

Analysis of the 2005 PST File Inventory Summary Report
Office of Administration
Office of the CIO
November 2007



FINAL REPORT

I. OVERVIEW:

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 the OCIO's former Director of Architecture and Engineering, [REDACTED]. [REDACTED] is no longer a member of the OCIO and therefore was not available to answer questions relating to the summary report he created, often referred to as the "Red and Yellow" spreadsheet or chart. [REDACTED] created and maintained this summary report without the assistance of any current member of the OCIO. It is therefore not possible for the OCIO to answer questions today on why [REDACTED] chose to create the formulas or represent the data as he has done in this summary report. This inventory summary report was created in Microsoft Excel and is a workbook consisting of fifteen pages (tabs). At the time of its creation by [REDACTED] the summary report was to provide a trend analysis which [REDACTED] used to determine specific days by component which, according to the CMDFI tool, appeared to have had low 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 inventory summary created and maintain solely by [REDACTED] in Microsoft's Excel spreadsheet tool. This report details the findings of that review. The review was conducted by the OCIO's Data Architect. The Data Architect examined the formulas contained in the Microsoft Excel workbook to identify the source of key data elements which determined the following:

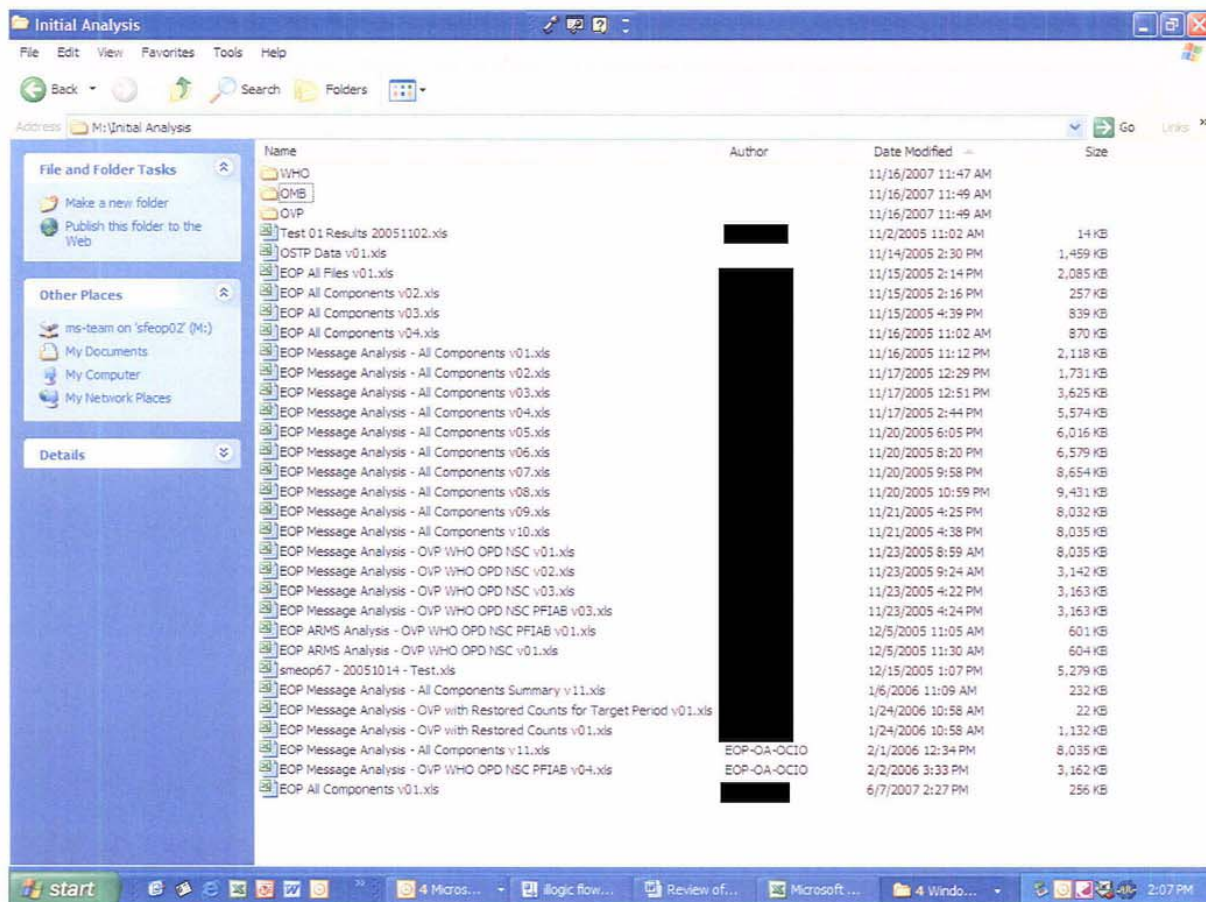
1. The dates (per component) which were subject to the trend analysis
2. The actual daily message count, as captured by the flawed inventory tool (CMDFI)
3. The expected daily volume of email per component
4. The email volume listed under "Issues"
5. The identification of "low" and "zero" days

Screen shots from the final version of the actual inventory summary workbook will be provided throughout this report to better explain the formulas that were embedded in the spreadsheet by its author, Mr. McDevitt. This report will not be 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.



Analysis of the 2005 PST File Inventory Summary Report
Office of Administration
Office of the CIO
November 2007

II. DESCRIPTION OF THE MICROSOFT EXCEL WORKBOOK:



The OCIO has carefully retained the original and all subsequent drafts which lead up to the final version of the inventory summary workbook which was titled “EOP Message Analysis – All Components v11.xls”. This report will focus solely on this final version created by [Redacted] on 01/06/2006. Note: the workbook titled “EOP Message Analysis – All Components Summary v11.xls” created by EOP-OA-OCIO on 2/1/2006 contains the summary page only and therefore could not be used for the purpose of this review.

The subject workbook contains 15 pages. They are located left to right and named as follows. A brief description of each page is included below:

1. **Summary-** Daily totals per component, including actual message count (determined from the flawed inventory tool, CMDFI), expected message count, issues, the predicted daily values per component, and Y/N indicators which determined what days were to be considered in the analysis per component. Please note three views of this page are needed in this report in order to provide a view of all of the columns contained in the Summary page which are utilized in formulas throughout the workbook.
2. **WHO** (White House Office) the column headings are as follows
 - Daily actual count of WHO email only (based on the flawed inventory tool, CMDFI)



Analysis of the 2005 PST File Inventory Summary Report
Office of Administration
Office of the CIO
November 2007

- Daily predicted WHO message count
 - Work Day count
 - Work Day Average
 - Non-work Day Count
 - Non-work Day average
 - An indicator Y/N to determine if the date should be included in the analysis
 - Zero count
 - Low count
 - High count
3. **OVP** (Office of the Vice President)
Daily count of OVP email presented in the same format as described above for WHO
 4. **OPD** (Office of Policy Development)
Daily count of OPD email presented in the same format as described above for WHO
 5. **NSC**(National Security Council)
Daily count of NSC email presented in the same format as described above for WHO
 6. **PFIAB** (President's Foreign Intelligence Advisory Board)
Daily count of PFIAB email presented in the same format as described above for WHO
 7. **CEA** (Council of Economic Advisors)
Daily count of CEA email presented in the same format as described above for WHO
 8. **CEQ** (Council on Environmental Quality)
Daily count of CEQ email presented in the same format as described above for WHO
 9. **OMB** (Office of Management and Budget)
Daily count of OMB email presented in the same format as described above for WHO
 10. **ONDCP** (Office of the National Drug Control Policy)
Daily count of ONDCP email presented in the same format as described above for WHO
 11. **OSTP** (Office of Science and Technology Policy)
Daily count of OSTP email presented in the same format as described above for WHO
 12. **USTR** (United States Trade Representative)
Daily count of USTR email presented in the same format as described above for WHO
 13. **OA** (Office of Administration)
 14. Daily count of OA email presented in the same format as described aCMDFI tool because of format or storage issue, presented in the same format as described above for WHO

Note: It is not known how [REDACTED] was able to determine the count of emails in the pst files he listed on the "Issues" page of the workbook.

15. **Placeholder**

- Comp (Component abbreviation)
- File Name
- Received Date
- Email Count



Analysis of the 2005 PST File Inventory Summary Report
Office of Administration
Office of the CIO
November 2007

III a. SUMMARY PAGE ANALYSIS: View one (Main Summary section)

The screenshot shows an Excel spreadsheet with the following data tables:

Summary - Messages per Day
Last Update: 21-NOV-2005 16:38

	WHO	OVP	OPD	NSC	PFIAB	CEA	CEQ	OMB	ONDCP	OSTP	USTR	OA	Total
Total Days with Zero Msg	12	16	11	47	20	103	81	59	20	15	73	16	473
Total Days with Low Msg	28	30	7	9	14	29	5	10	24	39	10	24	229

Date	Day	Day Type	WHO	OVP	OPD	NSC	PFIAB	CEA	CEQ	OMB	ONDCP	OSTP	USTR	OA	Issue Files	Actual Msg Count Total	Predicted Msg Count Total
01-Jan-03	Wed	FH	864											1,125	0	1,989	1,989
02-Jan-03	Thu	WD	10,542											8,583	0	19,125	19,125
03-Jan-03	Fri	WD	8,004											6,987	0	14,991	14,991
04-Jan-03	Sat	WE	1,470											1,173	0	2,643	1,989
05-Jan-03	Sun	WE	1,572											968	0	2,538	2,316
06-Jan-03	Mon	WD	1,586											8,605	0	10,371	17,058
07-Jan-03	Tue	WD	7,824											9,441	0	17,265	14,829
08-Jan-03	Wed	WD	12,280											11,241	0	23,521	15,438
09-Jan-03	Thu	WD	12,640											10,635	0	23,275	17,055
10-Jan-03	Fri	WD	14,860											1,935	0	16,595	18,091
11-Jan-03	Sat	WE	1,168											0	0	1,168	2,390
12-Jan-03	Sun	WE	2,292											0	0	2,292	2,085
13-Jan-03	Mon	WD	15,912											2,148	0	18,060	17,878
14-Jan-03	Tue	WD	17,480											9,897	0	27,357	17,900
15-Jan-03	Wed	WD	9,308											10,914	0	20,220	18,951
16-Jan-03	Thu	WD	9,850											10,539	5	20,189	19,078

Page Title Information indicates November 21, 2005 at 16:38 as the date and time of the last update. However the metadata for this workbook indicates [REDACTED] created this final version of the workbook on January 6, 2006. The significance of the November 21, 2005 date for this workbook is unknown at this time.

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.

Total Days with Zero Msg – The number of days, per component, for which no messages were found using the Microsoft developed tool, CMDFI. The combined total of all zero days is also given (sum of cells D6 thru O6 is the formula embedded in the total column). The number of zero days listed on this summary page is the sum of the zero days found on the individual component pages. (How the zero days were determined will be described under the WHO page analysis later in this report). The “zero days” number for each component is the result of a formula that takes the sum of the total of days listed in column K “Zero Count” of each



Analysis of the 2005 PST File Inventory Summary Report
Office of Administration
Office of the CIO
November 2007

component page of the workbook. The sum is determined by a formula that adds the total number of zero days in column K and places the total in cell M2 of each component page of the workbook. The formula that is embedded here on the summary page simply brings the value listed in cell M2 from each component page to the cell on this page that for each component (cells D6 thru O6).

Total Days with Low Msg – the number of days, per component, for which the number of messages found using the CMDFI tool were less than 30% of the predicted amount. A combined total of all low days is also given (sum of cells E6 thru E6 is the formula embedded in the total column). The number of low days listed on this summary page is the sum of the low days found on the individual component pages. (How the low days were determined will be described under the WHO page analysis later in this report). The “low days” number for each component is the result of a formula that takes the sum of the total of days listed in column L “Low Count” of each component page of the workbook. The sum is determined by a formula that adds the total number of low days in column L and places the total in cell M3 of each component page of the workbook. The formula that is embedded here on the summary page simply brings the value listed in cell M3 from each component page to the cell on this page that for each component (cells E6 thru E6).

Note: It is unknown why [REDACTED] did not provide a summary table for the dates that exceeded the 300% High Threshold. Therefore it is unknown what the impact of the high message volume days had on the analysis [REDACTED] performed.

Date – A listing of the actual calendar date from January 1, 2003 through October 31, 2005. There are no table references or indication of a copy. Why the analysis end date of October 31, 2005 was selected is unknown. The Message Store Team Charter (see appendix A of this report) did not indicate any timeframe for the inventory project.

Day – A listing of the abbreviation for the actual day of the week for each calendar date. There are no tables referenced. The value in this column for all rows appears to be a manual entry.

Day Type – One of three choices are found in this column: FH assumed to be Federal Holiday, WD assumed to be Work Day, and WE assumed to be Week End. There are no tables referenced. The value in this column for all rows appears to be a manual entry.

Actual Message Count – Each component is listed across a series of 12 columns (D thru O). The value found in each cell, for each component, for each day is pulled from the component page of the workbook. For example the value “864” listed for WHO on January 1, 2003 is the result of a formula embedded in cell D13 which is using the value in cell D6 (864) from the WHO page of the workbook as the source for the information presented in the Summary page.

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



Analysis of the 2005 PST File Inventory Summary Report
Office of Administration
Office of the CIO
November 2007

Issue Files - The source location, per the formula embedded in the cells in the column titled "Issues Files" of the Summary page, are the values found in the Issues page of the workbook in column D titled "Actual Issues MSG per Day". The values listed in column D of the Issues page appear to be manually entered. There is no formula, nor any table reference on the Issue page for this data.

Actual Msg Count Total - The values listed in this column for each date are the sum of all of the components actual messages listed in 12 columns (D thru O). The formula is a simple sum action of columns D thru O for each row.

Predicted Msg Count Total - The values listed in this column for each date are the sum of all of the components predicted message counts found in columns T thru AE of the Summary page. How these predicted values were derived will be provided under the second view of the Summary page below. The formulas in column R "Predicted Msg Count Total" are a simple sum action of Columns T thru AE for each row.



Analysis of the 2005 PST File Inventory Summary Report
Office of Administration
Office of the CIO
November 2007

IIIb. SUMMARY PAGE ANALYSIS: View two (Predicted Daily Values section)

	WHO	OVP	OPD	NSC	PFIAB	CEA	CEQ	OMB	ONDCP	OSTP	USTR	OA	Issues
13	864											1,125	
14	10,542											8,583	
15	10,542											8,583	
16	864											1,125	
17	1,167											1,149	
18	9,273											7,785	
19	6,704											8,125	
20	6,984											8,454	
21	8,043											9,011	
22	8,809											9,282	
23	1,302											1,088	
24	1,269											816	
25	9,645											8,232	
26	10,429											7,472	
27	11,210											7,741	
28	11,019											8,059	
29	10,895											8,284	
30	1,473											853	
31	1,393											855	
32	1,403											979	
33	11,184											8,503	
34	11,941											8,918	
35	12,025											9,385	
36	12,630											11,787	
37	1,663											1,210	
38	1,683											1,266	
39	12,919											13,981	
40	13,048											13,793	
41	13,316											13,665	
42	13,592											13,574	

Predicted Daily Values (columns T thru AF) – The source of the information listed for each date and each component is derived from two places in the workbook. Note the formula for cell T13 is =IF(AI13="Y",WHO!\$E8,"") in the screen shot above. Cell AI13 is contained in the Summary page and will be described in detail in the third view of the Summary page below. Essentially this part of the formula is stating that if the date for this component is to be included in the analysis proceed with bringing forward the predicted message count. The predicted message count for the date is found on the component page of the workbook. In this example the "WHO!\$E8" portion of the formula is taking the value found in cell E8 of the WHO page of the workbook. If the cell AI13 of the Summary page had contained an "N" the formula would end there and no information would be brought forward from the component page.

Note: It is not known why [REDACTED] 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.



Analysis of the 2005 PST File Inventory Summary Report
 Office of Administration
 Office of the CIO
 November 2007

IIIc. SUMMARY PAGE ANALYSIS: View three (Included in Analysis section)

	WHO	OVP	OPD	NSC	PFIAB	CEA	CEQ	OMB	ONDCP	OSTP	USTR	OA	Issues
13	Y	N	N	N	N	N	N	N	N	N	N	Y	
14	Y	N	N	N	N	N	N	N	N	N	N	Y	
15	Y	N	N	N	N	N	N	N	N	N	N	Y	
16	Y	N	N	N	N	N	N	N	N	N	N	Y	
17	Y	N	N	N	N	N	N	N	N	N	N	Y	
18	Y	N	N	N	N	N	N	N	N	N	N	Y	
19	Y	N	N	N	N	N	N	N	N	N	N	Y	
20	Y	N	N	N	N	N	N	N	N	N	N	Y	
21	Y	N	N	N	N	N	N	N	N	N	N	Y	
22	Y	N	N	N	N	N	N	N	N	N	N	Y	
23	Y	N	N	N	N	N	N	N	N	N	N	Y	
24	Y	N	N	N	N	N	N	N	N	N	N	Y	
25	Y	N	N	N	N	N	N	N	N	N	N	Y	
26	Y	N	N	N	N	N	N	N	N	N	N	Y	
27	Y	N	N	N	N	N	N	N	N	N	N	Y	
28	Y	N	N	N	N	N	N	N	N	N	N	Y	
29	Y	N	N	N	N	N	N	N	N	N	N	Y	
30	Y	N	N	N	N	N	N	N	N	N	N	Y	
31	Y	N	N	N	N	N	N	N	N	N	N	Y	
32	Y	N	N	N	N	N	N	N	N	N	N	Y	
33	Y	N	N	N	N	N	N	N	N	N	N	Y	
34	Y	N	N	N	N	N	N	N	N	N	N	Y	
35	Y	N	N	N	N	N	N	N	N	N	N	Y	
36	Y	N	N	N	N	N	N	N	N	N	N	Y	
37	Y	N	N	N	N	N	N	N	N	N	N	Y	
38	Y	N	N	N	N	N	N	N	N	N	N	Y	
39	Y	N	N	N	N	N	N	N	N	N	N	Y	
40	Y	N	N	N	N	N	N	N	N	N	N	Y	
41	Y	N	N	N	N	N	N	N	N	N	N	Y	
42	Y	N	N	N	N	N	N	N	N	N	N	Y	

Included in Analysis (columns AI thru AU) – The values listed in each cell for each component indicate whether or not the calendar date (which is not shown but is the same as the first view of the Summary page and begins on row 13 with January 1, 2003) should be included in the message analysis or not. A “Y” indicates the date is to be included and an “N” indicates it is not to be included. The source of the “Y” or “N” listed for each date and each component is the value that was manually entered on the component pages of the workbook. For example, the formula that returned a value of “Y” in cell AI13 (WHO for January 1, 2003) is derived from the manually entered “Y” in cell J8 of the WHO page of the workbook. The formula for AI13 is =WHO!\$J8.

Note: 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 [redacted] did not include those dates which list a value of some number of actual messages in the analysis for those days or those components.



Analysis of the 2005 PST File Inventory Summary Report
Office of Administration
Office of the CIO
November 2007

IV. COMPONENT SPECIFIC PAGE ANALYSIS:

(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	(J)	(K)	(L)	(M)
Date	Day	Day Type	Actual VHO Msg per Day	Predicted VHO Msg per Day	Work Day Count	Work Day Avg	Non-Work Day Count	Non-Work Day Avg	Include in Analysis	Zero Count	Low Count	High Count
01-Jan-03	Wed	FH	864	864		10.542		864	Y	0	0	0
02-Jan-03	Thu	WD	10.542	10.542	10.542	10.542		864	Y	0	0	0
03-Jan-03	Fri	WD	8.004	10.542	8.004	10.542		864	Y	0	0	0
04-Jan-03	Sat	WE	1.470	864		9.273	1.470	864	Y	0	0	0
05-Jan-03	Sun	WE	1.572	1.167		9.273	1.572	1.167	Y	0	0	0
06-Jan-03	Mon	WD	1.566	9.273	1.566	9.273		1.302	Y	0	1	0
07-Jan-03	Tue	WD	7.824	6.704	7.824	6.704		1.302	Y	0	0	0
08-Jan-03	Wed	WD	12.280	6.884	12.280	6.884		1.302	Y	0	0	0
09-Jan-03	Thu	WD	12.640	8.043	12.640	8.043		1.302	Y	0	0	0
10-Jan-03	Fri	WD	14.660	8.809	14.660	8.809		1.302	Y	0	0	0
11-Jan-03	Sat	WE	1.168	1.302		9.645	1.168	1.302	Y	0	0	0
12-Jan-03	Sun	WE	2.292	1.269		9.645	2.292	1.269	Y	0	0	0
13-Jan-03	Mon	WD	15.912	9.645	15.912	9.645		1.473	Y	0	0	0
14-Jan-03	Tue	WD	17.460	10.429	17.460	10.429		1.473	Y	0	0	0
15-Jan-03	Wed	WD	9.308	11.210	9.308	11.210		1.473	Y	0	0	0
16-Jan-03	Thu	WD	9.650	11.019	9.650	11.019		1.473	Y	0	0	0
17-Jan-03	Fri	WD	14.364	10.895	14.364	10.895		1.473	Y	0	0	0
18-Jan-03	Sat	WE	990	1.473		11.184	990	1.473	Y	0	0	0
19-Jan-03	Sun	WE	1.464	1.393		11.184	1.464	1.393	Y	0	0	0
20-Jan-03	Mon	FH	3.486	1.403		11.184	3.486	1.403	Y	0	0	0
21-Jan-03	Tue	WD	21.027	11.184	21.027	11.184		1.663	Y	0	0	0
22-Jan-03	Wed	WD	13.115	11.941	13.115	11.941		1.663	Y	0	0	0
23-Jan-03	Thu	WD	21.093	12.025	21.093	12.025		1.663	Y	0	0	0
24-Jan-03	Fri	WD	17.268	12.630	17.268	12.630		1.663	Y	0	0	0
25-Jan-03	Sat	WE	1.842	1.663		12.915	1.842	1.663	Y	0	0	0

Component Subtotals (and their descriptions)

Low Threshold (set at 30% of predicted message count per day. Reason unknown why 30% was selected for this value)

High Threshold (set at 300% of predicted message count per day. Reason unknown why 300% was selected for this value)

Zero Day Count – the subtotal value for each component is derived from a simple formula which is =SUM(K8:K1042). This formula adds the value listed in each row for column K titled “Zero Count”. The subtotal listed in cell M2 on each component page is brought forward to the Summary page and listed in the table at the top of the page. (See description provided in the first view of the Summary page above).

Low Day Count - the subtotal value for each component is derived from a simple formula which is =SUM(L8:L1042). This formula adds the value listed in each row for column L titled “Low Count”. The subtotal listed in cell M3 on each component page is brought forward to the Summary page and listed in the table at the top of the page. (See description provided in the first view of the Summary page above).



Analysis of the 2005 PST File Inventory Summary Report
Office of Administration
Office of the CIO
November 2007

High Day Count - the subtotal value for each component is derived from a simple formula which is =SUM(M8:M1042). This formula adds the value listed in each row for column M titled "High Count". The subtotal listed in cell M4 on each component page.

Note: *It is not known why [REDACTED] 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.*

Column Headers (and their descriptions):

Date - the sequential calendar dates beginning January 1, 2003 entered directly into the spreadsheet

Day - abbreviated day of the week entered directly into the spreadsheet)

Day Type - It is assumed that FH stands for "Federal Holiday", WD stands for "Work Day" and WE stands for "Weekend"

Actual WHO Msg per Day – Column D

Rows: There is no indication of a formula or table reference used to populate each cell with the actual number of messages per day. The Message Store team inventory process is detailed in Appendix B of this report. The information that resulted from the inventory was added to a database created and maintained by the Architecture and Engineering Directorate of the OCIO.

Note: *It is not known what method (manual entry, cut and paste, etc) [REDACTED] used populated the cells in column D for each of the component pages. It is not known if he consistently used the results of the inventory which were given to the A&E Directorate by the Message Store team. It is not know how he 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.*

Predicted WHO Msg per Day – Column E

Rows: =IF(C8="WD",G8,I8) where the number "8" is incremented in a series by 1 for each subsequent row.

The formula noted above, checks the value of the contents in each row of column C for the indicator "WD" (OCIO assumes this stands for Work Day). If column C contains the value of "WD" for the day being a Week Day then the predicted message count is used as the value from the "Work Day Avg" column. If column C does not contain "WD" then the value from "Non-Work Day Avg" column is used. See below for the description on how formulas in the "Work Day Avg" and "Non-Work Day Avg" columns function.

Note: *It is not known why [REDACTED] 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.*

Note: *It should be noted that on [REDACTED] summary report, the total number of actual messages is 70,977,120 and the total number of the result of his formula to predict what the total message count should is 69,838,614. This indicates that the OCIO has 1,138,508 messages more than his "predicted" total.*

[REDACTED]

Analysis of the 2005 PST File Inventory Summary Report
Office of Administration
Office of the CIO
November 2007

Work Day Count – Column F

Header cell does not contain a formula

Rows: =IF(\$C8="WD",D8,"")

If the "C" column contains the manually entered value of "WD" then copy the value in the "D" column and if not then put a missing (empty cell) in the value of the cell.

Work Day Avg – Column G

Row 8: =G10 (which equals the actual message count for WHO on January 2, 2003)

Row 9: =G10 (which equals the actual message count for WHO on January 2, 2003)

Row 10 thru 35: =AVERAGE(F8:F9) increment the last value "F9" by 1 for each subsequent row of the prior work day message counts (Averaging January 1 and January 2 to January 28 to use up to the prior 27 days)

Row 36: =AVERAGE(F8:F35) note the change in the formula, results are an average of the prior 27 days with no exceptions for non-work days or holidays)

Note: It is not known why [REDACTED] 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.

Non-Work Day Count – Column H

Rows: =IF(\$C8<>"WD",D8,"") formula remains the same for the entire column just as the Work Day Count column.

Non-Work Day Avg – Column I

Row 8: =AVERAGE(H7:H8) which takes the average of the Header and the first row of this column.

Row 9 thru 35: =AVERAGE(H8:H9) increment the last value "H8" by 1 for each subsequent row

Row 36 =AVERAGE(H8:H35) note the change in the formula, results are an average of the prior 27 days with no exceptions for work days)

Note: It is not known why [REDACTED] would use the header in his formula. Or why he selected 27 as the number of consecutive days to base the Non-Work Day Average. The H column only includes actual message counts from days coded as WE or FH. Example, since January 2, 2003 was a work day there is no value listed in column H for that date. The date, even though it is included in the 27 days range, adds nothing to the subtotal of messages which is then divided by 27 to produce the Non-Work Day Average value in column I.

Include in Analysis – Column J

There is no table reference or formula embedded in the cells located in this column. It is unknown what method [REDACTED] used to populate these cells with the values "Y" and "N". The value found in this column for each date on each component's page is brought forward to the Summary page (see third view of the Summary page described above). An "N" entered here on the component page would therefore be transferred to the Summary page and the cell for that component on that date in the first view of the Summary page would be color coded grey with no values provided for analysis.



Analysis of the 2005 PST File Inventory Summary Report
Office of Administration
Office of the CIO
November 2007

Note: 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 [REDACTED] 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.

Zero Count – Column K =IF(\$J8="Y",IF(\$D8=0,1,0),0)

This formula would result in a "1" (indicating the date is a zero count date) only if the J column (Include in the Analysis) contains a Y and there is a value of zero in the D column (Actual Msg per Day). The resulting "1" in the Zero Count column is then summed for total number of Zero Days for each component.

Low Count – Column L =IF(\$J8="Y",IF(\$D8<>0,IF(\$D8<\$E8*\$H\$2,1,0),0),0)

This formula would result in a "1" (indicating the date is a low count date) only if the J column (Include in the Analysis) contains a Y and there is a value in the D column (Actual Issues Msg per Day) other than a zero and less than the low day amount (E column times the 30%). The resulting "1" in the Low Count column is then summed for total number of Low Days for each component.

High Count – Column M =IF(\$J8="Y",IF(\$D8>\$E8*\$H\$3,1,0),0)

This formula would result in a "1" (indicating the date is a high count date) only if the J column (Include in the Analysis) contains a Y and there is a value in the D column (Actual Issues Msg per Day) other than a zero and greater than the high day amount (E column times the 300%). The resulting "1" in the High Count column is then summed for total number of High Days for each component.

Note: It is not known why [REDACTED] 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.

Analysis of the 2005 PST File Inventory Summary Report
Office of Administration
Office of the CIO
November 2007

V. ISSUES PAGE ANALYSIS

Date	Day	Day Type	Actual Issues Msg per Day	Predicted Issues Msg per Day	Work Day Count	Work Day Avg	Non-Work Day Count	Non-Work Day Avg	Include in Analysis	Zero Count	Low Count	High Count
01-Jan-03	Wed	FD	0	0	0	0	0	0	N	0	0	0
02-Jan-03	Thu	WD	0	0	0	0	0	0	N	0	0	0
03-Jan-03	Fri	WD	0	0	0	0	0	0	N	0	0	0
04-Jan-03	Sat	WE	0	0	0	0	0	0	N	0	0	0
30-Apr-03	Wed	WD	2,097	2,097	27	2,097	27	0	N	0	0	0
01-May-03	Thu	WD	8,550	8,550	107	8,550	107	0	N	0	0	0
02-May-03	Fri	WD	7,802	7,802	532	7,802	532	0	N	0	0	0
03-May-03	Sat	WE	916	916	912	916	912	0	N	0	0	0
04-May-03	Sun	WE	568	568	912	568	912	0	N	0	0	0
05-May-03	Mon	WD	9,326	9,326	914	9,326	914	235	N	0	0	0
06-May-03	Tue	WD	10,182	10,182	1,381	10,182	1,381	235	N	0	0	0
07-May-03	Wed	WD	9,782	9,782	1,860	9,782	1,860	235	N	0	0	0
08-May-03	Thu	WD	17,574	17,574	2,379	17,574	2,379	235	N	0	0	0
09-May-03	Fri	WD	11,091	11,091	3,277	11,091	3,277	235	N	0	0	0
10-May-03	Sat	WE	1,266	1,266	3,832	1,266	3,832	235	N	0	0	0
11-May-03	Sun	WE	1,108	1,108	3,941	1,108	3,941	235	N	0	0	0
12-May-03	Mon	WD	11,489	11,489	3,832	11,489	3,832	532	N	0	0	0
13-May-03	Tue	WD	11,976	11,976	4,405	11,976	4,405	532	N	0	0	0
14-May-03	Wed	WD	10,894	10,894	5,004	10,894	5,004	532	N	0	0	0
15-May-03	Thu	WD	10,028	10,028	5,549	10,028	5,549	532	N	0	0	0
16-May-03	Fri	WD	12,003	12,003	6,050	12,003	6,050	532	N	0	0	0
17-May-03	Sat	WE	1,705	1,705	6,650	1,705	6,650	532	N	0	0	0
18-May-03	Sun	WE	1,788	1,788	6,650	1,788	6,650	532	N	0	0	0
19-May-03	Mon	WD	14,717	14,717	6,650	14,717	6,650	969	N	0	0	0
20-May-03	Tue	WD	15,529	15,529	7,386	15,529	7,386	969	N	0	0	0
21-May-03	Wed	WD	15,479	15,479	8,183	15,479	8,183	969	N	0	0	0

The Issues page of the workbook follows the same format as the component pages. All columns and rows contain the same formulas as described for the WHO page above. The value found in the “Actual Issues Msg per Day” column is copied from the Issues page of the workbook to the Summary page of the workbook.

The difference in the use of the formulas contained on this page results from all rows on the Issues page for column J (Include in Analysis) equals “N”, therefore all Zero, Low, and High rows report a “0”.

Zero Count – Column K =IF(\$J8="Y",IF(\$D8=0,1,0),0)

This formula would result in a “1” (indicating the date is a zero count date) only if the J column (Include in the Analysis) contains a Y and there is a value in the D column (Actual Issues Msg per Day) other than a zero. All rows on this page of the workbook contain an N in the J column (include in the Analysis). This results in no days on the Issues page being identified as “Zero” days.

Low Count – Column L =IF(\$J8="Y",IF(\$D8<>0,IF(\$D8<=\$E8*\$H\$2,1,0),0),0)



Analysis of the 2005 PST File Inventory Summary Report
Office of Administration
Office of the CIO
November 2007

This formula would result in a “1” (indicating the date is a low count date) only if the J column (Include in the Analysis) contains a Y and there is a value in the D column (Actual Issues Msg per Day) other than a zero. If other than a zero and less than the low day amount (E column times the 30%) then it will be counted as a low day. All rows on this page of the workbook contain an N in the J column (Include in the Analysis). This results in no days on the Issues page being identified as “Low” days.

High Count – Column M =IF(\$J8="Y",IF(\$D8>\$E8*\$H\$3,1,0),0)

This formula would result in a “1” (indicating the date is a high count date) only if the J column (Include in the Analysis) contains a Y and there is a value in the D column (Actual Issues Msg per Day) other than a zero. If other than a zero and greater than the high day amount (E column times the 300%) then it will be counted as a high day. All rows on this page of the workbook contain an N in the J column (include in the Analysis). This results in no days on the Issues page being identified as “High” days.

Note: It is unknown why [REDACTED] created this Issues page. The only data from the Issues page utilized elsewhere in the workbook is on the Summary page in column P (Issues File). Column P of the summary page includes a formula that brings forward the value from column D (Actual Issues Msg per Day) of the Issues page.

Note: OCIO is uncertain what the exact process was to determine if any given pst file should be handled as an “Issue File”.

Note: OCIO is uncertain at this time what analysis was made to address the files identified as being “Issue” files.

Note: 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.

Note: It is not known why [REDACTED] would attempt to determine an amount of predicted issues messages. The predicted amount of issues is not used anywhere else in the workbook.



Analysis of the 2005 PST File Inventory Summary Report
 Office of Administration
 Office of the CIO
 November 2007

VI. PLACEHOLDER PAGE ANALYSIS

Comp	File Name	Received Date	Email Count
DP	Date Placeholder	1/1/2003	0
DP	Date Placeholder	1/2/2003	0
DP	Date Placeholder	1/3/2003	0
DP	Date Placeholder	1/4/2003	0
DP	Date Placeholder	1/5/2003	0
DP	Date Placeholder	1/6/2003	0
DP	Date Placeholder	1/7/2003	0
DP	Date Placeholder	1/8/2003	0
DP	Date Placeholder	1/9/2003	0
DP	Date Placeholder	1/10/2003	0
DP	Date Placeholder	1/11/2003	0
DP	Date Placeholder	1/12/2003	0
DP	Date Placeholder	1/13/2003	0
DP	Date Placeholder	1/14/2003	0
DP	Date Placeholder	1/15/2003	0
DP	Date Placeholder	1/16/2003	0
DP	Date Placeholder	1/17/2003	0
DP	Date Placeholder	1/18/2003	0
DP	Date Placeholder	1/19/2003	0
DP	Date Placeholder	1/20/2003	0
DP	Date Placeholder	1/21/2003	0
DP	Date Placeholder	1/22/2003	0
DP	Date Placeholder	1/23/2003	0
DP	Date Placeholder	1/24/2003	0
DP	Date Placeholder	1/25/2003	0
DP	Date Placeholder	1/26/2003	0
DP	Date Placeholder	1/27/2003	0
DP	Date Placeholder	1/28/2003	0
DP	Date Placeholder	1/29/2003	0

Comp: Unknown what this column was used for in the workbook. No reference found in any formula on any other page. All rows contain “DP” in this column

File Name: Unknown what this column was used for in the workbook. No reference found in any formula on any other page. All rows contain “Date Placeholder” in this column

Received Date: Unknown what this column was used for in the workbook. No reference found in any formula on any other page. Listings are in the mm/dd/yyyy format of all calendar dates from January 1, 2003 through October 31, 2005.

Email count: Unknown what this column was used for in the workbook. No reference found in any formula on any other page. All rows contain “0” in this column



Analysis of the 2005 PST File Inventory Summary Report
Office of Administration
Office of the CIO
November 2007

Appendix A: Message Store Team Charter

MISSION: To organize the existing pst files currently located on [REDACTED] and to draft a set of procedures for maintaining all pst files.

All process and technical decisions for the work described below are to be proposed, reviewed, and approved by all of the Oversight Team prior to executing any changes to PST files.

Task 1: Manage the copying of pst files from [REDACTED] and the placement of these copies on [REDACTED].
Reference ICCP #0943 for details.

- A. Confirm that the pst file content or names remain completely unchanged.\
- B. Confirm and validate that all of the files have been successfully copied and properly located on SFEOP01

DELIVERABLE: Memo of certification

Task 2: Analyze the pst files on [REDACTED]

Various naming conventions have been used when pst files were created. An analysis of the file names must be performed to determine if there are any potential issues in the time series for each component.

- A. Create and populate inventory of all pst files on [REDACTED]

DELIVERABLE: Inventory.

- B. Draft Technical procedures to analyze pst files on [REDACTED]

DELIVERABLE: Draft Procedures.

- C. Analyze pst files which appear to be duplicates to determine if they are in fact identical.
 - a. If a file is identified as being a true duplicate, note this on the inventory.
 - b. If a file is identified as not being a duplicate, note this on the inventory.

DELIVERABLE: Annotated inventory.

- D. Analyze pst files which appear to be non-duplicates to determine if they are in fact unique.
 - a. Identify and record on the inventory the time span and number of messages for each pst file (Earliest timestamp, latest timestamp, and number of messages)
 - b. Identify and record on the inventory any potential gaps in a time series by component from the point in time the component began their Exchange migration.

DELIVERABLE: Annotated inventory.

Team Members: [REDACTED]

Analysis of the 2005 PST File Inventory Summary Report
Office of Administration
Office of the CIO
November 2007

Lead: [REDACTED]

Task 1: [REDACTED] inventory db
[REDACTED] analysis
[REDACTED] analysis
[REDACTED] validation
[REDACTED] analysis

Task 2: [REDACTED] analysis
[REDACTED] analysis
[REDACTED] analysis

Task 3: [REDACTED] procedures
[REDACTED] procedures

Oversight Team: [REDACTED]

Resource Requirements:

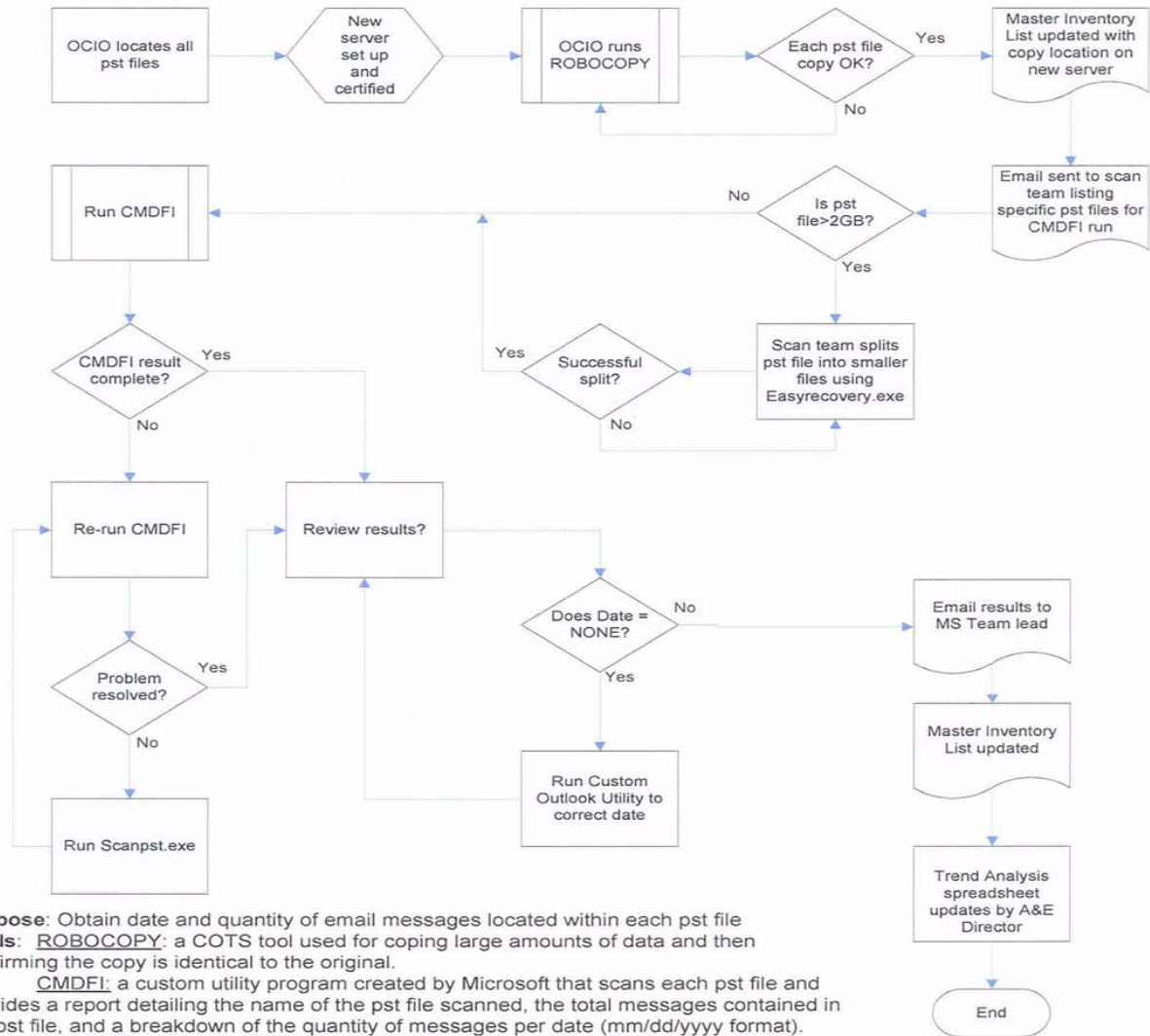
- 1800 G OCIO Bullpen
- MS Utility ? (cost & feasibility)
- Phase 2 manual help if no utility
- Records Management Contractor, 6 month
- Workspace (relocation) for [REDACTED]

[REDACTED]

Analysis of the 2005 PST File Inventory Summary Report
Office of Administration
Office of the CIO
November 2007



Analysis of the 2005 PST File Inventory Summary Report
 Office of Administration
 Office of the CIO
 November 2007
Appendix B:
2005 PST Inventory/Scan Process



Purpose: Obtain date and quantity of email messages located within each pst file
Tools: ROBOCOPY: a COTS tool used for coping large amounts of data and then confirming the copy is identical to the original.

CMDFI: a custom utility program created by Microsoft that scans each pst file and provides a report detailing the name of the pst file scanned, the total messages contained in the pst file, and a breakdown of the quantity of messages per date (mm/dd/yyyy format).

SCANPST: a standard Microsoft utility program used in this process to repair pst files so that they could be opened and scanned by CMDFI.

EasyRecovery: a standard COTS program created by OnTrack which was used to split pst files that were larger than 2GB so that CMDFI could be run. This program created a report which detailed the results of the split to confirm all email from the original large pst file was included in the new smaller pst files.

Custom Outlook Utility: Developed by OCIO to resolve the Date = NONE error returned in some cases from CMDFI.