

## **DECLARATION OF STEPHEN M. EVERETT**

I, Stephen M. Everett, declare as follows:

I am Stephen M. Everett, Chief Information Officer, Office of Administration (OA), Executive Office of the President (EOP). I have held the position of Chief Information Officer (CIO) since September 22, 2008. The statements in this Declaration are based on my personal knowledge or documents or on information provided to me by members of my staff.

OA was created by Reorganization Plan No. 1 of 1977 and Executive Order 12028. Its primary function is to provide common administrative and support services for EOP components. The services provided include information, personnel, and financial management; data processing; library services; records management; and general office operations, such as mail, messenger, printing, procurement, and supply services. OA offices include the Office of the Director; Office of the Chief Facilities Management Officer; Office of the Chief Financial Officer; Office of the Chief Information Officer; Office of the Chief Operating Officer; Office of the Chief Procurement and Contract Management Officer; Office of Security and Emergency and Preparedness; Office of the General Counsel; and the Office of Equal Employment Opportunity.

One of the areas of responsibility of OA relates to the unclassified network ("EOP Network") that is used by the twelve components of the EOP (including, for example, the White House Office, the Office of Management and Budget, the Office of National Drug Control Policy, etc.) to send and receive e-mail messages. The Office of the Chief Information Officer (OCIO) is the division of OA that controls the EOP Network.

### **I. 2005 Review**

Based upon discussions I have had with OCIO staff and documents presented to me, I understand that in late 2005 OCIO performed an inventory of the "EOP e-mail message archive," attempting to identify the number of e-mail messages archived by the various EOP components for dates ranging between January 1, 2003 and August 10, 2005 (hereinafter "the First Inventory Period").

The “EOP e-mail message archive” is the repository of archived emails from the EOP Network for the Microsoft Exchange email system. Prior to use of the Microsoft Exchange email system, the components of EOP used Lotus Notes to send and receive emails. However, a determination was made to migrate EOP components to Exchange, and the components were migrated over (through various pilot periods) through the course of two years. When Exchange was first deployed at the EOP, emails were archived through “Exchange Journal Mailboxes,” which contained a duplicate copy of every email sent or received by EOP components on the EOP Network. When a Journal reached its storage capacity, a .PST file was then manually created by contractors within OA to archive the messages contained in the Journal. A .PST file therefore contained multiple e-mail messages in its archived form, and was stored in the EOP e-mail message archive. The names of each .PST file contained the name of the component from which the messages were supposed to have been Journalled.

In late 2004, OCIO developed an operating procedure for the inventory of .PST files in the email message archives. In addition, a .PST file database was developed, and the process for moving files from Journal to .PST files was automated with a program called “Mail Attender.” As with the process used above, Mail Attender relied on the creation of a duplicate copy of every email sent or received by EOP components on the EOP Network. Thus, the EOP email process is “bifurcated,” resulting in two identical messages on the Exchange server for every message sent or received on the EOP Network. One message is placed into the Journal Mailbox for the component sending or receiving the email, and the other message is contained in the user’s mailbox. As before, this email bifurcation process is automated, and no end user may control it. Mail Attender then automatically moves emails from the component Journal Mailbox into .PST files in the appropriate component directory. Those .PST files constitute the email message archives. Attached as Exhibit 1 is a true and correct copy of a document entitled “Current PST Creation Process”, which explains the archiving functions.<sup>1</sup>

---

<sup>1</sup> Beginning January 20, 2009, an “Email Extender” program will be added to the PST file creation process.

It is my understanding that OCIO in 2005 attempted to inventory the .PST files in the EOP e-mail archives. OCIO presented the results of the inventory process in a spreadsheet referred to internally as the “Red/Yellow Chart,” and more widely as the “2005 Chart.” The 2005 review identified EOP components for which message levels were considered low for certain days (“yellow” days); it identified a total of 229 low or “yellow” component days.<sup>2</sup> The 2005 review also identified component days for which there appeared to be zero messages archived (“red” days); it identified a total of 473 “red” component days, that is, days in which it appeared that a particular EOP component had zero e-mail messages preserved in the e-mail message archive.

According to OCIO documents, the original e-mail review conducted in 2005 had a number of flaws and limitations. First, the 2005 effort evidently attempted to identify the number of e-mail messages preserved by the various EOP components on specific dates by counting the number of email messages contained in .PST files. As described above, .PST files are “personal storage table” files in which Microsoft Outlook email messages are saved. One .PST file contains many individual email messages. The 2005 effort assumed that all email messages counted within a .PST file were assigned to the correct components in all cases based on the name of the .PST file. As we discovered, however, .PST files could contain messages for multiple components, and all messages within a .PST file accordingly could not be counted for the component named in the .PST file name. In addition, the “low day” determinations in that effort were based on statistical averages using a 27-day cycle of e-mail counts to determine if the count for a particular day was statistically low. This method, however, did not account for factors which could vary counts within a short period of time such as weather-related closures and holidays, when “low counts” or “zero days” might be expected, and especially for low population components (like the Council of Economic Advisers) that were not issued Blackberry devices.

---

<sup>2</sup> The term “component days,” as opposed to “calendar days,” refers to days for specific components that were identified as “low” on one particular calendar day. Therefore after the migration process from Lotus Notes to Exchange was completed, each calendar day contained 12 component days. But because the 2005 review showed some “low” counts for certain components on a specific day, but not for other components on that same day, we do not refer to a “calendar day” as low, but instead rely on component days here.

The 2005 effort failed to account for seemingly “low” email counts that were attributable to the pilot program process used to deploy Microsoft Exchange. For example, the 2005 review appears to have assumed that any use of Exchange for e-mail by certain employees meant that the component had entirely migrated to Exchange. As we ultimately learned, however, that assumption was incorrect. There were breaks in use of Exchange where components would return to Lotus Notes use (owing, for example, to budget constraints), and certain components required extended periods of time before the majority of its employees used Exchange, rather than Lotus Notes. Accordingly, for “pilot periods” in 2003 through mid-2004 for some components, not all employees within a component had been transferred to Microsoft Exchange from Lotus Notes, leading to artificially “low” and “zero” counts for the archives for that component during the pilot period days.

We also learned that the tool used to “count” messages in the 2005 effort was limited. For example, the tool stopped counting messages when a .PST file for a single day contained more than 32,000 e-mail messages. As we learned, .PST files did contain more than 32,000 e-mail messages. As a result, the 2005 effort failed to count all of the e-mail messages within large .PST files, reporting “low” numbers when the e-mail messages in fact existed in the archive.

## **II. Three-Phase Email Restore Process**

Below, I describe the steps taken to address the concerns raised by the 2005 inventory and review. Based on documents I have seen and discussions with my staff, in addition to my own firsthand experience at OA, I understand that OA has taken significant steps to address these concerns. I describe these steps in the following paragraphs. I also describe the results of OA’s efforts, identifying the numbers of individual EOP e-mail messages accounted for after a process of re-inventory and back-up take restoration. Attached as Exhibit 2 is a true and correct copy of a chart identifying the number of e-mail messages accounted for each of the components between January 1, 2003 and August 10, 2005.

Based on discussions with OCIO staff in the course of my official duties, as well as my review of relevant documentation related to the e-mail recovery effort, I understand today that the following took place before my arrival at OA:

To address the issue raised by the 2005 review, the previous CIO, Theresa Payton, assembled a team from various members of the OCIO staff to perform an analysis of the e-mail message archive and, if needed, a recovery effort. The team pursued a three-phase approach to address the possibility that there may have been a substantial number of e-mails missing from the EOP e-mail message archive.

### *Phase I*

The primary purpose of Phase I was to study the 2005 review and to recreate the inventory underlying that review, but with better technology. The OCIO team conducted a new inventory of the .PST files created in connection with e-mail messaging using Microsoft Exchange. In creating the new inventory, the team relied on the name on the .PST file in which the messages were located to associate the messages with specific EOP components, as had been the case in the 2005 effort. This approach reflected the methodology used in the 2005 review, enabling the team to achieve an “apples to apples” comparison between the Phase I results and the 2005 inventory.

However, the Phase I process differed from the 2005 review in important ways:

First, OCIO determined that a number of “low” or “zero” days from the 2005 review were not the result of “missing” emails, but due to the limited number of messages that were actually sent or received on the Exchange system during the migration of the EOP e-mail system from Lotus Notes to Microsoft Exchange. The 2005 effort evidently counted days in “pilot periods” when a component had not entirely migrated to Microsoft Exchange as “low,” even though the component employees were on both Lotus Notes—where emails were archived in a separate system, “ARMS”—and some on Exchange. OCIO was able to determine the dates of migration, as well as the “pilot periods” of use of Exchange, by reviewing Exchange migration schedules that were available, and by reviewing the Exchange email volume and the volume of emails

captured on ARMS for specific components. These reviews enabled OCIO to determine, for example, that the Office of Management and Budget (OMB) used Exchange during a three-month pilot program, then entirely terminated use of Exchange in favor of Lotus Notes for a two-month period owing to “budget season” considerations, and then engaged in another extended eight month pilot period while some staff used Lotus Notes and other Exchange. This discovery addresses the “zero” days from November 1, 2003 through December 29, 2003 on the 2005 review for OMB, as an example. Thus, in Phase 1, OCIO confirmed that the suspected “anomaly” of zero-message days in the 2005 review for that period for OMB was not the result of “missing emails.” Similar analyses confirmed that the pilot periods for the migration from Exchange to Lotus Notes was not accounted for in the 2005 review. See Exhibit 2 (orange columns reflecting pilot periods for use of Microsoft Exchange).

The team also discovered and eliminated a “counting” limitation found in the tool used in 2005 for the inventory of messages in .PST files. That tool (called “CMDFI”) had a message count limit of 32,000 e-mail messages per day in a .PST file. Because some .PST files did contain more than that number of message objects, and due to this flaw, the 2005 inventory resulted in inaccurately low counts for certain .PSTs. During Phase 1, the team was able to count with the tool without the limitation and achieve accurate message counts for large .PST files.

Based only on this first level of analysis alone, OCIO determined at the conclusion of Phase 1 the following: (1) That the “zero” message component-day count had dropped from 473 to 293; and (2) that the total message count for the electronic inventory had increased from the approximately 81 million counted in the 2005 review to a new total of approximately 94 million.<sup>3</sup> The OCIO team concluded that the 2005 effort had failed to account for approximately

---

<sup>3</sup> In the 2005 chart, the 81 million total number of messages listed included messages counted during the pilot periods. When the flaw in methodology regarding the pilot periods was discovered in Phase I, message counts attributable to pilot periods were no longer included in analyses of “low” days because they were not consistent series that represented all messages of a component for that time. As confirmed using the more accurate PIVIT tool in Phase II, nearly 9 million messages are in the archives for the pilot periods. The 94 million messages counted in Phase I does not include those 9 million messages, which means a total of approximately 103 million messages were counted in Phase I.

13 million messages that existed in 2005 in the EOP email message archives, and that effort had not succeeded in counting and associating those messages with components in the 2005 review.<sup>4</sup>

I have been made aware of additional Phase I discoveries that bear mention: In 2005, there were approximately 10 million messages that could not be allocated to a component based on the PST file name (thus, of the approximately 81 million messages identified in the 2005 effort, only 71 million were attributed to components); Phase I identified approximately 14 million such messages (thus, of approximately 94 million messages identified in Phase 1,<sup>5</sup> approximately 80 million were attributed to components). To effectively allocate these 14 million messages, the OCIO team concluded that it would be necessary to investigate the messages at the message-header level—i.e., the portion of the message indicating sender, recipient, date, time, etc. By reading message-header information, the team could find out more precisely what EOP components should be associated with a given e-mail message.

Phase I revealed another significant limitation in the 2005 approach: the use of a 27-day rolling average to determine statistically “low”-day counts—an approach that did not account for seasonal variations in the time-series data. The team quickly realized it needed a more sophisticated statistical approach for determining what days had “low” message counts. It is my understanding that Phase II was developed with an eye toward addressing these concerns.

### *Phase II*

In Phase II, the OCIO team analyzed the .PST file inventory by using a new scanning and indexing tool that was able to read message-header information and then associate individual messages with EOP components on that basis. Thus, rather than rely on the name of the .PST file to attribute all messages within the file to the component in the .PST file name, OCIO was able to allocate messages within .PST files to the appropriate component with significantly more

---

<sup>4</sup> Including the pilot period message counts means that nearly 22 million messages that existed in 2005 in the EOP email message archives were not counted and associated with components in the 2005 review.

precision. Due to the unavailability and/or uncertainty of commercially available products, the new tool was developed in-house. The resulting tool—known internally as the *PST Inventory Verification & Investigation Tool* (PIVIT)—allowed OCIO to allocate messages to the appropriate EOP component. PIVIT was reviewed by a third party contractor, and independently verified as an accurate tool in both methodology and process.

While PIVIT was initially developed to associate each of the 14 million unallocated messages found in Phase 1 with a particular component, the new tool also enabled the OCIO team to more precisely allocate all of the messages in all of the .PST files in the e-mail message archive for the First Inventory Period. In effect, each of the e-mail messages was appropriately allocated to a component in Phase 2.

PIVIT counted approximately 94 million unique e-mail messages from the First Inventory Period, excluding the pilot period message counts. See supra notes 2 and 3. Including the approximately 9 million messages counted in the email message archives for pilot periods, the message count totaled approximately 103 million messages. In addition, OCIO was able to locate other repositories of e-mail messages that were not accounted for in the email message archives, such as PSTs created as a result of searches or mailbox restorations due to file corruption, which added approximately one million unique messages, for a total of approximately 95 million messages, excluding the pilot periods. (A total of approximately 104 million including the archived messages from the pilot periods). Among these 95 million messages, more than 83 million were successfully associated with EOP components on the basis of message-header information with PIVIT. (The roughly 12 million messages that could not be associated with a component were either “system” messages or “undeliverable” messages, categories of messages that by definition do not have a component of origin or receipt.) Thus, compared to the 71 million message associated by component in the 2005 effort, OCIO confirmed that 83 million messages were in the email message archives in Phase 2, excluding pilot periods. Thus, 12 million emails that were not associated by component or not counted at all in 2005 were counted and associated to components in Phase 2. Including the approximately

---

<sup>5</sup> As noted in footnotes 2 and 3, the actual number of messages counted in the email archives was approximately 103 million, including messages for the pilot periods.

9 million messages in the email archives for pilot periods, nearly 92 million messages were attributed to components at the conclusion of Phase II.<sup>6</sup>

In addition, whereas the 2005 review identified 473 “zero” message component-days, and Phase I counted 293 component days, the results of Phase II using the PIVIT tool established that only 7 component-days had zero messages. By allocating the previously unallocated messages, and allocating messages within .PST files to the appropriate component with more precision and accuracy using PIVIT, Phase II reduced the 2005 count of 473 component-days with “zero” messages to 7.

Having assigned roughly 83 million messages to specific component-days, excluding pilot periods, and having identified only 7 “zero” component-days after that process, OCIO then confronted how to analyze whether any message counts for component days was statistically low.<sup>7</sup> In order to address that question, the OCIO team believed that it needed to replace the 27-day rolling average used in the 2005 review with a more sophisticated time-series approach for assessing the e-mail message counts produced by PIVIT. OCIO engaged a recognized authority in this methodology, Dr. Nancy J. Kirkendall, to recommend a more accurate statistical approach. Dr. Kirkendall recommended the analysis of the e-mail counts using an “Auto-Regressive Integrated Moving Average” (ARIMA) model, an approach used widely for large data pools with time-series components.<sup>8</sup> It is my understanding that the ARIMA model was used to identify days or groups of days with statistically significant low e-mail counts for further

---

<sup>6</sup> On the chart attached as Exhibit 2, adding up the sum of emails from the row entitled “2008 Counts incl. pilot” (approx. 92 million) with the approximately 12 million in the Final “system” and “undeliverable” columns on the right add up to 104 million total email messages in the archives. Similarly, adding the “Total” on the right side table in light blue (approx. 95 million) with the row entitled “2008 pilot counts” (approx. 9 million) total to the 104 million messages.

<sup>7</sup> As explained above, only message counts from non-pilot periods were analyzed.

<sup>8</sup> For example, ARIMA models are used by the U.S. Bureau of the Census to analyze time series census data, by the Institute of Electrical and Electronics Engineers to predict electricity prices, and to forecast sugar cane production in India. Dr. Kirkendall’s paper, Time Series Analysis of Daily Email Counts, fully explains the technical approach that was used to develop and apply the ARIMA model that was used for the e-mail analysis. The description herein only describes the analysis methodology used for the model’s output.

investigation and possible restore from disaster recovery back-up tapes. Attached as Exhibit 3 is a true and correct copy of the “Time Series of Daily Email Counts” explaining the ARIMA methodology.<sup>9</sup>

It is my understanding from discussions with OCIO staff and access to relevant documentation that, at the conclusion of Phase II, the ARIMA model identified 76 potentially “low” component days in the e-mail message archive. As indicated in charts presented to me, Phase II accounted for approximately 95 million e-mail messages in the First Inventory Period, excluding the pilot periods. Based in part on these results, OCIO determined that the disaster recovery back up tapes should used to in order to recover any potentially missing messages. This final phase of the e-mail recovery effort was known as Phase III.

### *Phase III*

The final stages of Phase II were nearly complete prior to my arrival at OA, with the results stated in the previous paragraph. Phase III was also underway when I arrived. OA awarded a restoration-process contract to a third-party vendor in August 2008 to help OA complete the work of Phase III. In the first step of Phase III, the contractor created a copy of approximately 26,000 disaster recovery back-up tapes which were last written on dates between January 1, 2003 through August 10, 2005. The contractor created the copies in order to create a possible cache of tapes to use for any restore process.

---

<sup>9</sup> It is important to note, however, that the report was issued for application of the ARIMA model to message counts at the conclusion of Phase II. As explained below, however, additional email messages (approximately a few hundred thousand) restored at the beginning of Phase III were not counted in this report. The ARIMA model was run over the results of the message counts including the messages restored in Phase III, though Dr. Kirkendall did not produce another formal report. The report is therefore most helpful in explaining the methodology of the ARIMA model, though the numbers reported are not accurate. As could be expected in a statistical model seeking outliers, some days that had additional e-mail allocated were no longer indicated by ARIMA as low days, and new days that had not been previously indicated in the ARIMA model as low days were now indicated as such due to a change in counts for other days. In short, because the message counts had changed when the messages from the 116 .PST files were added, as described below, the statistical model predicted that some days that were

In the next step of Phase III, a subset of the backup tapes was used to restore 125 PST files, which had been identified in previous work as existing at one point, but which the team could not locate during Phase II. Specifically, based on a prior analysis, there was information that led OCIO to believe that up to 125 .PST files were not in the .PST file stores, but that existed at one time; with a number of those believed to be duplicates of existing PSTs with different file labels. All 125 PSTs were located and restored, and 9 were determined to be duplicates. The 116 resulting PSTs were indexed and de-duplicated against the index created during the Phase II work. The messages recovered from the disaster recovery back-up tapes from these 116 .PST files (approximately a few hundred thousand unique messages) were added to the e-mail message archive, and the de-duplicated counts of email messages including those restored emails was again analyzed using the ARIMA statistical model to determine if any component day counts could be considered statistically low.

Running the ARIMA model on the number of messages assigned to components resulted in 106 potentially “low” component days; 40 of these potentially “low” component days were in Federal Records Act components. There was no movement in the total number of “zero” message component-days found after the 116 PST files were restored; it remained at 7. Of those 7 component-days with “zero” message counts, the ARIMA model highlighted 4 component-days as potentially “low.” After reviewing the 106 component days that were potentially low, and considering whether explanations like snow days, the day after Thanksgiving, population of component, or other calendar impacts that were not accounted for in the ARIMA model could explain the potentially “low” counts, OCIO in consultation with counsel and OA staff selected 21 calendar days to be restored from the disaster recovery tapes by the contractor. Those 21 calendar days covered 48 of the 106 component days determined to be statistically low in the ARIMA model. Of those 48 component days, 18 were FRA component days. Also included among the 48 restored component-days are the 4 “zero” message component-days that ARIMA identified as potentially problematic. See Exhibit 2 (reflecting the 21 calendar days restored from the disaster recovery back up tapes).

---

previously “low” were no longer low, and that some days that were previously statistically “sound” were now considered low.

OCIO has now restored emails for the selected 48 component-days from the copy set of the disaster recovery back-up tapes. OCIO and the contractor are completing the analysis of the messages to determine what, if any, messages might have been found in the restored files from tapes that were not previously accounted for in the e-mail message archive. This process is time consuming because of the need to de-duplicate the data restored from the back-up tapes and to allocate the large volume of results from this restoration to the previous index of messages. In addition, quality assurance work is being done by the contractor as specified in the contract to ensure accuracy of the work. All results of this work, including all of the de-duplicated restored e-mail, and all indices made of the restored e-mail, will be transferred to NARA.

As noted in previous declarations by Theresa Payton, it is impossible to state categorically whether every e-mail ever generated or received by the EOP unclassified e-mail system is available in the disaster recovery back up tapes. More fundamentally, even after all the relevant tapes have been restored, one cannot say whether “all” the e-mails are present and accounted for without knowledge, which no one possesses, of how many and what precise e-mails there should in fact be.

Nonetheless, OA has engaged in a deliberate effort – costing over \$10 million dollars, many thousands of staff hours and the energy of OA personnel – to address the concerns initially flagged in the 2005 review. OA has completed all that reasonably may be done to restore any records that may exist that are potentially not contained in the e-mail message archives by using the disaster recovery back up tapes. That effort has resulted in a reduction in problematic “low” days from 702 component-days (2005 review) to 48 component-days. The number of “zero” message count component-days has dropped from 473 (2005 review) to 7, of which only 4 were identified by our statistical model as potentially problematic. Based on these results, OA decided to engage in the restoration of e-mail messages from back-up tapes for 48 component days. The process of adding these messages to the overall inventory and associating them to the proper components is ongoing. When complete, OA believes that it will have addressed the concerns raised by the 2005 chart and taken appropriate, reasonable steps to recover any potentially missing e-mail messages.

I declare under penalty of perjury, pursuant to 28 U.S.C. § 1746, the foregoing to be true and correct.

Executed the 20<sup>th</sup> day of January, 2009.

A handwritten signature in cursive script, reading "Stephen M. Everett", written over a horizontal line.

Stephen M. Everett  
Chief Information Officer  
Office of Administration  
Executive Office of the President