Best Practices for **Governing Social Media Content**

John T. Phillips, CRM, CDIA, FAI

The personal and professional use of social media is continuing to proliferate in homes and enterprises including across international boundaries. According to the Statista website, Facebook alone was estimated to have about 1.94 billion monthly active users as of the second quarter 2015, and many online social networks, such as Google+, Twitter, and Instagram, had several hundred million users each as of August 2015.

A recent study by the Pew Research Center, according to its July 14 online article "The Evolving Role of News on Twitter and Facebook," found that 63% of both Twitter and Facebook users get a major portion of their news from these sources.

Social Media Challenges

As these online communications environments slowly replace e-mail as the preferred electronic communications medium, they are creating a variety of digital records types – YouTube and Pinterest, for example, are specifically designed to share video and image format data – and they are storing records for information-sharing and collaborating.

The impact of this change in communications mode and information-

sharing architecture creates unique challenges for ensuring that personal and workplace-originated electronic records are created, stored, and retained in a responsible manner.

These challenges can be met with an effective information governance (IG) program, the enterprise-level strategic information management perspective that most thoroughly encompasses records retention guidelines, regulatory compliance mandates, information privacy concerns, intellectual property protection, and litigation document production requirements.

IG for Social Media

IG – which ARMA International defines as "A strategic framework composed of standards, processes, roles, and metrics that hold organizations and individuals accountable to create, organize, secure, maintain, use, and dispose of information in ways that align with and contribute to the organization's goals" – is rapidly developing. Its concepts are embedded into the Generally Accepted Recordkeeping Principles® (Principles), which are available at *www.arma.org/principles*.

The implementation of the Principles and IG practices depends on the state of an organization's information management policies, IT systems architecture, and business environment. This means that the best way to approach IG will vary based on the organization's recordkeeping mandates, litigation profile, risk management priorities, and social media systems architectures.

Because IG is an evolving arena, it currently has few examples of extensive enterprise-wide implementation, so there is a need for sharing new perspectives and lessons learned, especially with regard to incorporating social media records into IG plans. This article provides technical information about social media platforms and suggestions for governing their use that should be of value to organizations that are struggling to extend IG controls over their social media-based information assets.

Social Media Technologies

Social media applications like Facebook and Twitter usually reside on multiple servers hosted by third parties in the Internet "cloud." They also can be hosted and maintained internally by organizations wanting to reap the rewards of browser-based collaboration functionality while maintaining access control and security themselves. But, most organizations today want to benefit from lowering their IT systems infrastructure maintenance costs through taking advantage of cloud-based applications that intrinsically offer worldwide access, minimal IT infrastructure installation costs, and little or no internally required help desk or training challenges.

This approach to outsourcing IT support and maintenance infrastructure means that organizations do not directly control the recording and retrieval of their electronic records that are being stored on third-party systems. This means they must ensure that their end user license agreements (EULAs) and service level agreements (SLAs) are designed or modified to be in compliance with their IG program plans.

EULAs and SLAs

EULAs and SLAs form legally binding contractual relationships between software users and vendors of the software. The EULAs that organizations must sign with cloud vendors rarely address data management; vendors make their own design decisions to maximize their software platform's responsiveness, ease of maintenance, and robustness of security.

Typically, software vendors employ EULAs to limit the use of their software by prescribing the software buyers' rights with respect to operation, archiving, sale, and backup. SLAs are arranged by both software buyers and software vendors to specify mutual expectations for system performance, data ownership, and service support levels.

EULAs are generally created by software vendors, whereas end users often initiate and negotiate SLAs. Although both are contractually binding, they may need to be enforced through legal system actions, thus creating business cost and operational hazards for those organizations trying to exert their rights under these agreements.

These types of agreements present opportunities for proactively addressing issues that will arise during IG initiatives, especially with respect to the ownership and management of electronic records.

[EULAs and SLAs] present opportunities for proactively addressing issues that will arise during IG initiatives, especially with respect to the ownership and management of electronic records.

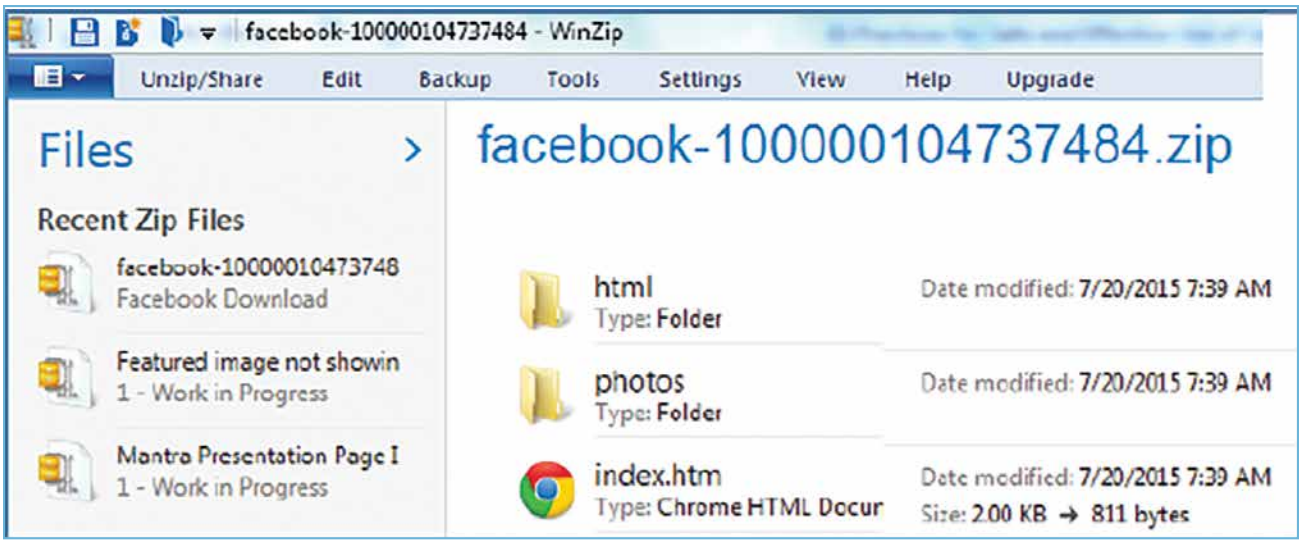


Figure 1 - WinZip software view of downloaded Facebook files

Use of Software

EULAs will often contain clauses that limit the use of the software, potentially causing problems with retaining multiple copies of the software for business continuity or data archiving reasons.

Liability Limitations

EULAs also contain clauses that limit the liability of the software vendor if the customer experiences loss of data while using the software. SLAs focus more on a characterization of services, expectations of performance, problem resolution, customer responsibilities, and limitations of the scope of the agreement.

Although historically, SLAs were mostly process- and services-focused, the need to deliver potentially vast quantities of data to customers for electronic discovery exercises during litigation has become a feature of some proactively designed SLAs.

Ability to Negotiate

Unfortunately, most EULAs are agreed to during the standard processes for installation of software by end users or information technology (IT) systems personnel, thus they are vendor-prescribed and largely non-negotiable. When logging into many social media systems to create a user account, the new system users simply click acceptance to the EULA with their mouse or the software does not install. Unless IT and IG personnel intervene in this process by prior arrangement with the software vendor, there will be no opportunity to decline and negotiate this agreement.

In contrast, SLAs are often subjected to contractual negotiations, especially during procurement business processes, and present more opportunities to insert IG policy that will support an organization's IG program.

Data Delivery

SLA deliberations allow specification in advance of a need for data deliveries, data formats, and database system reports with metadata and data content expected during litigation or audits. Having these parameters designed in advance into an SLA with a vendor can preclude them from telling a customer to just "log into the system and print it yourself."

Access and Preservation

For IG policies to work in an organization, advanced planning will be required to ensure that the organization can access and preserve any information stored on social media by an employee or contractor when that action is taken during a "normal course of business."

For instance, Facebook does not agree to share all information on all parties with whom an individual may have communicated or shared data. As a best practice for any IG program, the data map that describes the information repositories to be subject to organizational retention policy must include a description of the information an individual may store on Facebook or other social media, as well as the means and methods of preserving it.

Social Media Records Capture

Capturing records outside of the organization's control has unique challenges that will need to be addressed. Some social media vendors allow users to obtain their own data in limited formats, and they make the process straightforward, but obtaining data from others sometimes requires special expertise.

Following is information about two of the most prevalent social media platforms, Facebook and Twitter. It makes it clear that even though information can be obtained from both fairly easily, interpreting and analyzing the results can be challenging.

NEXT LEVEL™
information governance assessment

Your business data is doubling by the year. And all this new data can either help you or hurt you. Find out what it's doing for your company with the Next Level Information Governance Assessment. You'll discover areas of strength and opportunities for improvement. In the end, you will be empowered to increase organizational transparency and data integrity while decreasing risk.

Start turning information into an asset by visiting **arma.org/nextlevel**.

Cloud Database Management Systems Architecture

There are different considerations for storing data in social media applications than for single server content management applications. Due to geographically dispersed users, social media application data must be distributed across many geographically dispersed servers to maximize data retrieval speed and promote data redundancy. Following are examples from three broadly used database management systems.

Amazon's Dynamo

Amazon's Dynamo is a database system used to store large data volumes that span many servers for data-intensive applications requiring fast read/write activities. Based on NoSQL, it is a non-relational database management system that does not need fixed table schemas and can scale without complex join operations. According to Amazon's web services documentation website, it uses consistent hashing to partition data across multiple servers. Data replication is supported for increased data availability, and data versioning is used to ensure consistency in data recall.

Google's Bigtable

Another example of a database system used to store social media data is Google's Bigtable. By using proprietary technology, such as Google File System (GFS), Chubby Lock Service, and Stored String Table (SSTable), Google has created its own schema-free database system built for high performance with respect to the special requirements of storing social media data.

In Bigtable, data and log files are stored on top of the GFS in SSTable format. Additional software technology constructs that empower Bigtable are the use of a master server, tablet servers, and a client library. This type of database management system architecture is much more complex than simple relational database architectures.

Apache's Cassandra™

Similar to both Dynamo and Bigtable, Apache's Cassandra™ database management system uses elements of both, including a Bigtable data model and a Dynamo-like architecture for servers. Cassandra is used by Facebook, Twitter, and other social media applications for large scale data storage on distributed systems.

Instead of using a set of relational tables, Cassandra uses a keyspace – a container for application data – where a table column family consisting of different rows and columns does not have to share the same columns. This provides an ability to store frequently accessed columns in separate files located closely together to enhance speed of information recall. For this reason, information about a Facebook user's friends, for instance, can be more easily collocated, speeding up retrieval of frequently accessed information.

System Performance Benefits

These specialized database architectures improve overall system performance and maintenance activities by providing increased capabilities with respect to information replication, partitioning, and querying. They also allow managing the "social locality" of data by enabling the storage of a user's data closer to its neighbor user's data store, thus increasing system performance.

Unfortunately, this means that data stores for social media applications are sufficiently complex that simple data dumps or backups to ASCII text, comma delimited text, spreadsheets, or simple database formats are unlikely to provide meaningful data for recordkeeping purposes.

Facebook

Facebook has specific informational pages, such as "Accessing Your Facebook Data" (*www.facebook.com/help/405183566203254*), that direct users regarding downloading information about shared posts, pictures, and videos, as well as information on messages, conversations, and profiles.

Specific information is available for requests from law enforcement agencies and parties to civil litigation. But where an individual has agreed to share his or her information with those entities, Facebook prefers that users authenticate the appropriateness of this third-party access, including by signing an authorization for access.

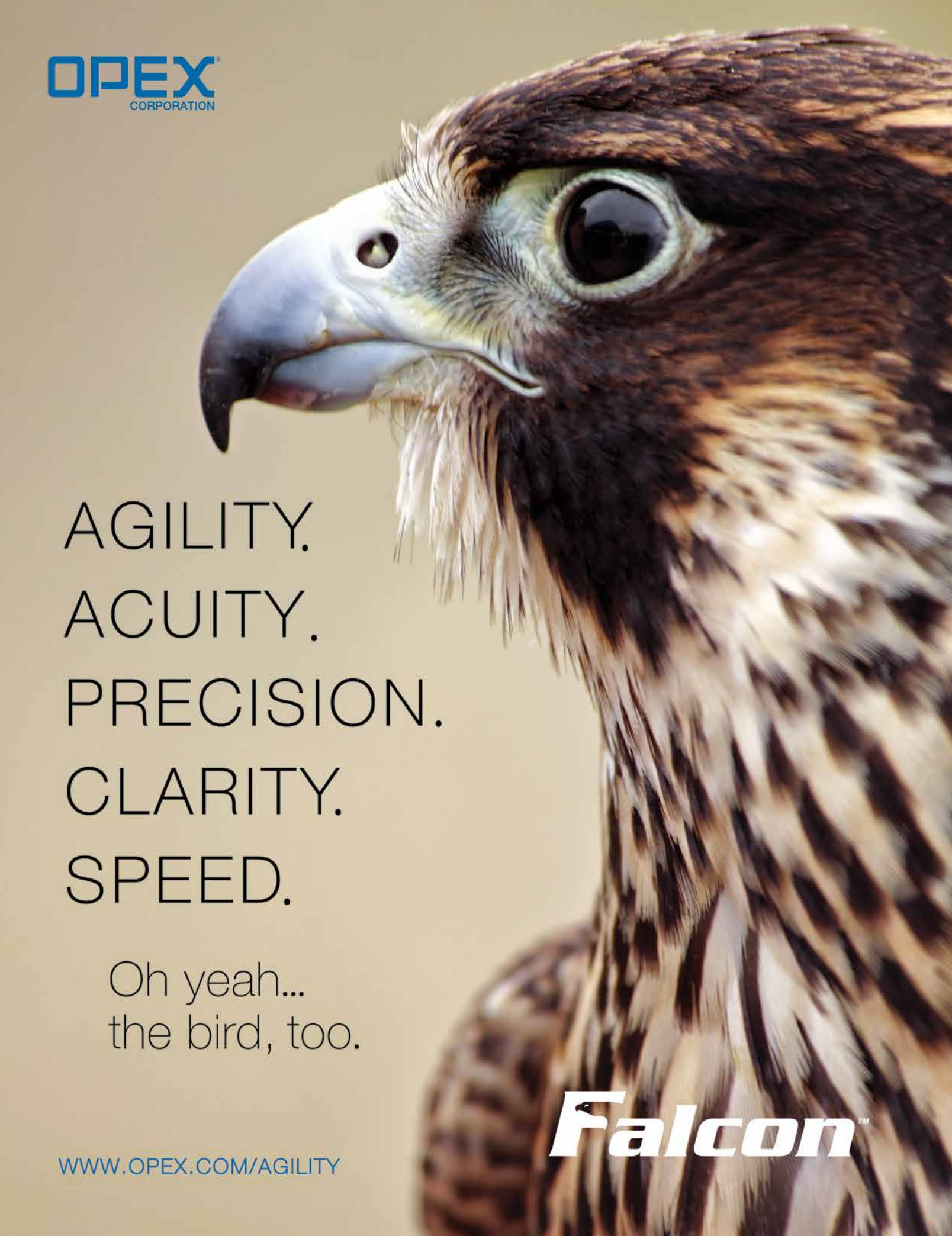
Using Facebook's download option can provide an extensive variety of data, including logs of activities such as posts, administrative settings, messages sent/received, login histories, and personal information posted. This information does not replicate the Facebook application interface most users are familiar with, though, so it will take some interpretive analysis to make use of it as a record of Facebook content or interactions.

It also is not immediately available online and must be collected by Facebook and sent to the user through an e-mail interaction, unlike many corporate computer system reports that are available online. The data comes as a compressed "zipped" file with HTML and JPG data format components in different directories that must be viewed separately. (See Figure 1.)

Twitter

Unlike Facebook, Twitter does not offer a directly available download for users to obtain information about their accounts' records and prefers that users make requests for information through help desk queries.

Twitter specifically warns that it does not validate the authenticity


OPEX
CORPORATION
AGILITY.
ACUITY.
PRECISION.
CLARITY.
SPEED.
Oh yeah...
the bird, too.
Falcon™
WWW.OPEX.COM/AGILITY

of users' identities or metadata and does not validate the information they post online. Twitter also states that retention of the information varies depending on its perception of the information's value to users or for varying system administration purposes.

Twitter information will be produced as text files in word processor compatible format, and users are cautioned that some images seen on Twitter may be hosted on third-party systems. The company also cautions that cost reimbursement may be requested for some types of information.

Twitter, like Facebook, has a well-developed policy for law enforcement organizations to request information. Users who are the subject of a law enforcement request for information will be notified.

Decision to Outsource

There are different considerations for storing data in social media applications than for single server content management applications. Due to geographically dispersed users, social media application data must be distributed across many geographically dispersed servers to maximize data retrieval speed and promote data redundancy.

There also are no universally accepted data formats or processes for social media records retention due to the tremendous variety and scale of the technologies and data stores used to operate social media applications. (See the sidebar "Cloud Database Management System Architecture" for technical information about three of the prevalent database management systems.)

Due to the technical and procedural complexities involved in capturing, storing, and retaining social media records, many organizations prefer to enlist the aid of third-party organizations with experience and expertise in the arena.

As discussed in the ARMA International *Newswire* article "Scheindlin Issues Landmark Opinion on Custodian Self-Collection," there is considerable professional debate and discussion as to how advisable it is to employ "custodian self-collection" in e-discovery activities during litigation. While records custodians may be the best informed about the nature, content, and locations of electronic records, they are often only marginally prepared to accurately collect records and could have inherent conflicts of interest in doing so.

The challenges of capturing social media records may indicate the best solution in many cases is to retain third-party vendors of software, hardware, and data collection services. Vendors of such services can operate on behalf of users to make requests for records, capture records, and store them in a dedicated repository for archiving and retention.

This can be performed occasionally or through a social media application style interface so users' data can be accessed in a manner that creates the look and feel of the original social media system. Though these services come at a price, the costs of collection, storage, and retrieval may be more easily managed than the costs of self-collection and local data storage.

Best Practices for Using Social Media

Social media use by an organization's employees must be subject to the same policies as for the use of internally hosted systems. However, as mentioned earlier, information stored on externally hosted and administered computer systems will be managed based on the dictates of the social media system owner unless there are specifically negotiated EULAs and SLAs that ensure compliance with expected IG policies.

The best practice for any organization today is to take initiative in arranging for technology solutions to preserve social media online records. This will ensure that the records' content, metadata, and formats to be preserved are designed in advance to meet the needs of specific records for the IG program.

It is best for most organizations to employ these practices:

1. Enforce IG policies that distinguish employees' personal and work-related use of social media.
2. Ensure that IG issues are encompassed by contracts with vendors of cloud-based services. Design or modify EULAs and SLAs to be in compliance with IG program plans.
3. Create data maps for e-discovery initiatives to incorporate social media applications.
4. Consider outsourcing the collection, storage, and production of social media records for e-discovery.

Above all of these, the best practice for effective IG of social media is to do comprehensive advanced planning for its governance before the organization begins using it. **END**

John T. Phillips, CRM, CDIA, FAI, can be contacted at john@infotechdecisions.com.
See his bio on page 47.

Due to geographically dispersed users, social media application data must be distributed across many geographically dispersed servers to maximize data retrieval speed and promote data redundancy.