

[REDACTED]

From: [REDACTED]
Sent: Friday, August 29, 2008 4:42 PM
To: [REDACTED]
Subject: Updates

PIVIT

As [REDACTED] has explained to me, the 43 pst files that were recently ingested into PIVIT are currently processing through the most time consuming portion of the tool; counting every recipient of every message. This step has processed approximately 90 million of the 250 million messages. This time consuming step is expected to complete Saturday. [REDACTED] will be remotely accessing the tool over the weekend to check status and confirm all is working well.

Once the number of recipients are counted the de-duplication itself begins. Barclay is confident we will meet the stated delivery deadline of September 5.

To give everyone a perspective on the level of effort being done, these are the steps PIVIT goes through for the 250 million messages in the database. Since we did not program PIVIT to produce a multiple databases for each separate component or year we must follow the full process anytime we ingest new pst files. The messages they contain must be processed against all of the other 250 million messages in the inventory. Therefore whether or not we are ingesting 50 or 150 pst files the effect is nearly the same to the entire inventory. We are adding less than 1% of the total 250 million messages into the tool and the steps below have to be run against all 250 million messages. This may not be the most elegant system but it was the best we could build and it has served us well.

1. Load all data into 4 index servers (done)
2. Update the base tables (done)
3. Insert lookups (done)
4. Clean-up email addresses to determine component (done)
5. Date and Time calculations (done)
6. Count recipients per message (40% complete)
7. De-duplication
8. Sender analysis
9. Recipient analysis
10. Run SAS and ARIMA
11. Generate reports

IV&V vs. Quality Check:

I may just be arguing semantically here, but in the IT world an Independent Verification & Validation (IV&V) would include review of requirements, system design, actual coding, and results. Given the fact that we could not find a COTS product to do the Phase 1 and Phase 2 work (not to mention the time it would have taken to procure such a system if one existed) we had to rely on our contracted design and development contractors, InfoReliance (IR), to develop the PIVIT tool.

IR has a standard business practice which requires a peer review of the system code they develop before it is handed-off to clients. The IR peer review was done. However, when we sought another independent review of this tool none of the vendors or government offices we contacted were interested in this work. The potential notoriety in the media was likely more than they wanted to deal with and I can't say as I blame them.

We then looked to the contract we use to IV&V search results. The IV&V contract with [REDACTED] was in place and we looked at it to see if they could perform the level of review on PIVIT we desired. Unfortunately NAID did not have the skill set to review the actual coding and other technical aspects. They could, however, check the quality of three very critical aspects namely 1) all of the .pst files for the targeted time frame included in the project, 2) the

8/30/2008

item count in each .pst file matches in PIVIT and the source .pst files, and 3) the number of messages by component per day reported in the project spreadsheets match the number of messages in the PIVIT database.

NAID's quality check deliverable is a report detailing what they found when they looked at those 3 critical aspects of Phase 2. [REDACTED] and I reviewed the draft version and then met with [REDACTED] to gain clarification and address errors we found in their report. I believe [REDACTED] has turned in its final report to [REDACTED] and I will complete the OCIO response to [REDACTED] findings next week. As reported earlier, due to the [REDACTED] we discovered 43 .pst files which had not been ingested into PIVIT. While we do not expect any significant impact on the Phase 2 numbers we can at least now report that all known .pst files for the targeted timeframe have been ingested into PIVIT.

Clearly the lack of any attempt at a quality check or IV&V on the 2005 inventory report has taught the OCIO a very valuable lesson. We have said from the beginning of the email re-baselining project that we would take the time to check our work before we can, with confidence, report our findings. Had vendors or other government offices been willing to take on this critical review we may have been able to report sooner, but given the options at our disposal I am very pleased with the progress to date.

Please let me know if you have any further questions.

[REDACTED]

*Director, Federal Records Management
Office of Administration
Executive Office of the President*

8/30/2008