Information Governance Initiative

TAKING CONTROL OF EMAIL

LEARNING FROM THE U.S. FEDERAL GOVERNMENT'S CAPSTONE STRATEGY

information governance initiative

COMPLIMENTS OF OPENTEXT™

Information Governance Initiative • 1271 Avenue of the Americas • Suite 4300 • New York, NY 10020 • iginitiative.com

About the Information Governance Initiative

The Information Governance Initiative (IGI) is a cross-disciplinary consortium and think tank dedicated to advancing the adoption of Information Governance practices and technologies through research, publishing, advocacy, and peer-to-peer networking. The IGI publishes research, benchmarking surveys, and guidance for practitioners that is freely available on its website. Join the IGI Community, a place for practitioners from all facets of IG to come together and learn from each other. The IGI was founded by recognized leaders in the field of Information Governance, and is supported by leading providers of Information Governance products and services.

About this Publication

This publication was written by the Information Governance Initiative as part of our ongoing series exploring issues, strategies, and techniques related to information governance.

This publication was made possible by OpenText's support of the IGI. OpenText is an IGI Charter Supporter. More information about OpenText is available at www.opentext.com

Taking Control of Email

What Information Governance Practitioners Can Learn From the U.S. Federal Government's Capstone Strategy

Executive Summary

Email has been an IT headache for at least two decades. It continues to push messaging infrastructure to its technical and budgetary limits. As CIOs have seen their budgets flat line, the importance of better email management has only increased. To make matters worse, many email messages today are clearly business records, and courts and regulators are losing patience with organizations that do not treat them as such.

The U.S. Federal government recently realized that they had to deal with their email problem. Manual approaches to governance (such as "print to retain") were not working and did not scale. The Capstone approach was created to provide Federal agencies with a simple solution. While not perfect, Capstone does improve on the arbitrary retention and deletion of emails that is prevalent in both the Federal government and private sector.

As Capstone gains steam and the private sector watches to learn if they can successfully leverage the same approach, email auto-classification is maturing into a viable option. Based upon much of the same science as accepted e-discovery tools, auto-classification offers a chance for organizations to proactively classify and manage all emails based upon their importance to the business.

When used as a starting point for email auto-classification, Capstone is a solid strategy. Capstone helps organizations today as they prepare to implement auto-classification technologies, improving their confidence level in their governance of email management.

The Email Challenge

Email is how business gets done in today's world. No matter how many people claim that email is dead and is ready to be replaced, email continues to grow in volume. This growth has reached the point where some people have declared email bankruptcy and deleted all of their emails. As appealing as this may be, it cannot be safely done in most workplaces.

Collaboration systems and social networks have been developed to address the email problem, even to replace email. Many of these potential replacements, while useful, have forced people to visit yet another website in order to get work done. To avoid missing important information, the systems send email notifications to the users, thus merely compounding the problem that the system was trying to fix in the first place.

Email is simply too ubiquitous to be replaced. People can communicate with each other over email without regard to the system on the other side of the *send* button. Email can be checked on every personal computing device, making email the best way to be sure an important business message gets through.

However, the flood of email taxes storage, bandwidth, and the core software itself. While organizations increasingly look to cloud-based email solutions to solve these headaches, not all organizations are willing or able to migrate to the cloud.

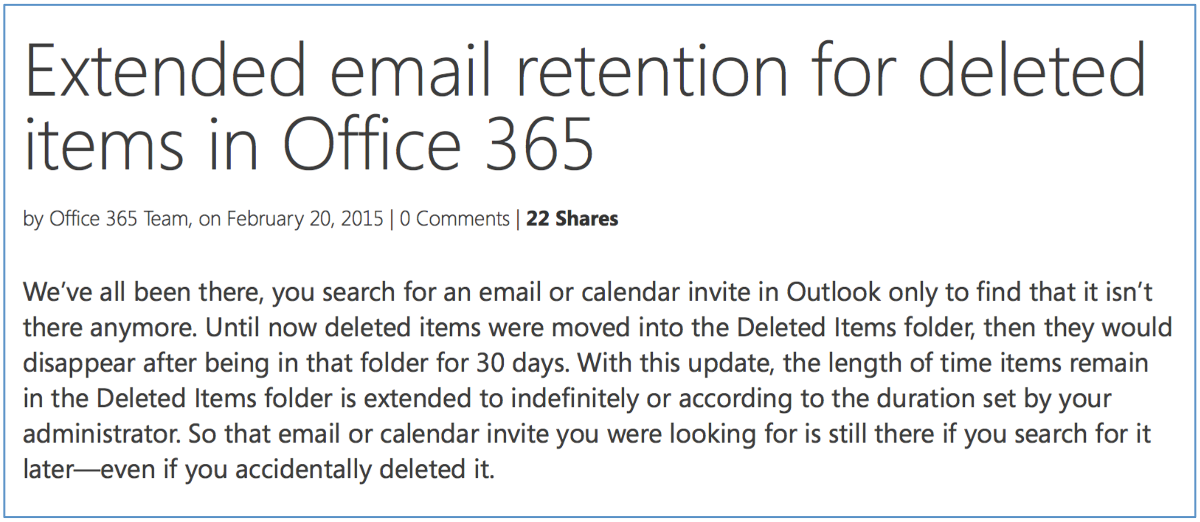
This has hindered the ability of organizations to manage and retain emails to meet legal requirements. When simply keeping a system running and efficient is an effort, anything that adds to the complexity and the scale of the problem is a challenge. For most organizations, efforts to effectively govern email and meet retention and e-discovery requirements have become a mix of shortcuts and oversights.

Early Days of Email Governance: Mailbox Quotas

The first solution that many organizations tried was to create mailbox size quotas. By limiting the size of a mailbox, users were forced to more actively manage their email. However, this approach has many unintentional side effects, such as:

- Premature deletion of email,
- Accident deletion of email,
- Personal exports of email for storage outside of enterprise systems, and
- Use of personal email accounts to conduct business.

Users are forced to regularly delete email to make room for yet more email. They delete (or export) email they *personally* do not need, regardless of the needs of the organization. This leads to information being lost that might later be needed for business or legal purposes. Also, under this system users often delete by mistake – a mistake they do not notice until the deleted items folder is emptied. Requests to the IT department to restore a deleted email from backup tapes are not uncommon in this system.

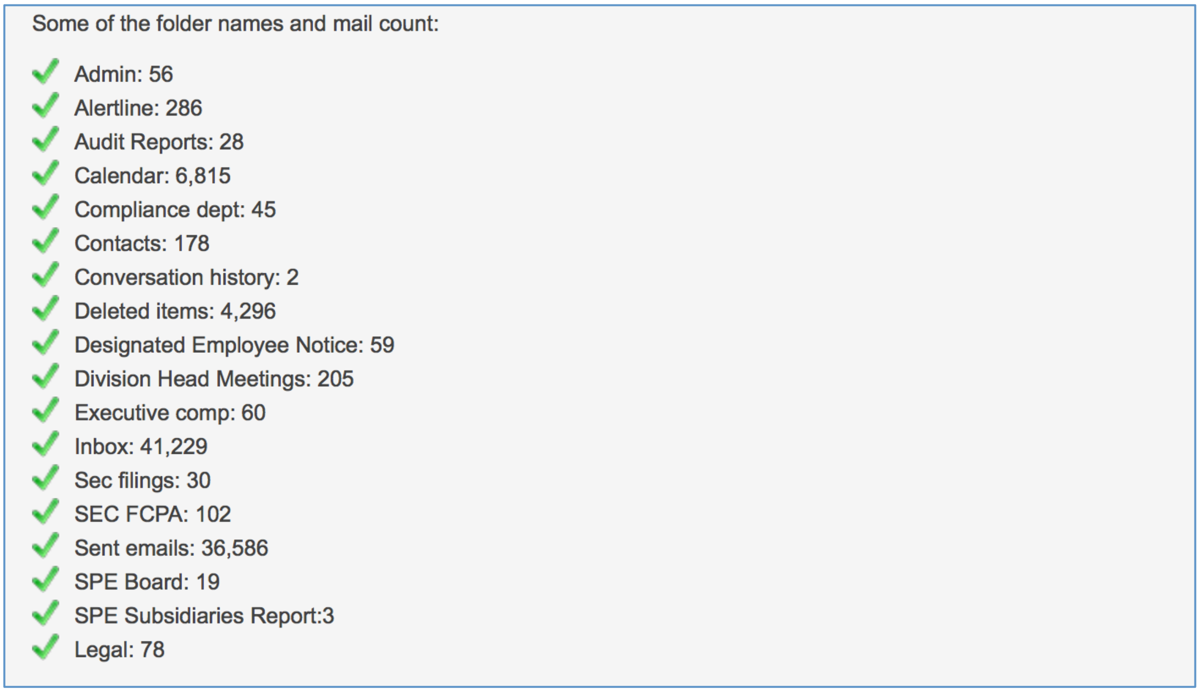


Extended email retention for deleted items in Office 365

by Office 365 Team, on February 20, 2015 | 0 Comments | **22 Shares**

We've all been there, you search for an email or calendar invite in Outlook only to find that it isn't there anymore. Until now deleted items were moved into the Deleted Items folder, then they would disappear after being in that folder for 30 days. With this update, the length of time items remain in the Deleted Items folder is extended to indefinitely or according to the duration set by your administrator. So that email or calendar invite you were looking for is still there if you search for it later—even if you accidentally deleted it.

Figure 1: "Deleted Items" Can Exist Forever: Does Email Ever Go Away?[1]



Some of the folder names and mail count:

✓ Admin: 56
✓ Alertline: 286
✓ Audit Reports: 28
✓ Calendar: 6,815
✓ Compliance dept: 45
✓ Contacts: 178
✓ Conversation history: 2
✓ Deleted items: 4,296
✓ Designated Employee Notice: 59
✓ Division Head Meetings: 205
✓ Executive comp: 60
✓ Inbox: 41,229
✓ Sec filings: 30
✓ SEC FCPA: 102
✓ Sent emails: 36,586
✓ SPE Board: 19
✓ SPE Subsidiaries Report:3
✓ Legal: 78

Figure 2: 4296 Email "Deleted Items" Were Published As Part of the Sony Hack[2]

The constant need to delete emails typically leads people to one of two methods to work around the quotas: exporting to a personal file (e.g. a PST file) or forwarding to a personal email account. This gives the employee a copy that they can manage themselves and find when needed. It is not uncommon for employees to set up a rule to automatically move or forward the email to their preferred location.

A third, more radical, bypass of quotas is the use of personal email accounts for conducting business. Rather than deal with the complexities of routing emails or saving them offline, some people only minimally used their assigned email account and worked primarily in their personal email accounts. This practice was illustrated by former U.S. Secretary of State Hillary Clinton's use of her personal email account exclusively.[3] This allowed her to work unencumbered by imposed security and compliance restraints. This meant that Clinton's email existed in a less secure environment, leaving it open to hackers or deletion without being able to preserve the email properly as a record.

This creates a lot of risk for an organization as email is now being stored in an unknown number of places, many of which are out the control of the organization. PST files began appearing on file stores and in content management systems. They were taken home and kept on an employee's personal computer. Emails that were forwarded to personal email accounts could end up anywhere.

Evolving to Arbitrary Email Retention Periods

Email has been used as a "smoking gun" in many highly publicized trials over the past few years.[4] Many organizations now see it as a liability and have determined that instead of quotas, and sometimes in addition to quotas, all email should be deleted after a predetermined amount of time has passed. A typical time frame many organizations use is 30, 60, or 90 days.

This 30-60-90 approach has two major flaws.

1. Employees still save email in external systems, and
2. Not all email is created equal.

Employees often work around the system to save the emails that they think are important. While an email may not be of utmost importance, losing email after 90 days during a six-month project was something people were not willing to accept. Organizations began restricting the ability of employees to create PST files, which only accelerated the forwarding of emails to external email accounts.

The other flaw is that the 30-60-90 approach treats all email as equal in importance. Email is not all equal just as not all spreadsheets are equal in importance. If the suggestion was made to protect all spreadsheets in the same manner for retention, the courts and regulatory agencies would likely come down hard. Email is a format and ignoring the content of the email in making a determination for retention will inevitably lead to the disposition of information that otherwise has value.

To resolve this problem, many organizations deployed records management solutions that enabled employees to declare emails as records. This permitted employees to make the decisions necessary as to how relevant an email was for the organization. All other email would be purged according to the 30-60-90 rule.

This approach has been met with mixed success. Employee email capture and classification by employees was seen not only as a way to meet the organization's goals but to help employees be more efficient by retaining emails that were important. When given the proper training and support, staff greatly improved the efficiency of managing emails as records.

The problem is that the manual classification of email records is prone to errors. People would declare any email that they wanted to keep past the deadline as a record even if it was not considered a record by the organization. Emails would be improperly classified and stored in the wrong location and retained for the improper amount of time.

Having employees declare important emails as a record forced every email user to become familiar with the records policies of the organization. In many systems, declaring email as a record would also remove an email from the email system either immediately or when the 30-60-90 rule took effect. This made finding emails challenging for people. Once again, they saved email in their offline stores and would not classify email into the proper records categories unless it was an explicit part of their job description.

This has left many organizations either unsure about where all of their email is living or has led to extremely large and expensive archiving systems to keep all email. All in order to satisfy both the real and perceived needs of their employees and the organization as a whole.

Understanding Capstone

> *"It is very difficult to conceive of a scenario — short of nuclear winter — where an agency would be justified in allowing its cabinet-level head officer to solely use a private email communications channel for the conduct of government business," said Jason R. Baron, a lawyer at Drinker Biddle & Reath who is a former director of litigation at the National Archives and Records Administration."*
>
> New York Times, March 2, 2015[5]

The scale of the email governance challenge in the U.S. Federal Government is staggering. The Department of the Interior alone is managing over 70 million emails a month.[6] For the Federal government, email is a major channel of communication with both constituents and the large world. Many of these communications are important to keep as records for periods measured in years, not days.

In August of 2012, the "Managing Government Records Directive" was released jointly by the U.S. Office of Management and Budget (OMB) and the U.S. National Archives and Records Administration (NARA). The Directive set a deadline of December 31, 2016 for Federal agencies to "manage both permanent and temporary email records in an accessible electronic format."[7] In support of this effort, NARA was tasked to devise methods for managing, disposing, and transferring email by the end of 2013, including an investigation into economically feasible methods of automating the management of email.

Capstone is Born

NARA recognized the challenges that Federal agencies were facing on a daily basis and that they would need time to plan and implement any recommendations that they made prior to the 2016 deadline. Consequently, in August 2013 NARA introduced its "Capstone" strategy as an alternative approach to manual classification of emails.[8] Instead of depending on individuals identifying which emails were important and keeping those only those specific emails, all the email for "Capstone" email accounts would be kept.

The Capstone approach is simple and pragmatic. Each agency identifies email accounts that would be most likely to contain information that would be construed as a permanent record. These accounts are picked based upon the role they play within the agency. Senior officials are obvious Capstone accounts. Others include key decision makers like contracting officers, heads of key initiatives, and those whose job called for frequent communication with constituents and other external entities.

All emails for these Capstone accounts are to be kept as permanent records, regardless of content. All non-Capstone email accounts are determined to contain temporary records. If an email was important, it was determined that the information was likely to be captured within either a Capstone account or in other systems. Owners of non-Capstone accounts can still manually declare an email as a permanent record should they feel that the information was critical to capture.

The length of time to keep these temporary email records is up to each agency, subject to any minimum retention period prescribed by future General Records Schedules (under the current draft proposal, the minimum retention period for agencies adopting Capstone would be 3 years)[9]. The goal was consistency and simplicity without having to keep all email or make people act as records managers.

When created, NARA envisioned Capstone to provide several benefits to Federal agencies.[10]

- Reduced reliance on manual filing by staff.
- Optimizing information requests responsive to discovery or FOIA requests by storing everything digitally.
- Preserving permanent email record automatically for eventual transfer to NARA
- Easing the burden of managing email on the end-user.
- Simplifying disposition of temporary and permanent email records.
- Leveraging existing technologies that were already successfully deployed.
- Reducing the risk of unauthorized destruction of email.

Capstone provides federal agencies with a clear and straightforward path to comply with the Records Directive by the end of 2016 and has provided an approved means to safely manage email as records in a consistent manner.

According to Jason R. Baron, one of the primary architects of the Capstone strategy, and Co-Chair of the Information Governance Initiative,

> *"NARA's Capstone policy presents Executive branch agencies with a clear path forward to comply with new Federal policies for the management of email and electronic records. Under the deadline set in the Archivist's 2012 Managing Government Records Directive, Federal agencies are to be managing all e-mail records in electronic form by December 31, 2016. Moreover, by the end of the decade (December 31, 2019), agencies must be preserving all of their permanent electronic records in a manner so as to ensure transfer in digital or electronic form to the National Archives – including permanently appraised email records. Given how long government procurement processes generally take, these policy mandates – and especially the 2016 date – are approaching rapidly."*

Applying Capstone to the Private Sector

As Capstone gains momentum in the Federal government, the private sector needs to ask itself, "Will Capstone work for us?" Private sector CIOs (and General Counsels) face the same daunting email challenges as the Federal government, complete with the same shrinking IT budgets.

Capstone, in many ways, is easier to implement in most private companies than it is in the Federal government. Roles are much more defined and the key positions through which important decisions flow are well defined. Executives, directors, and the finance and human resources departments are a good starting point for any private firm. While there would still be the same challenges of managing email of non-Capstone accounts, the organization can be more confident that important emails are being captured.

Moving Beyond Capstone

> *"In an era when critically important government activities and decisions are conducted via email, a plan to delete the majority of emails at any agency should raise great concern."*
>
> *Senators Patrick Leahy and John Cornyn*[11]

In December of 2014, Senators Patrick Leahy and John Cornyn sent a letter to NARA expressing concern over the Central Intelligence Agency's (CIA) Capstone approach.[12] The CIA had proposed deleting the email of all non-Capstone employees and contractors after they had been departed for three years. Only 22 senior officials were identified as Capstone accounts whose email would be kept permanently, approximately 0.1% of the CIA's employees.[13]

The letter raised two critical issues with the deletion of any records in an indiscriminate matter. The first is that critical information related to activities conducted by the CIA, or any Federal agency, would be deleted prematurely. The second was a concern about losing "a piece of American history."

While the plan presented by the CIA is a dramatic improvement on their existing print-and-file approach to email management, the letter hints at the imperfections of Capstone if not implemented appropriately.

By not requiring a significant level of technology investment by organizations or require people to take actions to capture the records, Capstone is a very pragmatic approach to capturing email records. Everything is done behind the scenes, removing human error from the equation. However, Capstone is also a solution that works because the bar for success is set so low.

Capstone's Weaknesses

Arbitrary deletion of email may be an IT best practice, but it is not an IG best practice.

The factors that make Capstone work are the same factors that diminish it as a long-term solution to email governance. In fact NARA itself proposed it as an interim approach that is expected to be supplanted by more sophisticated solutions in the future.

The most obvious problem with Capstone is that all email for each Capstone account is captured. That means that for every "budget discussion" email, every "cookies are in the kitchen" email will also be captured as a permanent record. This leads to messages being kept as permanent records that barely qualify as temporary records.

The other side of this concern is that some conversations may ultimately be missed. While Capstone specifically recommends capturing the email of key personnel that are not part of the senior leadership, it only takes a special project or a key person delegating a task for a large chunk of important email being missed. The larger the organization, the more likely that a critical email will be missed.

So where do IG professionals draw the line? In the case of the CIA, 22 of the top officials were listed as the key accounts. Why not the emails for the 23rd highest position? What about staff in the contracting office? HR? Email from the staff in charge of intelligence in each region will likely be of great interest to historians in 50 years.

As for emails that do not fall into any Capstone email accounts, Capstone recommends determining a minimum date for keeping those emails and purging them at that point (as noted above, under a current NARA proposal, the minimum retention period would be 3 years).

Arbitrary deletion of email may be an IT best practice, but it is not an IG best practice. Keeping all emails from Capstone accounts permanently or for extremely long periods of time has the same weakness as same as the 30-60-90 day problem: it is a decision not based on the value of the content itself. Treating all emails as the same by assigning the same retention, regardless of the content of the email, challenges the very records management doctrine that retention decisions should be based on record content, not record format. By assigning all emails from a Capstone account to the same retention period, some email will be kept too long. More importantly, some email in non-Capstone accounts might not be kept long enough.

Auto-Classification of Email

Email auto-classification technology has advanced significantly since the initial release of Capstone in 2013. Often referred to as "predictive coding" in the legal space, email auto-classification for the purposes of e-discovery has been gaining acceptance rapidly. In 2014, the U.S. Tax Court ruled,

> *Predictive coding is an expedited and efficient form of computer-assisted review that allows parties in litigation to avoid the time and costs associated with the traditional, manual review of large volumes of documents.*[14]

The question that has been invariably asked, what if the same technology used to determine if an email was responsive to a discovery request could auto-classify the email on day one of its existence? Could this same technology determine which emails to keep and how long to keep them automatically?

Many vendors in the e-discovery and content management space began asking those same questions. After advancing the space to a level of maturity and trust within the courts system,[15] many providers began to adapt their technology to the auto-classification of emails and documents for the purposes of automated records management. The logic was *why sort through all the emails sent in the past 5 to 10 years when you can search through only the relevant email records*?

The difference in execution is slight but important. For e-discovery, the classification engines are generally only placing emails into two categories, responsive and nonresponsive. For proactive email records classification, email must be sorted into potentially hundreds of classifications, including the temporary email record category.

Teaching the auto-classification engine which email belongs in which category can be a long process, though much shorter than sorting all email manually. This is where the benefit of existing records management programs comes through. While there may be hundreds of record categories, many organizations likely have hundreds, if not thousands, of existing records already identified as belonging to each category. These existing records can be used as a starting point to training the auto-classification engine how to correctly assign email to the proper category.

If an organization has already adopted a modern, streamlined "Big Bucket" records schedule, the auto-classification process is more accurate and efficient. The broader record categories decreases the precision needed for the auto-classification engine to achieve high levels of accuracy. The reduced number of records categories also reduces the learning burden for the auto-classification engine given the fewer categories and more existing records per category.

Many of the advanced tools use statistics to show natural groupings of emails and can estimate confidence levels of the emails that it has classified. If the engine has determined that an assigned group appears too loosely related, it can request additional emails to further refine its auto-classification techniques. As the engine learns and becomes smarter, it can reevaluate already classified email and either improve the confidence level in the classification or move an improperly classified email. This can also apply to emails that were filed manually, leading to a much cleaner and more accurate email records repository.

A Holistic Information Governance Approach

Looking past 2016, the Managing Government Records Directive set a 2019 deadline for managing all permanent records electronically. Emails are records of business decisions and actions but they are not the only dimension of business information. It is equally important to retain and manage documents, spreadsheets, case data, and other information artifacts as records to capture the full picture. The ultimate goal is to ensure that all information – regardless of its format – is managed consistently and according to the same rules and controls.

This holistic vision for IG is one that the public and private sector have been working towards for years. Applying Capstone, auto-classification, and automation to this challenge can improve the accuracy and reliability of preserving not just email records, but all business records. Using complimentary technologies, such as enterprise content management (ECM) also allows for emails, documents, and all information to be retained and managed in the proper context. NARA recently discussed the importance of a integrated approach to the "automated management of email, social media, and other types of digital record content" in its Automated Electronic Records Management Report/Plan.[16]

Start Planning Now

Email is not going away. If the past two decades of email growth has taught the legal and technology worlds anything, it is that email is where work gets done. Email is self-perpetuating and has entrenched itself as permanent part of the business landscape.

Using Capstone plus auto-classification technologies has several advantages. Capstone is easy to implement and can be done today by most organizations without the purchase of new technology. Instead of randomly deleting all email after a set date, Capstone accounts can be kept in their entirety.

In parallel, organizations can start investigating email auto-classification technologies to determine the feasibility of deploying the technology. Once brought-in, auto-classification technology can be used to process all archived email, including Capstone accounts, to determine which emails can be safely disposed and which ones need to be treated as a business record. Eventually all email can be evaluated in the same manner and retained only as long as they have both business and record benefits.

This approach meets the needs of all stakeholders. By minimizing employee involvement in determining which email to retain, and for how long, errors and situations such as former Secretary Clinton's use of personal email are avoided. Email is not deleted because nobody took the time to classifying the email. People can go back to focusing on their jobs and records managers can rest assured that all important email is captured and preserved for future use.

The IGI is a member of the newly formed Coalition for Public Sector Information Governance Leadership, formed to assist NARA and private industry to meet challenges in the public recordkeeping space, including how agencies can best comply with the Archivist's 2016 and 2019 digital mandates. The IGI will be working with partners in the Coalition on education and training on how to integrate Capstone policies into industry provider solutions.

Endnotes

[1] “Extended Email Retention for Deleted Items in Office 365.” *Office Blogs*. Microsoft, n.d. Web. 27 Feb. 2015. <http://blogs.office.com/2015/02/20/extended-email-retention-deleted-items-office-365/>.

[2] “A Breakdown and Analysis of the December, 2014 Sony Hack.” *Risk Based Security*. N.p., n.d. Web. 26 Feb. 2015. <https://www.riskbasedsecurity.com/2014/12/a-breakdown-and-analysis-of-the-december-2014-sony-hack/>.

[3] Gearan, Anne. "Hillary Clinton Used Private E-mail for Government Business at State Dept." *Washington Post*. The Washington Post, 3 Mar. 2015. Web. 04 Mar. 2015. <http://www.washingtonpost.com/politics/hillary-clinton-used-private-e-mail-for-government-business-at-state-dept/2015/03/02/275d13d8-c156-11e4-9271-610273846239_story.html>.

[4] Craig, Susanne, and Ben Protess. “Former Trader Is Found Liable in Fraud Case.” *DealBook*. The New York Times, 01 Aug. 2013. Web. 27 Feb. 2015. <http://dealbook.nytimes.com/2013/08/01/former-goldman-trader-is-found-liable-in-mortgage-deal/>.

[5] Schmidt, Michael S. “Hillary Clinton Used Personal Email Account at State Dept., Possibly Breaking Rules.” *New York Times.* 2 Mar. 2015. Web. 03 Mar. 2015. <http://www.nytimes.com/2015/03/03/us/politics/hillary-clintons-use-of-private-email-at-state-department-raises-flags.html>

[6] Montel, John; “Super Buckets and Auto-Classification: How to Manage 70 Million Documents,” AIIM Conference 2014, Orlando.

[7] “Managing Government Records Directive”, OMB M-12-18, 24 Aug. 2012.

[8] NARA. “NARA Bulletin 2013-02.” 29 Aug 2013. Web. 27 Feb 2015 < http://www.archives.gov/records-mgmt/bulletins/2013/2013-02.html>.

[9] NARA. “NARA GRS 6.1, Email Managed Under a Capstone Approach (Draft).” 21 July 2014. Web. 27 Feb 2015. <http://www.archives.gov/records-mgmt/grs/grs-6.1-review-package.pdf>

[10] Rosen, Don, “Transforming Federal Records Management”, AIIM Conference 2014, Orlando, FL.

[11] “Leahy & Cornyn Press For CIA To Retain Email Records." *Senator Patrick Leahy*. United States Senate, 1 Dec. 2014. Web. 27 Feb. 2015. <http://www.leahy.senate.gov/press/leahy-and-cornyn-press-for-cia-to-retain-email-records>.

[12] "Leahy & Cornyn Press For CIA To Retain Email Records." *Senator Patrick Leahy*. United States Senate, 1 Dec. 2014. Web. 27 Feb. 2015. <http://www.leahy.senate.gov/press/leahy-and-cornyn-press-for-cia-to-retain-email-records>.

[13] Gellman, Barton, and Greg Miller. “‘Black Budget' Summary Details U.S. Spy Network's Successes, Failures and Objectives.” *Washington Post*. The Washington Post, 29 Aug. 2013. Web. 26 Feb. 2015. <http://www.washingtonpost.com/world/national-security/black-budget-summary-details-us-spy-networks-successes-failures-and-objectives/2013/08/29/7e57bb78-10ab-11e3-8cdd-bcdc09410972_story.html>.

[14] Fruchter, Joshua, Esq. “Tax Court Okays Use of Predictive Coding to Review Documents.” *The National Law Review*. N.p., 10 Dec. 2014. Web. 27 Feb. 2015. <http://www.natlawreview.com/article/tax-court-okays-use-predictive-coding-to-review-documents>.

[15] Cohen, Akiva. “Evolving Judicial Attitudes Towards Predictive Coding Suggest It May Soon Be Time To Retire The Defensibility Question.” *ITLex Technology Law*. N.p., 10 Apr. 2013. Web. 27 Feb. 2015. <http://it-lex.org/evolving-judicial-attitudes-towards-predictive-coding-suggest-it-may-soon-be-time-to-retire-the-defensibility-question/>.

[16] Office of the Chief Records Officer for the U.S. Government. “Automated Electronic Records Management Report/Plan.” National Archives and Records Administration. 19 Sept. 2014. Web. 27 February 2015. < http://www.archives.gov/records-mgmt/prmd/A31report-9-19-14.pdf>.