NATIONAL ARCHIVES

OFFICE *of*
INSPECTOR GENERAL

Date   : January 5, 2011

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

Subject : Management Letter No. 11-08, Electronic Records Archive Lacks Ability to Search Records' Contents

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

The Electronic Records Archives (ERA) program is critical to the future of the National Archives and Records Administration (NARA) and the nation. From inception, ERA was envisioned as the primary way our nation's electronic records will be preserved and accessed. Thus, ERA is situated to become NARA's flagship system in a world dominated by born digital records, and there is no alternative venue to which the public can turn for comprehensive access to these records of our democracy. However, other than for select records, system limitations will not allow users to conduct a content search of the comprehensive inventory of electronic records ERA has and will continue to ingest. Instead, ERA will only allow users to locate records by searching through metadata generated about the records, and not the text of the records themselves. We believe this constraint has not been adequately communicated to NARA stakeholders.

The inability of all American citizens to fully search the content of records ingested into ERA will have a profound and adverse impact upon this nation. We believe this looming deficit in capacity for American citizens and others to access the records of our democracy has not been effectively communicated outside of this agency. Further, the need for resources which could secure the equipment and staff to process and house the tsunami of records heading to NARA in a manner which might facilitate full-text search capability has not been communicated to OMB, or our Congressional oversight committees. This is akin to not calling for reinforcements when our positions are surely about to be overrun. Figuratively speaking, there is no cavalry on the horizon, and NARA opted not to send out an SOS.

In 2005 NARA awarded Lockheed Martin a design contract to build the foundation of ERA, and it has been announced the contract will end on September 30, 2011. As of January 5, 2011 the

actual costs of ERA totaled approximately $430 million.[1] This office has taken an active role in providing audit coverage to ERA. In reports and testimony we have identified a troubled program. ERA has experienced failed deliverables, numerous changes in requirements, cost overruns, material key staff turnover, uncertain funding, technological challenges, miscommunication between vital stakeholders etc. Throughout the saga this office has asked one question repeatedly:

> At the end of the contract, when the contractor turns in their keys and badges, what exact functionality will ERA provide to the most important stakeholder, the American citizen?

From program inception, research papers commissioned by NARA and published by the National Academy of Sciences extolled the benefits of full-text content searching. Indeed, NARA contractual documents defined that ERA shall be able to search assets based on their contents and be able to perform keyword, exact phrase, proximity and other types of searches. Simply speaking, this means that a document such as this very letter would be ingested into ERA, and anyone now or in the future would be able to locate it by searching the body of this text. This is consistent with the manner in which users navigate through existing publicly available search engines such as Google, BING, Yahoo etc. In fact, the commercial search engine applied to ERA not only has full-text content search capability, but we are advised it is the actual default option.

However, once publicly available, the base ERA system, other than for a select population of records, will not support full-text search of the contents of individual records.[2] Instead, users will generally only be able to search through metadata generated about the records, with limited records being made full-text searchable only after being identified as high-request records. This shortcoming will become ever more significant as billions of records begin to flow into NARA in the coming years. Indeed, a current key ERA staff member defined "we expect that the percentage of electronic records that are full text searchable will decrease over time, depending on 1) volume of records coming in and 2) resources available."

Compounding this situation is the fact NARA has deferred the capability for automated generation of records descriptions beyond the end of the ERA contract. In a computer system which does not search the content of records, the record descriptions take on additional importance as the only searchable narrative of the record's contents (presuming the descriptions are made part of the searchable metadata). However, as ERA has now been set-up, such descriptions will not be automatically generated by the system, but instead must be manually generated. Considering the massive amount of data expected to be put into the system, such a manual process will invariably create substantial, perhaps insurmountable, bottlenecks. Without full content searching, this potential delay in generating records' descriptions will degrade

---

1 According to the Office of Management and Budget's Federal IT Dashboard at
http://it.usaspending.gov/?q=content/cost-summary&buscid=799.

2 We realize not all records will have text to search (i.e., photos), and this Management Letter is limited to the significant portion of anticipated ERA-housed records which will contain text (i.e., e-mails, reports, databases, etc.).

ERA's usefulness in providing timely and full access to electronic records now and into the future.

It is unclear when, and by whom, the decision not to pursue full text searching of record contents was made. ERA's previous director advised the OIG in April of 2010 that this was a policy decision that would be made in sometime around the end of 2010 or the beginning of 2011. Other senior NARA staff stated there was never a requirement to search the full text of record contents, and thus this decision was made at the very beginning of the contract. When questioned about the reason for ERA's general lack of search capacity, senior NARA officials have had a variety of responses. Some adamantly stressed they believed the ERA requirements never included the capacity to search the full text of record contents. Others blamed the technical requirements, combined with reduced funding for the program. Still others claimed they believed researchers would not want full text search capability, or that being able to completely search record contents is not desirable for an electronic archive. Many of the "architects" of ERA have since left the program and Federal service.

We look forward to your prompt response to these concerns. Should you have any questions, please contact me at (301) 837-1532.

Paul Brachfeld
Inspector General