

LJN's

LEGAL TECH

Newsletter®

An **ALM** Publication

Volume 29, Number 3 • July 2012

Social Media Data Preservation

Tools and Best Practices

By Diana Fasching, Staci Kaliner and Tamara Karel

ocial media is no longer a new phenomenon. Yet, the development of business rules governing the long-term retention of social media content in the ordinary course of business as well as the legal precedent governing the preservation of social media for litigation remain in their infancy. (For convenience, we will refer to the concept of saving social media for both purposes as "preservation.") Businesses are only now beginning to explore the value and need to retain their social media interactions. Courts are likewise just starting to assess the interplay of privacy rights and regulatory laws, authentication of social media data, the waiver of privileges, and issues related to the preservation and collection of social media. See, Romano v. Steelcase, Inc., 2010 WL 3703242 (N.Y. Sup. Ct. Sept. 21, 2010) (privacy); EEOC v. Simply Storage Management LLC, 270 F.R.D. 430 (S.D. Ind. 2010) (privacy); Cripsin v. Christian Audigier, Inc., 717 F.Supp.2d 965 (C.D. Cal. 2010) (regulatory laws); Lorraine v. Markel American Insurance Co., 241 F.R.D. 534 (D. Md. 2007) (authentication); Griffin v. State of Maryland, 419 Md. 343 (Md. 2011) (authentication); Ledbetter v. Wal-Mart Stores, 2009 WL 1067018 (D. Colo, Apr. 21, 2009) (waiver); Tener v. Cremer, 2011 N.Y. slip op. 6543 (N.Y. App. Div. Sept. 22, 2011) (accessibility).

Diana Fasching (dfasching@redgravellp.com) and Staci Kaliner (skaliner@redgravellp. com) are senior advisors in the Minneapolis and Washington, DC, offices of Redgrave LLP, respectively. Both Fasching and Kaliner help solve the legal and technical challenges associated with information management, litigation readiness, and electronic discovery. Tamara Karel (tkarel@redgravellp.com) is an attorney in the Washington, DC, office of Redgrave LLP. Karel focuses her practice in the areas of Information Law, which include electronic discovery, records and information management, data protection, and data privacy. The authors gratefully acknowledge Redgrave attorney, Michael Kearney, for his assistance in the preparation of this article.

The emerging requirements for preservation solutions have created a two-fold need: identifying process solutions and the technology to support them. Fortunately, the market has seen growth in offerings aimed at meeting social media preservation strategies. This article explores some of these developments and provides a set of basic considerations to evaluate when assessing the technology.

EMERGING SOLUTIONS FOR PRESERVING SOCIAL MEDIA DATA

There is no "best" solution for preserving social media data. There are manual processes such as printing Web pages, using tools such as SnagIt to save screen captures electronically, using the Web browser "save as" option to save the Web page as a Web archive (MHT) file, or using the data export features provided by platforms like Facebook. But these manual processes can quickly become overwhelming and impractical. Imagine that you must preserve LinkedIn accounts for 10 individuals. Assuming that you can access the account, preservation is not a simple matter of printing one Web page for each account. Rather, each account is a complex set of pages with links to other pages and content. Preserving this content manually requires mapping, methodical processes and documentation to ensure you collect everything and you can defend your approach later. Now imagine that you must preserve this data on a daily basis. As you can see, the approach can quickly become an overwhelming task — and it ignores the collection of associated metadata.

Fortunately, there are now more automated preservation tools that, for example, take a searchable screen snapshot of the Web page or leverage the Application Programming Interface (API) of the site to capture updates from an account and capture some associated metadata. Akrovi, Aleph Archives, Hanzo Archives, Iterasi Archives (offered through partner Reed Technology), NextPoint (via its Cloud Preservation solution), PageFreezer, Perpetually and X1 Social Discovery (X1) are all examples of currently available preservation solutions. Likewise, players such as Autonomy, Exterro and Symantec/Clearwell have or are developing solutions for preserving social media that work

with their existing services.

The relative immaturity of social media preservation technologies, however, leads to a variance in approach and sophistication among vendors. The process of finding the "best" solution can seem daunting unless you take a logical and methodical approach.

First, you must identify your organization's business requirements — what, when, where, why and how much questions are important to answer before you try to find a solution to meet your needs. Is your organization seeking to preserve social media for discovery; to fulfill regulatory requirements; to archive social media as business records; to monitor social media for appropriate use; or some combination of these goals? How often will you need to preserve your social media content? What social media platforms do you need to preserve? In what format will you need this data? What are the data security or privacy concerns? Do you have access to logins for private content? As you can see, there are a significant number of questions that you should answer to define your social media preservation needs.

Second. your organization should evaluate the different vendor solutions and their abilities to meet your organization's requirements, as well as their maturity, stability, current capabilities and future development roadmap. As a starting point, consider that you may already have a technology solution that has capabilities to preserve social media (e.g., if you have existing solutions from Autonomy, check first regarding their current capabilities). If your existing solutions will not meet your needs, target additional vendors to evaluate against the business requirements you previously identified.

Third, once you have assessed your business needs as well as identified the entities you want to investigate further. You will want to explore the following 10 starter questions that will help frame the right solution for your circumstances.

10 STARTER QUESTIONS TO ASK WHEN EVALUATING SOCIAL MEDIA PRESERVATION TOOLS

1. Does the tool support the capture of more than just social media sites?

Solutions such as Hanzo Archives and PageFreezer began by tackling the challenge of LJN's Legal Tech Newsletter July 2012

archiving Web content and have expanded into capturing social media sites. Other solutions, such as X1, have focused squarely on preservation of social media. Choose a tool that is best suited to your particular needs — whether it is Web content, social media or both.

2. What social media sites does/will the solution support?

Vendors typically develop solutions to support widely used social media sites (such as FaceBook, LinkedIn, Twitter and YouTube) before developing solutions for other social media platforms (such as foursquare and Flickr). Vendors must be nimble to keep up with new sites that gain popularity at warp speed and it appears that many are rising to the challenge by continually enhancing their products. Try to choose a vendor that is continuously being proactive and not reactive to the market.

3. How many users will need to use the solution?

Social media preservation solutions are offered in both cloud-based and single-user local installation varieties. If your organization plans to have multiple consumers of the preserved data, a solution like X1 — which is a single-user license - may not fit your needs, while a cloudbased solution like NextPoint, Reed Technology/ Iterasi, or Aleph Archives may be better suited for this environment. Cloud-based solutions often make new functionality and fixes available more quickly than those with a locally installed client application, which relies on your limited IT resources to handle installation of software updates. Single-user licenses are desirable when you want full control of where your data is stored.

4. How many collections can run simultaneously?

Solutions such as X1 allow one case to be actively collecting data at any one time; therefore, for example, if you want to prospectively collect data from a Twitter account on an ongoing basis, you can only do so for one case at a time for each licensed copy of the tool.

5. How does the vendor collect and preserve the data?

Vendors take different approaches to the collection and preservation of data. Vendors like Aleph Archives, Hanzo Archives and PageFreezer use automated crawlers to identify and capture data; most other vendors such as Arkovi, NextPoint and Reed Technology/Iterasi leverage the APIs that social media sites make available. Use of the APIs may be preferable because they: enable the collection of all data and metadata that the social media provider makes publically available; may allow the collection of private content, if user credentials are supplied; and generally provide notification when the API changes (whereas websites may just change whenever).

6. What does the solution capture and how is data stored and exported?

Many of the differences between solutions appear to be in: 1) how the data is displayed; 2) the ability to define the depth of data captured;

3) what metadata is captured; 4) authentication of the collected data; 5) ability to capture public and private data; and 6) the capabilities for review and export. Some vendors, such as Aleph Archives and Hanzo Archives, store preserved data in the WARC Web archive format; others, such as X1, store preserved data in the MHT Web archive format, and those like NextPoint store it as PDF, HTML and PNG files.

Export of the data is another key differentiator between solutions. NextPoint supports export of pages as PDF and HTML, or batch exporting into popular file formats such as Concordance and Summation load files, EDRM XML, Trial Director and more. NextPoint also enables seamless import into its cloud-based review platform. X1 offers export to Concordance, CSV and HTML, while other vendors, such as Arkovi, may offer fewer export options such as Excel or XML. The bottom line is to understand what social media data your organization needs to preserve and how your organization wants to be able to later access, search, export, review, and produce data.

7. What is the vendor's approach to subsequent captures?

Social media content changes frequently. Frequency of collection (e.g., hourly, daily, weekly, etc.) and deduplication features vary among solutions. Solutions such as NextPoint capture incremental changes while other solutions may recapture the entire site. Lack of deduplication features can lead to a significant amount of data being accumulated over time; however, capturing changes only can introduce potential challenges in viewing the content as of a point in time, depending on how the vendor reconstructs the changed parts. Either approach can be effective, but it is important to ensure that you can successfully navigate, search and identify preserved information at different points in time, and, because many products are priced by the amount of data you preserve, you should understand the impact the chosen approach will have on the cost of using the solution.

8. How will your organization search and use the data?

Most vendors will tell you that you can search and browse your preserved content at any point in time and that their solution offers robust search capabilities. The differences among the solutions will be found in the ease of use of these search capabilities, the intuitiveness of the user interface, the abilities of the product to deduplicate collected data and to display the unique data in a meaningful way. While asking questions to the vendor about these items or seeing a demonstration may give you some initial perspective on these features, evaluating the product using a free trial over a period of time will best enable you to get a feel for the product. Most vendors are willing to set up a trial period for your evaluation purposes. We strongly encourage you to take advantage of these opportunities and to give some dedicated attention to testing your preferred solutions.

9. How long will the preserved data be kept?

Some vendors, such as Arkovi, will store your preserved social media (archives) for as long as you are an active customer, allowing you to offload the archives to a third-party on a regular basis. Other vendors, such as Hanzo Archives and Reed Technology/Iterasi, can apply records retention periods and legal holds against archived content to more systematically manage the life of the preserved information. As in most other aspects of preservation and review (as well as information governance as a whole), there are costs associated with the longterm preservation and storage of data in these environments. We recommend that you identify these costs up-front when evaluating the best solutions for your long-term needs.

10. What is the cost of the vendor's solution?

Pricing models offered by vendors in this space are still immature and vary from one another such that an apples-to-apples comparison may not be easy. Pricing approaches range from price-per-page, price-per-page and capture frequency, price-per-number of target sources and capture frequency, price-per-number of target sources and amount of storage, number of users and amount of storage, and so on. There may also be consulting or professional services fees associated with configuring and implementing the solution. Many vendors are willing to negotiate pricing. Keep in mind that you may need to capture more data than you initially estimate, so be sure to understand how pricing changes as your needs expand to avoid any surprises in the future.

CONCLUSION

While case law and solutions applicable to the preservation of social media data will continue to evolve rapidly, the fundamental need for a preservation strategy will not change. This puts a premium on forward-thinking approaches that allow organizations to effectively and efficiently capture content that is needed or wanted for future retrieval and use. That said, perfection cannot become the enemy of the good: Organizations must analyze their needs against the available solutions to define the best social media preservation strategy and execution for the organization.

Note: The views expressed in this article are those of the authors and do not represent the views of their employer and/or any clients. In addition, the authors selected and evaluated select social media solutions in the market prior to and during the course of drafting this article. The solutions discussed in this article are in no way a comprehensivwe list of available solutions, and inclusion of a solution in this article is not an endorsement by the authors or Redgrave LLP.

Reprinted with permission from the July 2012 edition of the LAW JOURNAL NEWSLETTERS. © 2012 ALM Media Properties, LLC. All rights reserved. Further duplication without permission is prohibited. For information, contact 877.257.3382 or reprints@alm.com. #055081-07-12-08