

# 10 Things to Know About Preserving Social Media

Rakesh Madhava



Governments, corporations, and organizations of every kind are adopting Internet-based technologies that radically impact the most essential of human activities – the act of communicating with one another. Internet channels are many and varied, but in total, the purview of any organization communicating with internal and external audiences must include not only websites, but also blogs, video outlets, and social media efforts – most notably Facebook, LinkedIn, and Twitter, not to mention rapidly evolving geo-located services, such as Groupon, Foursquare, and Yelp. Employing a “website only” online presence has become antiquated as quickly as having a website became a necessity.

While most organizations have repeatable processes in place for preserving e-mail and other types of electronically stored information

(ESI), most do not have a process in place to preserve, archive (securely maintain in an indexed and searchable database), and research social media. Undoubtedly, failing to capture and preserve social media activities risks violating any number of compliance, regulatory, and legal requirements, and in particular, it leaves organizations woefully unprepared for producing social media data during e-discovery.

## Social Media Content's Role in Litigation

The courts are beginning to catch up with the deluge of social media, and judges are displaying less patience with organizations that haven't properly managed their archives. In one case, *Arteria Prop. Pty Ltd. v. Universal Funding V.T.O., Inc.*, the judge ruled in no uncertain terms, “This court sees no reason to treat websites differently than other electronic files.”

In that particular case, negotiations for a loan had collapsed, and the plaintiff asked for paper copies or “snapshots” of the defendants' website as it looked at the time of the negotiations. When the defendants could not produce copies in any format, the plaintiff claimed spoliation. The judge found in favor of the plaintiff, ruling that “... the Court finds that Defendants still had the ultimate authority, and thus control, to add, delete, or

modify the website's content. There is no evidence to the contrary.”

And, in a sign of the rapidly emerging landscape, the American Academy of Matrimonial Lawyers released a study in 2010 indicating that 81% of its 1,600 members had seen an increase in the number of cases that had relied on information taken from social networks in the previous five years.

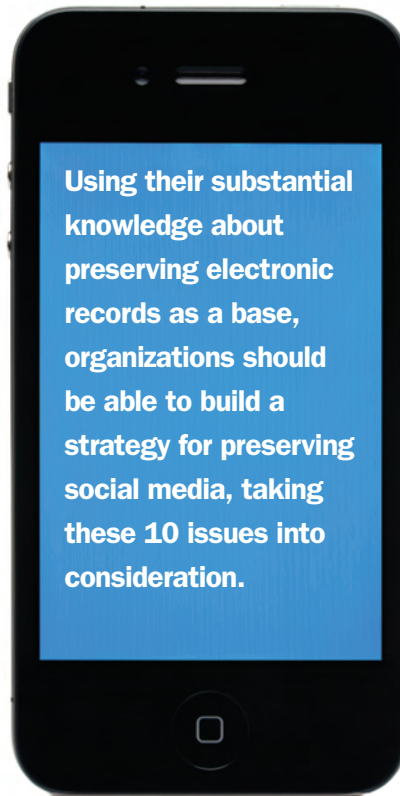
### It's Time to Act!

As more organizations and their employees begin to use social media, and more types of social media develop, the time to act is now. As daunting as the task may seem today, it will only become increasingly difficult. Using their substantial knowledge about preserving electronic records as a base, organizations should be able to build a strategy for preserving social media, taking these 10 issues into consideration.

#### 1. Organizations need to use social media.

From the point of view of records managers and the legal department, it would certainly be easier to simply ban social media like Twitter and Facebook. But ignoring the enthusiastic embrace of social media by millions of people is simply not a realistic approach, and marketing departments and salespeople will most certainly be clamoring in opposition of any such ban. One only needs to look at the failed efforts of dictatorial regimes to shut down social media activities in their own countries for proof of the inevitable failure of a ban.

Much like e-mail, social media is quickly becoming an essential aspect of communications and marketing within organizations. A 2010 *Harvard Business Review* report, “The New Conversation: Taking Social Media from Talk to Action,” reported 79% of 2,100 organizations surveyed were using or planning to use social media (58% were using it, and 21%



were preparing to launch initiatives).

And, according to a 2010 report by the University of Massachusetts, “The Fortune 500 and Social Media: A Longitudinal Study of Blogging, Twitter and Facebook Usage by America’s Largest Companies,” 60% of the *Fortune* 500 had a Twitter account with a Tweet in the 30 days previous to the survey; this is dramatically up from 35% in 2009. Fortunately, principles currently exist around how to manage social media, unlike the situation when e-mail use first began to explode.

#### 2. Organizations need to preserve social media – and websites.

Organizations have a clear obligation to preserve and archive all social media. *Federal Rules of Civil Procedure* Rule 26 requires organizations to be able to produce all potentially responsive information for e-discovery purposes.

According to the recently released Gartner report “Social Media

Governance: An Ounce of Prevention,” by the end of 2013, half of all companies will have been asked to produce material from social media websites for e-discovery. According to the report, “... in e-discovery, there is no difference between social media and electronic or even paper artifacts. The phrase to remember is “if it exists, it is discoverable.”

The leading think tank on electronic document retention, The Sedona Conference®, includes as the first principle in *The Sedona Principles: Best Practices Recommendations and Principles for Addressing Electronic Document Production*:

Electronically stored information is potentially discoverable under Fed. R. Civ. P. 34 or its state equivalents. Organizations must properly preserve electronically stored information that can reasonably be anticipated to be relevant to litigation.

Organizations should treat social media as they would any other ESI and assume it is potentially discoverable. Under Rule 34 of the *Federal Rules of Civil Procedure*, litigants can request “any designated documents or electronically stored information – including writings, drawings, graphs, charts, photographs, sound recordings, images, and other data or data compilations – stored in any medium from which information can be obtained either directly or, if necessary, after translation by the responding party into a reasonably usable form ...”

There are also regulatory requirements. Some regulatory authorities, including the United States’ Financial Industry Regulatory Authority, the Securities and Exchange Commission, and the Food and Drug Administration, require social media to be preserved. State and federal freedom of information laws may also require organizations to retain and produce social media postings,

Tweets, and the like.

All laws, regulations, and requirements should drive a social media archiving policy, and those policies should complement the organization's existing records and information management protocols. Organizations should carefully consider the underlying business reasons; considering not only the reasons why they should preserve, but under what rationale there is not a need to preserve. If an organization decides not to preserve some or all of its social media, it needs to be able to point to the law or regulation that says there is no requirement to do so.

### **3. Social media files often involve more than posts.**

Social media doesn't exist in a vacuum. When posting on Facebook, an employee may link to a YouTube video. Or an organization's website may hyperlink to a PDF of a white paper from another site.

An archival solution has to include the original Facebook post or website record, but it also has to be able to follow and capture the YouTube link or third-party source's white paper. An organization will need to be able to preserve embedded native files, whether those are PDF files, Word documents, Excel spreadsheets, or PowerPoint presentations. Otherwise, it's akin to archiving e-mails without saving attachments.

### **4. Use APIs to capture, archive, and review data from the web.**

Compared to preserving the Word files employees create and share, accurately preserving social media can be extremely complicated. It requires knowing how application programming interfaces (APIs) work.

Simply stated, software programs communicate with each other through APIs. When archiving social media, consider whether the solution is actually pulling data from an API

or simply taking a "screen shot" of what can be viewed in a browser. A screen shot won't include metadata or other information that can't be "seen," but which may be critically important in a lawsuit or regulatory hearing.

When it comes to archiving, some may think printing a file or saving to PDF using a browser's print function

will suffice. But it's not possible to print a YouTube video or save it as a JPEG or PDF. And social media sites don't display all of the content available in a single interface. Facebook, for example, uses an algorithm to display the content the application believes the user is most interested in.

Through Web 2.0 architecture,

APIs can be accessed by outside applications. At this time, the most accurate way to preserve web data is to make use of APIs available from social media properties. Facebook, for example, has no reliable methodology for its displaying of data to individual users. The amount and types of data displayed varies greatly from session to session and from user to user. Mapping preservation applications to the Facebook API allows full access to the entire population of data for any targeted user's Facebook profile.

## 5. Social media archiving solutions need to be customized.

While guidelines and suggestions exist around social media archiving, an organization can't simply grab another's social media archiving policy and duplicate it exactly. One size does not fit all. Facebook, Twitter, Flickr, and other types of social media have unique API structures.

Archiving solutions should involve an open authorization (oAuth) approach, which provides an open standard for authorization that simplifies

API. OAuth allows third-party site access to information stored with another service provider without sharing their access permissions or the full extent of their data, thus allowing organizations to preserve protected employee accounts.

Implementing a preservation strategy is not as easy as simply flipping a switch. An organization will need to make decisions about the social media data it wants to capture and preserve. It should consider an employee's Facebook page – when someone responds to a post on that page, does that response need to be included in the preservation? Or, if an organization Tweets about another organization's business, does the organization being Tweeted about need to preserve that Tweet?

## 6. Consider setup or installation requirements.

Once an organization understands the need for a social media archiving policy, it needs to start considering technical issues. Will it require and benefit from an applications service provider (ASP), software-as-a-service (SaaS), or a behind-the-firewall network installation? What provider can offer the needed solution?

Many archiving products/services require a system that must be installed onto networks, and data is then saved to on-premise servers – at the provider's or the client's location. This approach can be costly (buying and maintaining servers for ever-expanding data stores) and may be vulnerable to breaches and outages.

With today's evolving technology, other options exist, such as ASP or cloud-based SaaS solutions. An ASP model simply means that a software instance is hosted off-site and accessed remotely, as with many website content management systems. Cloud-based solutions, theoretically, should be better situated to scale as data increases. And with these types of solutions, overhead costs are often much

lower than that of on-premise servers, and protection against outages are mitigated. Very large organizations can save every one of their employees' Tweets without much worry about reaching the end of their available stores.

### 7. Robust search capabilities of the preservation are necessary.

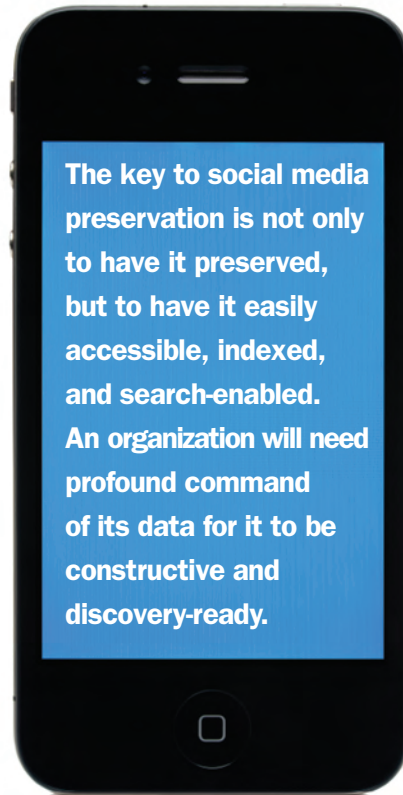
Any archival system deployed needs to include highly sophisticated search capabilities that will keep pace with the exploding use of social media and give the organization command over the preserved data. Though organizations may have few social media feeds now, they may have many more in the near future. For example, Hyatt Hotels now has a unique Twitter account for each of its 451 properties worldwide.

Advanced search tools need to be a minimum ante for any archival service or product to be a viable and worthwhile solution. The key to social media preservation is not only to have it preserved, but to have it easily accessible, indexed, and search-enabled. An organization will need profound command of its data for it to be constructive and discovery-ready.

### 8. Review and production from the preservation is critical.

An organization has to assume that further down the road it will need to be able to manage its archive and easily produce data from it. Huge amounts of social media data will not be useful if it can't be searched or data can't be produced from it in a reasonably timely manner and on a budget. At the outset of the process, an organization should be developing the workflows that allow the data to be copied and made available for relevance review in e-discovery platforms.

Of course, being able to search and review the data isn't quite enough. An organization needs to be able to do so without altering the data's original



state in any way, or it could face legal sanctions.

It also needs to consider how to restrict access to the preserved data. If specific data can't be put in a silo, anyone with access to the archive will be able to view every piece of information in the database. This can compromise the confidentiality, privilege, and privacy of the information. The right workflows will limit access of potentially relevant information to those who have the proper clearance.

### 9. Aim for capturing data in real time.

For most social media outlets, it will not be sufficient to preserve on a monthly basis. Real-time capture and preservation is the standard to ensure the most forensically sound and complete archive. Sending a 140-character Tweet can take mere seconds, and then it is out in the universe, even if later deleted or lost by Twitter. A real-time capture solution provides the best chance of archiving everything. An organization needs to consider how

often it will "crawl" (deploy web spiders to extract data) for its archiving purposes.

At this point, Twitter makes it relatively simple to capture feeds in real time, while LinkedIn and Facebook, because of their inconsistent presentation of data, are a bit more challenging.

### 10. Think about how content for legal holds will be locked or excluded from normal RIM schedules.

Like other types of data storage, social media may be subject to legal holds, or an organization may otherwise need to deviate from the usual retention schedule. It will need to develop a strategy for executing legal holds, and it will also need to have the technology to execute this.

Much like organizations must be able to silo specific data, as mentioned before, they also need to be able to sort out the data that should be deleted according to regular schedules and keep the data that may be potentially responsive in case of a lawsuit.

### The Revolution Will Be Televised ... on YouTube

These are the beginning stages of the social media revolution. Denying this or delaying a preservation strategy will only compound the difficulty to execute. Despite the massive amounts of data that exist today, the quantity is still manageable. Putting it off will only make the mountain of data that much higher when it must finally be climbed.

For records management professionals who survived the sea change that e-discovery brought, the situation with social media archiving should feel familiar. When organizations need to create policies on the fly, it's extremely difficult to catch up. Take action now to be sufficiently prepared. **END**

*Rakesh Madhava can be contacted at rmadhava@nextpoint.com. See his bio on page 54.*