

Summary of Bush 43 Missing White House E-mail Controversy

(updated December 14, 2009)

The record-setting levels of government secrecy during the Bush Administration have been a threshold barrier to accomplishing serious accountability for the actions of that Administration. When whistleblowers with inside knowledge of the White House computer systems came forward in Spring 2007 claiming that the unclassified White House email system had not archived email systematically since 2002, and that at least 5 million email messages were missing from the period of March 2003 through October 2005, the National Security Archive responded as it had during three prior presidential administrations – the Archive brought a lawsuit under the Administrative Procedures Act and the Federal Records Act challenging the White House’s failure to preserve its records.

Consider the time period of the missing email. Early 2002 was the peak of the Afghanistan war, just as special forces troops were being shifted to plan for what ultimately became the war in Iraq. Roundups and renditions after 9/11 were reaching their peak as well. March 2003 was the month of the invasion of Iraq. The following two and a half years would include the revelations of Abu Ghraib, the rise of the insurgency in Iraq, the 2004 elections, Hurricane Katrina, an investigation into the disclosure of the covert CIA status of Valerie Plame, and the allegedly political firing of United States Attorneys.

As the National Security Archive knew from its prior experience litigating over missing White House emails, just because the email were missing from the archiving system did not mean that the email were unrecoverable. Our prior White House email lawsuit took six years against Presidents Reagan, Bush 41 and Clinton just to establish the principle that email qualify as records and must be appraised and preserved like other records, not tossed like telephone message slips.

As part of the negotiated settlement of the Archive’s litigation, the White House Executive Office of the President (EOP) and the White House Office of Administration (OA) have released thousands of pages of documents that shed light on the inadequacy of the Bush Administration email archiving system, the discovery that emails were missing, and the deficient response to that discovery.

The End of ARMS and the Beginning of the Journaling Archive System

In 1994, EOP implemented the Automated Records Management System (“ARMS”), an electronic records management system that worked with Lotus Notes to automatically receive emails and then preserve and categorize them. Both federal and presidential records were stored in ARMS and ARMS contained controls to prevent unauthorized deletion of emails. *See* Defendant’s Answer; Congressional Testimony of Theresa Payton.

In 2002, EOP decided to replace its Lotus Notes email system with Microsoft Exchange. The migration of the EOP email system from Lotus Notes to Microsoft Exchange occurred over a two-year period from 2002 through 2004. Payton Test.; Declaration of Stephen M. Everett.

Because ARMS purportedly could not be effectively integrated with Microsoft Exchange, the EOP did not use the ARMS' automatic archiving system with Microsoft Exchange.

In place of the automatic ARMS system, the Office of the Chief Information Officer (“OCIO”) inside OA developed a manual archiving process that used the journaling function inherent in Microsoft Exchange. In her testimony, Theresa Payton, head of OCIO, explained how this journaling archive system was supposed to work:

Under that process, and in very general terms, whenever email is sent or received by an EOP Exchange customer, a copy of that email is automatically created and stored on a journal to which the customer should not have access. Journalled emails are then archived on a separate server in what is referred to as a Personal Storage Table or “PST” file.

Thus, when Exchange was first deployed at EOP, the .PST file was *manually* created by contractors within OA to archive the message contained in the Journal. Everett Decl. This .PST file therefore contained multiple email messages in its archived form, and was stored in the EOP email message archive. This manual process was the sole method of email preservation at the time.

The Journaling Archive System Failed to Properly Archive White House Emails

The journaling archive process, never intended to be a permanent archival solution, was at best an inadequate archival solution, and at worst posed a grave danger to the integrity of the very records it was supposed to be preserving. Specifically:

- The Memorandum to Majority Members of the Committee on Oversight and Government Reform said that the journaling archive process “posed serious dangers, such as the risk of data loss, the risk of tampering, and the inability to verify that systems were working properly.”
- One NARA official wrote: “I refer to it as a ‘message collection system,’ even though we all understand that it hardly qualifies as a ‘system’ by the usual IT definition.”
- Carlos Solari, Theresa Payton’s predecessor as Chief Information Officer, described the journaling process as a “temporary” solution that was not considered by the White House as a “good long-term solution.”
- A White House Discussion Document stated: “There is operational risk in current email storage management processes. Lost or misplaced email archives may result in an inability to meet statutory requirements . . . Standard operating procedures for email management do not exist . . . The current version is prone to failure.”
- In 2005, Steven McDevitt, then-Director of Architecture and Engineering for OCIO, warned that, “[t]he current email archive process depends on manual operations and monitoring, standard operating procedures do not exist, automated tools that support the email archive process are not robust, and there is no dedicated archive storage location.”
- As early as 2004, NARA warned the White House that it was “operating at risk by not capturing and storing messages outside of the email system.”

In written responses to questions from Congress, Steven McDevitt provided an extensive description of the potential problems with the journaling archive system. First, he stated that data might be missing because “[t]he process by which email was being collected and retained was primitive and the risk that data would be lost was high . . . [and] the risk was compounded by the fact that there was no mechanism to reconcile the messages that were retained in the .PST files and the messages that had been processed by the Exchange system.” *Id.* Second, he identified a problem with “data reconciliation” because “[t]he use of .PST files for warehousing email records does not provide a mechanism to reconcile against what was originally retained by the system. This is there [sic] is no way to guarantee that all records are retained in their complete and unmodified state.” *Id.* Third, there was little “user accountability” in the system because “[t]he approach of simply storing email message in .PST files provides no mechanism or audit trail that tracks changes to data files or the activities performed by users or system administrators.”

Documents produced during the litigation demonstrate that the White House had knowledge of these problems and state that “the manual processes were applied inconsistently,” and that additional training and monitoring was necessary. The manual PST file management and search processes were “resource intensive” and “time consuming.” The “operating routines did not reflect the complex environment leading to inconsistent management of the process” and there was “limited written documentation.”

The White House Was Aware of Serious Problems as Early as February 2004, Yet Failed to Fix the Problem

When the Office of Administration (OA) had problems locating emails in response to a January 22, 2004 subpoena from Special Counsel Patrick J. Fitzgerald, – who was investigating the leak of Valerie Plame Wilson’s covert CIA identity – it called in Microsoft for assistance. Microsoft wrote a “post-mortem” analysis in February 2004 that described several serious problems with preserving and searching for emails:

- No mechanism to transfer Exchange mail into ARMS. OA’s current tool “is not yet a stable and consistent solution to transfer data from Exchange into ARMS.”
- Inadequate backup strategy. The server drives for the EOP’s Record Management (EOPRM) “crashed” on February 5, 2004, the day before the EOP’s response to the January 22 subpoena was due. The only backup available was two weeks old, and Microsoft was forced to completely rebuild EOPRM on a new server.
- Journal mailboxes are too large, causing incomplete search results. In October 2003, after the White House had received an earlier subpoena from Special Counsel Fitzgerald, OA stopped moving the journal mailboxes into PST files. This decision left all journaled email to accumulate in the journal mailboxes for some unknown length of time. These mailboxes were very large and searching them “does not provide a complete set of results.”
- Lack of adequate search utilities. The journal archiving system did not easily allow for printing and paginating search results, requiring the creation of a manual workaround by Microsoft. Further, EOP has no way of determining how many

messages had attachments and determining the total number of attachments in all messages.

Millions of Emails Go Missing

Libby Subpoena

In 2005, in response to a subpoena issued by Special Counsel Patrick Fitzgerald in the Valerie Plame leak investigation, EOP discovered that neither the email message archives, nor the backup tapes, contained any journaled emails or .PST files from September 30 through October 6, 2003 for the Office of the Vice President, an EOP Presidential Records Act component. The only emails that could be recovered and provided to Fitzgerald were emails the White House was able to restore from the personal email accounts of officials in the Vice President's Office.

By January 23, 2006, Mr. Fitzgerald tells Scooter Libby's defense counsel "we have learned that not all email of the Office of Vice President and the Executive Office of the President for certain time periods in 2003 was preserved through the normal archiving process on the White House computer system." Publicity about this letter apparently triggered NARA officials to contact the White House Counsel's Office to remind them that NARA should be contacted prior to destruction of Presidential records. In a response on February 6, 2006, OA Counsel informs NARA that White House "believed that the emails existed and could be accounted for."

Internal Analysis Confirms a Problem

When it became apparent in October 2005 that OCIO staff and contractors were not effectively managing the .PST files used to retain the email records for the EOP, a team of OCIO employees undertook an analysis of the .PST file storage process led by Steven McDevitt. The goal of the 2005 Statistical Analysis was to organize and inventory the .PST files used for EOP email records retention and to put in place a formal process to manage these files. In this analysis, OCIO found that "[i]n addition to there being hundreds of days for which specific components had no email retained, there were a number of days for which it was clear that the number of emails retained was lower than expected." Letter from McDevitt to Rep. Waxman. Specifically, a spreadsheet revealed that there were 473 days during this period in which a component of the EOP had no emails preserved on the servers, as well as 229 in which a component had an unusually low number of emails preserved on the servers. The number of emails deleted from servers, and recoverable only from backup tapes, was estimated at approximately five million. Documents obtained during the litigation reveal that OA believes that "Email messages were not properly captured on certain days." OA estimated that for at least 473 days, "information was not captured."

Despite Missing Emails, the White House inexplicably cancelled a comprehensive electronic records management system it had spent millions of dollars to develop

The government spent several years—not to mention millions of dollars—designing, developing, and testing the Electronic Communications Records Management System (ECRMS), in order to, in their words, “provide an efficient and effective solution to meet the current Electronic Records Management requirements of the EOP.” Letter from McDevitt to Rep. Waxman. In May 2006, after four years of development, a member of the OCIO called it “the most important system we have implemented in a long time.” Yet six months later, the government cancelled it. By late 2006, the Government had still failed to implement an appropriate records management system. Instead, “many procedures still entail manually intensive tasks.” “Due to competing demands and the complexity of the processes [the email database] was not completely maintained; the database provides incomplete inventory.”

OA explained to NARA on February 6, 2007, that it decided not to install ECRMS and that the White House’s emails were not being preserved in a formal electronic record system (and NARA would thus likely receive emails in multiple formats). OA did not give NARA any indication that there is a concern about missing emails. The government’s purported reason for getting rid of ECRMS was that the “system offered users no option to distinguish between Presidential records and political or personal materials.” Yet, the records make clear that the White House was aware of this limitation throughout the development of ECRMS. Later CIO Theresa Payton described concern over the predicted 18-month “ingestion” of backlogged messages (although the ingestion would have been complete by the end of Pres. Bush’s second term,) and the inability for staff to distinguish records subject to the PRA and FRA from political or personal email (despite the fact that OA Counsel had previously approved this aspect of the ECRMS system.)

After CREW publishes Whistleblower Report, White House Admits Emails Are Missing

In April 2007, Citizens for Responsibility and Ethics in Washington (“CREW”) released a report based on confidential information that publicized the fact that emails were missing at EOP. At least two spokespersons from the White House admitted the missing email problem. White House Spokesperson Dana Perino acknowledged in a press gaggle that there was a problem with missing White House emails and stated that she was “not taking issue” with CREW’s conclusions that there were hundreds of days for which emails were missing. Apr. 13, 2007 Press Briefing by Dana Perino at 2. Neither did Ms. Perino dispute CREW’s estimate that there were over five million missing emails. *Id.* Around this time, NARA meets with White House Officials and the Archivist of the United States sends a letter to the White House Counsel urging restoration of the missing email. NARA meets with White House officials several times over the next several months and repeatedly requests details on the email situation, but NARA is not permitted to see the 2005 Statistical Analysis until October 2007 – six months after the White House admitted the problem and almost a year after the Analysis was compiled.

White House Begins to Deny Missing Emails

The White House’s discussion of the state of its email shifts back and forth over several months. On December 4, 2007, White House spokesman Scott Stanzel states, “We are aware that some emails may not have been automatically archived in the past, but they may be available on backup tapes. ... The Office of Administration at the White House has been

maintaining and preserving backup tapes for the official email system.” On January 15, 2008, the White House, in a declaration by CIO Theresa Payton, refuses to acknowledge any missing emails but states that it “has undertaken an independent effort to determine whether there may be anomalies in Exchange email counts.” Later, White House Spokesperson Tony Fratto said that an uncertain number of emails were deleted: “It was a problem we announced, admitted to and will remedy.” Still later, however, on January 17, 2008, White House spokesman Tony Fratto states in a press briefing that “we have no reason to believe than any email at all are missing.” Fratto said the correctness of the data in the 2005 Statistical Analysis could not be verified.

The Archive Initiates Litigation to Preserve These Vital Documents

The National Security Archive brought this suit in September 2007 to protect our nation’s history. The Complaint alleges that an unknown problem caused emails to be deleted from White House servers and put them at risk of permanent destruction. To prevent this unlawful loss of federal records, the Archive invoked the remedies of the Federal Records Act — seeking an order compelling Defendants to seek legal action from the Attorney General to preserve or recover emails and review of Defendants’ recordkeeping guidelines. Later CREW filed a virtually identical lawsuit and the cases eventually were consolidated.

Defendants Resisted Every Effort to Preserve the Emails at Issue

When Plaintiffs sought early on to preserve backup tapes for the emails at issue in the case while the legal issues were resolved, Defendants implored the Court to rely solely on counsel’s assurances that everything would be preserved and, as Magistrate Judge Facciola noted, “balk[ed] at entering into any stipulation that would in effect serve as the premise of an order requiring it to do so.” The Court wisely chose to order complete preservation of the backup tapes.

When Magistrate Judge Facciola recognized in January of 2008 that “a small amount of information not currently in the record may have a large [e]ffect on the resolution of [the discovery] motion and the direction of the lawsuit,” he posed a series of targeted questions to Defendants regarding preservation. In response, Defendants submitted a series of declarations that were so vague and incomplete, and that so utterly failed to answer the Magistrate Judge’s questions, that Magistrate Judge Facciola decried the “lack of precision in EOP’s responses,” and declared that Defendants’ “response was inadequate.” After Defendants’ *third* attempt at responding to his inquiries, Magistrate Judge Facciola found that Defendants “once again fail[ed] to provide anything more than generalizations,” and “once again failed to describe the potential costs [of proposed relief] in precise terms.”

When the House of Representatives Committee on Oversight and Government Reform held a February 24, 2008 hearing on the issues at the heart of this case, it became clear that the Court’s existing preservation order was not broad enough to preserve the emails at issue in this case because, among other things, it was made public that Defendants had recycled email backup tapes for several of the months at issue. When the Archive moved to extend the preservation order, and submitted evidence from the congressional hearing, , Defendants again strenuously resisted this motion, as well as efforts to gain further discovery, even though it was Defendants’

own failure to preserve backup tapes — or to mention this fact to the Court — that had necessitated the motion. After several unsuccessful attempts to extract information from Defendants, Magistrate Judge Facciola recommended that the Court grant the Archive's motion in part, recommending that the Court order EOP to search employee workstations for .PST files from the March 2003 through October 2005, and to issue a preservation notice to employees ordering them to surrender any and all media that might contain emails from the relevant period.

Meanwhile, Although They Did Not Tell the Court or the Plaintiffs, Defendants Had Commenced a Costly Analysis and Partial Restore

After the 2005 Email Statistical Analysis concludes that millions of email are missing across the EOP's components, OA staff recommended a series of steps to recover this missing email. It appears that rather than conduct a full restore of email from the backup tapes, once the April 2007 whistleblower report exposed the loss of emails, the White House instead decided to first focus on undermining the validity of the 2005 Statistical Analysis. The White House established a lengthy and complex three-phase process to decide which days of missing email to restore.

Phase I involves searching for missing PST files by recreating the inventory underlying the 2005 Statistical Analysis with better technology. Phase I was conducted at the PST file level. Phase II involved use of a PST Inventory Verification and Investigation Tool (PIVIT) to allocate messages to the appropriate EOP component by reading message header information. PIVIT was run several times as OA continued to discover previously undetected PST files. Phase II also involved use of an Auto-Regressive Integrated Moving Average Model (ARIMA) to identify statistically low email days. ARIMA modeling identified 76 statistically low component days and OA decides to restore 21 calendar days comprising 48 component days. Phase III involved a restoration of the selected days from emergency recovery backup tapes and a compare of the emails in existence before the restoration and those available after the restoration for the specific component days that were restored. The White House first informed the court and plaintiffs about these processes on January 21, 2009, when it filed a second motion to dismiss claiming that, although many email had been mislabeled and categorized, OA had now located 22 million additional email and had completed a new analysis, statistical modeling, and partial restoration and the case should be dismissed because it was now moot. Despite White House representations on January 21, 2009, this partial restoration process continued well into 2009 and the comparison process has not yet taken place.

Ultimately, it appears that the partial restoration cost over \$11 million and countless man hours at OA.

Conclusion

There is not yet a final tally on the extent of the missing email problem and the record suggests that there may never be final numbers. We do know, however, that 22 million presidential and federal record email that were previously mislabeled and effectively lost have now been found and properly categorized so that they can be searched. In addition, 94 calendar days will have been restored and preserved at NARA for future research. Although the numbers are not definitive, at least some sources suggest that between dozens and thousands of email have

been recovered on several of the component days in questions. Importantly, the White House now has an appropriate email archiving system in place and improved emergency recovery backup practices.