1. The verification and validation review of Exchange record keeping process

There was a gradual loss of confidence in the 2005 chart depicting component / daily message data created in the OCIO. During a rotation of duties in the Director of IS&T position in November 2006, the new Director wanted a full understanding and review of the procedures involved with Exchange record keeping. The decision was made to allocate a dedicated resource to review the processes and technologies.

A plan was developed that included performing a full review and update of the database that stored the inventory information about the email storage files. The plan also was developed to review and improve all processes related to creation, inventory, and tracking of email storage files. Through this process, technical deficiencies were discovered in the scanning tool (CMDFI) and the database (PFMS) that impacted the accuracy of the information about email storage files. Those deficiencies included count limitations, inadequate error logging, and database record issues.

Part of this overall review included the re-evaluation of the structure and documentation of the email storage inventory database. During this analysis the information within the database was compared to the 2005 chart depicting component daily message data. It was in this evaluation that the OCIO noticed inconsistencies in message counts and began losing confidence about the daily message data within the 2005 chart. Because the software used for determining the date and component information about email storage files was utilized in the creation of the database information as well as the 2005 chart, it became obvious to the leadership that there possible problems with the accuracy of the data being represented in the 2005 chart.

At this point, the decision was made to design and develop a better and more accurate tool for daily message counts per component. Because the raw data for all historical message traffic is contained within the email storage files, this new tool would provide the confidence to the leadership needed to ensure consistency in the archive.

Factors:

- The 2005 spreadsheet chart lacked adequate documentation on how it was created.
- A comparison of the 2005 chart against information obtained from the PFMS database in a documented process provided different results.
- Database used to track email storage files (PFMS) was analyzed in early 2007
  - Lacked adequate documentation
  - Had record relationship disconnects that could impacted daily message counts
  - Inconsistent, incomplete or redundant data from 2005 scan
  - Database issues prevented accurate counts of PSTs scanned
- A review of the PFMS database that contains per day message counts found what appeared to be a message count limit of 32K, while PST's sometimes contain many more.
  - Problem detected during 2007 analysis and believed to only show up on days with larger message counts.
  - Verified in source code as a limitation
- CMDFI used an output file to record what files it analyzed.

  - Manual process to import and export data
  - Import process of raw count data into the PFMS database is susceptible to human error
  - Export process of raw count data to allow analysis is susceptible to human error
  - Log files did not always accurately show error conditions (corrupt PST, skipped PSTs, etc)
  - CMDFI unable to provide daily message counts for all messages. Requires manual remediation when this issue occurs.
- CMDFI Written by Microsoft in C# and Visual Basic 6.0
  - C# chosen by developer due to rapid deployment requirements. Developer also used Visual Basic application as part of solution due to rapid deployment requirements.
  - Older technology with known limitations or bugs that could impact the accuracy of count data or error checking.
- Microsoft Access database created by contractor staff.
  - MS Access chosen due to rapid deployment requirements.

Timeline

- **November 2006:** The OCIO embarked on a project to:
  - Review and update the PST inventory tracking database.
  - Review and update inventory information for email storage files.
  - Review and improve all processes related to storage, inventory, and tracking of email storage files.

- **November 2006 – March 2007:** The OCIO began to update the tracking database, began to update inventory information for email storage files, and began to update and revise documentation to reflect improved processes for storing and tracking email storage files. During this project, OCIO staff identified the following deficiencies in the tracking database:
  - The tracking database was originally designed to identify all email storage files, not to support searches. Therefore, the database's design makes querying for information related to searches a fairly complex process.
  - Although the database contains information about all PST files, some information is incomplete or redundant, which leads to inefficiencies in the search process. (Ensuring all correct storage files are searched falls within the searching and Independent Verification and Validation stages of the search process.)

- **April 2007:** OCIO staff discovered deficiencies in the software for determining date and component information about email storage files.
  - Within a single email storage file, the software cannot provide counts of more than 32,767 individual email messages per day (a storage file may contain more messages).
  - Older email storage files may contain email messages from multiple days. In those messages, the software cannot provide a total message count of more the 65,535 individual email messages for multiple days (older storage files may contain more messages).

- For some email messages, the software does not provide a "Received" date. (OCIO staff has developed a workaround solution to determine Received dates, but this workaround introduces inefficiency to the process.)

- **April 2007:** OCIO management decided to bring in a contractor to resolve the deficiencies in the software and tracking database.

  Why OCIO decided to contract for resolution of these issues:

  - Staffing constraints would prevent the OCIO from completing this project in a timely manner.
  - Management believed it would benefit from a third-party perspective in resolving deficiencies and inefficiencies.

- **May – August 2007:** OCIO staff, with input from the OA Procurement Office, drafted a statement of work and related artifacts for the contract. These artifacts included a Limited Source Justification.
  - - OA solicited proposals in July 2007 based on the statement of work.
- **August 10, 2007:** OA awarded the contract to InfoReliance Corp. with a period of performance from August 13, 2007, to November 12, 2007.
- **September 2007:** Cleared InfoReliance staff began working on the project (mostly requirements gathering).
- **October 2007:** Project went on hold due to delay in required security clearance (name check) process.
- **December 2007:** Name checks for key contract staff completed. Project to develop new scanning tool commences.

Goals:

The OCIO intends to increase efficiency through an improved tracking database and improved tool for gleaning inventory information about email storage files.

Current processes entail several manual steps, which the contractor can automate. For example: providing statistical information about email storage files currently requires staff to take query results from the database, place the results into an Excel spreadsheet, and then verify no errors were introduced. This process is extremely time-consuming.

In addition, leveraging contractor expertise to improve the tools ultimately will lead to quicker searches and the capability to conduct more searches concurrently, regardless of complexity.

More specifically, goals include the following:

- Provide new CMDFI tool (tool for gleaning information about email storage files — provides date and quantity of information by date within a PST file) to obtain date and quantity information from newly created PST files (email storage files), and store that information in a database on a daily basis.
- Provide a new database that is designed to support searches, has a one-to-one correspondence with every existing PST file, and contains the ability to recall previous search queries.

- Provide database reporting, including daily and weekly validation and trending and quantity reports by component or components.
- Complete procedures and training on procedures and new tools.

Original deliverables and estimated delivery dates:

Deliverable	Est. Date
Acceptance testing procedures	10/16/2007
Production software code and documentation	11/07/2007
Begin scans of all PST files and population of database	11/07/2007
Production implementation and support	11/19/2007

2. The review of the 2005 inventory summary spreadsheet (November 2007)

In the fall of 2005, the Office of Administration, Office of the CIO (OCIO) performed an inventory of the files containing the record copies of the electronic mail created or received in the Microsoft Exchange environment from 2003 thru October 2005 for all of the components of the Executive Office of the President. The inventory process, using a tool developed by Microsoft (CMDFI) for the OCIO and performed by the Message Store Team (see appendix A of this report for the Team Charter) is detailed in the attached flow chart (Appendix B).

The results of the inventory process were recorded in a Master Inventory List and a summary of those results was prepared by a member of the OCIO staff who is no longer an employee and therefore was not available to answer questions relating to the "Red and Yellow" spreadsheet or chart. The 2005 chart was created and maintained without the assistance of any current member of the OCIO. There is no formal documentation and therefore it is not possible for the OCIO to answer questions today on rational for formulas or data representation. This inventory summary report was created in Microsoft Excel and is a workbook consisting of fifteen pages (tabs). At the time of its creation, the 2005 chart was to provide a trend analysis used to determine specific days by component which, according to the CMDFI tool, appeared to have had abnormal email volume.

The OCIO later determined that the tool used to perform the actual count of the messages contained in the stored email files was flawed and therefore the volume of email counted per day for each component can not be considered as accurate.

The OCIO has reviewed the 2005 chart and examined the formulas contained in the Microsoft Excel workbook to identify the source of key data elements which determined the following:

1. The dates (per component) which were subject to the trend analysis
2. The actual daily message count, as captured by the flawed inventory tool (CMDFI)
3. The expected daily volume of email per component
4. The email volume listed under "Issues"
5. The identification of "low" and "zero" days

The review was not able to answer questions as to why the author chose to use the logic employed in the creation of these formulas. It will merely report on the method used via an examination of the formulas embedded in the summary spreadsheet's cells which resulted in specific dates per component being reported as problematic.

Factors:

- It is not known how the author was able to determine the count of emails in the PST files listed on the "Issues" page of the workbook.
- Thresholds
  - Low Threshold 30%: Why this percentage was determined as the low threshold is unknown.
  - High Threshold 300%: Why this percentage was determined as the high threshold is unknown.
- The source number 864 in cell D6 (and all subsequent values in this column) of the WHO page appears to be a manual entry. There is no formula or table reference.
  - The Message Store Team provided the results of the CMDFI tool to the Architecture and Engineering (A&E) Directorate. A contractor in A&E created

and maintained a database (PFMS) where, per the understanding of the Message Store Team, the results they provided from the CMDFI tool were stored.
  - It is not known how the author took the data from the A&E database and populated each source cell on the component pages. It is not known if the data the author used to populate the summary report routinely matches the results delivered by the Message Store Team using the CMDFI tool.

- It is not known why the author included a column titled "Issues" under the Predicted Daily Values since it would not be possible to predict the number of messages that would be the result of issues for any specific date and component combination.
- It is not known why the author did not bring the number of days each component had a high count (greater than 300% of the predicted message count) forward to the Summary page of the workbook.
  - For example, the WHO had a total number of 21 days which exceeded 300% of the predicted number of messages however this information is not presented in the table at the top of the Summary page.
- It is not known what method (manual entry, cut and paste, etc) was used populate the cells in column D for each of the component pages. It is not known if the author consistently used the results of the inventory which were given to the A&E Directorate by the Message Store team. It is not known how the author extracted the actual message counts by day and component from the database the A&E Directorate created and maintained to store the results of the inventory project.
- It is not known why the author chose to base the predicted value on the average message count from work days or non-work days rather than a statistically derived value that takes the day of the week into account.
- It is not known why the author selected 27 as the number of consecutive days to base the Work Day Average. The F column only includes actual message counts from days coded as WD.
  - Example, since February 1, 2003 was a Saturday there is no value listed in column F for that date. The date, even though it is included in the 27 day range, adds nothing to the subtotal of messages which is then divided by 27 to produce the Work Day Average value in column G.
- As noted in the section describing the third view of the Summary page, in some cases for non-WHO components there are actual message counts listed on the component pages of the workbook, but the value in the "Include in Analysis" column (J) is "N". It is not known why the author did not include dates which list a value of some number of actual messages in the analysis for those days or those components.
  - For example see the OVP page cell J52. An "N" is entered in J52 (Include in Analysis) but there is an actual message count of 64 listed in cell D52 (Actual OVP Msg Count) for February 14, 2003.
- It is not known why the author did not bring the number of days each component had a high count (greater than 300% of the predicted message count) forward to the Summary page of the workbook.

  - For example, the WHO had a total number of 21 days which exceeded 300% of the predicted number of messages however this information is not presented in the table at the top of the Summary page.

- OCIO is uncertain what the exact process was to determine if any given pst file should be handled as an "Issue File".
- OCIO is uncertain at this time what analysis was made to address the files identified as being "Issue" files.
- OCIO has determined that the Actual Issue Message Count was not included in the daily message count which was then compared to the predicted message count on the Summary page. Therefore any messages contained in the Issues Files did not impact the "red/yellow" day flags.