Effective Social Media Discovery

William Hamilton and Brett Livingood

Social media houses increasingly important litigation data. However, the format of data maintained by social networking service (SNS) providers presents e-discovery accessibility and functionality challenges. SNS account holders (the "user") cannot access all data related to their own SNS accounts. Users also cannot access their information as it actually exists in an SNS database. Social media collection is often reduced to capturing screen shots of SNS pages or making serial captures of web page HTML code. Application Programming Interfaces (APIs), though, offer access to otherwise inaccessible metadata and solve collection and search issues by providing workable access to SNS data similar to its true database format. SNS API deployment is limited. Some SNS providers offer a robust publically available API, some public APIs are in construction, and some SNS providers lack such access altogether. SNS providers are seen as neither obligated nor required to offer API access to a user's own data. SNS providers make APIs available to access user data only for strategic business reasons, not with actual or anticipated litigation in mind.

This calculus may change under the evolving rules of e-discovery spoliation and the duty to preserve relevant information. Revoking or not offering API access could prove damaging to SNS providers under a variety of legal theories related to the withholding of evidence. First, if SNS providers are presented with a demand for preservation notice of litigation related to specific user accounts, parties to the action could seek damages if a public API is revoked or modified. The revocation deprives the user of access to "native" data and causes the user, at best, to resort to downgraded web capture methods. Second, in situations where an API is unavailable to the user, SNS providers may be ordered to provide data in reasonably usable formats. Absent compliance, parties may seek sanctions in the form of damages covering the additional cost of manual reviews of SNS web pages.

The collection of social media data typically proceeds under one of two scenarios. Project teams may learn of highly relevant social media data and seek to collect that data in a sound evidentiary manner. Or, project teams may be required to respond to discovery production requests for social media data. The first scenario present an easier case. The project teams may capture the limited target social media data through HTML or screenshot capture. The second scenario, however, is more complex. How does one search for relevant SNS information? The problem does not derive from securing access to social media accounts. Litigators responding to a discovery request may secure client custodians' SNS passwords, or, if needed, courts may compel password disclosure.[1] The problem lies in the exponentially expanding volume and diversity of SNS data driving review costs. Manually reviewing page after page of custodian's Facebook® and other social media accounts could easily overwhelm a

[1] *E.g. Zimmerman v. Weis Markets, Inc.,* 2011 Pa. Dist. & Cnty. Dec. LEXIS 187 (Pa. County Ct. 2011).

project budget. SNS platforms generally do not allow effective search of user data. Live viewing is often limited to a set number of objects. Automated litigation support (ALS) tools and related workflows are required to provide an efficient means to search SNS data.

While ALS tools solve the search problem, they nonetheless depend on APIs to effectively collect and manage SNS data.[2] SNS data exists in databases maintained by SNS providers. To add functionality such as search, an ALS tool requires a robust API to access and collect metadata stored in SNS databases, link objects from one ALS user account to objects linked to another user account, or otherwise maintain a structure similar to that found in the SNS while adding litigation-oriented functionality.[3]

APIs are not foreign or magical SNS tool. SNS providers themselves must construct a variety of APIs to operate SNS platforms. The internal SNS API translates data from the SNS database to the SNS webpage. Publicly available APIs similarly allow external applications to interact with an SNS database when given user permission. An API is like an SNS's loadfile format. However, unlike loadfiles, API forms are neither standardized nor consistent. APIs change as technology, SNS goals and offerings, and programming schemes change. In other words, APIs constantly evolve requiring applications (including ALS tools) interacting with SNSs to regularly update their processes. It is a perpetual cat and mouse game, which can abruptly end if the SNS removes some functionality in the API or the API altogether.

Since APIs are necessary to effectively collect and search SNS data, SNS providers thus may owe a duty to maintain API access to user's relevant data and information during litigation.[4] Tweaks or modifications to an existing API could eliminate access to specific metadata or objects within an SNS database, making API access inadequate for a legally

---

[2] Two social-media focused ALS tools' user agreements illustrate the perceived dependence of ALS tools on unreliable SNS public API offerings: "APIs published by leading social media sites are subject to change and differing levels of access without notice in the sole discretion of those social media sites. . ." *X-1 Social Discovery End-User License Agreement*. Retrieved November 22, 2013, from http://www.x1.com/terms/eula.html#X1SD_EULA; "Cloud Preservation uses Facebook, Flickr, LinkedIn, Twitter, Tumblr and YouTube public APIs to collect data for your archive. Because these social networks have the option of modifying or terminating access to their APIs at any time, Nextpoint can't guarantee the current performance or future availability of this feature or any of its functionality, including access to the data that is currently provided or to any particular social network as a whole." Retrieved November 1, 2013, from http://www.cloudpreservation.nextpoint.com/ (specific language accessed during a software trial).

[3] The Facebook download tool provides a perfect example. While Facebook offers a robust API (FQL) with substantial support, the download tool only allows a download of a specific snippet of database information associated with a user account. So, for instance, when we downloaded Facebook data via the Facebook download tool in November of 2013, the messages.htm page contained all messages sent or received by the user, was structured according to sender and recipient (then date within that group of sender/recipients), did not show any attachments to messages, and even listed some sender/recipients not by name, but by user id--e.g. 1000#####@facebook.com.

[4] For instance, cloud providers including Facebook and other social media sites must comply with litigation holds on user data. *See* Harris, James P. (March 19, 2012). Grabbing Hold of a Cloud: Litigation Holds for Users of the Cloud. *Business Law Insights*. Retrieved November 22, 2013, from http://blog.sheehan.com/index.php/cloud-computing/grabbing-hold-of-a-cloud-litigation-holds-for-users-of-the-cloud/.

defensible capture. Changes to a public API during litigation that causes the parties to resort to manual review can also substantially hinder ediscovery. Specifically, tweaks could essentially downgrade data from a rich, accessible or online format similar to "native" to something analogous to near-line or offline data similar to TIFF or PDF.[5] While SNS providers often do not have duties to disclose user data due to exemption under the Stored Communications Act, the Act does not apply when the user requests access to his or her own data.[6] In most civil disputes, the user generally provides the SNS password to project teams--the user consents. Whether an SNS provider owes a duty to a party requesting SNS data from another party is irrelevant. The SNS provider's duties relate to the user as the "owner" of the data and to the court or tribunal to preserve relevant information to facilitate the truth seeking judicial process.

The SNS provider and user relationship is memorialized in user agreements.[7] Thus, providing the SNS provider with a litigation hold notice involving social media data and a specific request by the user to maintain API access during litigation arguably lays a foundation for a tort claim of spoliation or similar remedy where an SNS provider subsequently renders its public API unworkable. While tort of spoliation claims have been rejected in many jurisdictions, the reasoning of those holdings demonstrates that other sufficient remedies were typically available against spoliators.[8] SNS providers, as third parties with direct, sole access to relevant data in its most usable format, are a unique class. SNS providers have the sole ability to hide, obscure and obstruct the litigation process by an API downgrade.

SNS user agreements may require SNS providers give users consistent access to user data via a robust API even absent active litigation. One could argue SNS contractual duties offering users access to their data (absent user violation of service agreements) implies access

---

[5] While the Sedona Conference commentary on accessibility stresses media-based factors, data complexity factors such as search complexity may also solely determine relative accessibility. So, an SNS provider revoking API access and consequently limiting review to the SNS platform instead of more robust search mechanisms may be analogous to converting native documents to TIFF format. *See* Allman, Thomas (ed.), et al., *The Sedona Conference Commentary on Preservation, Management and Identification of Sources of Information that are Not Reasonably Accessible* (2008), 10-12.

[6] Gigante, Evandro C., & Len, Jacklina A. (Oct. 7, 2013). New Developments in Social Media Discovery in Employment Cases. *New York Law Journal*. Retrieved November 22, 2013, from http://www.newyorklawjournal.com/PubArticleNY.jsp?id=1202621942422&thepage=2.

[7] SNS providers generally acknowledge users own their data. Facebook's user agreement clarifies user's own their data subject to a royalty-free license on pictures on videos, Statement of Rights and Responsibilities. (2013, November 15). *Facebook*. Retrieved November 21, 2013, from https://www.facebook.com/legal/terms, whereas other social media sites like Reddit extend a license to all content, Reddit. (2013, November 20). *User Agreement*. Retrieved November 22, 2013, from http://www.reddit.com/wiki/useragreement.

[8] *See generally* U.S. Law, *Spoliation of Evidence Compendium*. Retrieved November 22, 2013, from http://www.uslaw.org/files/public/Spoliation_Compendium.pdf. Additionally, one of the strongest decisions rejecting a tort of third-party spoliation, Temple Cmty. Hosp. v. Sup.Ct., 20 Cal.4th 464, (1999), depended on assumptions not applicable to social media API issues--namely the inability to end a cycle of litigation due to inability to prove harm or damages in API spoliation. Since social media API issues relate to the form rather than the content of social media, harm and damages could be readily proved by viewing data in the social media platform itself, albeit at potentially substantial cost.

to *all* data created by the user in readily available formats, not simply data as it appears within the SNS platform. Proving damages may be challenging generally. However, users involved in litigation and unable to access, collect, and effectively review their data without incurring substantial costs of manual review could clearly demonstrate harm. In that case, an SNS may be found to owe a duty to offer API access even when a public API is not already accessible.

SNS providers could most easily fulfill duties to users by offering robust APIs and supporting these APIs for reasonable time periods. SNS providers would presumably only be required to maintain robust APIs for users involved in litigation. As such, SNS costs would be limited to adjusting already-existing APIs and allowing access through adjusted APIs to discrete sets of users.