



ARMA INTERNATIONAL
**EDUCATIONAL
FOUNDATION**
RESEARCH · EDUCATION · SCHOLARSHIP

***Social Media Systems
Records and Information
Governance Challenges***

By

John T. Phillips

September 24, 2015

Project Funded by:

ARMA International Educational Foundation

© 2015 ARMA International Educational Foundation
www.armaedfoundation.org



Table of Contents

ACKNOWLEDGEMENTS.....	4
INTRODUCTION.....	5
SOCIAL MEDIA NETWORKING SYSTEMS.....	8
Facebook 	9
Background.....	9
Functionality.....	9
Metadata and Privacy	10
Twitter 	10
Background.....	10
Functionality.....	11
Metadata and Privacy	11
LinkedIn 	12
Background.....	12
Functionality.....	12
Metadata and Privacy	13
YouTube 	14
Background.....	14
Functionality.....	14
Metadata and Privacy	15
SOCIAL MEDIA TECHNOLOGIES	16
Data Storage.....	16
Facebook.....	17

Twitter.....	18
LinkedIn.....	19
YouTube.....	19
SOCIAL MEDIA RECORDS CAPTURE.....	21
Facebook.....	22
Twitter.....	23
LinkedIn.....	25
YouTube.....	26
Summary.....	30
AGREEMENTS AND CONTRACTS.....	31
EULAs.....	31
SLAs.....	33
Summary.....	34
SOCIAL MEDIA ARCHIVING TECHNOLOGY TOOLS.....	36
Solutions Design.....	36
Solutions Alternatives.....	38
Social Media Technology Business Case.....	39
INFORMATION GOVERNANCE FOR SOCIAL MEDIA.....	41
APPENDICES.....	45
Categories of Facebook Data Available:.....	45
CONTACT.....	54
END NOTES.....	55

ACKNOWLEDGEMENTS

The author would like to thank the ARMA International Educational Foundation, its Board of Trustees, and officers for the opportunity to contribute this research to the professional literature on Social Media Information Governance. In addition, the author would like to thank the numerous individuals who were interviewed to build the base of knowledge presented in this work for their insights into the issues and principles that impact the management and retention of Social Media records.

INTRODUCTION

Social Media (SM) is fast becoming a primary conversational environment for many individuals, as well as a growing means of communicating business information, sharing data and storing electronic records. Sometimes termed social networking services (SNS) or online social networks (OSN), these leading technology enabled information sharing architectures, such as Facebook and Twitter, are often used as a replacement for electronic mail services due to their information broadcast and sharing capabilities, including business or personal data, text content, and images. These new communications technologies have increased the personal enjoyment and productivity of many businesses, organizations, and individuals. It is now possible to easily share thoughts, concepts, data, documents, pictures, and videos among friends and families, as well as across enterprises, both nationally and internationally.

The use of Social Media in personal lives and professional workplaces is continuing to proliferate in homes and enterprises, including across global boundaries. Facebook alone is estimated to have about 1.44 billion monthly active users as of the first quarter of 2015¹ with many online social networks such as Google+, Twitter, and Instagram having several hundred million users each.² In a recent study by the Pew Research Center, 63 percent of both Twitter users and Facebook users get a major portion of their news from these sources.³ And, as these online communications environments slowly but surely replace email as a preferred electronic communications medium, it becomes clear that they are being used to create a variety of digital records types, as well as, actually store records for information sharing and team collaboration. Some social media, such as YouTube and Pinterest, are specifically designed to share video and image format data.

The expanding use of Social Media applications over the last decade has created an obvious need for Social Media policies to better manage this information intensive environment. For that reason, the creation of Social Media “policies” has proliferated among most large corporate users of this environment and many public organizations now have Social Media use policies. There are a great many examples of these that can be found through Internet searches or from various online “databases” of Social Media policy.⁴ Such policies often cover extensively how to responsibly use Social Media without ever mentioning the need to preserve the data that is being stored in that environment.

The tremendous utility of the Social Media communications environment lends itself naturally to the deployment of technology tools in support of business workflow and information management. Such tools include collaborative software that enables virtual workgroups to post discussion documents to Internet Web sites, image content repositories that enable businesses to post informational advertising, and blogging sites that encourage professional discussion and concept development. Many of these technology driven “electronic environments” have historical

counterparts in more traditional office environments, such as presentations in office meetings, publication of advertising in traditional print journals, or conferences where professionals can exchange ideas for collaborative professional development.

However, the business models of traditional office environments usually have workflow “nodes” where records are captured for official consideration regarding organizational retention and disposition actions. During typical physical processing of correspondence, reports, or meeting presentations in the past, it was apparent that such hard copy items needed individuals and departments to preserve and retain documents for future reference. In addition, informational items were being created, dispersed, and utilized mostly in-house, at least when the information was originally generated, making capture of these potential official records more likely. Most individuals would take the initiative to reserve copies of the records for their own purposes, and many organizations subjected selected records to official records retention policy at some point in the records workflow. In fact, many organizations staffed and assigned resources to a Records Management Department to lead this effort and assure compliance with organizational policies.

Today’s drastically altered business models take advantage of computer based technologies but have simultaneously created new opportunities and dangers for managing information. Laptop computers were probably the first mobile information repository with challenges in assuring resident information was well governed, but many cell phone users also quickly migrated to “smart” phones capable of participation in email and text message exchanges that can store messages, documents, contacts and images locally on the phone itself. These transitions to creating and sharing information collaboratively through Internet servers and networks occurred during an increasing global acceptance of using Internet based Web pages for information storage as well as distribution. The cumulative effect has been that many individuals and organization now rely on “cloud” based communications infrastructures to both create and store electronic records.

Information created, distributed, and stored on SM Web servers, in externally hosted software applications, and on publicly accessible communications networks needs to be protected, secured, and managed according to an organization’s accepted Information Governance (IG) policies. These IG policies incorporate elements of records retention rules, information technology (IT) systems security procedures, litigation hold compliance processes, and information privacy mandates. Unfortunately, the design, adoption, and implementation of IG policy is a major strategic undertaking for an organization’s internal systems and data regardless of recording format. However, the extrapolation of these internal governance expectations to externally hosted computer systems and another organization’s business processes presents even more daunting operational challenges.

It is not usually possible to expect compliance with internally mandated policies on the part of external independent self-managing organizations. They have their own IT and IG challenges and their business processes serve their own unique business models. Most inter-business operating

relationships are conducted under contractual frameworks that generally do not take into consideration the respective IG goals and IT architectures of each organization. Extrapolating IG policy frameworks to encompass the larger management umbrella covering both internal and external business processes and data stores often requires negotiations instead of mandates.

A typical contractual relationship that might exist between an externally hosted application such as Facebook and a user of the service would be an End User License Agreement (EULA) or a Service Level Agreement (SLA). These agreements are usually a part of software installation procedures (for EULAs) or services subscription processes (for SLAs) and often non-negotiable. As will be seen from the research in this study, it is possible to extend IG practices into data sets stored externally, if some compromises are accepted in data formats, capture processes, and retention periods for electronic records. The specific nature of these compromises will vary with each vendor of the Social Media services offered.

However, considering the magnitude of the electronic records now being “outsourced” to “cloud” services storage devices and applications, well managed organizations must create comprehensive IG strategies that address data stored in SM applications if they are to be fully comprehensive in managing their information assets.

SOCIAL MEDIA NETWORKING SYSTEMS

Social Media (SM) applications have become a computer-mediated information sharing community that encompasses both personal lives and professional workplaces. In homes and enterprises today many individuals now spend ever-larger portions of their days and evenings communicating with friends and business associates through remotely hosted software systems such as Facebook, Twitter, YouTube and other applications. With increasing numbers of individuals using Social Media for personal and business news as well as information on companies, individuals, organizations, products, and intellectual concepts, the utility of these information gathering tools is only expected to increase.

In addition, access to these data resources is now a global phenomenon, creating large audiences of users, enormous virtual libraries of informative content, and many different technical and social challenges for creators of software applications, application interfaces, and data storage repositories. For many users of Social Media, text oriented electronic mail and text messaging systems are less preferable than a graphical interface to somewhat menu-driven “tabbed” systems with entertaining and informative images, audio, and video content. It is now common for organizations, both public and private, to have Facebook “storefronts” to share information about their companies, products, and services. Social Media interfaces are especially used by younger generations of people as many have grown up using these interfaces on computer games for personal entertainment, as well as, using Internet Web sites to perform research for educational challenges in secondary education and university settings. Text based email services can seem less interesting and appealing for this reason, though simple fast text messaging is still used extensively by this group of individuals.

However, the interaction interfaces, resident application functionality and data types stored vary tremendously between Social Media systems. This variety of technical and functional capabilities makes it very important for both individuals and organizations to assure that their information is governed reliably and consistently with their online electronic records properly preserved and protected. For this reason it is important to understand the differences between some typical Social Media applications and to investigate how those differences may affect plans to govern that digital information.

The following brief overviews of several types of Social Media systems are not intended to present a complete discussion of their extensive utility and functionality for users. These overviews are intended to provide an introduction to the types of data and metadata that might be retained in these systems online that could be of interest to system users, researchers, legal counsel, or investigators.

Facebook

Background

Facebook (<https://www.facebook.com>) is an online social network created in 2004 by Mark Zuckerberg and some Harvard roommates primarily to serve college students in the Boston area and some other New England universities with a goal of sharing messages, personal information, videos and other content through postings to the Web site. Users log in to create an account (user profile) and declare other individuals to be “friends” thus letting them access the information posted online. Users can also join groups with similar interests to post messages, share photos, and videos and also receive notifications when “friends” or members of the various groups change their postings or status. Facebook often sends notifications to users of potential friends with whom they may not be currently linked by taking the initiative of searching a user’s past workplace, educational, or family data and links. The goal is to promote maximum sharing of information among users. Facebook had 1.44 billion users internationally as of March 2015.

Facebook’s business model is based on revenues received from advertising. It offers reaching customers by associating content to user’s interests within an appropriate social context, and offers communication based on age, gender education, location, or work experiences. Advertisers can set up their own Facebook pages to facilitate engagement with customers and promote dialog about products and services.

Functionality

Facebook provides for users a number of functions, including a user’s Home Page, Photo galleries, Timeline, News Feed, Email connections, Chat sessions, Group discussions, Lists of Events and Places, as well as, notifications to others of postings. Facebook provides a collaborative information sharing environment that precludes individuals and organizations from the need to build their own Web sites, while providing a networking infrastructure for communications and links to information and experiences. It would be possible to replicate much of the functionality one gains from becoming a Facebook user with respect to the ability to post information online on a self-created Web site.

However, the access provided within the Facebook application to searching online for other individuals and interests on Facebook and then setting up regular communication links to them would be extremely difficult and time consuming for an individual Web site creator to duplicate. It is much easier to share information about oneself on Facebook and then share experiences as others do the same. Many users take advantage of the “poke” function to reach out to other users or the “friend” function to declare a positive relationship with other users. In addition, one can create picture portfolios for others to view and comment on the postings of others, as well as, see postings to comments by others users.

Metadata and Privacy

Facebook wants to create a very personalized communication and information sharing environment to maximize the ability of individuals to identify others with personal similarities or interests, as well as, provide a fertile base for advertisers to grow customer contacts. For this reason, there is considerable information that is accepted from application users, once an initial electronic mail (email) address is provided. This email address is confirmed by sending an email from Facebook to the user for verification of the email address along with a CAPTCHA ⁵test to keep out spammers and hackers. This is the only required evidence of authenticity that a user must actually submit. Subsequently Facebook offers an opportunity to provide birth date primarily for determination of age-appropriateness of content and to send Facebook “friends” notification of one’s birthday.

Some of the other metadata that Facebook will ask for and store, however, can become pretty comprehensive. Access can be provided to one’s email contact listing, likes and interests, schools attended, educational attainment, workplaces, friends, school mates, relatives, personal photographs, spoken languages, religious sentiments, political views, cultural interests (music, art, etc.), and sports or other activities. For this reason, a user’s profile privacy settings are an important filter that can serve to regulate who can or cannot have access to information.

One of the advantages or disadvantages of using Facebook, depending on one’s perspective, is that entering data about oneself such as a High School or College attended will create a spontaneous system notification to the user that there are other individuals on Facebook with similar backgrounds and who have similar or related friends. The goal is to promote communications with these individuals, but illustrates clearly that Facebook can be the source of much more information about an individual than they would normally share with the public.

This collection of data could make a privacy invasion of one’s Facebook account far more revealing than even breaking into one’s home. It is for reasons such as these that it is critical for users of Social Media to not share account passwords or other login security metadata that could vastly increase the impact of a security failure. It is also critical that one review one’s profile and settings occasionally. It is possible to allow complete public access to all information, or limit access to Friends, Only Me, or create a Custom access profile.

Twitter 

Background

Twitter (<https://twitter.com>) is an online network created for social communications and commentary exchange. Each text message sent on Twitter is referred to as a “tweet” and can be 140 characters in length. Anyone can read these messages, but only registered users can create them. Twitter can be accessed through the Twitter Web site, short messaging services (SMS), or

mobile applications. It began in March of 2006. Jack Dorsey, Evan Williams, Biz Stone, and Noah Glass created Twitter in March of 2006 and the service rapidly expanded to being one of the Internet's most visited Web sites with over 100 million users posting 340 million tweets per day by 2012. As of May 2015, Twitter has more than 500 million users with 302 million active users.⁶

Twitter's business model is to offer small work group style communications with tremendous speed and group networking capabilities. Though often referred to as a social network or "microblogging" platform, its primary value added for most individuals is typically the ability to transfer small bits of information quickly. Social groups are a part of its functional architecture; however, the limitation of text to 140 characters per tweet focuses most commentary into timely but brief sentences and only modest amounts of information content per tweet. Twitter news and alerts can be extremely useful since they can be quickly widely broadcast and refer the reader to Web sites or other more informative content. In addition, the ability to focus one's interests to specific content types, sources, or individuals creates a customizable information service.

Functionality

Twitter is very easy to set up and configure, by simply going to the twitter.com Web site and filling out name, email address and password, and then selecting a Twitter handle. Twitter encourages one to start following individuals and organizations at that time, by selecting from among the most commonly followed individuals on Twitter so that their tweets will be included in a customized Twitter feed. One's Twitter feeds can be customized by interests, as well. A good example of the Twitter social media environment is a part of the Twitter interface known as "Trending Topics" that identifies some of the most commonly posted topics or individuals that may be of interest to Twitter users to follow.

One initial "opportunity" provided by Twitter is a request to view one's email address book, instant messages, and LinkedIn connections to identify potential individuals that may be interested in following their tweets. Of special interest to users of Twitter is the ability to adjust settings to concentrate communication with either the public at large or focused workgroups. For persons interested in public debates and discussions, the public setting may suffice. However, many individuals find a special utility in Twitter by using account settings to protect tweets to only being visible by followers.

Metadata and Privacy

A positive response to a request for allowing the application to search one's email address, instant messages, or LinkedIn account will send this data to a Twitter controlled server, but is also an example of the potential privacy issues that come to bear if external cloud vendor computer sites are hacked. With Twitter, however, there is less data used by the application (than as is the case with a Facebook profile or timeline) that could be considered personal or business sensitive stored to use by the application itself. Most of the identifiable information regarding an individual or organization comes from the commentary posted in actual tweets rather than extensive context or

background.

As with Facebook, it is possible to associate an “avatar” or personal photo with tweets. This is a good example of the dilemma of either being openly communicating by revealing one’s appearance or choosing to remain more anonymous by using only a pictorial caricature that resembles something with which one identifies or feels is somehow representative of one’s persona. Unlike Facebook where most uses of the communication platform concentrate on building positive “friendships” and “likes”, Twitter users often want to more openly broadcast their tweets to larger less well known audiences, thus making revealing personally identifiable information a bit more risky. Privacy control settings can moderate these issues though there is less opportunity with Twitter to engage in content development. For example, one’s bio is limited to 160 characters.



Background

LinkedIn (<https://www.linkedin.com>) is a Facebook-like social network founded in 2002 and launched in 2003 by Reid Hoffman, Allen Blue, Konstantin Guericke, Eric Ly, and Jean-Luc Vaillant. It is used by professionals and businesses for exchanging information about themselves to other interested individuals and organizations. LinkedIn quickly grew to approximately 20 million users by 2006 and as of May of 2015 is estimated to have had about 364 million users in 200 countries.

This rapid rise in users is attributed to the premise of LinkedIn that information exchanges about professional accomplishments and interests enhance careers and offer a means of electronically reaching out to other professionals for sharing ideas and exploring potential work opportunities. Professional business relationships form the basis for information sharing on LinkedIn, in contrast to the personal information exchanges that occur on Facebook. The existence of LinkedIn also allows individuals to very conveniently separate their information exchanges about professional activities from information sharing about their personal lives, as is often expected by their employers or customers.

Functionality

A LinkedIn user typically uploads information to the LinkedIn online Web application that resembles a professional resume. This includes academic and professional accomplishments, as well as, contact information and a professional timeline of career positions and accomplishments. Along with this personal professional data is an opportunity to become “linked” to other professionals that one knows or with whom one has had work experiences. LinkedIn pages are increasingly used by Human Resources staff or other organizations seeking professional services to identify candidates of potential interest. In recent years the extensive use of LinkedIn contacts

has created a situation where personnel professionals increasingly look at the type and nature of contacts rather than the volume.

With respect to value added to system users, LinkedIn provides a useful means of forming a network of similarly engaged professionals that share information about work experiences, as well as, creates a platform for showcasing their contacts and credentials. This can result in work or job referrals for those individuals with occasionally reported updates to their LinkedIn profiles and their skill sets. It also reduces for some individuals the need to create and maintain their own professional Web sites. However, the utility of this information for both job seekers and potential employers is greatly impacted by the currency of the information maintained in the LinkedIn profiles. Similar to other social media, LinkedIn users can publish posts or upload photos.

Metadata and Privacy

Due to the resume style of the LinkedIn profile and ease of seeing their fellow professionals' contact information online, privacy settings are an especially important aspect of using LinkedIn. It is common for instance to get requests to be "linked" to someone that an individual has never met in person. Each individual must balance the appearance of a huge volume of contacts with the assumed utility of actually being in communication with all of those individuals. In addition, it can be argued that exposing some of this information publicly can be adverse in that it enables this information to be gleaned by potential competitors.

Each individual user must make these decisions based on the actual added value to their careers that they will derive from any new linked contact. It is also possible for instance to determine the alleged credentials of an individual and for a competitor and then either claim similar credentials or attempt to duplicate them. It is important to remember that LinkedIn does not typically verify claims by individuals regarding the information in their profiles. It is important for LinkedIn users to make these verifications of profile information by research or background checks conducted independently of claims made simply on LinkedIn. Many of an individual's contacts or employers may also place restrictions of the details of work experiences documented and discussed publicly.

One unique aspect of LinkedIn that is a powerful asset for an individual's presence there is the concept of "recommendations". Similar to a Facebook "like" but more directed, a LinkedIn user can request recommendations from other users in their contacts list with whom they may have had particularly involved and successful experiences. These "endorsements" can carry great weight in decisions regarding employment and can be detailed in some instances. However, it is another example of revealing online some potentially very private information, and must be done carefully to avoid overexposure or risking violation of an organization's privacy policies. In addition, anyone using LinkedIn to search for new employment must carefully control their privacy settings with respect to their activity broadcasts and preserving some anonymity in their public profile if they want their job searches to have some privacy.



Background

The most well-known video sharing site on the Internet is undoubtedly YouTube. It was founded in 2005 by Chad Hurley, Seven Chen, and Jawed Karim, early employees of PayPal. By mid-2006 YouTube was reporting the uploading of videos at the rate of 65,000 per day.⁷ In November 2006, it was bought by Google for \$1.6 Billion. YouTube is said now to have more than 1 billion users with more than a million advertisers using the Google ad platform. A majority of these are small business users.⁸

Many YouTube video viewers do not sign up for an account, though this is required for uploading videos. The social impacts of YouTube as a means of communicating educational documentaries, entertainment videos, on-the-scene captures of news events, and government press announcements has been extensively documented. YouTube is widely acknowledged as having an international impact on news, education, and entertainment. However, the ability of almost anyone or any organization internationally to upload videos for viewing also creates issues with respect to the privacy of individuals, organizations, or events being covered, in addition to potential copyright issues. YouTube is used by both individuals and organizations, especially for advertising purposes.

Functionality

YouTube is configured to serve as a video distribution mechanism, much like commercial television. However, rather than commercially generated “programming” YouTube enables almost anyone to “broadcast” videos about themselves, family, friends, or any topics of interest. Videos can be educational, entertaining, or just humorous. In fact, when a video is particularly entertaining and is so extensively viewed it is said to have “gone viral,” it is common for commercial cable and broadcast television media to display them on their own news and entertainments shows, as well as, occasionally post some of their own content on YouTube. YouTube videos can be viewed by most anyone with a modern Internet browser that is HTML5 capable. However uploading a video requires becoming a registered user and providing some personal data. One unique aspect of YouTube is the concept of channels. Each individual can configure specific channels, similar to traditional television viewing, that display videos in designated subject areas on their home page. They may include the videos the account user has uploaded or videos found on YouTube. Based on this body of preferences, it is possible to determine many of the interests of an individual that are important for marketing purposes to many organizations. YouTube regularly determines whether subject areas someone has been viewing and customizes the content they will see on their home page.

One important aspect of YouTube is the ability to comment on uploaded videos. Video posters can add their comments as well as, encourage viewers to post their own comments in a “add a public comment” box that appears below the video. Public postings appear in reverse chronological order

below the video. YouTube reviews these comments for appropriate and inappropriate content due to some very controversial and sometimes abusive comments made over the years about some videos.

Metadata and Privacy

YouTube stores less metadata for an account user than many other Social Media systems as it is less oriented toward enhancing communication between individuals than it is oriented toward enabling individuals to view videos and showcase content (primarily videos) to audiences. It does store the usual login information, such as username, password, and email address for authentication, as well as, information on connected accounts to other social media systems. However, where other Social Media systems focus on enhancing connections between individual users, YouTube maximizes a user's ability to upload their own videos, and to view videos created by others.

A primary difference between YouTube and other Social Media systems is the concept of channels, where an individual can create and/or select channels of content to display based on the subjects of the videos. In fact, some full length television shows are available through YouTube. It is also possible to receive regular notifications from other YouTube content creators by "subscribing" to their posted content. Standard channels available for subscription such as Popular on YouTube, Music, Sports, and Gaming, also exist. A History function allows seeing what an individual has been viewing over time, though it can be easily edited.

There are numerous privacy and copyright issues of special interest with YouTube use. It is possible to violate both privacy and copyright laws by posting inappropriate content or content not owned by the person uploading it. It is also possible to place inappropriate content in comments on videos and to place copyrighted works in uploads for public viewing. For this reason, YouTube has often been the center of litigation with respect to the authorship of some content, as well as, the number of view of any particular content to establish impact and potential financial damages. It is very important to understand the mechanisms of obtaining Social Media use data.

SOCIAL MEDIA TECHNOLOGIES

Social Media networking applications take advantage of many technologies and business processes that can make capturing electronic records, practicing records retention policies, and implementing Information Governance strategies very complex. Social Media systems are by definition engaging in the exchange and sharing of information across Internet network architectures and computer servers that are external to an organization's in-house IT systems and business operations. In addition, Social Media applications are seldom created by the organizations they serve, but rather are generated by Social Media software creators and utilized through subscription, leased, or other forms of "outsourcing" an application. For these reasons, it is critical to understand the technological opportunities and business limitations with respect to retaining control of the information that an organization creates and shares within Social Media application frameworks.

Social Media application development tools, data storage technologies, and networking communications infrastructure can vary drastically from those tools, technologies, and infrastructures that are normally supported and maintained by an organization's internal IT personnel. Most organizations create, procure and support internal IT systems that can be defined and developed based on a concise set of "functional requirements" derived from the business goals and processes originating from the organization's Line of Business. Financial institutions, law firms, engineering companies, and government agencies for instance all have management control and direct influence over the internal business applications and organizational networking infrastructures that they develop and support.

Performing system backups, creating disaster planning solutions, and capturing electronic records for discovery for auditing or electronic discovery during preparation for litigation are well within the sphere of influence of an organization's IT staff. Information Governance personnel can establish and implement enforceable information management policies, procedures, and records retention rules. In contrast, once an organization decides to outsource an application such as occurs with the use of a Social Media service, it typically becomes a "user" of a system created with applications development technologies generally outside of the usual skill set of its own IT support personnel. Social Media applications support business models with different priorities from the business models of less technology driven enterprises. For this reason, they often use different technologies to store, retrieve, and transmit data.

Data Storage

Storage of data serving Social Media applications must take into consideration factors not experienced in single server content management applications. Because of the many geographically dispersed users that are expected, data must be distributed across many servers located in a geographically dispersed manner to maximize data retrieval speed and promote data redundancy. As an example, 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. 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 recall⁹.

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. 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.

Similar to both Dynamo and BigTable, Apache's Cassandra database management system uses elements of both including a BigTable data model and a Domino-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" 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.

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 their 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

The Facebook Web site is accessible by all commonly used Internet browser software including Microsoft Internet Explorer, Google Chrome, Mozilla FireFox, and Apple's Safari. Anyone with one of these Internet browsers, Internet access, and a verifiable email address can use Facebook. The Facebook application is written in PHP, a server side scripting language, though it is then "compiled with HipHop for PHP, "a source code transformer" built by Facebook engineers that turns PHP into C++".¹⁰ Though this information is of little interest to most users of the Facebook application it illustrates the sophistication level required of any hacker that would attempt to subvert the system.

From an Information Governance perspective this also illustrates clearly that few if any natively stored Facebook files will be of use to anyone wanting a download of their data stored on Facebook. Without an application interface to Facebook that allows downloading information in standard useable data formats, such as text, comma delimited, spreadsheet, HTML, or a conventional database format, such as MS Access or Oracle, the data will be both unintelligible and unusable. It is for this reason that most individuals attempting to preserve snapshots of their Facebook data will need to use the Facebook provided online application found under the heading “Accessing Your Facebook Data”¹¹ This Facebook accessible function directs users to procedures for downloading information on posts, pictures, videos, and content or metadata covering messages and profiles.

Unless an individual or an organization has sophisticated IT support to enable utilization of data dumps more sophisticated than those provided by the Facebook data offload application, they will need to assume that it will not be as useful as the data provided from the native Facebook interface.

Twitter

Open Source software is a primary component of the Twitter technology infrastructure. The interface to Twitter was originally written in a Web application framework language known as Ruby on Rails. This environment provided a development framework for database, web service and Web page development. It encouraged the use of Web programming standards such as HTML, XML, CSS, and JavaScript. Although initially tweets were stored in MySQL, the enormous volume of tweets that began to be created over time caused redesign of the architecture away from the use of a single consolidated platform.

Between 2009 and 2013 the Ruby on Rails search stack was replaced by a Java server named Blender. Message handling is now done by software written in Scala. In addition, a new services architecture is now based on remote procedure calls to improve distributed performance across servers. Twitter takes some in-house developed applications and combines those with other open source technologies to maximize the ability to send messages that may reach over a hundred thousand tweets per second.¹² In addition, as with many Social Media application platforms, Twitter does not employ a traditional relational database model or software for data storage. Instead, proprietary database technology is used that allows dispersing data components across geographically distributed servers to enable maximizing information searching and messaging system communications performance.

For these reasons, the best way to obtain content and metadata from a Twitter account is to use the Twitter provided interfaces for downloading data about activities. A user can download this data from the Twitter application itself.

LinkedIn

The LinkedIn application is programmed in Java Virtual Machine, an abstract “computing machine” that enables running a Java program. It also uses JRuby, Scala, Groovy, and GRAILS as well creating a dynamic and powerful development environment. Data storage is split into a three tiered structure of on-line, near-line and off-line storage that disperses such content as profiles, messages, skills, activity streams, messages, people you may know and recommendations into appropriate storage mechanisms.¹³ It uses Hadoop and a Teradata warehouse to handle batch job processing, data analytics, and near line system management.¹⁴

As with many Social Media applications, it can be seen with LinkedIn that there is a preference for in-house development of software where product teams and development teams coordinate to plan future system improvements. The result is high performance systems that maximize the technology infrastructure to meet data search and retrieval performance objectives. However, this can create instances where the only rigorously authoritative custom data offloads must be performed by the Social Media applications developer’s internal IT personnel.

Once again we see that raw data files from a Social Media application would be virtually useless for most users, auditors, or investigators. Only data offloads that use a Social Media platform’s data download application interface to provide application-specific data are really useful. It is possible to document actions within Social Media applications by screen captures or printing from an application but these will be generally useless with respect to assessing historical use of a system.

YouTube

YouTube, like many of the Social Media application environment has some development uniqueness due to the development environment of its originators and their preference in tools that served them well. YouTube was developed in Python, and it is still the most common code being executed. Other core software components of the YouTube stack are Apache for requests, Linux for operational monitoring and MySQL for data storage. MyVitess often serves as a proxy in data requests, while Zookeeper, Wiseguy and Spitfire are used for other components of YouTube applications.¹⁵

Playback and file format technologies are particularly important for the YouTube application environment. Historically, the Adobe Flash Player plugin for browsers was required for viewing YouTube videos. However, in January 2010 YouTube transitioned to use of the HTML5 standard with supported browsers including Internet Explorer 11, Google Chrome, and Safari 8. The streaming solution in current use is Adobe Dynamic Streaming for Flash. Most commonly users are limited to uploading 15 minute videos, but verified users that have a history of compliance with the Community Guidelines may qualify for uploads up to 12 hours long. Files sizes can be permitted up to 20 GB. Video formats accepted for upload include most

file formats such as .AVI, .MKV, .MOV, .MP4, .DivX, .FLV, .ogg, .ogv, and others. Video formats accepted include MPEG, MPEG-4, VOB, and WMV. Videos today can playback at a rate of 720p or higher.¹⁶

This diversity of upload issues, including file sizes, formats, and data standards indicates that any production of user or system data from YouTube for use outside of the application can be extremely complex requiring the user of such data to be capable of handling a variety of content management, display, and data storage challenges. This could be a good example of where one might “regret getting what you asked for!”

SOCIAL MEDIA RECORDS CAPTURE

Modern organizations with thoroughly planned Information Governance (IG) programs typically have well considered policies and procedures for retaining and preserving electronic records in all data formats. Such data formats include electronic mail, database data in applications, and office files on shared drives, as well as, items stored in more sophisticated content management systems such as Microsoft SharePoint. IG programs that encompass a single enterprise can be implemented with efficiency and expediency due to the direct management control possible over the organization's own computer applications and networks. However, Social Media applications such as Facebook, Twitter, and others are designed by their creators for public accessibility by anyone that logs in, creates a user account, and then starts using an application. Such systems were designed primarily for high volume user performance and specialized functionality when accessed from nothing more than an Internet browser.

Ease of use, sophisticated functionality for system users, and interconnectivity to other Social Media technology platforms is the primary goal of companies offering Social Media software for public use. As with the development and generation of many IT systems, electronic records preservation and archiving is typically an afterthought. And, for the Social Media software vendor's immediate goals, an ability to back up data and perform occasional offloads for internal audits or externally generated investigations is sometimes considered sufficient for most business activities. However, over time, the users of these Social Media application began to have their own needs to produce or archive the information they had been putting into the Social Media system. That "recorded" information was perceived to be "their" records and became valued for their own business purposes or to respond to audits and investigations.

Most initial attempts to "retrieve" records from Social Media applications consisted of screenshots and printing of pages of content. As such cloud based applications became more complex it became more apparent that traditional approaches to records production would not suffice. It was not possible to produce truly representative accurate records due to the usual loss of dynamic links between pages of documents when printing to paper or to Adobe Acrobat Portable Document Format (PDF) files. Attempting to save pages as Hypertext Markup Language (HTML) or other similar format files often resulted in volumes of pages that were only somewhat accurate, and the display of those files later would vary depending on the Internet browser used to display them. Such "renditions" of Social Media records were often viewed with skepticism by investigators, regulators, lawyers, and judges as they were not regularly captured during a "normal course of business" but were created later for specialized purposes.

For these reasons, vendors of Social Media systems today offer a variety of solutions that allow system users to download some of the content and metadata they have been putting into the

systems they are using. There are more successful methods of acquiring electronic records for business, legal, and compliance purposes.

Facebook

For some initial displays of records that exist in the Facebook application, it may be permissible in some courts or with some auditors to simply print screenshots of records. However, as we have seen over the years with maintaining electronic records for federal recordkeeping purposes or for use during litigation and e-discovery, the lack of metadata that accompanies printed or imaged records does not provide sufficient information to verify the authenticity and accuracy of the records. Early attempts to request electronic data from Facebook to present evidence of what was occurring in an application often resulted raw data dumps that needed to be reconstructed into intelligible records and were thus subject to interpretation, consumption of time, and costs.

For that reason, Facebook created a user accessible method for accessing data that users have entered into the Facebook application or created within the Facebook environment by using it. This information can be accessed through the “Accessing Your Facebook Data”¹⁷ application that directs users regarding downloading information about shared posts, pictures and videos, as well as, information on messages, conversations, and profiles. The idea is to let Facebook users initially serve their own needs for Facebook content and data without requiring an intermediary IT professional at Facebook to intervene. It empowers Facebook users to take charge of their own records to some degree as often as they are interested, depending on the frequency and timeliness of their needs for records. This method of accessing one’s own data yields this menu screen:

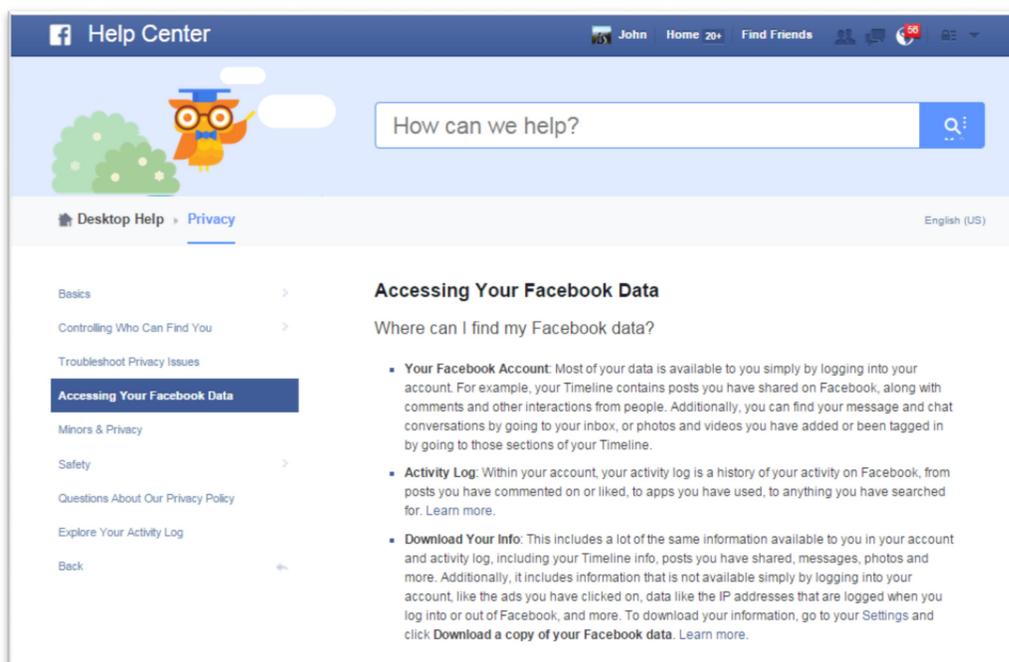


Figure. Facebook Help Center

By taking advantage of the Facebook Download option a variety of data including logs of activities such as posts, administrative settings, messages sent/received, login histories, and personal information posted can be obtained. However, this information does not duplicate the Facebook application interface most users are familiar with. It will require some interpretive analysis to make use of it as a record of Facebook content or interactions. In addition, it does not immediately make information available online. Once the request is submitted, the data must be collected by Facebook and sent to the user through an email interaction; unlike many IT generated corporate computer system reports that are available online with an organization’s internal IT applications. The user’s Facebook data comes as a compressed “Zipped” file with HTML and JPG data format components in different directories. Each file must be viewed separately and some interpretation applied to get meaningful use of the content.

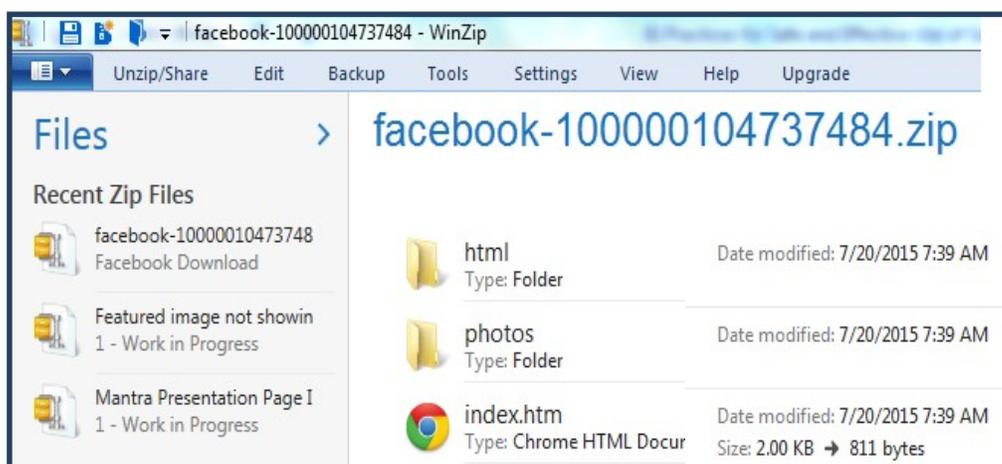


Figure. WinZip software view of downloaded Facebook files

Facebook, on special occasions, will provide information about users to third parties. However, Facebook prefers that the user authenticate the appropriateness of any third parties to access information including participating in the process by signing an authorization for those parties to access any Facebook data. Law enforcement agencies and parties to civil litigation can make requests for information where an individual has agreed to share their information with those entities.

Twitter

Each Social Media application has a different set of content and metadata to support the business model in which it specializes. For that reason, the informational records they are willing to provide will vary. Twitter does not provide a direct interface for users to access their own electronic records regarding account data or communications. Twitter prefers that a user take advantage of the Help Desk function to request information about their communications and activities.

Support issues can be addressed by accessing one’s “account profile and settings” menu to select

“help.” This selection takes one to the Twitter Help Center screen at <https://support.twitter.com> and ways to access one’s own Twitter data:

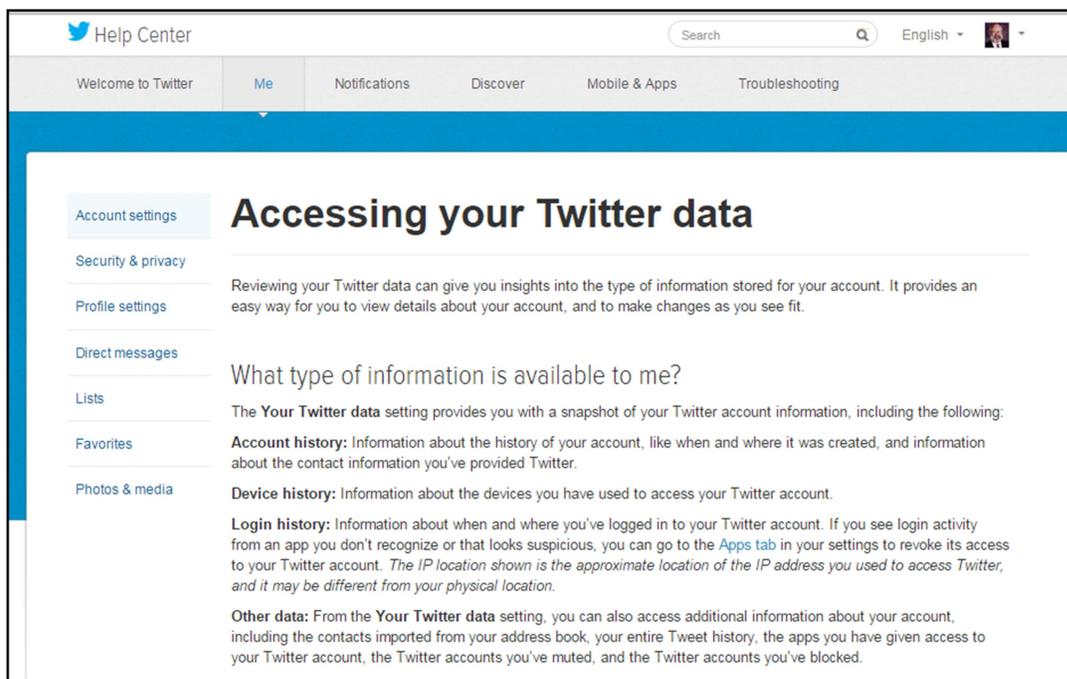


Figure. Accessing Your Twitter Data

One is directed to take the following actions:

Where can I find my Twitter data?

You can access your account’s data by logging in to your account on the web:

- 1. Log in at twitter.com.**
- 2. Click your profile icon in the top right navigation bar.**
- 3. Select Settings from the drop-down menu.**
- 4. From the menu on the left, click Your Twitter data.**
- 5. Enter your password and click Confirm.**

Figure. Twitter Data Retrieval Actions

This action leads one to a screen that provides general account history settings such as Account Creation Date, User Name, Email Address, Phone Number, and recent Log-In History with respect to IP addresses. Under another section of the page entitled Other Data, one can access and review one's Contacts, Connected Applications, Muted Accounts, and Blocked Accounts, as well as, download one's entire Tweet History by clicking on Twitter Archive.

Subsequently an introduction is provided to a method for accessing Twitter data that leads to more specific directions regarding procedures at <https://twitter.com/settings/account> :

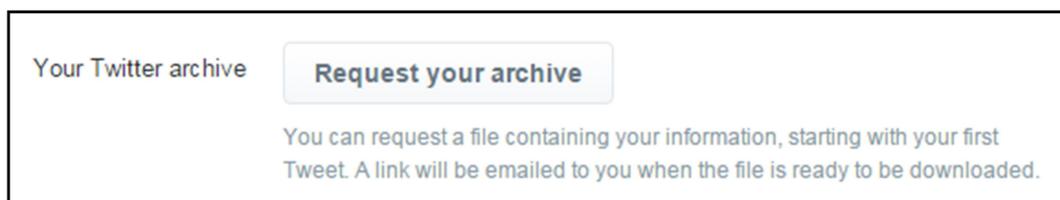


Figure. Your Twitter Archive

Twitter will provide text files that can be used in a Word processor to review and analyze the historical content of tweets.

LinkedIn

For access to the metadata and content that a user inputs to and stores in their Social Media account, LinkedIn offers information on Accessing Your Account Data in the Help Center at https://help.linkedin.com/app/answers/detail/a_id/50191/kw/download+account+data.

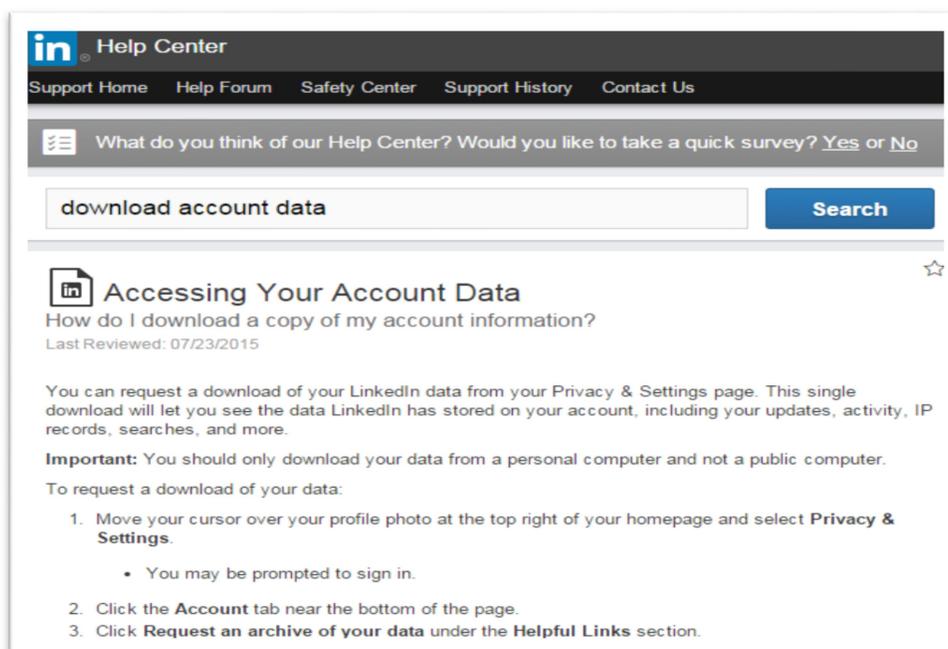


Figure. LinkedIn – Accessing Your Account Data

Information provided from this download includes account registration information, histories of log-ins with associated IP records, email addresses, and account status changes. Operational information that is available includes connections to others, photos uploaded, endorsements, skills identified, recommendations from others, contributions to groups, content posted, likes posted, and mobile applications used, as well as, ads on LinkedIn that have been viewed. The information will be made available within 24 hours by sending the user a link in an email. However, this information is about the account user, not the people that have viewed their profile or they may know. Additional information about LinkedIn use and interactions may be available by a direct query through the LinkedIn Data Consent form.

As with many Social Media systems, when one is interested in LinkedIn data about account usage that could compromise the privacy of someone with whom one interacted, additional data may be available by using forms evident in Privacy settings. In addition, there are restrictions on the data that can be share external applications that request information from a LinkedIn user's account. Some types of metadata that can be made available in a more restricted manner but the overall data offered to other providers of applications that integrate with LinkedIn include the LinkedIn account user's Profile Overview (with name, headlines and current position), Detailed Profile (with experience, education, skills, and recommendations), email address, connections, and personal contact information.

Of particular note, other applications outside of LinkedIn may request through their end user agreements that they be able interact with respect to Network Updates, Group Discussions, and the sending of invitations and messages as being identified by the LinkedIn account holder. Many of these refinements in the metadata and content shared and managed by LinkedIn actually are specified in the End User Agreements of both LinkedIn and the integrating software applications. For this reason, a complete understanding of the content and metadata associated with each users' account requires considering the End User agreements for all involved applications.

YouTube

YouTube's publicly accessible application interface appears to have no intrinsic means of requesting user contributed content or metadata. Most YouTube Account Settings and the Help Desk interface do not have menu items that indicate a manner through which a YouTube user can initiate personal downloads of uploaded videos; user contributed video descriptions or other electronic records that may be needed during audits, investigations, or litigation. This could be due to the large data volumes that could be requested or simply that YouTube would prefer a more direct interactive request from a user personally, though this is not clear. YouTube does have an extensively develop YouTube Help Center interface that provides information on Account Settings management and methods for addressing Legal, Safety, and Copyright issues.¹⁸ However, no direct method of downloading a user's content or metadata is evident.

A direct search of Internet access to a means of obtaining user content and metadata records from YouTube revealed the following informative references.¹⁹

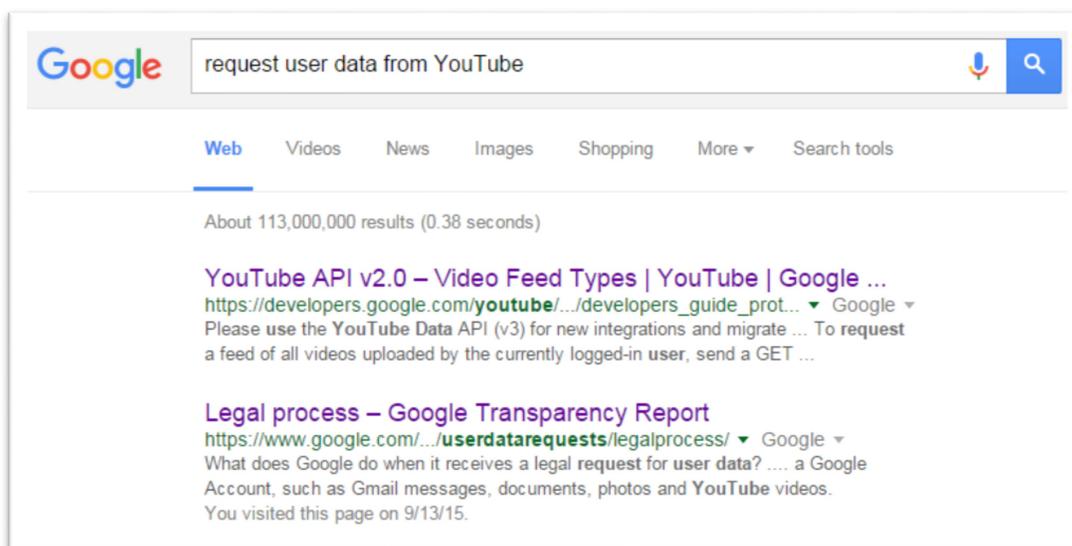


Figure. “Request user records from YouTube”

The first reference indicates that user data can be accessed through the Google Developers interface.²⁰ However this approach would be beyond the technical skills of most users and would be best performed by an organization with extensive IT skills.

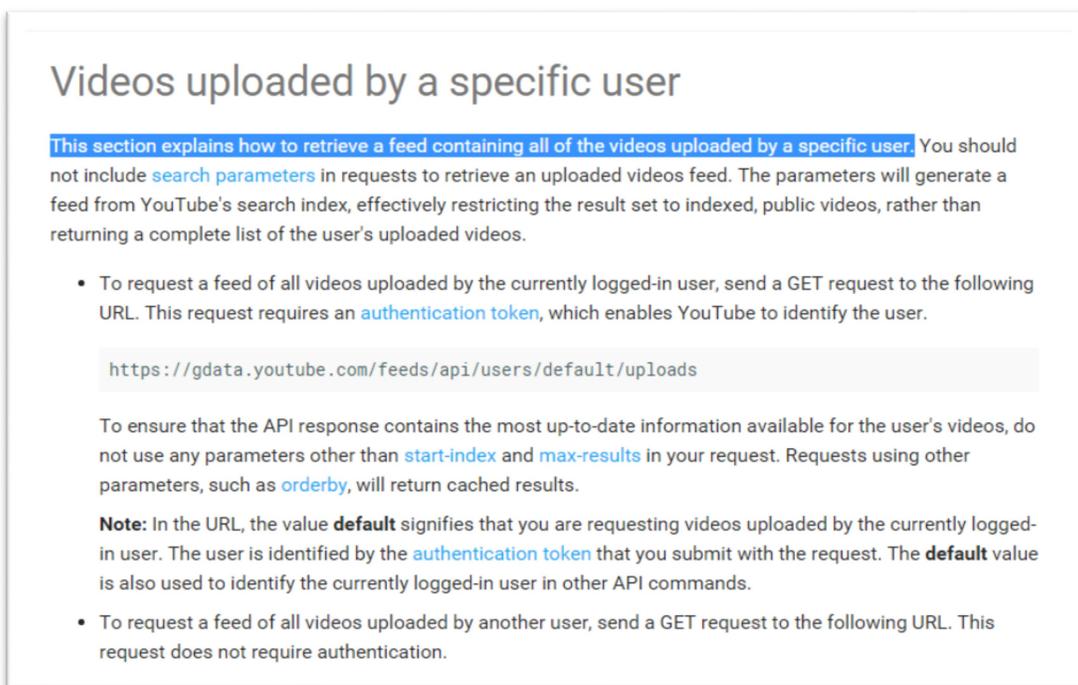


Figure. YouTube Google Developers interface.

However, YouTube is a subsidiary of Google, and as such it could be assumed that the overall Google corporate policies would apply to its subsidiary. And, interestingly, submitting the search question “request user records from YouTube” to the Google search engine also indicates in the second reference that there is policy available directly from Google on this subject.²¹

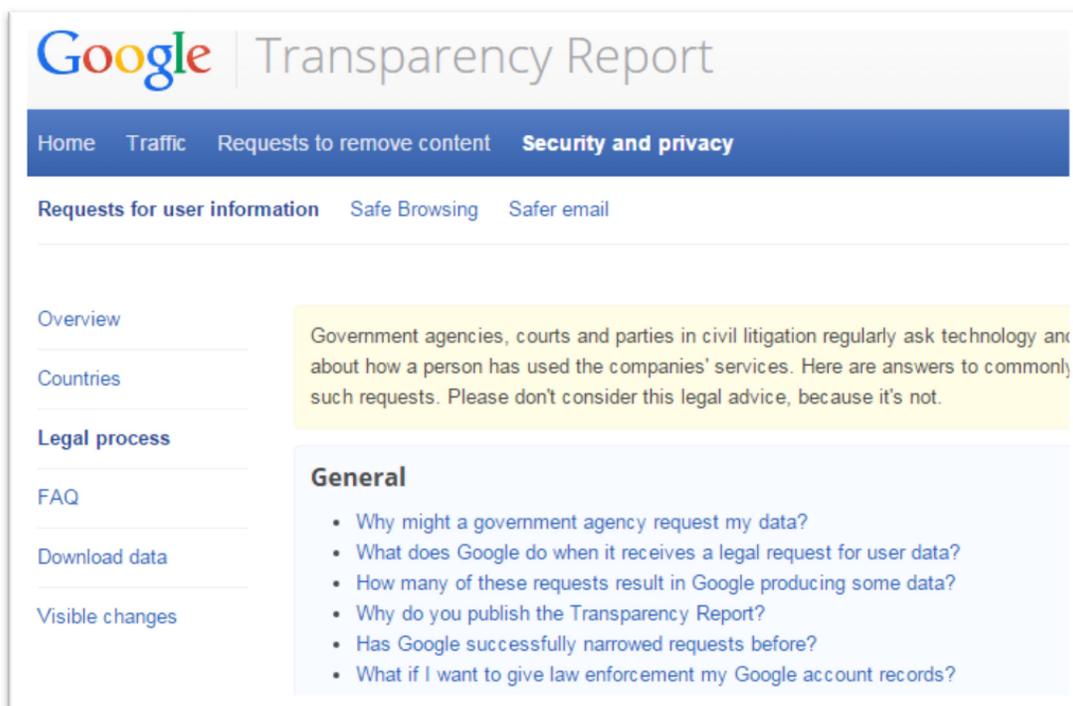


Figure. Google Transparency Report

The Google Transparency Report gives information to Google users about the criteria, policies, and processes regarding the release of data about users, their accounts, and their activities. It is intended as a public relations outreach mechanism to answer general questions about Google’s responses to government agencies, law enforcement agencies, and requests for information during litigation. However it does contain considerable information about the kinds of records released and the requirements for doing so, which vary with the source and nature of a request. It specifically outlines for instance, that under an Electronic Communications Privacy Act (ECPA)²² request typical YouTube data that might be released includes:

1. Subpoena:
 - a. Subscriber registration information
 - b. Sign-in IP addresses and associated time stamps
2. Court Order:
 - a. Video upload IP address and associated time stamp

- b. Information obtainable with a subpoena
- 3. Search Warrant
 - a. Copy of a private video and associated video information
 - b. Private message content
 - c. Information obtainable with a subpoena or court order

What is discoverable during this Transparency Report is that there is a method for Google application users to download their own data. A question is posed on the Web page as follows:

What if I want to give law enforcement my Google account records?

You can. We provide tools that allow you to download your content from many of our services using [Google Takeout](#). Google, however, requires valid legal process before we will produce data in response to a request from a government agency (even if the request is being made on your behalf), absent an emergency situation.²³

Google Takeout is a method for a user to download their own data from google applications:

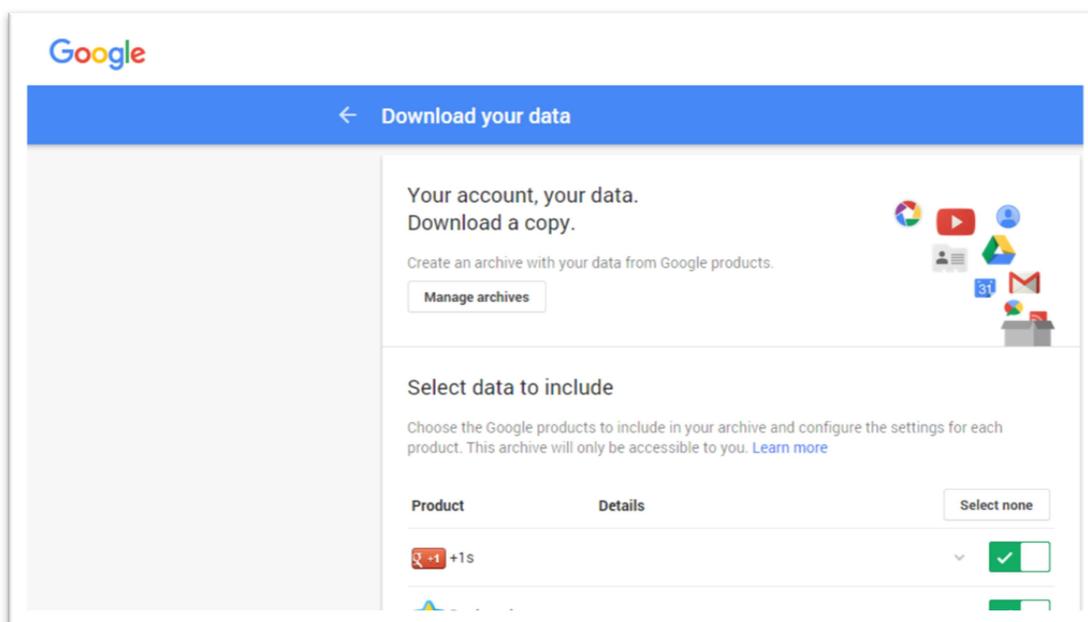


Figure. Google Takeout application interface.

There are 20 google products for which data can be requested, including such applications as Bookmarks, Calendars, Maps, Mail, and YouTube. For the YouTube application there are a number of parameters that can be modified to include or exclude downloading of histories, playlists, subscriptions, or videos in their original formats. That interactive screen appears as follows:

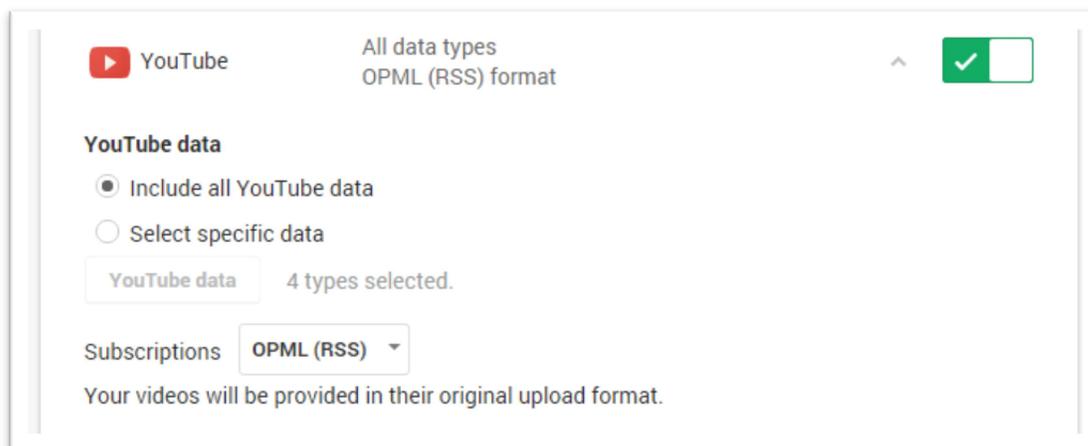


Figure. YouTube Download Selection Screen.

Summary

It is very evident from this discussion of four major vendors of Social Media services that each organization views its responsibilities to provide access to user supplied data and associated metadata very differently. In some cases, the information available to a user can be accessed through what is offered as a service related to litigation or law enforcement, and the process is very dependent of the identity and source of the requestor. This is understandable in that these are undoubtedly the first types of queries for this information that they received. However, in recent years, as users have become more empowered with respect to managing their own applications and data, expectations have grown in assuring that their data is properly archived and retained. Probably for those reasons, as well as to reduce their own involvement and costs in providing this data to users, some Social Media companies now offer more robust interfaces to users.

Unfortunately, the routes to finding these resources within Social Media applications vary with each application and can be convoluted. Social Media applications for the most part do not offer ways to download files or data that mimic closely the “look and feel” of the interfaces used to input the data. Rather than directly access the historical screens used to input and display the data, one often must request a data download in a format that must be used to interpolate what the interaction with the application was originally. Informational downloads can be in a variety of digital formats, including text files, spreadsheets, HTML, Zip files or other data formats. For this reason, it is important to “experiment” with a download process to be ready ahead of time for any critical need for meaningful and useable data. Advanced planning will be very important if, for instance, the downloaded Social Media records are to be referenced as a component of a Data Map or an Information Governance strategy. And it must be remembered that this data changes daily.

AGREEMENTS AND CONTRACTS

Social Media system vendors frequently use End User License Agreements (EULAs) and Service Level Agreements (SLAs) to govern the use of their software applications. Rarely do we see today instances where software is actually sold to end users. The software user model seen most commonly today in personal computers, network servers, cell phones, and other software driven digital devices is a licensing model wherein one in a sense “rents” the use of the software, rather than actually owning it. This allows the software creator to retain intellectual property control of the products they create and to assure those products are operated, maintained, and utilized as they intended. Both EULAs and SLAs are legally binding.

The major operational difference between EULAs and SLAs is that EULAs protect the software vendor and are usually agreed to by users during the software installation process. They are for the most part, non-negotiable. In contrast, SLAs protect both vendors and users. They are typically negotiated between software vendors and buyers and/or users to assure a mutually expected degree of operational capability, level of maintenance delivery, and timeliness of support. Both EULAs and SLAs can greatly impact an ability of a software user to access, manage, produce, and archive their data that exists within a software application. This is especially important when the software is a “cloud” based system where the actual physical residence of the data and application are external to the using organization. Both EULAs and SLAs can have a major impact during audits, litigation, and regulatory compliance initiatives, with respect to the legitimate or expected use of software and access to data. Consequently both EULAs and SLAs can impact Information Governance plans and a general knowledge of their attributes should be evident to Information Governance plan creators.^{24 25}

The legalities and applicability of EULAs and SLAs can certainly be debated, but that is not really the issue. The challenge is fundamentally that once an organization “outsources” IT services to organizations and architectures outside of the direct command and control purview of its management, the relationship that the organization has to those IT services and its “own” data have changed drastically. Management cannot direct and reprioritize the actions of the IT staff in charge of their Social Media data, because there is no direct reporting relationship. They can contact an account manager, a Help Desk or possibly a procurement organization to resolve issues. Or they may have their attorney talk to the Social Media organization’s attorney. This type of “organizational buffer” can take time that may not be readily available. For this reason, it is critical that organizations incorporate specific Information Governance related actions in advance of a critical need for data or reporting services.

EULAs

An End User License Agreement is a contract that protects the software owner’s copyright while

granting certain rights of use to the software user. Typically, there are limitations on the ability of the software buyer to replicate or copy the software with respect to the others using the software, while allowing that buyer to copy, archive, and backup the software as needed for normal business operations. For most publicly accessible versions of software, there will be a standard software vendor generated EULA that everyone accepts or declines to accept. In some cases, when a larger organization is buying a large number of licenses, the EULA may have inserted specific clauses that modify the EULA to meet the needs of that organization.

Common aspects of EULAs include:

1. Limitations of liability regarding performance – the software vendor gains agreement from the user to limit them from liability should the software malfunction and cause computer systems problems,
2. Limitations in liability regarding injury from use – the software vendor cannot be held responsible for misuse of the software,
3. Limitations regarding the manner in which the software may be used once it is installed, and
4. Limitations on the work products that can be derived from using the software with respect to illegal content creation in certain legal jurisdictions, etc.

Many of these limitations on software use rely on the resolution of disputes in legal systems, potentially internationally. And there are often clauses in EULAs that may be attempts to extend copyright control that is not necessarily legally applicable in some legal settings. These limitations are typical of what is often termed “proprietary” commercially generated software in contrast to a “free software license” where the user is granted much wider use of the software with respect to software use, modification, and distribution of generated works.

Unfortunately, in their speed to quickly use the software after buying and probably downloading it, the software user may accept a “click-through” agreement wherein they click an “Accept” button on the computer screen to a EULA presented during the installation process. These EULAs may be hundreds of pages long and have been often criticized as abusive of the user with respect to their actually understanding the legal obligations and operational limitations to which they are agreeing.

A typical example of the kinds of statements and clauses that can be found in a EULA is this example for Adobe Acrobat X Pro:

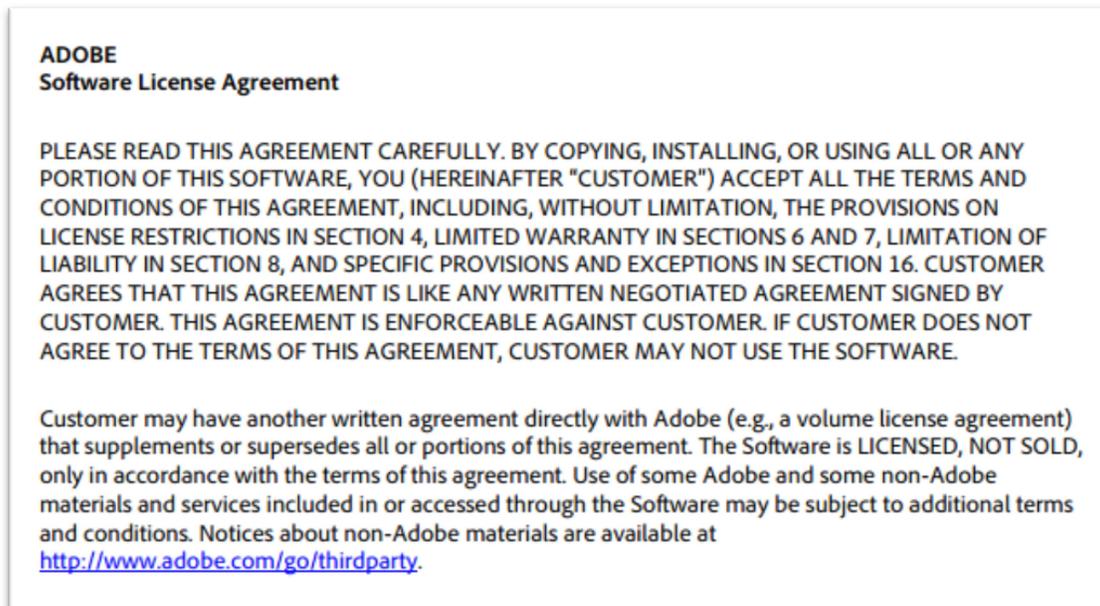


Figure. Adobe Software License Agreement

Additional examples of these agreements can be found on the Adobe Web site at <http://www.adobe.com/legal/licenses-terms.html>.

SLAs

Service Level Agreements are contracts for computer services that assure specified services will meet performance and delivery expectations. In contrast to EULAs, SLAs are created with mutual input from both the services provider and the buyer. Scope of agreement, relative responsibilities, performance metrics, and contracted delivery time are typical general aspects of SLAs.

SLAs greatly vary in scope, nature, and approach depending on the vested interests of the primary SLA generator. Customer oriented SLAs can itemize the specifications that a customer wants from a single or multiple vendors of a particular service. Service oriented SLAs can focus on a particular set of services offered by a services vendor to multiple customers. The actual contracts can be between organizations or departments within an organization and the services provider. To assure that there is mutual understanding and communication between the SLA vendor and the customer, there are often specified contacts, team members, and meeting milestones included in an SLA.

Major areas to be addressed in most SLAs are:

1. Scope of Services – what is the entire spectrum of services to be provided by the vendor?

2. Performance Monitoring – what communications architecture and means are available, as well as, the specific factors to be monitored?
3. Performance Issues Resolution – What are the policies and processes for creating and implementing solutions to issues?
4. Responsibilities - What are the relative responsibilities of the vendor and the buyer of the services?
5. Security - How is security addressed, including facility and system physical security, as well as, software/hardware information security and privacy controls?
6. Business Continuity and System Recovery – What is the vendors’ plan(s) for continuity during a business disaster, and how will the system and the buyer’s data be protected?

Each major area has issues that can affect the ability of the buyer of the services to implement their own Information Governance programs. More specific factors that can be addressed in an SLA that would directly impact an ability to access one’s data would include:

1. Dates of contract duration with an itemization of data retention rules,
2. Itemization of specific hardware, software, and networks for data access,
3. Help Desk organization response time,
4. Issue reporting mechanisms and issue resolution time,
5. Acceptable “mean times between failures”,
6. Acceptable “mean time to recovery”,
7. Application data throughput metrics reports,
8. Data retrieval metrics reports,
9. Data security and protection impacts on user access to data, and
10. Data production parameters on file download availability, sizes, formats, and time frames.

SLAs can be developed based on a number of SLA development industry examples.^{26 27 28} One must assure that there are operational parameters imbedded in the SLA policies, processes, and overall architecture that assure Information Governance plans and policies can be implemented.

Summary

The fast adoption of Social Media use by both public and private sector organizations is creating a growing interest in understanding the nature of the contractual agreements that are entered into once the use of a Social Media application begins. The working relationship with the vendor of a publically accessible Social Media service can be best managed through a mutual understating of the terms and conditions already in use in both EULAs and SLAs. Everyone that uses a Social Media service or a “Cloud based” business application typically “accepts” a EULA when they install the software. And, before any well managed organization buys, leases, or begins to use a Cloud-based application of any kind, they try to negotiate an SLA that meets their organization

needs for cloud computing support.

Unfortunately, in many cases, the size of an organization’s pool of potential users or the size of the budget of the buyer of the services may impact the receptivity of a vendor of cloud services to negotiation of the contents of the SLA, especially with Software as a Service (SaaS) providers. Small and Medium size Businesses (SMBs) may have additional challenges:

“In general, the larger the customer deployment, which translates to higher subscription and upfront fees, the more power customers exert in negotiating stringent SLAs. Even in the case of SaaS providers, large customers are successful in negotiating a stronger agreement, where none may be offered to SMB customers. This does point to the trend that as cloud deployments proliferate to larger enterprise customers, the demand for stronger SLAs will intensify. As competition increases at all levels, better SLAs will inevitably become a competitive factor. At that time, large enterprises and SMBs, alike, will be able to choose based on more flexible and more favorable SLA terms and, in general, these terms will improve.”²⁹

EULAs and SLAs are intended to protect users and vendors of computer services and often focus on liability limitations. Unfortunately, most all EULAs and SLAs are process, services, copyright, and litigation oriented and do not contain clauses to protect the data. However, they do offer business mechanisms for implementing Information Governance plans with respect to Social Media applications. Both are contractually binding, though they may need to be enforced through legal system actions, thus creating costs and operational hazards for organizations trying to exert their rights under those agreements. And, both agreements offer opportunities related to prescribing the processes from which Information Governance goals can be leveraged. SLAs are particularly important in the realm of Social Media vendor relationship management, such as when organizations begin to depend on Facebook or Twitter for an assumed communications and content posting infrastructure.

However the most important features of EULAs and SLAs will be those clauses that add protection to the information to be governed and also offer a verifiable testable means of obtaining the information at a reasonable cost and in a manner that supports Information Governance Program initiatives. In political relationships an often heard phrase is “Trust, but verify.” In software user and software vendor relations this is also good advice. The “trust” would be mutually generated EULAs and SLA so that both sides could understand and make a mutual commitment to the relationship. The “verify” would simply be to have occasional testing of data downloads to assure the computer systems performed as expected regarding backups, data downloads, and ability to produce records during audits and litigation related information emergencies.

SOCIAL MEDIA ARCHIVING TECHNOLOGY TOOLS

Successfully retaining and preserving Social Media system electronic records requires more than issuing Information Governance and records retention policies. Information Governance strategies must be actually implemented to be more than “words on paper”, or in today’s world - “digital communications and proclamations.” In fact, since Social Media records are really only another data format of records created “in the normal course of business,” they are subject to many of the same Information Governance issues as other electronic records, including electronic mail, electronic files on network file shares, and data in databases. However, there is an additional challenge intrinsic to this environment in that Social Media records are typically stored on “cloud” based servers in external networking architectures making capture, preservation, and retention a much greater challenge.

Social Media records must be:

1. Located and identified,
2. Captured and retained,
3. Classified to an extent that the application of retention rules is possible, and
4. Stored to preserve the records’ completeness, accuracy, and authenticity.

Solutions Design

It is possible to identify records within an existing Social Media application and retain them within the original application. However, as has been discussed previously, Social Media systems vendors do not guarantee the retention of records within their own architectures. A more typical solution is to capture the records and move them to an actual archival recordkeeping application. It is in these dedicated Social Media records archives where one is more likely to have available a file plan with an organization’s custom developed metadata and a retention schedule that can be used to apply the organization’s retention rules to the captured and classified records. Having an application that is built for information retrieval and retention that incorporates a specific organization’s information classification and retention policies is usually the best practice.

This has been the historical experience in the more commonly seen challenges with respect to electronic content management and electronic records management in the last decade or so. Electronic content management systems originally were able to manage content such as electronic mail messages and word processing or other office documents, as well as images, graphics files, and other data formats. They could index the files, apply in-house generated metadata, and apply primitive retention rules. However, they did not have sufficient robustness in classification and retention attribution functionality to enable full implementation of an organization’s data and records retention policies, as did fully developed electronic records management systems. It was

often necessary to use two such systems to achieve complete compliance with retention rules until many software industry mergers occurred that enabled the integration of full scale content management and electronic records management functionality into one integrated software system.

Because Social Media applications were not developed with sophisticated retention rules application in mind, today's most successful practice is going to usually be to extract at least occasional downloads of Social Media records and capture those electronic records into more specialized software capable of applying Information Governance policies including data retention and preservation rules. The most cost effective means of capturing and preserving those records could be by using an existing in-house content management or electronic records management software to capture and preserve Social Media records. Social Media records could be captured by screen "snapshots" as HTML, XML, or PDF format files. This is theoretically possible in many cases, but much of the interactivity and functionality of the original Social Media records will be lost. And the "normal course of business" for performing these functions would need to be well documented and demonstrably consistent.

Due to the organizational resources needed to credibly create and maintain in-house records capture and preservation routines, as well as to have additional specialized functionality, there are now software systems that are specifically developed to capture and manage Social Media records. Unfortunately, there are still many challenges to capturing and preserving Social Media records, even for dedicated software tools:

1. Social media files can be complex with links to files or graphical components stored on other applications and servers,
2. Viewing the contents of social media files may require displays of both video and audio components thus demanding "plug-ins" for certain browsers,
3. Use of Applications Programming Interfaces will probably be required to dynamically real-time capture metadata from the Social Media application, and
4. Robust real-time Social Media capture solutions will often need to be custom developed due to the need for different solutions for different social media systems.

It will be expected that robust real-time Social Media 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 permission or the full extent of their data, thus allowing organizations to preserve protected employee accounts." And complex rules must be developed for selecting the content to be captured. Should a response to a Facebook posting also be captured? Do references to items mentioned in a Tweet also need to be captured? ³⁰ Many detailed business driven technology decisions need to be made in advance of beginning to successfully capture Social Media records in real-time system integrations.

Solutions Alternatives

Capture, preservation and retention of Social Media records can be performed in a variety of manners, including hands-on records capture. This is not an uncommon approach until a sophisticated and capable software solution can be procured. In many cases, a low-technology solution is a good starting point for beginning records capture until a complete set of system requirements can be defined. As an example, in May 2013, the U.S. National Archives and Records Administration (NARA) created a “white paper” on Best Practices for the Capture of Social Media Records.³¹ In addition to other insights this study revealed that many U.S. Federal government organizations were struggling to implement policies with currently available low-technology solutions. There were numerous examples cited of “work-in progress” as agencies attempted to implement policy. Among the strategies employed and in use were such partial solutions as:

1. Backup tools,
2. Ask vendor directly for content,
3. Copy and paste into a word document,
4. Export from social media platform in CSV format,
5. Print and file,
6. PDF/A, and
7. Application Programming Interfaces (APIs) to create a customized tool to download into a database.³²

Although these (initially) low cost approaches may provide evidence of an attempt to follow Information Governance policy, actually using the results of these low technology efforts will be time consuming and costly in organizational resources at some point. For such efforts to be of use in responses to audits and litigation, for instance, these partial solutions would need to be extensively documented to prove reliability and replicability in their execution. For this reason, many organizations are considering commercial software solutions to capturing and preserving records.

There are numerous commercially available software systems today that can capture Social Media records while offering a basic level of access functionality. As an example, the previously mentioned NARA white paper lists approximately 60 software candidates that can perform some extent of Social Media records capture.³³ However, those listed range in sophistication from API toolkits to full featured Electronic Content Management Systems. Increasingly, there is an expectation in the technology market place that Social Media capture software should offer solutions with direct connectors to the Social Media technology platforms. Some examples of software platforms that are focused on having the utilities needed for actively capturing Social Media records include (but are not limited to):

1. ArchiveSocial - <http://archivesocial.com> ,
2. Smarsch, <http://www.smarsh.com> ,
3. X1 Social Discovery, http://www.x1.com/products/x1_social_discovery ,
4. NextPoint, <http://www.nextpoint.com>,
5. WebPreserver, <https://webpreserver.com> ,
6. RegEd, <http://www.reged.com> , and
7. DSicoverly, <http://dsicoverly.com/dsicollect/social-media-2>.

Products like these are focused on archiving large scale volumes of Social Media records for powerful access and potential production of records in large volumes. In contrast, there are other numerous products in the marketplace that focus on small personal individual solutions to archiving records such as Jolicloud, Memolane, TimeHop, TwInBox, Facebook Timeline, and others. In addition, there are also a number of electronic discovery software tools to support litigation case development that may be adequate for occasional collections of Social Media data but cannot actually archive large volumes of data while applying formal retention rules based on Information Governance program mandates. Additional studies of Social Media relevant software systems include reports by corporate-focused technology consulting firms such as Forrester³⁴ and Gartner³⁵. The players in this market change constantly due to new entrepreneurs entering the market, mergers, and acquisitions. Any organization expecting to make system selection decisions will need to revise those activities every three to six months based on market changes.

Social Media Technology Business Case

As with many technology driven environments there exists a large variety of options from which to choose. When deciding on technology solutions to support implementing Information Governance programs, these factors will be critical:

1. What are the primary drivers for selecting and implementing software – an e-discovery crisis, general volume of litigation or compliance, long term business strategy to implement Information Governance?
2. What is the level of skills and sophistication of in-house IT personnel?
3. Will the solution be used by individuals from many different locations?
4. What impact will different levels of cost have on existing IT and general budgets?
5. Is the organization ready for another cross enterprise multiyear IT project?

First, assessing the primary drivers for this engagement will allow the organization to establish immediacy of need which will determine an approximate timeframe for completion. Project drivers can also determine the organizational scope of the project and give a sense of prioritization regarding the needed functionality from the system. This will in turn cause an assessment of available financial and personnel resources.

Second, as previous discussion in this report has indicated, implementing an IT solution that must utilize data stored on external applications in the “cloud” requires considerable technical expertise and contract negotiation skills. A project manager and associated project team will need to be created with internal IT architecture expertise, social media technology systems expertise, general IT systems integration expertise, and a formally chartered requirements development group. Third, if the solution is to be used across several organizational departments or functions, representatives from those units will need to be a part of the requirements development, thus expanding the scope and resources required for the project. Is the solution targeted to be just used by Legal Counsel and their staff to respond to litigation or is the solution seen as an enterprise wide solution to manage all Social Media content for everyone?

Fourth, all IT projects must be negotiated for prioritization considering other organizational budgets and priorities. This may dictate a “phased implementation” and may affect the initial software solutions purchased and their implementation timing. In addition, it may be tempting for a solution to be selected that takes advantage of an existing budgeted in-house archiving system primarily developed for email archiving. Such solutions often require converting Social Media records to email data formats for storage, thus losing much of the complex metadata and content associated with Social Media records. Search and retrieval capabilities may be more limited and associated linked files may not be retrievable. This technology “compromise” may not meet the needs of some compliance regulatory mandates.³⁶

Fifth, cross enterprise IT projects are one of the most difficult undertakings of any organization and will require widely communicated executive level sponsorship to be successful. A critical component of the business case for a technology solution will be support from the organizational leaders of Legal Counsel, Information Governance, Business Organizations, Finance, and Procurement.

It can be readily seen that technology solutions are a critical aspect of bringing Information Governance concepts to Social Media records. It is also apparent that technology solutions can only target, incorporate, and manage those records that have been identified as being subject to Information Governance program criteria.

INFORMATION GOVERNANCE FOR SOCIAL MEDIA

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 is Information Governance. Information Governance is a rapidly developing professional perspective today though currently with few examples of extensive enterprise-wide implementation. For that reason, it is an evolving arena with a need for sharing new perspectives and lessons to be learned, especially with regard to incorporating Social Media records into Information Governance plans.

The newness of Information Governance principles contributes to a fluidity of concepts regarding core best practices. These will undoubtedly change somewhat over time due to the wide variety of professional organizations seeing the value of this strategic information management perspective. A primary leader in Information Governance concepts today is ARMA International where Information Governance concepts are embedded in the Association's Generally Accepted Recordkeeping Principles³⁷. Due to of the comprehensive scope of Information Governance concepts it is also being embraced by many other professional organizations including the purveyors of the highly respected Electronic Discovery Reference Model (EDRM)³⁸, the US National Archives and Records Administration³⁹, and the American Health Information Management Association⁴⁰.

However the implementation of Information Governance principles and practices depends on an organization's current state of information management policies, IT systems architecture, or business environment. The best means to approaching Information Governance within an enterprise will vary based on an organization's recordkeeping mandates, litigation profile, risk management priorities, and Social Media systems architectures. However, due to the fast changing pace of Social Media technologies and systems, it is critical to encompass Social Media records from the beginning of use of any new Social Media application:

“Identify records during the social media planning stage. Both a social media policy and the records and information policy should refer to a form to be completed by the person or unit proposing a new social media initiative. The person completing the form should indicate if records will be created and, if so, how they will be managed”.⁴¹

The information used and stored in Social Media applications by an organization's employees must be encompassed by the same overall Information Governance policies that govern the use of electronic records housed within that organization's in-house IT systems. There will be an expectation during regulatory audits, litigation related electronic discovery actions, and disaster planning efforts, that the information stored in Social Media systems will be accessible, producible,

and retained according to standard information management policies. Otherwise, the use of Social Media applications can pose operational and compliance risks for the organization. For that reason, when enterprise-wide Information Governance programs are developed, Social Media records must be specifically addressed in the standard programmatic mechanisms used in the development of any Information Governance Program:

1. Policy Development,
2. Information Gathering and Analysis,
3. Identifying Required Policy Elements,
4. Capturing and Preserving Records,
5. Tagging and Metadata, and
6. Creating Data Maps.⁴²

As can be seen from the examples in previous sections of this report, Social Media applications and data present a number of challenges with respect to incorporating these electronic records within the management scope of an Information Governance Program. These information management challenges span a number of programmatic, organizational, and technical arenas but are most commonly encompassed by these high level issues:

1. Most Social Media applications are “outsourced” Internet “cloud” based applications not under the direct management control of the user or the organizations using the application,
2. The Social Media users organization’s data is stored in software systems that cannot be easily altered due to the existence of vendor created EULAs and SLAs that impact access,
3. The data in the Social Media Applications is frequently not available in a readily useable data download format for most end user organizations,
4. There are major conceptual and procedural challenges to implementing an organization’s Information Governance policies and procedures in an effective manner in outsourced applications not under the direct management control of the organization.

Many Social Media applications provide user interfaces that can be employed to obtain some raw data about someone’s use of an application. However, these interfaces only provide an indication of the overall use of the system, and it is not possible to “roll-back” the application interface to re-experience how a user input data and visually interacted with the software. The application user interaction logs and associated data downloaded from Social Media download applications are useful, but activity and history logs do not replicate the “look and feel” of a software application interface. Those data downloads and application interfaces can provide an initial capture method for organizations to begin to incorporate Social Media records into their own archiving solutions.

Of equal concern, any vendor of Social Media services can alter access to the data within its application, sometimes with little warning. In early 2015 Facebook altered rules for allowing

outsiders access to data causing some applications developers dependent on Facebook data to close their business or drastically change business operations. This is said to evolve from Facebook's changing attitude toward allowing third parties to use what is increasingly seen as extremely valuable Facebook data. Some of the data that used to be available by default but now requires specific Facebook approval includes:

1. Likes,
2. Photos,
3. Listings of Facebook friends,
4. Birthdays,
5. Education history,
6. Relationship status,
7. Religious and political affiliations,
8. Work history, and
9. Hometown.⁴³

In addition to the challenges associated with changing Social Media vendor policies, there are two other very important aspects of using this type of data as formal records where they fail to meet the usual definition of a good record:

1. There is often a warning such as that found with Facebook, where it states "You will not find information or content that you have deleted because this is deleted from Facebook servers. For that reason, these data downloads do not constitute an "archive" of all the data input into Facebook over time, but rather, a download of currently resident data.
2. Also lost is the "context" of a complete record in many cases. If making some metadata and content related to an interaction with the Social Media interface impinges on the privacy or security of another system user, the Social Media application vendor will generally not provide that information with a simple download, thus losing a major portion of the completeness of the electronic records. They will likely insist on a court order or regulatory agency mandate to provide that level of records detail.

Thus, for Information Governance policies to work in an organization, advanced planning will be required to assure 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, 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. Implementing Information Governance principles in this manner will often require extending the practices into the Social Media software application *where cooperation can be achieved* through modification of EULAs and SLAs.

Fortunately, there are some already designed best practice models for managing data in the cloud that recommend such data be subject to itemization in SLAs. For instance, in the 2011 Cloud Standards Customer Council report **Practical Guide to Cloud Service Level Agreements v1.0**, recommendations are made regarding Data Policies:

1. **Data Preservation**
Cloud users should ensure the service supports their data preservation strategy that includes sources, scheduling, backup, restore, integrity checks, etc.
2. **Data Redundancy**
Users should ensure they have an appropriate data preservation strategy that addresses redundancy with the system.
3. **Data Location**
Users should carefully consider as part of their data management strategy how the SLA will complement where their data will reside, where it will be processed, and how this meets regulatory requirements.
4. **Verification of New Data Location**
Clients should ensure that when a provider elects to provide its service from another location it will be required to notify its clients of the new location and provide a means for the client to independently verify where the data will be relocated.
5. **Data Seizure**
Consumers should also ensure there are arrangements in place to make a user's data available in the event that their SLA provider goes out of business and the data provider locks access to its systems pending payment of the outstanding account.
6. **Data Privacy**
The declared data privacy policy from the provider should be included in the SLA.⁴⁴

As Social Media vendor cooperation may be limited, other practices that can be initiated solely by the data owner may be required, such as performing snapshots of the Social Media applications.

At a basic level one would expect these practices to be addressed to assure Social Media records are properly accessible, protected and preserved:

1. Assure Information Governance policies distinguish between employees personal and work-related use of social media,
2. Assure Information Governance issues are encompassed by any contracts with vendors of Cloud based services.
3. Design or modify EULAs and SLAs to be specifically in compliance with Information Governance Program plans,
4. Create data maps for e-Discovery, retention compliance, and other Information Governance initiatives to incorporate Social Media applications and records.

For Information Governance programs to be credible and reliably implemented, comprehensive advanced planning is the key.

APPENDICES

Categories of Facebook Data Available:

This user supplied content and associated metadata could be downloaded by users from an application interface on Facebook as of September 1, 2015. It is provided as an indicator of the nature of the information stored in a Social Media application and database system, providing insights into the need for introspection and interpretation when an actual Social Media software user-oriented interface is not used to view Social Media data. This data is useful but activity and history logs do not replicate the “look and feel” of a software application interface. In addition, notice that it says “You will not find information or content that you have deleted because this is deleted from Facebook servers. “ For that reason, it does not constitute an “archive” of all the data input into Facebook over time, but rather, a download of currently resident data.

Internet Reference September 8, 2015 -From - <https://www.facebook.com/help/405183566203254>

What categories of my Facebook data are available to me?

These are the categories of Facebook data that are available to you either in your activity log or your downloaded data, or in both places. We have provided a short explanation of what each data category is and where you can find it. We store different categories of data for different time periods, so you may not find all of your data since you joined Facebook. You will not find information or content that you have deleted because this is deleted from Facebook servers.

Remember, most of your Facebook data is available to you simply by logging into your account (ex: all of your messages and chats are available in your inbox.) Also note that the categories of data that we receive, collect, and save may change over time. When this happens, this list will be updated.

What info is available?	What is it?	Where can I find it?
About Me	Information you added to the About section of your Timeline like relationships, work, education, where you live and more. It includes any updates or changes you made in the past and what is currently in the About section of your Timeline.	Activity Log Downloaded Info

Account Status History	The dates when your account was reactivated, deactivated, disabled or deleted.	Downloaded Info
Active Sessions	All stored active sessions, including date, time, device, IP address, machine cookie and browser information.	Downloaded Info
Ads Clicked	Dates, times and titles of ads clicked (limited retention period).	Downloaded Info
Address	Your current address or any past addresses you had on your account.	Downloaded Info
Ad Topics	A list of topics that you may be targeted against based on your stated likes, interests and other data you put in your Timeline.	Downloaded Info
Alternate Name	Any alternate names you have on your account (ex: a maiden name or a nickname).	Downloaded Info
Apps	All of the apps you have added.	Downloaded Info
Birthday Visibility	How your birthday appears on your Timeline.	Downloaded Info
Chat	A history of the conversations you've had on Facebook Chat (a complete history is available directly from your messages inbox).	Downloaded Info

Check-ins	The places you've checked into.	Activity Log Downloaded Info
Connections	The people who have liked your Page or Place, RSVPed to your event, installed your app or checked in to your advertised place within 24 hours of viewing or clicking on an ad or Sponsored Story.	Activity Log
Credit Cards	If you make purchases on Facebook (ex: in apps) and have given Facebook your credit card number.	Account Settings
Currency	Your preferred currency on Facebook. If you use Facebook Payments, this will be used to display prices and charge your credit cards.	Downloaded Info
Current City	The city you added to the About section of your Timeline.	Downloaded Info
Date of Birth	The date you added to Birthday in the About section of your Timeline.	Downloaded Info
Deleted Friends	People you've removed as friends.	Downloaded Info
Education	Any information you added to Education field in the About section of your Timeline.	Downloaded Info
Emails	Email addresses added to your account (even those you may have removed).	Downloaded Info

Events	Events you've joined or been invited to.	Activity Log Downloaded Info
Facial Recognition Data	A unique number based on a comparison of the photos you're tagged in. We use this data to help others tag you in photos.	Downloaded Info
Family	Friends you've indicated are family members.	Downloaded Info
Favorite Quotes	Information you've added to the Favorite Quotes section of the About section of your Timeline.	Downloaded Info
Followers	A list of people who follow you.	Downloaded Info
Following	A list of people you follow.	Activity Log
Friend Requests	Pending sent and received friend requests.	Downloaded Info
Friends	A list of your friends.	Downloaded Info
Gender	The gender you added to the About section of your Timeline.	Downloaded Info
Groups	A list of groups you belong to on Facebook.	Downloaded Info
Hidden from	Any friends, apps or pages you've hidden from your News	Downloaded Info



News Feed	Feed.	
Hometown	The place you added to hometown in the About section of your Timeline.	Downloaded Info
IP Addresses	A list of IP addresses where you've logged into your Facebook account (won't include all historical IP addresses as they are deleted according to a retention schedule).	Downloaded Info
Last Location	The last location associated with an update.	Activity Log
Likes on Others' Posts	Posts, photos or other content you've liked.	Activity Log
Likes on Your Posts from others	Likes on your own posts, photos or other content.	Activity Log
Likes on Other Sites	Likes you've made on sites off of Facebook.	Activity Log
Linked Accounts	A list of the accounts you've linked to your Facebook account	Account Settings
Locale	The language you've selected to use Facebook in.	Downloaded Info
Logins	IP address, date and time associated with logins to your	Downloaded Info

	Facebook account.	
Logouts	IP address, date and time associated with logouts from your Facebook account.	Downloaded Info
Messages	Messages you've sent and received on Facebook. Note, if you've deleted a message it won't be included in your download as it has been deleted from your account.	Downloaded Info
Name	The name on your Facebook account.	Downloaded Info
Name Changes	Any changes you've made to the original name you used when you signed up for Facebook.	Downloaded Info
Networks	Networks (affiliations with schools or workplaces) that you belong to on Facebook.	Downloaded Info
Notes	Any notes you've written and published to your account.	Activity Log
Notification Settings	A list of all your notification preferences and whether you have email and text enabled or disabled for each.	Downloaded Info
Pages You Admin	A list of pages you admin.	Downloaded Info
Pending Friend Requests	Pending sent and received friend requests.	Downloaded Info

Phone Numbers	Mobile phone numbers you've added to your account, including verified mobile numbers you've added for security purposes.	Downloaded Info
Photos	Photos you've uploaded to your account.	Downloaded Info
Photos Metadata	Any metadata that is transmitted with your uploaded photos.	Downloaded Info
Physical Tokens	Badges you've added to your account.	Downloaded Info
Pokes	A list of who's poked you and who you've poked. Poke content from our mobile poke app is not included because it's only available for a brief period of time. After the recipient has viewed the content it's permanently deleted from our systems.	Downloaded Info
Political Views	Any information you added to Political Views in the About section of Timeline.	Downloaded Info
Posts by You	Anything you posted to your own Timeline, like photos, videos and status updates.	Activity Log
Posts by Others	Anything posted to your Timeline by someone else, like wall posts or links shared on your Timeline by friends.	Activity Log Downloaded Info
Posts to Others	Anything you posted to someone else's Timeline, like photos, videos and status updates.	Activity Log

Privacy Settings	Your privacy settings.	Privacy Settings Downloaded Info
Recent Activities	Actions you've taken and interactions you've recently had.	Activity Log Downloaded Info
Registration Date	The date you joined Facebook.	Activity Log Downloaded Info
Religious Views	The current information you added to Religious Views in the About section of your Timeline.	Downloaded Info
Removed Friends	People you've removed as friends.	Activity Log Downloaded Info
Screen Names	The screen names you've added to your account, and the service they're associated with. You can also see if they're hidden or visible on your account.	Downloaded Info
Searches	Searches you've made on Facebook.	Activity Log
Shares	Content (ex: a news article) you've shared with others on Facebook using the Share button or link.	Activity Log
Spoken Languages	The languages you added to Spoken Languages in the About section of your Timeline.	Downloaded Info

Status Updates	Any status updates you've posted.	Activity Log Downloaded Info
Work	Any current information you've added to Work in the About section of your Timeline.	Downloaded Info
Vanity URL	Your Facebook URL (ex: username or vanity for your account).	Visible in your Timeline URL
Videos	Videos you've posted to your Timeline.	Activity Log Downloaded Info

NOTE: As of early 2015, Facebook changed many of the rules regarding allowing outside access to the data stored online. Thus this table is subject to change at the discretion of Facebook at any time.⁴⁵

CONTACT

For more information about the content of this report contact:

John T. Phillips
Information and Technology Decisions
1803 Nantasket Rd
Knoxville, TN 37922
865-966-9413; john@infotechdecisions.com
www.informationtechnologydecisions.net

END NOTES

-
- ¹ Statista Internet Web site, July 15, 2015, <http://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide>
- ² Statista Internet Web site, July 15, 2015, <http://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users>
- ³ Pew Research Center Internet Web Site, July 15, 2015, <http://www.journalism.org/2015/07/14/the-evolving-role-of-news-on-twitter-and-facebook>
- ⁴ Social Media Governance Web Site, Internet Reference September 10, 2015, <http://socialmediagovernance.com/policies/#axzz1mwYxW1pc>
- ⁵ Captcha Web Site, Internet Reference August 12, 2015, <http://www.captcha.net>
- ⁶ Wikipedia Internet Reference, August 10, 2015, <https://en.wikipedia.org/wiki/Twitter>.
- ⁷ John K. Waters, The Everything Guide to Social Media, F+W Media, 2010, p. 121.
- ⁸ YouTube Web Site, Internet Reference August 23, 2015, <https://www.youtube.com/yt/press/statistics.html>
- ⁹ Amazon Web Services Documentation Internet Web Site, <http://docs.aws.amazon.com/amazondynamodb/latest/developerguide/Introduction.html>.
- ¹⁰ Wikipedia Internet Reference, August 10, 2015, https://en.wikipedia.org/wiki/Facebook#Technical_aspects .
- ¹¹ Facebook Web Site, August 10, 2015, www.facebook.com/help/405183566203254.
- ¹² Wikipedia Internet Reference, July 18, 2015, <https://en.wikipedia.org/wiki/Twitter>.
- ¹³ <http://www.slideshare.net/amywtang/webscale-analyticsfb-24031929> , Internet Reference August 25, 2015.
- ¹⁴ <https://gigaom.com/2013/03/03/how-and-why-linkedin-is-becoming-an-engineering-powerhouse> , Internet Reference August 25, 2015.
- ¹⁵ High Scalability Web Site, Internet Reference September 1, 2015, <http://highscalability.com/blog/2012/3/26/7-years-of-youtube-scalability-lessons-in-30-minutes.html>.
- ¹⁶ Wikipedia, Internet Reference, September 1, 2015, <https://en.wikipedia.org/wiki/YouTube> .
- ¹⁷ Facebook Internet Web Site, July 18, 2015, <https://www.facebook.com/help/405183566203254>.
- ¹⁸ The YouTube Help Center, September 13, 2015, <https://support.google.com/youtube/?hl=en#topic=6151248>.
- ¹⁹ Google Search Internet Reference, September 13, 2015 for “request user data from YouTube”, https://www.google.com/search?q=requesting+youtube+metadata&rlz=1C1SFXN_enUS545US598&oq=requesting+youtube+metadata&aqs=chrome..69i57.28554j0j8&sourceid=chrome&es_sm=122&ie=UTF-8#q=request+user+data+from+YouTube
- ²⁰ YouTube Web Site, Accessing user data through the application interface, Internet Reference, September 12, 2015, https://developers.google.com/youtube/2.0/developers_guide_protocol_video_feeds?hl=en.,

- ²¹ Google Transparency Report, Internet Reference September 13, 2015, <https://www.google.com/transparencyreport/userdatarequests/legalprocess>.
- ²² Electronic Communications Privacy Act, Internet Reference September 13, 2015, https://en.wikipedia.org/wiki/Electronic_Communications_Privacy_Act.
- ²³ Google Transparency report Internet Reference, September 12, 2015, <https://www.google.com/transparencyreport/userdatarequests/legalprocess>.
- ²⁴ Wikipedia, End User License Agreement Overview, Internet Reference September 15, 2015, https://en.wikipedia.org/wiki/End-user_license_agreement.
- ²⁵ Wikipedia - Services Level Agreement Overview, Internet Reference September 15, 2015, https://en.wikipedia.org/wiki/Service-level_agreement.
- ²⁶ Service Level Agreements: A Legal and Practical Guide, Jimmy Desai, IT Governance Publishing, 2010, Internet Reference, <http://www.itgovernanceusa.com/shop/p-1115-service-level-agreements-a-legal-and-practical-guide.aspx>.
- ²⁷ SLA Information Zone, Internet Reference September 15, 2015, <http://www.sla-zone.co.uk/disaster.htm>.
- ²⁸ Practical Guide to Cloud Service Level Agreements v 1.0, Cloud Standards Customer Council, 2011, Internet Reference September 16, 2015, http://www.cloud-council.org/2012_Practical_Guide_to_Cloud_SLAs.pdf.
- ²⁹ Ibid, p. 6.
- ³⁰ Rakesh Madhava, 10 Things to Know About Preserving Social Media, **Information Management**, ARMA International, September/October 2011, p.36.
- ³¹ Best Practices for the Capture of Social Media Records, Internet Reference August 20, 2015, <http://www.archives.gov/records-mgmt/resources/socialmediacapture.pdf>.
- ³² Ibid, p. 9.
- ³³ Ibid. p. 11-20.
- ³⁴ Nick Hayes and Chery McKinnon, Market Overview: Social Media Archiving, Forrester, March 9, 2015, <https://www.forrester.com/Market+Overview+Social+Media+Archiving/fulltext/-/E-RES117533>.
- ³⁵ Jie Zhang and Garth Landers, Magic Quadrant for EDiscovery Software, Gartner, May 18, 2015.
- ³⁶ Nick Hayes, Ibid. p. 4.
- ³⁷ ARMA International Internet Web Site, <http://www.arma.org/r2/generally-accepted-br-recordkeeping-principles>.
- ³⁸ Electronic Discovery Reference Model Internet Web Site, <http://www.edrm.net>.
- ³⁹ US National Archives and Records Administration Internet Web Site, <http://www.archives.gov/records-mgmt/prmd/automated-erm.html>.
- ⁴⁰ AHIMA Internet Web Site, <http://www.ahima.org/topics/infogovernance/igbasics?tabid=overview>.
- ⁴¹ Dr. Pat Franks and Robert F Smallwood, Chapter 13, Information Governance for Social Media, *Information Governance: Concepts, Strategies, and Best Practices*, Wiley CIO), Kindle Edition.
- ⁴² Using Social Media in Organizations, ARMA TR 21-2012, ARMA International, August 2012, Overland Park,

Kansas, P. 13-17.

⁴³ Facebook Restricts Access to Its Data Trove, **Wall Street Journal**, September 22, 2015, p. B1.

⁴⁴ Practical Guide to Cloud Service Level Agreements v 1.0, Cloud Standards Customer Council, 2011, p. 15, Internet Reference September 16, 2015, http://www.cloud-council.org/2012_Practical_Guide_to_Cloud_SLAs.pdf.

⁴⁵ Facebook Restricts Access to Its Data Trove., Ibid.