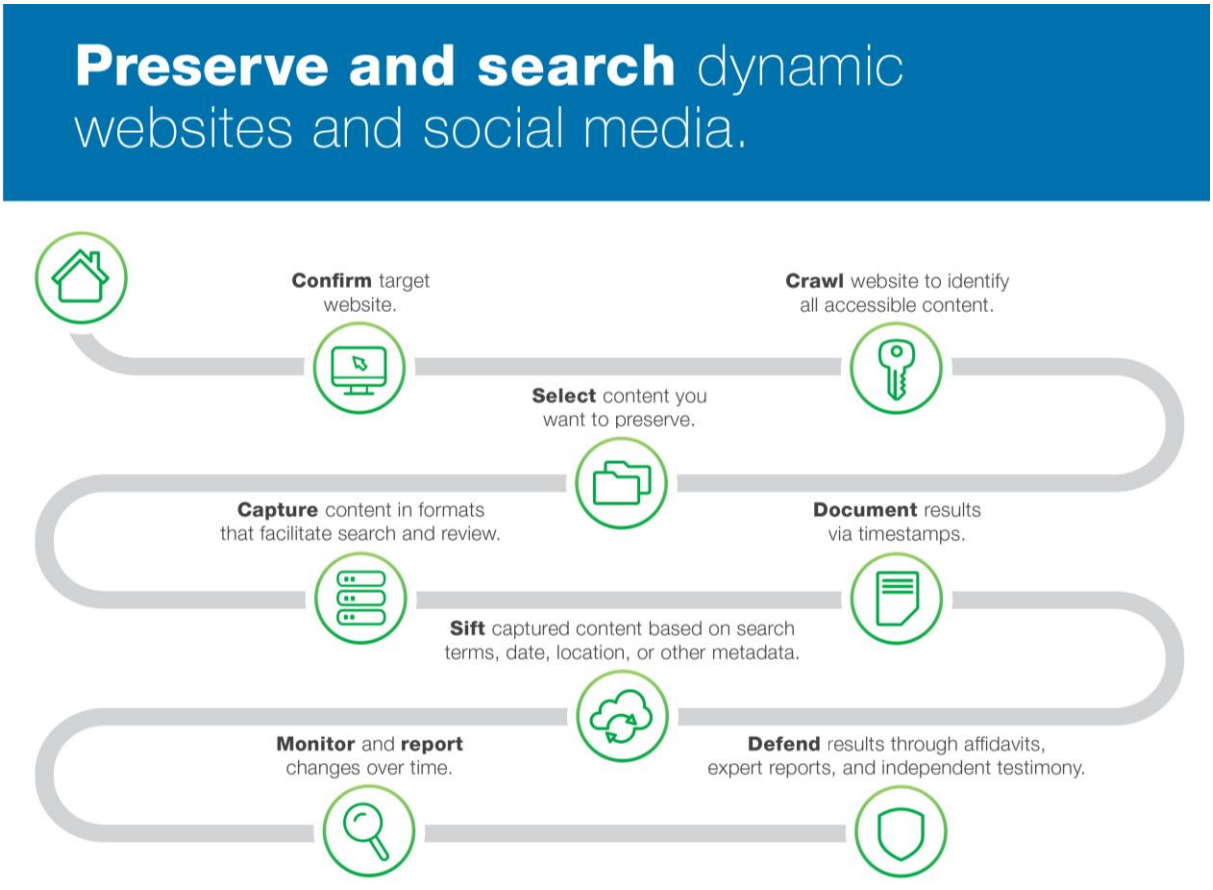
December 2017

Strategic discovery considerations: preservation and analysis of web and social media content

While the Internet can provide a rich store of information that may be critical for litigation and investigation matters, it can also be a significant challenge to collect, search, and produce the information in a forensically-defensible way.

The methods used to capture Internet-hosted information can vary; however, there are preferred methods for capture that can help ensure the authenticity and completeness of the information captured.



Capture, sift, document

A web site is effectively a collection of files transferred from the hosting location, and assembled in a particular order by the accessing device. For small web sites, a screen capture program can be used to collect pictures or video of a user browsing pertinent web pages. This works if the site is small and the person collecting the data has the time to visit each page. It should be noted, however, that while this produces a record, it does not provide any information outside of a visual comparison. This may be sufficient for the requirements of an individual case, but for many this method is not sufficient, nor is it always cost effective.

A sophisticated vendor will have software at their disposal to automatically “crawl” or access every publicly accessible part of an Internet-hosted resource, so a forensic investigator doesn’t have to manually access and document each page. Typically, this type of software will “hash” the source and download files to ensure proper duplication and create a time stamp to record the date and time of access.

The digital fingerprint

The process to verify digital files is typically conducted using a value known as a “file hash.” A hash is a calculated value created using an algorithm which takes into account the value and location of each bit in a file and is an alpha-numeric character string representing a “digital fingerprint” of the file. Changing a single bit value will alter the hash value. This value is typically calculated on an original file and a copy to validate the copy is exactly the same as the original. This hash value can also be used to find specific files (if the hash is known) or copies of a file. This is helpful in cases where the identification of all copies of a file is important, as in intellectual property theft.

When capturing Internet resources, it is best to work with a vendor who respects the value of — and employs hash calculations in — their Internet resource collection methodology. This will provide validation that the copied web site or resource is an exact duplicate of what was hosted on the web page on the day of collection.

Additionally, some collection software can take screenshots of how a web page displays on screen as well as copying the underlying files so that the page is captured as it appeared to the user.

Social media discovery methods

The ability to capture, preserve, and analyze social media postings and messaging is critical to the investigation and discovery process. Similar to the web site capture information discussed earlier, technologies exist to record and search the information available on a social media account.

Occasionally, the continual monitoring of an account is preferred. A vendor should have the ability to grab regular snapshots of a social media account without duplicating what was already captured. This ensures that posts or pictures available in one capture, but deleted in a future one, will persist in the collection record.

The technological aspects of collecting, searching and producing web sites and social media accounts can be challenging enough. But for attorneys, it is often valuable to identify a professional who has the experience and expertise to also testify about the work, the authenticity of the evidence, and the method employed to collect the information. While many vendors can produce a copy of a web site, not as many have the experience to provide the testimony frequently needed in these matters.

To continue the conversation about digital discovery contact one of the individuals below.

Contacts

Kristofer Swanson, CPA/CFF, CAMS, CFE
Vice President and Practice Leader, Forensic Services
Chicago
+1-312-619-3313
kswanson@crai.com

Cuyler Robinson, CISSP, CIPT
Vice President
Chicago
+1-312-619-3394
crobinson@crai.com