Date : January 28, 2010

Reply to
Attn of : Office of Inspector General (OIG)

Subject : Advisory Report No. 10-03, Search Engine Analysis for Online Public Access to the
Electronic Records Archives

To : David Ferriero, Archivist of the United States (N)

The selection process for choosing the search engine to provide public access to the
Electronic Records Archives (ERA) was inadequate. An OIG review identified that ERA
program officials did not independently test or analyze the search engine prior to its
selection. Instead, NARA relied exclusively on the recommendation of Lockheed Martin
(LM) Corporation staff. NARA ERA program officials could not provide OIG auditors with
sound and transparent documentation to validate the selection of Vivisimo as the search
engine for the largest IT project ever undertaken by this agency, and one that will impact all
NARA stakeholders for the foreseeable future. NARA's *IT Architecture Systems
Development Guidelines* require that after evaluating commercial off-the shelf products, the
findings must be documented and justification for any particular item in the recommendation
process must be made. In this case, NARA accepted LM's selection of Vivisimo[1] to the
exclusion of other vendors and search engines based upon incomplete analysis, and devoid of
actual ERA technical staff hands-on testing. Thus, the selection process was flawed and the
impact upon future deployment of the ERA is unknown.

At a key decision point in the development of the ERA System, the public access search
engine was selected without any documentation prepared by the NARA ERA Program Office
to indicate who made the decision or the rationale used for the selection. According to the
Director of ERA's Systems Engineering Division, in March 2009 LM was tasked with
researching and analyzing commercial search engine products and making a recommendation
to NARA.

OIG auditors sought documentation defining the basis for the selection of Vivisimo.
Responsible ERA Program officials referred to a trade study of search engines prepared by
LM in 2007. We reviewed this study and identified that Vivisimo was not included in the
deliverable as a potential search engine candidate. The OIG staff was also provided a copy
of a 2009 trade report crafted by Gartner Inc. which identified leaders in the search-engine
field. In 2009, per the ERA Systems Engineering Director, LM brought in 10 to 12 vendors
who provided demonstrations, some of which were reportedly attended by NARA ERA
program staff but not the Systems Engineering Director or Program Director. After the
demonstrations, the list of potential vendors was narrowed to three. Then, LM developed a

---

[1] LM is defined as the selecting entity of Vivisimo as no documentation was provided to indicate when or in what
manner NARA ratified the selection.

list of questions for these three vendors that contained weighted criteria areas such as vendor viability and product functionality (See Attachment A for a description of the criteria areas and the assigned weights for each). Subsequently, LM selected one of the three vendors, Vivisimo. The Director told us that he is not aware of anyone at NARA expressing concerns with the selected vendor, and that his engineering staff talked with the vendor to clarify technical questions and to obtain a demo license. However, to reiterate, at the point of selection no evidence exists that ERA technical staff had actually tested Vivisimo to assess functionality and capacity specific to unique ERA requirements.

In August 2009, after ERA staff had reportedly validated LM's selection of Vivisimo we attended a demonstration at LM's facility in Greenbelt Md. LM officials responsible for the selection of Vivisimo presented an overview of Vivisimo search capabilities. We requested a demonstration of the actual testing that had been performed on NARA records encompassing those of the Revolutionary War that LM had reportedly ingested into the test bed. Keyword text searches we requested such as "Boston Tea Party", "Boston" and "George Washington" came back with no results. Thus the demonstrated ability of Vivisimo to function with sample ERA-like data failed. The LM official subsequently attributed this condition to the fact that he had just ingested the sample data into the system the night before and had not attempted to conduct queries such as the one we requested prior to the demonstration. Notwithstanding this fact, Vivisimo had already been recommended to the exclusion of other search engines by these very engineers.

OIG auditors asked the ERA Systems Engineering Director what internal documentation there was to support the search engine analysis and selection; he stated there may be some minutes of meetings with LM and an e-mail message notifying LM of the search engine selection. To date these minutes have not been produced and provided to OIG staff. Due to the fact the demonstration searches did not yield any results, we asked ERA officials what independent analysis, if any, they performed on the selected search engine. The ERA Systems Engineering Director told us he had a copy of the search engine which he reviewed for functionality, but had not tested it using actual ERA data.

On January 12, 2010 we met with Director of the ERA Program and the ERA Systems Engineering Director to determine if any further analysis of the selected search engine was performed after the demonstration we attended in August 2009. The ERA Systems Engineering Director stated no further testing or analysis was conducted due to lack of resources and his belief the selection of the ERA search engine was a responsibility assigned to LM under their contract and he had no standing in this regard. He further identified that a scenario under which NARA would have to move to a different search engine would likely be costly and technologically difficult. In our opinion, ERA officials should have exercised due diligence over the selection of this important system component by independently conducting their own tests and analyses prior to its selection, and not relying solely on the contractor's analysis. The rational for not doing so based upon inadequate internal ERA staffing and resources for a project of this scope is troubling.

As previously stated, according to NARA's *IT Architecture Systems Development Guidelines*, after evaluating commercial off-the-shelf products the recommendation process

should document the findings of the evaluation and provide the justification for a particular item. Of critical importance in the recommendation is to relate any intangibles that are not reflected in the assessment criteria. The final document should include the following:

- Executive summary briefly describing process, recommended product, and reason for recommendation;
- Additional changes to business processes and system architecture caused by recommended product;
- Summary analysis of each product with pros and cons;
- Product recommendation and justification; and
- Documented product and vendor evaluations.

Decisions involving key components of the ERA System, such as the search engine for online public access, should be formally approved and documented by senior NARA officials. The risk of not operating within these parameters is unnecessary and unwarranted.

It should be reemphasized this review is limited to the process initially used in selecting the ERA search engine. This review does not cover whether or not the best search engine for NARA's needs was chosen. We make no statements about the capabilities of the search engine, and specifically have not evaluated the Vivisimo product. It is our understanding that NARA and LM officials continue efforts to develop the ERA search engine. LM reported that successful prototype demonstrations[2] of the search engine were conducted in November and December 2009.

A beta release of the Online Public Access Instance, planned for March 2010, will include access to the Access to Archival Database (AAD) and the Archival Research Catalogue (ARC) assets. This beta release will be open to a limited number of users for evaluation. Online access to the public is planned for the fourth quarter of FY 2010, with access to AAD and ARC assets, as well as other holdings.
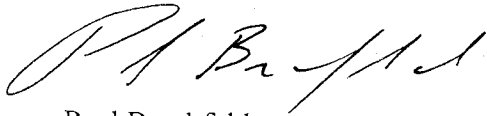
We will continue to monitor the development of the ERA search engine and report back to you on a periodic basis.

This project was part of our on-going effort to review NARA's development and implementation of the ERA. Our review effort consisted primarily of reviewing applicable ERA documentation such as the ERA FY 2010 Expenditure Plan, ERA Search Trade Study, Online Public Access for NARA Requirements Document: Search Functionality, ERA contract and modifications, contractor status reports, Congressional status reports, and interviews with responsible ERA Program Office officials. We conducted this performance audit in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objectives.

---

[2] In the meeting January 12, 2010 the OIG requested supporting documentation, however, none has been received to date.

We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

If you have any questions concerning the information presented in this report, please e-mail Mr. James Springs or me, or call us at extension 73000.

Paul Brachfeld
Inspector General

cc: NH (M. Morphy)

SUPPLEMENTAL INFORMATION

Search Engine Criteria Areas and Weights
Used by LM to Evaluate Each Search Engine Product Analyzed

Criteria	Description	Weighting Factor
Cost	The rough order of magnitude purchase cost for a component or system and whatever maintenance costs are required.	.075
Performance	Refers to the ability of the product to meet the defined performance objectives.	.050
Vendor Viability	Characterizes the size, history, responsiveness, and dependability of the vendor.	.025
Interoperability	Refers to how well the product integrates with our current system.	.125
Standards Compliance	Refers to how well the product supports the required software standards.	.105
Execution and Monitoring	Identifies the monitoring and event management capabilities of the system.	.125
Scalability	Refers to the time and cost to provide additional capacity and the capability to support clustering.	.075
Functionality	Identifies the specific search features required.	.320
Security	Identifies the security features of the product how well they support industry standards for information assurance.	.050
Usability	Refers to the quality of the product's graphical user interface.	.025
Implementation Complexity	An assessment of the technical factors that determine the degree of difficulty to implement a candidate solution.	.025