



NATIONAL
ARCHIVES

National Archives and Records Administration
White Paper on
Best Practices for the Capture of Social Media Records



May 2013

I. Background 3

II. Types of Social Media in use by Federal Agencies 3

III. Examples of Current Federal Agency Use..... 4

IV. Examples of social media capture tools and methods..... 9

V. Federal Records Council Study Findings..... 21

VI. Best Practices for Records Management..... 21

 Building a Foundation 22

 Tools and Methods 22

 Implementation 22

VII. NARA’s transfer guidance and best practices for formats..... 23

References..... 24

Additional Resources..... 24

I. Background

Federal agencies are increasingly using social media platforms to engage with their employees and the public. Such activity may result in the creation of Federal records that must be captured and managed in compliance with Federal records management laws, regulations, and policies.

GAO Audit 11-605, "Social Media: Federal Agencies Need Policies and Procedures for Managing and Protecting Information They Access and Disseminate," concluded that social media "can pose challenges in managing and identifying records, protecting personal information, and ensuring the security of Federal information and systems."¹ The audit specifically addressed security and records management issues concerning agency social media activity. In response, the Archivist of the United States committed NARA to developing "guidance on effectively capturing records from social media sites and ... best practices."² Additionally, through forums Federal agencies have frequently expressed the desire for additional guidance. This white paper is part of NARA's effort to provide guidance and best practices for capturing social media records.

It is not feasible for NARA to provide platform-specific guidance because it is difficult to predict which tools will be available and preferred in the future. NARA's Report on Federal Web 2.0 Use and Record Value concluded that agencies anticipate a great deal of change in the social media landscape in part due to increasing technological convergence. Altogether these changes will provide more challenges to records management staff and their ability to capture and preserve content.³

This paper reports on the current state of social media use in the Federal government. These examples offer building blocks for the creation of best practices for the successful capture of social media records. This white paper also provides an overview of available capture tools as of May 2013.

II. Types of Social Media in use by Federal Agencies

According to NARA's 2011 Records Management Self-Assessment Report, seventy percent of agencies currently use social media. Federal agencies employ a variety of social media platforms in their public outreach and internal communications. To reach targeted audiences on a large scale, agencies reported they most commonly used Facebook, Twitter, and YouTube among a multitude of platforms.⁴ GAO 11-605 reported that some Federal agencies have garnered a large base of followers through effective use of social media.⁵

Social media serves several purposes including reposting information available on agency websites, posting information not available on publicly accessible agency websites, soliciting and responding to

¹ U. S. Government Accountability Office, *Federal Agencies Need Policies and Procedures for Managing and Protecting Information They Access and Disseminate*, GAO-11-605 (June 2011), "What GAO Found," <http://www.gao.gov/new.items/d11605.pdf>.

² Ibid, 40.

³ National Archives and Records Administration, *A Report on Federal Web 2.0 Use and Record Value* (2010) 18-19, <http://www.archives.gov/records-mgmt/resources/web2.0-use.pdf>.

⁴ National Archives and Records Administration, *Records Management Self-Assessment Report* (2011), 23, <http://www.archives.gov/records-mgmt/resources/self-assessment-2011.pdf>.

⁵ U.S. Government Accountability Office, *Federal Agencies Need Policies and Procedures for Managing and Protecting Information They Access and Disseminate*, GAO-11-605 (June 2011), 4-5, <http://www.gao.gov/new.items/d11605.pdf>.

comments, and providing links to non-governmental websites. External communication can mean interacting with the public or collaborating with other Federal agencies or other organizations. Internal communication can be for business or social purposes. The U.S. Navy describes the importance of using social media in its Social Media Handbook as helping “fulfill your obligation to communicate with all of your stakeholders. It also provides another, often richer, means of sharing information with internal and external audiences. Your stakeholders are increasingly using social media, and you’re better off reaching them there than not at all.”⁶ Agencies have described social media tools as a “multiplier of resources” where “more audiences can be reached with fewer resources needed on the part of the agency.”⁷ The use of social media by agencies from across the Federal government points to a general consensus regarding its value as a powerful communication tool.

Despite the mass communication opportunities, GAO 11-605 identified several areas of concern regarding Federal agencies’ use of social media, including appropriate identification and management of Federal records, ensuring privacy for users, and security for government information systems.⁸ Reinforcing these concerns, NARA’s 2011 Records Management Self-Assessment Report indicated that only fifty percent of responding agencies had policies and procedures for capturing and managing official records created on social media platforms although many said that policies were under development.⁹ However, several agencies stated they had media neutral records schedules they believed made social media-specific schedules unnecessary. Others said that official social media records were not being captured or were being printed in hard copy form without contextual metadata. The above assumptions may be contrary to records management best practices and should be discussed with the agency’s Appraisal Archivist. Agencies must consider the preservation of content and context of Federal record material regardless of format.¹⁰

III. Examples of Current Federal Agency Use

A number of current social media recordkeeping and capture policies from Federal agencies were identified. The following chart provides details from selected policies, including excerpts addressing capture. The methods and tools included in the following chart are not endorsed by NARA but give a sense of the present landscape.

⁶ U.S. Navy, *Navy Command Social Media Handbook* (2010), 3, http://www.cnrc.navy.mil/PAO/socialnetwrk/soc_med_hnd_bk.pdf.

⁷ National Archives and Records Administration, *A Report on Federal Web 2.0 Use and Record Value* (2010), 15, <http://www.archives.gov/records-mgmt/resources/web2.0-use.pdf>.

⁸ U.S. Government Accountability Office, *Federal Agencies Need Policies and Procedures for Managing and Protecting Information They Access and Disseminate*, GAO-11-605 (2011), 8-9, <http://www.gao.gov/new.items/d11605.pdf>.

⁹ National Archives and Records Administration, *Records Management Self-Assessment Report* (2011), 26-27, <http://www.archives.gov/records-mgmt/resources/self-assessment-2011.pdf>.

¹⁰ National Archives and Records Administration, “Expanding Acceptable Transfer Requirements: Transfer Instructions for Permanent Electronic Records” (2004), <http://www.archives.gov/records-mgmt/initiatives/web-content-records.html>; National Archives and Records Administration, “NARA Guidance on Managing Web Records” (2005), <http://www.archives.gov/records-mgmt/policy/managing-web-records.html>.

Agency	Social Media Policies	Social Media in Use	Excerpts from Social Media Policies
Centers for Disease Control	CDC Social Media Tools, Guidelines and Best Practices Twitter Guidelines and Best Practices Facebook Guidelines and Best Practices	Facebook, Twitter, YouTube, LinkedIn	<p>"9. Establish a Records Management System: Set-up a system to keep track of your Twitter posts, @replies, retweets and mentions to comply with Federal guidelines for records management and archiving." (Twitter Guidelines p.4)</p> <p>"Records Management. All Facebook page administrators must establish a system to collect all Facebook posts, comments, fan posts, events, and hidden posts (spam) to comply with Federal guidelines for records management and archiving.</p> <p>Any comment removed for violating the CDC commenting policy must be recorded and archived prior to deleting." (Facebook Guidelines p.2)</p>
U.S. Coast Guard	Social Media Chapter 11 from the Public Affairs Manual Social Media Field Guide	Facebook, Twitter, YouTube, Vimeo, Flickr, blogs, RSS feeds	<p>"Because a social media site is conducting the communications and transaction on behalf of the Coast Guard, it shall be properly managed as a Federal record. All content and comments posted on official social media sites shall be preserved IAW the Information and Life Cycle Management Manual, COMDTINST M5212.12 (series)." (Chapter 11 p.11-4)</p> <p>"2. General Comments. Comments that do not affect or require any administrative action, policy decision or special compilation shall remain posted but are considered obsolete in 10 days. Delete after one year.</p> <p>3. Deleted comments. Comments that do not follow Public Affairs Social Media Policy will be deleted but retained for one year. The comments can be destroyed after one year if no other action has been taken regarding the deletion of the comment. The moderator shall take a screen shot...label the file with the date, title of Facebook post and the initials of the person who removed the comment; and store the file on a shared folder for easy retrieval if a Freedom of Information Act request or other action be necessary...</p> <p>4. IAW this records schedule, CG-0922 has identified a free online tool, Backupify (www.backupify.com) to archive content and comments on official Facebook pages...Coast Guard page managers shall establish a separate, non-personal Backupify account and maintain archived data of official page content and comments." (Social Media Field Guide p.8-9)</p>
Department of Agriculture	New Media Roles, Responsibilities and Authorities	Facebook, USDA Blog, Twitter, YouTube, Flickr, Storify, RSS, widgets	<p>"(2) Records must be maintained for original or unique content created or published to new media accounts or platforms, such as public comments or information posted by individuals acting on behalf of the Department.</p> <p>(3) Each Agency or program office is responsible for maintaining records related</p>

Agency	Social Media Policies	Social Media in Use	Excerpts from Social Media Policies
			<p>to their new media accounts or activities." (p.3)</p> <p>"(3) Working with the OC to ensure that new media technologies comply with established USDA IT security, archival and records management requirements...</p> <p>(5) Identifying and properly archiving record material in accordance with approved records disposition schedules.(6) Ensuring that records on new media web sites are copied or otherwise captured and maintained with related records, unless the site has a records management application that can manage the record throughout its lifecycle." (p. 5-6)</p>
Department of Housing and Urban Development	Departmental Policies and Procedures for Use of Social Media Sites by HUD Offices and Staff	Facebook, Twitter, Flickr, HUD Wiki, Looking into e-book publishing and formats	<p>"A statement of record retention should be posted 'This is the HUD page on Facebook. Comments posted on and messages received through HUD pages are considered Federal records and shall be archived.' Please see NARA's General Records Schedules, in particular Section 20 on electronic records http://www.archives.gov/records-mgmt/grs/." (p. 5)</p>
Department of State	Using Social Media	Facebook, Twitter, YouTube, Flickr, Tumblr, blogs, Google+, Pinterest, RSS feeds	<p>"(9) Records in social media sites must be copied or otherwise captured and maintained with related records, unless the site has a record management application that can manage the records throughout its lifecycle. Nonrecord content consisting of duplicate information which is maintained in other department recordkeeping systems (original recordkeeping copy is maintained in accordance with its records disposition schedule), and transitory records do not need to be archived and may be deleted when no longer needed." (p.11)</p>
Department of Veterans Affairs	VA Directive 6515 Use of Web-based Collaboration Technologies	Facebook, Twitter, YouTube, Flickr, blogs	<p>VA records officer shall "(2) Work with the Archivist of the United States and VA content owners to determine the most appropriate method(s) to capture and retain VA records on both Federal servers and VA activities hosted on non-Federal Web-based collaboration hosts." (p. 14)</p>
Environmental Protection Agency	Social Media Policy Using Social Media Internally at the EPA EPA Facebook Guidance	Facebook, Twitter, YouTube, Flickr, Challenge.gov, blogs, and a Yammer pilot	<p>"11. Maintain Records</p> <p>The laws, regulations and policies that govern proper records management (i.e., creation, maintenance/use and disposition) still apply when using social media...New content created with social media tools that qualifies as a Federal record must be captured and maintained in a recordkeeping system according to EPA's Records Management Policy...Please note that content created with social media tools may be subject to document production under FOIA or litigation" (Using Social Media Internally p.4-5)</p>

Agency	Social Media Policies	Social Media in Use	Excerpts from Social Media Policies
			<p>"Never simply delete comments and other fan-generated content! If content does not meet the comment policy, retain the entire item and as much detail as possible (fan name, the date and time of posting, etc.) in an offline format. For example, comments can be stored in a simple Word document. Store multimedia content in its native format and note the details in the same Word document. Capture enough about the post and the response so that someone reading about it later gets the context. Also note the reason why the content did not meet the comment policy. Then delete it from your page." (EPA Facebook Guidance)</p>
General Services Administration	Social Media Navigator: GSA's Guide to Official Use of Social Media	Facebook, Twitter, YouTube, LinkedIn, Flickr, blogs, Google+, RSS feeds	<p>"When using electronic media, whether it is a blog, a website, a wiki, email, or any other type of electronic communication, the regulations that govern proper management, archival, and release of records still apply. The National Archives and Records Administration (NARA) offers resources and guidance to agencies to ensure proper records management. Contact records@gsa.gov for questions pertaining to records management at GSA." (p.7)</p>
National Archives and Records Administration	Rules of Behavior for Using Web 2.0 and Social Media Web Sites and Responsibilities for Content Management, NARA Guidance 831-2, February 24, 2010.	Facebook, Twitter, YouTube, Flickr, Tumblr, Foursquare, Pinterest, Google+, Ustream, Historypin, Apps, blogs, wikis, Internal Collaboration Network (ICN), RSS feeds	<p>"(2) Records created and maintained in social media may be covered in the agency's Records Control Schedule and/or the General Records Schedules and should be managed in accordance with approved dispositions." (p.8)</p>
U.S. Navy	Navy Command Social Media Handbook	Facebook, Twitter, YouTube, Flickr, RSS feeds	<p>"Records keeping policy guidance for social media is being developed by USG/DoD. In the interim, it is up to COs and their PAOs to make a determination on when and what kind of information to archive. For example, if a unit is involved in an operation of historical significance then it would be prudent for the command to archive as much of the content of their social media presences as possible for the historical record. Some effective means of archiving information include ensuring the content posted on social presences is also available via a command website, archiving e-mail related to command social presences, taking screen captures of social presences and copying and pasting posted content into a text file or word document." (p.9)</p>
Smithsonian Institution	Electronic Records: Recom-	Facebook, Twitter, YouTube,	<p>For text, video, audio, and other files: "In accordance with best practices, SI Ar-</p>

Agency	Social Media Policies	Social Media in Use	Excerpts from Social Media Policies
	Recommendation for Preservation Formats The Smithsonian: Using and Archiving Facebook	Flickr, Pinterest, Virtual World, blog	chives prefers to preserve transferred electronic records in the formats described in the table below." (Electronic Records p.5) "We created PDF/A capture of Facebook pages." (The Smithsonian: Using and Archiving Facebook)

IV. Examples of social media capture tools and methods

Application of policies and techniques to capture Federal social media records in the Executive Branch is in its infancy. Agencies employ various methods to capture content created on social media platforms. In May 2012, the Office of the Chief Records Officer and NARA's Social Media Team held a forum with Federal records management staff and web managers to discuss best practices for social media capture. Participants confirmed they are using, testing, or considering a number of tools and techniques (see the table at right). They also asked questions regarding precisely what should be captured. Through presentations and discussions, only a few examples of successful implementation of capture tools and techniques emerged that could be classified as "best practices." For the purposes of this paper, best practices are defined as methods, techniques, or actions that allow agencies to capture complete social media records with associated metadata. The complete capture will allow agencies to manage records throughout their lifecycle while maintaining functionality and searchability to meet Federal records management requirements.

Capture Tools and Methods Currently in Use at Federal Agencies

- Backup tools
- Ask vendor directly for content
- Copy and paste into a word document
- Export from social media platform in CSV format
- Print and file
- PDF/A
- Application Programming Interfaces (APIs) to create a customized tool to download into a database

Social media content capture is an emerging topic that has not consolidated around standards for capture. Many of the available tools are focused on providing "backup" products to users and are not aimed at capture for recordkeeping purposes. These vendors often provide a free basic service and charge for additional functionality or number of information streams. Some tools focus on providing capture for e-discovery and regulations compliance (e.g. FINRA, SOX). These tools can assist in capturing content so that agencies have a copy of what their social media platforms contained at a given time.

Web crawlers are another potential tool that can assist in the capture of web content. They often require higher levels of expertise and specific knowledge to deploy effectively. Examples of web crawlers are listed at http://en.wikipedia.org/wiki/Web_crawler (as of May 2013). As stated in NARA Guidance on managing social media records, it is the responsibility of the agency to determine what kinds of content and metadata should be captured as records, weighing if these are adequate for preservation purposes.

The NARA Guidance on managing social media records addressed strategies agencies could take to manage records, including:

- Using web crawling or other software to create local versions of sites;
- Using web capture tools to capture social media content and migrate to other formats;
- Using platform specific application programming interfaces (APIs) to pull content;
- Using RSS Feeds, aggregators, or manual methods to capture content;

- Using tools built into some social media platforms to export content.

The following table lists available tools and software available at the time of publication (May 2013) that could assist Federal agencies in implementing the above strategies for capturing social media content. They were neither tested nor are they endorsed by NARA. It remains the responsibility of agency records officers to evaluate software functionality and compliance with recordkeeping requirements. The list represents the range of services available together with pricing information, product descriptions from their developers and available product reviews. The “Platforms Able to Capture” column illustrates that some tools are designed for a specific type of social media and others for many different kinds. Where the method of capture is divulged, it is included in its own column.

Provider	Paid Service	Product Description & Use Cases	Platforms Able to Capture	Method of Capture	Notes
Actiance Socialite		Socialite provides products to assist in the backup and management of social media content.	Facebook, LinkedIn, Twitter, YouTube		
Aleph Archives		Web archiving service uses CAMA tool for regulatory compliance and e-discovery aimed at corporations to capture, store, and sort web content for e-discovery and regulatory compliance. Provides a number of service plans.	Facebook, Twitter, LinkedIn, YouTube, websites	Web crawlers regularly crawl and capture web sites as complete snapshots and displays the content in its original form (no URL re-writing, no JavaScript injection, etc.) They use the Web ARChive (WARC) format (ISO 28500:2009). Content can be stored with them or on own servers.	
Alfresco		Provides content management within the tool. It is unclear if it captures content with related metadata.	Facebook, Twitter, YouTube, websites,	Content management system captures social media content when it publishes to the platform.	
Archify		Archify captures and organizes social media streams and browser activity and makes it searchable and accessible across all devices.	Facebook, Twitter, LinkedIn		
ArchiveFacebook		Mozilla Firefox plug-in saves content from Facebook accounts directly to hard drives, including photos, info, messages, activity stream, friends list, notes, events and groups.	Facebook	Prototype called Facebook Archiver uses a modified version of ScrapBook to perform specific AJAX requests in order to capture each page of a Facebook account. Modifying the internal linkage of the captured pages will make the archived collection easier to browse.	
Archive-It		Subscription service from the Internet archive allows institutions to build, manage and search their own web archive.	Facebook, websites, Twitter	Harvest web content according to subscriber's frequency preference for each URL they are capturing, including "on demand" capture request, such as the case of a historic event.	
ArchiveSocial		Automatically captures and archives social media content for compliance, records management, and e-	Facebook, Twitter, LinkedIn, YouTube		No software installation required.

Provider	Paid Service	Product Description & Use Cases	Platforms Able to Capture	Method of Capture	Notes
		discovery needs.			
Arkovi RegEd		Captures social media content to power your compliance and enable your marketing to expand.	Facebook, Twitter, LinkedIn, Google+, YouTube, RSS feeds	Arkovi accesses social networks directly through their APIs to capture content regardless of how it's created or published. They use RSS feeds to capture blogs and other social platforms that offer limited or no API access.	
Backup Buddy		Backup service for WordPress that allows users to create a backup of the entire installation and send to their server, Amazon S3, Rackspace Cloud, FTP, or e-mail. The backup can help users restore and migrate WordPress installations.	Wordpress	Backs up single or multipress Wordpress with widgets, themes, and plug-ins with scheduled backups.	
Backupify for Personal Apps		Pay as you go backup service with no contracts that provides fully-searchable automated backups of social media content and storage for the data.	Facebook, Twitter, Gmail, Google Drive, Google Calendar, Google Sites, Google Contacts, Flickr, Picasa and Blogger	Backupify queries the APIs of each online account to identify recently added and updated files and copies content to an encrypted archive in Amazon's high-availability storage cloud. It then makes copies available for download or restoration.	They do not capture Facebook Places, FML markup code, or Page Insights data. Backup is weekly for the free accounts and daily for paid accounts.
Convogence		A subscription service for continual capture and retention of social media content. It can be used for compliance with a records retention policy for data outside of a company's firewall.	Facebook, Twitter, blogs, RSS and ATOM Feeds, and Google Apps	Web crawler captures social media and subscribers can request an export of content at any time, but it is unclear what the formats are. They also provide an API for customers to build adapters to integrate with any legacy system.	
Downloadr		Windows app that allows users to download photos from Flickr to their computer.	Flickr	Writes EXIF and IPTC data so titles, tags and location are preserved. Can search by full text, user, tags, place, set, date, relevance, group, and favorites.	

Provider	Paid Service	Product Description & Use Cases	Platforms Able to Capture	Method of Capture	Notes
Erado		Offers email, social media, and instant message archiving to comply with FINRA, SEC, Sarbanes-Oxley, Gramm-Leach-Bliley, FERC, NERC, and HIPAA.	LinkedIn, Facebook, Twitter, and blogs	Captures content either directly from social media platforms or using Erado's platform-specific tools. Content is converted into Erado format and either hosted or delivered to a customer-preferred platform.	
Facebook Download service		Facebook provides a backup of a user's profile, including content posted to timeline, photos and videos uploaded to account, friend lists, user-created Notes, RSVP'd events, sent and received messages, any comments made on timeline posts, photos, and other timeline content. Users can also request an enhanced archive that contains additional information.	Facebook	Facebook sends a ZIP file to the email address associated with a particular Facebook account.	Only content associated with user's account can be backed up and accessible.
Flickr API		Flickr provides an open Application Programming Interface (API) so Flickr can communicate with other software or tools. Flickr also provides RSS feeds of updates to content.	Flickr		
FlickrEdit		Previously FlickrBackup, this open-source, Java-based desktop app allows users to download, edit, or upload photos to and from Flickr.	Flickr		
Free YouTube Download		Software allows download of single YouTube videos or a batch of all of the videos of a selected YouTube user or channel.	YouTube	Saves content in original YouTube format or converts to AVI, MP4, and WMV formats.	
freezePAGE		Service preserves web snapshots and automatically logs the date page was saved, IP address of the person who saved it, page size and more. Requires login every 3 days for unregistered	Web sites	Captures a snapshot of webpages and saves on freezePAGE server when a user enters URL (manual capture). Includes main web page and embedded elements such as images, stylesheets, and script files.	Webpage must be less than 3MB (or 10 MB for premium user accounts) with less than 500 embedded elements and retriev-

Provider	Paid Service	Product Description & Use Cases	Platforms Able to Capture	Method of Capture	Notes
		users and 31 days for registered users or account and pages may be deleted.			able within 120 seconds.
Global Relay		Global Relay Archive captures electronic messages in real time and creates a copy of each message, which is then indexed, serialized and time/date stamped. Users can access and search archived content.	Facebook, Twitter, LinkedIn		
Hanzo Archives		Offers commercial web archiving services for regulatory compliance, litigation-support, and e-discovery. Subscribers can manage their web archives according to their RM policies with associated metadata.	Web sites, social media	Uses proprietary tools to capture content from complex websites, including rich-media and interactive content. Content can be searched, reviewed and exported. Subscribers can define their capture policy and have it captured and organized by time.	
Hearsay Social		Hearsay Social's compliance module provides workflow management, monitoring, and capture of social media from a central dashboard.	Facebook, Twitter, LinkedIn, Google+, Four-square	Using APIs, data is archived within context, catalogued, and searchable. Users can export data to existing enterprise systems, including SiteMinder, Websphere, Autonomy, and Symantec Enterprise Vault.	
Hootsuite		Provides a social media dashboard for managing multiple accounts. Users can spread messages across networks, monitor keyword mentions in streams, and track results with built-in click-through stats and integrated Google Analytics.	Twitter	Archived Messages is an optional add-on for the HootSuite Pro Plan. Twapperkeeper, now part of Hootsuite, archives tweets.	First 100 messages are archived for free and starting at \$10/month for additional levels. Available from GSA's Apps.gov
If this, then that		Users create ittt tasks by putting one channel's trigger together with another channel's action. Tasks are executed every 15 minutes and can be turned on or off and shared with others.	Twitter, Facebook, weather forecasts, email, etc.		Limited channels

Provider	Paid Service	Product Description & Use Cases	Platforms Able to Capture	Method of Capture	Notes
Iterasi		Subscription service to create web archives for the corporate, legal and government industries. It includes data available requiring authentication such as direct messages on Twitter and messages on Facebook.	Twitter, Facebook, LinkedIn	Web crawlers capture entire sites or individual pages on-demand or on a regular schedule. Can also capture contents of RSS feeds (e.g., blog feeds, Twitter). The “Page Notary Tool” captures any webpage, even those password or firewall protected.	
LiveOffice		Social Archives service offered by Symantec is part of the larger LiveOffice software package.	Twitter, Facebook, LinkedIn	Captures social media content in a centralized repository	Only available with LiveOffice AdvisorMail.
Memento		Memento, an LC-funded project run by Los Alamos National Laboratory and Old Dominion University, proposes a technical framework for integrating current and past Web.	Web sites	Firefox plug-in retrieves web captures from the Internet Archive from a specified date and time.	
Cloud Preservation by Nextpoint		Cloud Preservation is a subscription service that provides automated, cloud-based capture of web content for marketing, compliance, and litigation-related needs.	Websites, blogs Twitter, Facebook	Uses Amazon’s Web Services to crawl the Web and archive sites, blogs and social media posts. Web crawler set to capture HTML source code and images at pre-determined intervals.	
Ohmygov		Social media monitoring and metrics service that allows users to track social media accounts and compare their news mentions and rankings against their peers.	Twitter, Facebook	Provides account tracking which captures the full content of tweets.	User’s comparisons are limited by which agencies the service tracks.
Ownbackup		Service that provides daily automated backups of social media with unlimited storage.	Facebook, Twitter, LinkedIn, Salesforce, Gmail	Provides daily snapshots of cloud data, encrypts data via AES 256-bit, and stores on Amazon's EBS.	
PageFreezer		A subscription service to archive, browse and search dynamic web content in compliance with records management laws and as legal evidence.		Uses web crawling software to take daily snapshots of websites. Only new web pages and changes to web pages are archived to save on storage. Subscribers can request a local copy of all their web content in their	

Provider	Paid Service	Product Description & Use Cases	Platforms Able to Capture	Method of Capture	Notes
				native formats (HTML, PDF, TXT, MS Office, OpenOffice, XML, CSS, Flash)	
Parallel-Flickr		Open-source tool for backing up Flickr photos and generating a database-backed website that matches the viewing permissions the user has chosen on Flickr.	Flickr	Downloads and stores a local copy of original photos and their "640x" versions along with the information retrieved via the API as a JSON file. It stores enough data about each photo in a database so that it can reconstruct your photostream and with a webpage for each photo. It uses the Flickr API as a single sign-on and validation service, which means that the site can retrieve and store your contact list and the relationship which each person in it.	Described as a work in progress.
Patrina		Captures, indexes, and consolidates social media feeds into a hosted archive as WORM optical, format.	Facebook, Twitter, LinkedIn, blogs		
Recollect		Backs up users pictures, tweets, and check-ins from multiple social media accounts.	Twitter, Flickr, Instagram, Foursquare		Users have the ability to download their data.
Reed Archives		Captures social media content on demand or on a schedule through social media APIs. Items can be tagged and searched. Users can also implement retention schedules on archived content.	Facebook, Twitter and LinkedIn, Web sites, RSS feeds		Users can export archives individually as PDFs or create bulk exports of entire websites and social media accounts into native format, eDRM XML and PDF.
Site Replay		Subscription service provides daily captures of screenshots with digital watermarks and signatures.	Websites	Captures screenshots of webpages and provides access to them on their secure server. Stored screenshots can be viewed online or downloaded monthly.	
SiteSucker for Mac OS X		SiteSucker is a Macintosh application that automatically downloads web	Web sites	Asynchronously copies a site's webpages, images, backgrounds, movies, and other files	

Provider	Paid Service	Product Description & Use Cases	Platforms Able to Capture	Method of Capture	Notes
		sites from the Internet.		to a local hard drive. Users enter a URL, press return, and SiteSucker downloads the web site.	
Smarsh		Subscription service provides cloud-hosted social media archiving with no installation. Content is captured, preserved and indexed in a Smarsh archive where posts can be searched and retrieved.	LinkedIn, Facebook, Twitter and Chatter	Social media activity is captured via proxy or through APIs in real-time and each archived page and object is time stamped, hashed and stored in native format. Service package includes a monthly copy of client data via encrypted DVD. Users control updates and deletion schedules and can export content in its native format.	Each archive is read-only with 3 copies stored across two physical locations. Content retains its original usability and live links.
SMC4 by Integritie		SMC4 enables automated capture, control, communication and compliance of social media. SMC4 workflow has all the features of standard with advanced case management.	Facebook, Twitter, LinkedIn, Google+, email		
SocialSafe		A downloadable application which will automatically download content when requested and store it locally on the user's hard drive. No additional storage of content is available elsewhere.	Facebook, Twitter, Instagram, LinkedIn, Google+, Viadeo	Saves all tweets (not just latest 3200) as a ZIP file for photos and CSV for Twitter.	
Socialware		The software platform provides a centralized access point for managing and capturing social media in compliance with organizational and legal policies.	Facebook, Twitter, LinkedIn		
Sonian		IM Archive claims that it is easy to deploy, requires no maintenance, and adapts to the evolving IM, SMS, and social media technology landscape.	Social media		

Provider	Paid Service	Product Description & Use Cases	Platforms Able to Capture	Method of Capture	Notes
Spredfast		Customers use the Social Media Management System to monitor, coordinate, and measure social media content.	Twitter, Facebook, Facebook applications, LinkedIn, YouTube, Flickr, SlideShare, blogs	Captures detailed records in the Enterprise Repository of every post made across each social media platform. It also captures an audit trail, internal comments and classifications, and all public engagement from a post.	
ThinkUp		An open source web application that captures activity on social networks.	Twitter, Facebook, Google+	Captures posts made to selected social networks (at the time of this writing, Twitter, Facebook and Google+). Provides a metrics dashboard and the ability to export posts made to those networks as a CSV files and a set of associated metrics.	
Total Discovery		Software supporting litigation, digital investigations and electronic policy consulting now includes data collection capabilities for social media.	Twitter, Facebook		Archives services are tied to e-Discovery services, not stand alone.
Tweet Archivist		Windows application that helps users archive tweets for later data-mining and analysis.	Twitter		
Tweetbook		Creates a PDF ebook of most recent tweets, replies, and favorites at user's request.	Twitter	Up to 3200 most recent tweets included due to Twitter API limitation. Allows an option for a backup file in XML.	
Tweet Library (Mac)		Creates a local and searchable archive of tweets, favorites, and retweets. Creates collections so that users can create timelines.	Twitter	Exports the archive, timeline, or collections to a text file for saving to your Mac or PC.	Downloads up to 3200 tweets on the first launch. Users can upload the .zip archive from Twitter for their complete tweet history.
Tweet Nest		Installs on web server to provide a backup of tweets that users can store, browse, and search. Users can also	Twitter		Users can follow creator Andy Graulund (@graulund) for script up-

Provider	Paid Service	Product Description & Use Cases	Platforms Able to Capture	Method of Capture	Notes
		customize the display.			dates.
Tweetstream		TweetStream provides simple Ruby access to Twitter's Streaming API using open authorization.	Twitter	Uses em-twitter, an EventMachine-based ruby client for the Twitter Streaming API, to connect to Twitter.	
TwInbox		Twitter add-on for Outlook email with searching and grouping capabilities as well as graphs of usage statistics.	Twitter	Downloaded to Outlook.	
Twitter API through manual backup		By submitting queries to the Twitter API, users can backup their data manually, including the data of all the people followed and user tweets.	Twitter	Users can save their tweets, the people they follow and their followers as XML files by manually saving each page.	With an active Twitter account, the manual backup can be a lengthy process.
Twitter Archive Download		Users can request a copy of their Twitter Archive from Twitter. The ZIP file includes user tweets and retweets.	Twitter		
Twitter Archiving Google Spreadsheet (TAGS)		Script that allows users to pull data from Twitter's API and save it in a Google Spreadsheet.	Twitter		The updated version of Twitter Archiving Google Spreadsheet (TAGS) works with the new Twitter API and authenticates access.
Twitter Backup		Downloadable software that captures all tweets and provides them in XML. Uses a document type identical to Twitter's API.	Twitter		
Twitterscribe		Provides a daily backup of user's last 200 tweets along with some metadata in the Twitterscribe database. Users can browse tweet by month and search by keyword, username, hashtag, or URL.	Twitter	When user's first sign up, they capture up to 3200 tweets (the limit of the Twitter API). Requests user's last 200 tweets from Twitter daily. Users can then login and export their tweets from Twitterscribe to CSV or PDF files.	Only captures tweets and retweets.
WARCreate		Google Chrome extension that allows users to create a Web ARChive (WARC) file from any browseable	Websites		WARCreate allows multiple archiving sessions to exist in a single WARC file

Provider	Paid Service	Product Description & Use Cases	Platforms Able to Capture	Method of Capture	Notes
		webpage.			automatically and integrate with Memento .
WP-DB Manager		WordPress plug-in allows users to optimize, repair, backup, restore, or delete a backup database, drop/empty tables and run selected queries. Supports automatic scheduling.	WordPress	WP-DB Manager uses mysqldump application to generate the backup and mysql application to restore them via shell while WP-DB-Backup uses PHP to generate the backup.	
X1 Social Discovery		X1 Social Discovery collects, authenticates, searches, reviews and produces content. MD5 hash values are calculated upon capture and maintained in native format. Content can be tagged, sorted and exported.	Facebook, Twitter, YouTube, LinkedIn	Data is collected and indexed from social media streams, linked content and websites through APIs and direct web navigation, aggregating data in real time. Metadata is captured through APIs provided by the sites.	
YTD Video Downloader		Software allows download of videos from YouTube, including HD and HQ videos.	Video from YouTube, Facebook, Google Video, Yahoo Video	Downloads and converts videos to MOV, MP4, 3GP, WMV, AVI, or MP3 files. Videos are played in Flash.	

V. Federal Records Council Study Findings

In 2011, the Social Media Subgroup of the Federal Records Council conducted a separate study on social media use in their agencies and evaluated the feasibility of various capture approaches. In particular, they examined how certain methods operated with existing Records Management Applications. They tested the following techniques for capture:

- 1) Copy/paste social media record content into a Microsoft Word document AND a PDF/A document and save into a Records Management Application (RMA).
- 2) Copy/paste social media record content into a Microsoft Word document AND a PDF/A document and save into a Share drive/Hard drive/any other non-RMA.
- 3) Utilize a Really Simple Syndication (RSS) feed into an RSS aggregator (i.e. Google Reader).
- 4) Utilize a RSS feed into an email account and save the record in an RMA.
- 5) Utilize commercial options embedded within the social media site or sold commercially (other than an RMA).

The Social Media Subgroup's study indicated that certain approaches are not advised for capturing social media records. Specifically, the study found that screenshots (images of how a web site appeared on a given day and time) were not a helpful tool for capture.¹¹ The President's Digital Government Strategy of 2012 reiterated that government must "[recognize] that simply publishing snapshots of government information is not enough to make it open, we need to improve the quality, accessibility, timeliness, and usability of our data and content through well-defined standards that include the use of machine-readable formats such as web APIs and common metadata tagging schemas."¹² Screenshots only creates a picture of content and do not preserve the metadata and functionality of the content, which does not comply with NARA's transfer guidance for permanent web content records.¹³

VI. Best Practices for Records Management

Most agencies have social media policies in place that cover various topics, such as what tools are being used, who is allowed to create content, and what can be said on social media platforms. Capture of social media records is often left out of these more general social media policies.

Agencies should establish a foundation for successful capture by ensuring appropriate policies and retention schedules are in place prior to capturing social media records. The NARA Bulletin Guidance on managing social media records provides guidance on records scheduling that can help agency records management staff determine if a new schedule is needed for their social media records. Through discussion with agencies and review of the literature, the following list of best practices was compiled:

¹¹ "Navy Official: 'Snapshots' Not Good Archiving Plan," *Information Management Journal* 46, no. 1 (Jan/Feb 2012): 16, ProQuest Research Library.

¹² Executive Office of the President, "Digital Government: Building a 21st Century Platform to Better Serve the American People" (2012), 10, <http://www.whitehouse.gov/sites/default/files/omb/egov/digital-government/digital-government-strategy.pdf>.

¹³ National Archives and Records Administration, "Expanding Acceptable Transfer Requirements: Transfer Instructions for Permanent Electronic Records" (2004), Item 4.2.3, <http://www.archives.gov/records-mgmt/initiatives/web-content-records.html>.

Best Practices for Capture

Building a Foundation

1. Institute a social media working group comprised of agency records management staff, web managers, social media managers, information technology staff, privacy and information security staff, and other relevant stakeholders to create policy and procedures around identifying, managing, and capturing social media records.
2. Apply the definition of a Federal record when making agency-specific determinations about social media content and what components form a complete record.
3. Review existing records retention schedules to determine if social media records are covered. In the absence of an appropriate records retention schedule, records must be treated as permanent until they are scheduled.

Tools and Methods

1. Evaluate and test the available tools for social media capture to determine if the tools meet your needs, budget, and skill level. Agencies may be able to find free or low-cost tools to capture content that can then be managed in a recordkeeping system.
2. Consider using tools that provide transparency related to processes and capture. For example, open source tools can be modified by the agency as necessary.
3. Use the export capabilities that certain platforms have built into their native systems. For example, the social networking platform Yammer converts content into CSV files and makes the files available for downloading.
4. Utilize open Application Programming Interfaces (APIs) to create customized tools that will allow appropriate export and download.
5. Ask the platform provider directly for social media content, such as Twitter.

Implementation

1. Incorporate specific language regarding capture and availability of content for export into the Terms of Service for each new social media platform used.¹⁴
2. Determine and delegate roles and responsibilities for capture on internal and external sites.
3. Provide training for staff on how and when to use automated and manual capture tools.
4. Provide centralized guidance for the agency from a team of records management and web staff so that there is standardization of processes across the agency.
5. Share your lessons learned so that others in the Federal government can benefit from your experience.

Agencies assert they want simple, effective, and low-cost solutions for capturing social media records. Although there are a number of tools that appear promising, agencies are encouraged to articulate their requirements directly to social media providers. NARA does not specifically endorse any tool, but agencies can review this white paper's compiled list of tools and identify ones that meet their specific business and records management needs. Because agencies use different platforms and have different needs, there is no one-size-fits-all tool for capturing social media records across the Federal

¹⁴ See the model Terms of Service agreement at: <http://www.howto.gov/sites/default/files/model-amendment-to-tos-for-g.doc>.

government.

VII. NARA's transfer guidance and best practices for formats

The current guidelines [Expanding Acceptable Transfer Requirements: Transfer Instructions for Permanent Electronic Records Web Content Records](#) specify the requirements for transferring permanent web content records to NARA. NARA is currently updating the transfer guidance for permanent electronic records which will result in the inclusion of additional acceptable formats. Even if the social media records are not permanent the use of sustainable formats can help guarantee accessibility for the life of the record (see [FAQs About Selecting Sustainable Formats for Electronic Records](#)). It is recommended that agencies use tools that capture content with associated metadata so that the content remains searchable and retains its functionality.

References

- 1) Government Accountability Office. *Federal Agencies Need Policies and Procedures for Managing and Protecting Information They Access and Disseminate*, GAO-11-605. (June 2011)
www.gao.gov/new.items/d11605.pdf.
- 2) National Archives and Archives Administration. *A report on Federal Web 2.0 use and record value*. (2010)
www.archives.gov/records-mgmt/resources/web2.0-use.pdf.
- 3) National Archives and Records Administration. "Expanding Acceptable Transfer Requirements: Transfer Instructions for Permanent Electronic Records Web Content Records." (2004)
<http://www.archives.gov/records-mgmt/initiatives/web-content-records.html>.
- 4) National Archives and Records Administration. "Frequently Asked Questions (FAQs) About Selecting Sustainable Formats for Electronic Records." (2012) <http://www.archives.gov/records-mgmt/initiatives/sustainable-faq.html>.
- 5) National Archives and Records Administration. "NARA Bulletin: Guidance on managing social media records." (2013)
- 6) National Archives and Records Administration. "NARA Guidance on Managing Web Records." (2005)
<http://www.archives.gov/records-mgmt/policy/managing-web-records.html>.
- 7) National Archives and Records Administration. *Records Management Self-Assessment Report*. (2011)
<http://www.archives.gov/records-mgmt/resources/self-assessment-2011.pdf>.
- 8) National Archives and Records Administration. NARA Regulations (Subchapter B of 36 Code of Federal Regulations Chapter XII), 2009. <http://www.archives.gov/about/regulations/subchapter/b.html>.
- 9) "Navy Official: 'Snapshots' Not Good Archiving Plan." *Information Management Journal* 46, no. 1 (Jan/Feb 2012): 16. ProQuest Research Library.
- 10) U.S. Navy. *Navy Command Social Media Handbook*. (2010): 3.
http://www.cnrc.navy.mil/PAO/socialnetwrk/soc_med_hnd_bk.pdf.

Additional Resources

- 1) American Council for Technology (ACT) and Industry Advisory Council (IAC). "Best Practices of Social Media Records Policies." (2011)
<http://www.actgov.org/knowledgebank/whitepapers/Documents/Shared%20Interest%20Groups/Collaboration%20and%20Transformation%20SIG/Best%20Practices%20of%20Social%20Media%20Records%20Policies%20-%20CT%20SIG%20-%202003-31-11%20%283%29.pdf>.
- 2) ARMA International. *Implications of Web-Based, Collaborative Technologies in Records Management* (ANSI/ARMA 18-2011). (2011) <http://www.arma.org/standards/index.cfm>.
- 3) Boudreaux, C. "Social Media Governance: Policy Database." (2012)
<http://socialmediagovernance.com/policies.php>.
- 4) Department of Defense. *Directive-Type Memorandum (DTM) 09-026 – Responsible and Effective Use of Internet-Based Capabilities*. (2010) <http://www.dtic.mil/whs/directives/corres/pdf/DTM-09-026.pdf>.
- 5) Federal CIO Council. "Guidelines for secure use of social media by Federal departments and agencies." (2009)
http://www.cio.gov/images/site_img/view_document.jpg.
- 6) Franks, P. C. *How Federal Agencies Can Effectively Manage Records Created Using New Social Media Tools*. San Jose State University on behalf of the IBM Center for The Business of Government. (2010)
<http://www.businessofgovernment.org/report/how-federal-agencies-can-effectively-manage-records-created-using-new-social-media-tools>.
- 7) General Services Administration. *CIO P 2106.2 GSA Social Media Handbook*. (2009)
<http://www.gsa.gov/graphics/staffoffices/socialmediahandbook.pdf>.

- 8) General Services Administration. "Center for Excellence in Digital Government." (2012) <http://www.gsa.gov/portal/category/101991>.
- 9) General Services Administration. "Social Media Apps." (2012) https://www.apps.gov/cloud/cloud/category_home.do?&c=SA.
- 10) GovLoop - Social Network for Government. (2012) <http://www.govloop.com/>.
- 11) Gregory, J. of Social Media Subcouncil. "Gov Social Media Wiki." (2012) <http://govsocmed.pbworks.com/>.
- 12) Herman, J. "Using Social Media in Government." (2012) <http://www.howto.gov/social-media/using-social-media>.
- 13) Kosar, K.R. *Congressional Oversight of Agency Public Communications: Implications of Agency New Media Use*. (2012) <http://www.fas.org/sgp/crs/misc/R42406.pdf>.
- 14) Krzmarzick, A. "Archiving Social Media-Comprehensive Resources on GovLoop and Beyond." (2010) <http://www.govloop.com/profiles/blogs/archiving-social-media>.
- 15) McClure, D. *Statement of Dr. David McClure Associate Administrator for Citizen Services and Information Technologies*, U.S. General Services Administration. (2010) www.gsa.gov/portal/content/158009.
- 16) National Archives of Australia. "Capturing records." (2012) <http://www.naa.gov.au/records-management/agency/create-capture-describe/capturing/index.aspx>.
- 17) National Archives of Australia. "Social media: Another type of Commonwealth record." (2012) <http://www.naa.gov.au/records-management/agency/digital/socialmedia/index.aspx>.
- 18) National Archives and Records Administration. "Implications of Recent Web Technologies for NARA Web Guidance." (2006) <http://www.archives.gov/records-mgmt/initiatives/web-tech.html>.
- 19) New York State Archives. "Records Advisory: Preliminary Guidance on Social Media." (2010) http://www.archives.nysed.gov/a/records/mr_social_media.shtml.
- 20) Social Media Subcouncil of the Federal Web Managers Council. "Web 2.0 Governance Policies and Best Practices." (2011) <http://govsocmed.pbworks.com/w/page/15060450/Web%20%200%20Governance%20Policies%20and%20Best%20Practices>.
- 21) Obama, B. *Presidential Memorandum - Managing Government Records*. (2011) <http://www.whitehouse.gov/the-press-office/2011/11/28/we-cant-wait-president-signs-memorandum-modernize-management-government->.
- 22) Office of Management and Budget. *Open Government Directive*. (2009) http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-06.pdf.
- 23) Office of Management and Budget. *Memorandum on Social Media, Web-Based Interactive Technologies, and the Paperwork Reduction Act*. (2010) http://www.whitehouse.gov/sites/default/files/omb/assets/inforeg/SocialMediaGuidance_04072010.pdf.
- 24) Office of the Law Revision Council. *Federal Records Act (44 U.S.C. Chapter 31)*. (2012) <http://uscode.house.gov/download/pls/44C31.txt>.
- 25) *Presidential Memorandum on Transparency and Open Government*. (2009) <http://www.whitehouse.gov/the-press-office/transparency-and-open-government>.
- 26) Sunstein, C. R. *Social Media, Web-Based Interactive Technologies, and the Paperwork Reduction Act*. (2010) http://www.whitehouse.gov/sites/default/files/omb/assets/inforeg/SocialMediaGuidance_04072010.pdf.
- 27) Wilkins, J. "Social Media Governance: The Policy Part." (2012) <http://www.aiim.org/community/blogs/expert/Social-Media-Governance-The-Policy-Part-1>.
- 28) Wilshushen, G.C. *Challenges in Federal Agencies' Use of Web 2.0 Technologies*, GAO-10-872T, Statement of Gregory C. Wilshusen, Director of Information Security Issues, Testimony before the Subcommittee on Information Policy, Census, and National Archives, Committee on Oversight and Government Reform, House of Representatives. (2010) <http://www.gao.gov/new.items/d10872t.pdf>.