




**Quality Control Report
for
PIVIT Database & Summary Spreadsheet**

**The Executive Office of the President
Assessment of E-mail Search Program and Results**

August 25, 2008

<p>Prepared by: </p>	<p>Submitted to: Executive Office of President Office of Administration 1800 G Street, NW Washington, D.C. 20503</p>
---	--

Sensitive But Unclassified – For Official Use Only

TABLE OF CONTENTS

1 Introduction..... 1

2 Results..... 1

2.1 Quality Control Check 1: Validate that all .pst files created before January 1, 2008 have corresponding entries in the PIVIT database..... 1

2.2 Quality Control Check 2: Validate the item counts for each .pst file in the PIVIT matches the item count in the corresponding .pst file on the source drive 3

2.3 Quality Control Check 3:..... 4

2.3.1 The first part of Quality Control Check 3 (QC3) is to examine issues discovered during Quality Control Checks 2 (QC2) to determine if there are reasonable technical explanations as to the cause of each issue case..... 4

2.3.2 The second part of QC3 is to evaluate the methodology used to determine message distribution to components to determine if it is technically sound and accurately reflects the distribution of messages to components..... 5

2.3.3 The third part of QC3 is to validate that the final master spreadsheet generated from PIVIT accurately represents the data contained in PIVIT. 6

3 Procedure 7

3.1 Quality Control Check 1: Validate that all .pst files created before January 1, 2008 have corresponding entries in the PIVIT database. 7

3.1.1 Obtain list of full pathname\filename from the PIVIT database..... 7

3.1.2 Obtain list of full pathname\filename from the source drives..... 7

3.1.3 Create spreadsheet to perform comparative analysis and summarize results. 8

3.1.3.1 Prepare PIVIT file data 8

3.1.3.2 Prepare drive file data 8

3.1.3.3 Compare Data 8

3.1.3.4 List Mismatches 9

3.1.3.5 Summarize the Errors 9

3.2 Quality Control Check 2: Validate the item counts for each .pst file in the PIVIT matches the item count in the corresponding .pst file on the source drive 9

3.2.1 Set up the population..... 10

3.2.2 Determine the sample size..... 10

3.2.3 Identify a random sample 10

3.2.4 Set up the comparison 10

3.2.5 Divide the random sample..... 10

3.2.6 Collect the item count from each .pst file 11

3.2.6.1 Merge data 11

3.2.6.2 Summarize results 11

Revision History

Author	Version Number	Revision Date	Revision Summary
██████████	Draft	08/20/2008	Initial Draft
██████████	0.1	08/23/2008	Added section QC3 (1,2) additional information on case resolution pending.
██████████	0.2	08/25/2008	Completed QC3 (3)
██████████	0.3	08/25/2008	Completed section QC3 (1). Submitting for customer acceptance.
██████████	0.4	08/26/2008	Correction made on page 5 changed 2208 to 2248
██████████	0.5	08/26/2008	Added table reference numbers
██████████	0.6	8/29/2008	Correct a transcription error on table 2.1 1 and corresponding references
██████████	1.0	8/29/2008	Final
██████████	1.1	9/17/2008	Added section 3.3

1 Introduction

OA-OCIO Directorate contracted development of the PST Inventory Verification and Inspection Tool (PIVIT) “to generate reliable data which will enable the leadership team to identify potential discrepancies in the e-mail archives.”¹ This system aids in analysis to provide an account of e-mail messages by day and component.

OCIO/IA was tasked to complete three quality control checks of the PIVIT database. NAID, Inc., was selected to perform the following tasks;

- Quality Control Check 1: Validate that all .pst files created before January 1, 2008 have corresponding entries in the PIVIT database.
- Quality Control Check 2: Validate the item counts for each .pst file in the PIVIT matches the item count in the corresponding .pst file on the source drive
- Quality Control Check 3: QC 3 includes three distinct parts; Part one is to examine issues discovered during Quality Control Checks 2 to determine if there are reasonable technical explanations as to the cause of each issue case. Part two is to evaluate the methodology used to determine message distribution to components to determine if it is technically sound and accurately reflects the distribution of messages to components. Part three is to validate that the final master spreadsheet generated from PIVIT accurately represents the data contained in PIVIT.

2 Results

2.1 Quality Control Check 1: Validate that all .pst files created before January 1, 2008 have corresponding entries in the PIVIT database.

The PIVIT database contains 27,393 entries for .pst files. The PIVIT database contains 185 duplicate .pst files with the same *PATHNAME\FILENAME* and different message counts. Details are included in the “PIVIT File Dupes” tab in the *QC1 PIVIT 2008-08-19-JLM.xls* spreadsheet. This leaves 27,208 unique .pst files.

Based on the path and filename, 27,159 of the 27,208 unique .pst files listed in the PIVIT database match .pst files on the source drives. The mismatch error rate is 1.06%. The following table summarized the results by drive:

¹ PST Inventory Verification and Inspection Tool (PIVIT) Vision Document, Version 1.2, Draft, dated 8/4/2008.

Drive	Source		Status	
	Drive	PIVIT	Match	Mismatch
e:\	851	850	850	1
f:\	2030	1859	1859	171
g:\	2397	2372	2372	25
h:\	9565	9565	9565	0
i:\	2550	2590	2543	54
k:\	10008	9972	9970	40
Totals	27401	27208	27159	291

Table 2.1 1

Mismatches were resolved for 3 files on the source drives as follows:

Drive	Resolved Mismatches On Drive	
	On Drive	Comment
e:\	1	1 directory on server with .pst extension
g:\	1	1 directory on server with .pst extension
i:\	1	1 directory on server with .pst extension
Totals	3	

Table 2.1 2

Mismatches were resolved for 9 mismatches for files identified in PIVIT as follows:

Drive	Resolved Mismatches In PIVIT	
	In PIVIT	Comment
i:\	8	8 files in PIVIT found on the server in the issues folder
k:\	1	1 file in PIVIT found on the server in the issues folder
Totals	9	

Table 2.1 3

This leaves 279 mismatches unresolved, distributed as follows:

Drive	Unresolved Mismatches	
	On Drive	In PIVIT
e:\	0	0
f:\	171	0
g:\	24	0
h:\	0	0
i:\	6	39
k:\	38	1
Totals	239	40

Table 2.1 4

Details are included in the "Mismatches" tab in the *QC1 PIVIT 2008-08-19-JLM.xls* spreadsheet.

2.2 Quality Control Check 2: Validate the item counts for each .pst file in the PIVIT matches the item count in the corresponding .pst file on the source drive

It was calculated that of the 27,393 .pst entries in PIVIT, 2,208 .pst files would be needed as the sample size to achieve a confidence level of 95% with a confidence interval of 2. A spreadsheet listing all files was created and the Random function used to identify the .pst files to include in the sample.

Using a confidence level of 95% and a confidence interval of 2 for a population of 27,393, the item counts in PIVIT match the item count obtained by opening the corresponding .pst file on the source drives 94% of the time. A summary of the matches for each drive follows:

Drive	Total	Match	Percent
e:/	90	71	79%
f:/	144	124	86%
g:/	180	165	92%
h:/	783	779	99%
i:/	210	196	93%
k:/	801	739	92%
Total	2208	2074	94%

Table 2.2 1

- The Drive column shows the source drive.
- The Total column shows the number of files in the sample size.
- The Match column shows the number of files whose item counts match between the data in the PIVIT *TOTAL_ITEMS_TOC* field and the actual count obtained by opening the .pst file on the source drive.
- The Percent column shows the percentage of files that whose item counts match between the data in the PIVIT *TOTAL_ITEMS_TOC* field and the actual count obtained by opening the .pst file on the source drive.

The item counts in the *TOTAL_ITEMS* and *TOTAL_ITEMS_TOC* fields match 99.85% of the time for the total population. This leaves 41 .pst files that show a mismatch between the two fields. Details of the mismatches are included in the "QC PIVIT Mismatch" tab in the *QC2 PIVIT 2008-08-20-JLM.xls* spreadsheet.

TOTAL_ITEMS vs TOTAL_ITEMS_TOC	
Population Size	27,393
Population Match	27,352
Population Match Percentage	99.85%

Table 2.2 2

The item counts in the *TOTAL_ITEMS* and *TOTAL_ITEMS_TOC* fields match 99.68% of the time for the random sample. This leaves 7 .pst files that show a mismatch between the two fields. Details of the mismatches are included in the "QC PIVIT Mismatch" tab in the *QC2 PIVIT 2008-08-20-JLM.xls* spreadsheet.

TOTAL_ITEMS vs TOTAL_ITEMS_TOC	
Sample Size	2,208
Sample Match	2,201
Sample Match Percentage	99.68%

Table 2.2 3

Further analysis showed that the item count identified by opening the .pst file matched the *TOTAL_ITEMS_TOC* field for all 7 files in the random sample. Thus, the *TOTAL_ITEMS_TOC* field is more accurate for reporting purposes.

2.3 Quality Control Check 3

2.3.1 The first part of Quality Control Check 3 (QC3) is to examine issues discovered during Quality Control Check 2 (QC2) to determine if there are reasonable technical explanations as to the cause of each issue case.

In examining the results of QC2 five distinct case classes were identified where message counts do not match between the PIVIT *TOTAL_ITEMS*, *TOTAL_ITEMS_TOC*, or QC2 counts. One additional case was identified where the three message counts were all zero.

- Case 1: Twenty nine PST files were identified where the *TOTAL_ITEMS*, *TOTAL_ITEMS_TOC* count is zero and the QC2 count is greater than zero.

Explanation: An evaluation of the data in concert with the PIVIT programmer shows the error occurred when a date field in one of the journaled messages was outside of the allowable range for .NET or SQL. The error in PIVIT is: "Error Reading PST: Year, Month and Day describe an un-representable date time." See "Recommendation on additional action" at the end of this section.

- Case 2: Six PST files were identified where the *TOTAL_ITEMS* equals *TOTAL_ITEMS_TOC* count but the QC2 count is zero.

Explanation: There are two probable explanations as to what could cause this issue. The PST message store is fully accessible by PIVIT but the message object view is corrupt so that messages are not visible to the user or the PST contains deleted messages accessible to PIVIT but not the user.

- Case 3: Twenty six PST files were identified where the *TOTAL_ITEMS* equals *TOTAL_ITEMS_TOC* count but the QC2 count is less than these counts.

Explanation: This case is similar to case 2. The PST message store is fully accessible by PIVIT but the view index is corrupt or requires update so that not all messages are visible to the user or the PST contains deleted messages accessible to PIVIT but not visible to the user.

- Case 4: Seven PST files were identified where the *TOTAL_ITEM* is zero but the *TOTAL_ITEMS_TOC* count and QC2 count are equal.

Explanation: The SQL operation in PIVIT which updates the TOTAL_ITEM count failed from a timeout.

- Case 5: One PST file was identified where the TOTAL_ITEMS equals TOTAL_ITEMS_TOC count but the QC2 count is greater than these counts.

Explanation: To determine the cause of this Case the PST file needs to be examined.

- Case 6: One Hundred Seventy Nine PST files have a zero message count for TOTAL_ITEMS, TOTAL_ITEMS_TOC, and QC2.

Explanation: The most probable cause for this case is that the PST was initialized, created, but never populated with messages or other objects (calendar, tasks, notes, etc.). As shown in Case 2 and 3 if the PST is corrupt or messages were deleted they would still be accessible by PIVIT. However, if messages were deleted then the PST compacted the deleted messages would be purged and all counts would be zero. It is recommended that an audit be conducted of the files which meet the criteria for this case. The objective of this audit would be to determine the creation origin (user, search team, journal) of each PST. The user and search team created PST files could be eliminated as they would be copies of already journaled messages and would not alter message count totals. If any files remain, which were created as a part of the journaling process, a forensics examination could be conducted to determine if the files were initialize but never populated, journaled but corrupt, or purged and compacted.

Recommendation on additional action: There are a total of 2248 records in the PSTIndexAll database TBL_PST table with a TOTAL_ITEMS and TOTAL_ITEMS_TOC count of zero. If the ratio of between Case 1 and Case 6 holds true for the 2248 PST files and the files are for zero or low volume days this would increase message volume totals for those days. It is recommended that a report be created from PIVIT which identifies all of these PST files and highlights any from zero or low volume days. A team should manually examine these files with priority on the zero and low volume days to determine if they can be classed as Case 1 or Case 6.

2.3.2 The second part of QC3 is to evaluate the methodology used to determine message distribution to components to determine if it is technically sound and accurately reflects the distribution of messages to components.

There are only two probable methodologies for calculating the distribution of messages by component.

The selected method not only results an accurate distribution of the senders intended component it is more efficient in terms of manpower and IT resources than the alternative. However, in cases where the local part (the text left of the @ sign) is invalid but the domain part (the text right of the @ sign) is valid it would classify a message as delivered to a component when there is not a valid user account for the message. In these cases PIVIT would count the message, which would exist in the journal, but the message would not have been deliverable to a user account mail box.

The second method that could have been employed would be able to identify non-delivery due to invalid local part. This method would require a cross reference with all of the current and his-

torical Active Directory person objects. This cross reference would need to account for every personnel reassignment across components, email address change, name change, and those persons with multiple accounts for every individual account created and removed during the inclusive dates. While this method would be able to provide accurate non-delivery information which would raise the total count of non-delivered messages it is doubtful that the component distribution percentages would change to a significant degree. Therefore, the conclusion is that using the component portion of the domain part of the email address to calculate the component distribution is accurate and efficient.

2.3.3 The third part of QC3 is to validate that the final master spreadsheet generated from PIVIT accurately represents the data contained in PIVIT.

Over one hundred data points were randomly sampled in PIVIT and compared to data elements contained in spreadsheets provided on a network drive² and data elements were also traced back through referenced spreadsheet³ or directly compared to PIVIT data. All of these data points are accurately presented in the spreadsheet. Based on the results of these data checks there is a high degree of confidence that data exported from PIVIT is accurately represented in the master spreadsheet "Phase 2 Message Counts - BAR-1.xls" at the point in time during which it was examined.

Recommendation: Due to the ease of data manipulation in Excel and potential complexities of the formulas used to take data and present it as information any validation is only certifiable at the point in time when it was examined. If it is the intent of EOP to present documents to third parties which can be considered as validated beyond a single point in time the production process used to create the master spreadsheet would need to be repeated under conditions in compliance with the "Electronic Discovery Reference Model⁴." The following is an example of such conditions.

Example: Each step in the creation process should be documented as to the Inputs, Processes, and Outputs (IPO). The IPO document(s) should be provided to a selected individual who would observe, monitor and record each step in the process. The observer would record information on all output files as to; location, name, size, creation date/time, and modified date/time. Access to the outputs should be tightly controlled and when the files pass from one individual to another access should be removed for the first and granted for the second. Chain of custody should also be documented. Once the spreadsheet is complete it should be converted to PDF file format, digitally signed, and locked. The record created by the observer including the IPO documents, output logs, and chain of custody information should then also be converted to PDF file format, digitally signed, and locked.

² Drive mapped as "W:\IVV\Phase 2 Documents\Statistical Analysis"

³ Data elements in the spread sheet referenced "\\ds\sharedir\OA\OCIO\Administration\Record Keeping\Phase 2 Documents\Statistical Analysis" from spreadsheet path "W:\IVV\Phase 2 Documents\Sarah's spreadsheets for analysis\The Ultimate Spreadsheet\Phase 2 Message Counts - BAR-1.xls"

⁴ <http://www.edrm.net/>

3 Procedure

3.1 Quality Control Check 1: Validate that all .pst files created before January 1, 2008 have corresponding entries in the PIVIT database.

NAID completed this assessment by using Microsoft Excel to sort and compare the full filename and path between a report from PIVIT and a directory dump from each source drive.

3.1.1 Obtain list of full pathname\filename from the PIVIT database.

OA-OCIO provided the spreadsheet, *PST Inventory-1 7-30-2008.xls*, which contained a listing of filename, pathname, and total item count for all .pst files in PIVIT. This spreadsheet is organized as follows:

- Intro: This tab contains a mapping of the PIVIT drives to the source drives. Following further investigation, we determined that the complete mapping is as follows:

PIVIT		SOURCE DRIVE
Tab	Drive	
Server 1:	P Drive	I Drive
Server 2:	P Drive	E Drive
Server 2:	Q Drive	F Drive
Server 3	P Drive	G Drive
Server 3	Q Drive	H Drive
Server 4	P Drive	K Drive

Table 3.1 1

- Server 1: This tab contains information mapped to I Drive.
- Server 2: This tab contains information mapped to E Drive and F Drive.
- Server 3: This tab contains information mapped to G Drive and H Drive.
- Server 4: This tab contains information mapped to K Drive.

3.1.2 Obtain list of full pathname\filename from the source drives.

OCIO/IA provided text files which contained the full pathname\filename for all files on each of the source drives. These files are:

- edrive.txt: This file contains directory information for the E Drive.
- fdrive.txt: This file contains directory information for the F Drive.
- gdrive.txt: This file contains directory information for the G Drive.
- hdrive.txt: This file contains directory information for the H Drive.

- idrive.txt: This file contains directory information for the I Drive.
- kdrive.txt: This file contains directory information for the K Drive.

3.1.3 Create spreadsheet to perform comparative analysis and summarize results.

The final spreadsheet is, *QC1 PIVIT 2008-08-19-JLM.xls*, contains the analysis. This section explains the steps for creating and working with the spreadsheet.

3.1.3.1 Prepare PIVIT file data

We created a “File Confirm” tab that contains a copy of the filename and pathname from the *PST Inventory-1 7-30-2008.xls* spreadsheet, which provides a starting point of 27,393 files to be analyzed.

- We sorted the files alphabetically.
- We used an IF statement to compare each file with its predecessor in the list, resulting in a 1 if the *PATHNAME\FILENAME* was identical and a 0 if the *PATHNAME\FILENAME* was different. This identified 185 instances where duplicate pairs of files exist.

We created a “PIVIT File Dupes” tab that contains all the duplicate files along with their *TOTAL_ITEMS* and *TOTAL_ITEMS_TOC* counts. In many, but not all instances the counts were different between the file pairs. No further analysis was performed.

We created a “PIVIT File De-Duped” tab to contain all the files that were unique. This tab was the source of PIVIT *PATHNAME\FILENAME* for comparison.

3.1.3.2 Prepare drive file data

For each drive, we created a “Drive” tab and imported the *PATHNAME\FILENAME* from the corresponding text file that contained this information. We used string functions to identify, sort and delete all lines that did not end with “pst” as the last three letters. This tab was the source of drive *PATHNAME\FILENAME* for comparison.

3.1.3.3 Compare Data

For each drive, we created a “Compare” tab to do the comparison.

- We copied the corresponding *PATHNAME\FILENAME* from the “drive” tab under the heading “drive path & filename” in column B and sorted alphabetically.
- We copied the corresponding *PATHNAME\FILENAME* from the “PIVIT File De-Duped” tab under the heading “PIVIT path & filename” in column C and sorted alphabetically.
- We used an IF statement in column A to compare the *PATHNAME\FILENAME* in the two columns, resulting in a 1 if they matched and a 0 if they did not match.
- When a mismatch occurred, we examined both columns to determine where the next match

would occur and moved the cells down in the appropriate column to create the next match. We repeated this process of aligning and comparing the files until we completed the entire list in both columns.

- We used a COUNTIF statement in cell A1 to obtain the count of files that matched.
- We used a COUNTIF statement in cell A2 to obtain the count of files that did not have a corresponding match.
- Cell B1 contains the number of files for the drive in the “PIVIT File De-duped” tab.
- We used conditional formatting to highlight blank cells (i.e. no corresponding match to the other column) in red.

3.1.3.4 List Mismatches

We created a “Mismatches” tab that contains a list of all filenames that either (a) existed on the source drives and was not listed in the PIVIT database or (b) was listed in the PIVIT database and was not on the source drives. We copied this information from the “Compare” tabs. We did further analysis to see if we could identify the cause of the mismatch.

- We found three directories with a .pst extension, which does not reflect a .pst file that should be included in the PIVIT database. Including a .pst extension on a directory should be discouraged.
- We found 9 files on the source drives that were in the PIVIT database but not in the text file containing the directory dump.
- We could not find 24 files on the G Drive that were listed in the text file containing the G Drive directory dump. None of these files were listed in the PIVIT database.
- We confirmed that the remaining files listed from the text file directory dumps were created prior to January 1, 2008 and should have been included in the PIVIT database.

3.1.3.5 Summarize the Errors

We created an “Error Summary” tab to summarize the following:

- Matches and mismatches by drive and source.
- Resolved mismatches for files on the drive.
- Resolved mismatches for files in PIVIT.
- Unresolved mismatches by drive and source.

3.2 Quality Control Check 2: Validate the item counts for each .pst file in the PIVIT matches the item count in the corresponding .pst file on the source drive

NAID completed this assessment by creating a random sample of .pst files to check, manually opening each selected .pst file on the source drive to obtain the item count, and comparing that number to the item count contained in PIVIT for that .pst file. We used Microsoft Excel to create the random sample and do the comparisons. The final spreadsheet is *QC2 PIVIT 2008-08-20-JLM.xls*.

3.2.1 Set up the population

We copied the PIVIT information from the *PST Inventory-1 7-30-2008.xls* spreadsheet into the *QC2 PIVIT 2008-08-20-JLM.xls* spreadsheet. Per the mapping in 4.1.1., we renamed the drive letter in each *PATH_NAME* to match the source drive. We copied the information from the individual “Server” tabs into a “Files & Count” tab to create a single list reflecting the entire population of 27,393 files.

3.2.2 Determine the sample size

Using a sample size calculator, we determined that we needed a random sample of 2,208 .pst files to achieve a confidence level of 95% with a confidence interval of 2 for a population of 27,393.

3.2.3 Identify a random sample

We copied the “Files & Count” tab to a “QC Sample” tab and added a *RANDOM* column, containing the formula *=RAND()* to randomize the data. Then, we sorted by the *RANDOM* column and selected the first 2,208 files for the sample and deleted the remaining files.

3.2.4 Set up the comparison

We copied the “QC Sample” tab to a “QC Breakout” tab and sorted the files into a logical sequence to facilitate collecting the item counts from the source drives. We added the following columns to set up the comparison:

- *QC_ITEMS*: To be populated manually based on the item count obtained from the file on the source drive.
- *QC_ITEMS - TOTAL_ITEMS*: Calculates the delta between the manual *QC_ITEMS* count and the PIVIT *TOTAL_ITEMS* count.
- *QC_ITEMS-TOC_ITEMS*: Calculates the delta between the manual *QC_ITEMS* count and the PIVIT *TOTAL_ITEMS_TOC* count.

3.2.5 Divide the random sample

We further divided the “QC Breakout” tab into four separate tabs to assign specific files to be checked by the person named on the tab.

3.2.6 Collect the item count from each .pst file

Each team member obtained a manual item count for each assigned .pst file using the following procedure:

- Open the .pst file.
- Obtained the item count by doing an advanced find (CTRL-SHIFT-F) at the root level and reading the number of items at the bottom left corner of the dialog box.
- Enter the item count in the corresponding *QC_ITEMS* cell.
- Repeat until item count has been obtained for all .pst.
- NOTE: In some instances, the results from advanced find differed from the item count listed on the status bar. For these instances, the results from the status bar were entered into the corresponding *QC_ITEMS* cell.

3.2.6.1 Merge data

We copied the information from the individual tabs into a single “QC Sample Analysis” tab and performed three sets of comparisons:

- PIVIT TI vs PIVIT TIT: We used an IF statement to compare *TOTAL_ITEMS* against *TOTAL_ITEMS_TOC*, resulting in a 1 if they matched and a 0 if they did not match.
- *QC_ITEMS* vs *TOTAL_ITEMS*: We used an IF statement to compare *TOTAL_ITEMS* against *QC_ITEMS*, resulting in a 1 if they matched and a 0 if they did not match.
- *QC_ITEMS* vs *TOC_ITEMS*: We used an IF statement to compare *TOTAL_ITEMS_TOC* against *QC_ITEMS*, resulting in a 1 if they matched and a 0 if they did not match.

We used the *COUNTA* and *COUNTIF* functions to determine the following:

- The total number of files by drive.
- Total number of files by drive that matched in terms of item count.

3.2.6.2 Summarize results

We created an “Analysis Summary” tab to provide the following summaries:

- *TOTAL_ITEMS* vs *TOTAL_ITEMS_TOC*: We summarized the matches between these two counts for both the total population and the random sample. The percentages were essentially the same (99.85% and 99.68%).
- *MATCHES BY DRIVE*: We summarized the matches by drive. Total matches were 94%, with a range of 79% to 99% by drive).

3.3 Quality Control Check 3

3.3.1 Results of Quality Control Check 2.1.1

The process used for this check was to compare the values within the 3 columns, PIVIT TOTAL_ITEMS, TOTAL_ITEMS_TOC, and QC2 counts and assign the case (see 2.3.1 for case definitions) of any found variances. Each variance was analyzed to determine a likely technical cause (PIVIT error, corrupted .pst, etc.). Recommendations for further analysis were then documented in the end of Section 2.3.1 following messaging best practices.

3.3.2 Results of Quality Control Check 2.1.2

The process used to determine if the message distribution is accurate was to compare the selected method to other possible methods. A comparison of the possible options was conducted and it was determined that the selected method would accurately provide distribution.

3.3.3 Results of Quality Control Check 2.3.3

For Quality Checks 2.3.1 and 2.3.2 see those sections on process.

Due to the large volume of messages and data and constraints on time to complete the analysis for Quality Check 2.3.3 it was not reasonable to attempt a sampling in quantities high enough to get statistical surety. The following process was used to complete QC 2.3.2.

- To ensure that the selected data points would provide sufficient volume for a cross reference between the spreadsheet and PIVIT no data points were used prior to January 1, 2003.
- Cell values were selected from the “7-25-2008 Pass2 results” tab of the spreadsheet. This tab referenced the external spreadsheets⁵ which were produced directly from PIVIT.
- Data points were cross referenced to PIVIT totals using simple SQL “SELECT” statements ran against the TBL_PST table.

⁵ Data elements in the spread sheet referenced “\\ds\sharedir\OA\OCIO\Administration\Record Keeping\Phase 2 Documents\Statistical Analysis” from spreadsheet path “W:\IVV\Phase 2 Documents\Sarah's spreadsheets for analysis\The Ultimate Spreadsheet\Phase 2 Message Counts - BAR-1.xls”