NATIONAL ARCHIVES

OFFICE *of* INSPECTOR GENERAL

Date : December 11, 2014

Reply to: Office of Inspector General (OIG)

Subject : Advisory Audit Report No. 15-04, Status Update of the Electronic Records Archives Census and Classified Instances

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

The purpose of this Advisory Audit Report is to update you on the status of the Electronic Records Archives (ERA) Census Data Storage (Census) and Classified Electronic Records Archives (CERA) Instances. The Census Instance, which stores the 2010 U.S. Census, cost over $1.4 million[1] and is functioning as intended. However, the National Archives and Records Administration (NARA) spent over $6 million on the development, software, hardware, and maintenance for CERA which did not meet its intended functionality and was never utilized. This was caused primarily by the ERA development contractor not being able to properly integrate the ERA Base System with CERA prior to the end of the contract. As a result, NARA continues to rely on legacy systems, external hard drives, and tapes to store and maintain classified records.

NARA built ERA to fulfill its mission in the digital age: to safeguard and preserve the records of our government, ensure that the people can discover, use, and learn from this documentary heritage, and ensure continuing access to the essential documentation of the rights of American citizens and the actions of their government. The total cost to develop the system was over $390 million[2]. The estimated annual cost to operate and maintain the ERA system is approximately $30 million[3].

One of NARA's primary challenges with ERA was to preserve different types of records along with the processes and documentation required for each type. Therefore, ERA was designed using separate subsystems, or instances, for each category of records. The initial three instances were the Federal Records Instance (Base ERA), deployed June 2008; the Executive Office of the

[1] The costs include development, hardware, software, migration, and maintenance.
[2] Source: NARA's ERA Exhibit 300 (5/24/2013 update) on the Office of Management and Budget's (OMB) Federal IT Dashboard.
[3] Source: NARA's ERA Exhibit 300 (5/24/2013 update) on OMB's Federal IT Dashboard. The estimated costs include the Operations and Maintenance contract, hardware/software licenses, technology refresh, government full time equivalents, and corrective and adaptive maintenance activities.

President Instance (EOP), deployed December 2008; and the Congressional Records Instance (CRI), deployed December 2009. Two additional instances, Census and CERA were developed in Fiscal Year 2011.

We continue our on-going effort to review aspects of the ERA system. Previous audits include "Audit of the Electronic Records Archives System's Ability to Preserve Records" dated February 15, 2013 (Report No. 13-03), and "Audit of the Base ERA System's Ability to Ingest Records" dated September 19, 2013 (Report No. 13-11) addressing weaknesses related to the preservation and ingest processes. We addressed issues in the EOP Instance in Management Letter No. 13-02, "Status of the Upgrade to the Electronic Records Archives Executive Office of the President System", dated October 18, 2012; Advisory Report No. 13-07, "Status Update of the Electronic Records Archives Executive Office of the President System Upgrade", dated January 31, 2013; and Re-issued Advisory Audit Report No. 14-14, "Status Update of the Electronic Records Archives Executive Office of the President Data Migration Project", dated June 25, 2014. This initiative focuses on the Census and CERA Instances.

Census Instance

The U.S. Census counts every resident in the United States. Mandated by Article I, Section 2 of the Constitution, the first Federal Population Census was taken in 1790, and has been taken every ten years since. The 2010 census contained ten questions about age, gender, ethnicity, home ownership, and household relationships. The data collected by the decennial census determines the number of seats each state has in the U.S. House of Representatives, and is also used to distribute billions in federal funds to local communities.

Due to the confidential nature of the census records, Title 13 of the United States Code specifies enhanced procedures NARA must observe to ensure the non-disclosure of this information and to segregate it from other ERA holdings. In addition, there is a 72 year restriction on access to this data.

The Census Instance was created to manage the storage and retrieval of records related to the 2010 Census. NARA spent over $1.4 million for development, hardware, software, migration, and maintenance of the Census Instance. In 2011, NARA received the 2010 census data, consisting of more than 300 terabytes.[4] The U.S. Census Bureau shipped the data on storage devices including 18 racks of equipment. In 2013, the census data was migrated to new storage devices as part of an effort to replace and consolidate storage hardware for various ERA instances. This was done to improve the availability and reliability of the ERA system, decrease risk of data loss, improve responsiveness of the ERA applications, lower the floor space allocation and operating and utility costs, and create greater storage capacity for additional records in the ERA system. According to a NARA official, the Census Instance is being maintained and functioning as intended.

[4] A terabyte is about 1 trillion bytes or about 1,000 gigabytes.

CERA Instance

With the understanding that NARA will be receiving classified data, the need for a classified instance of ERA was established, as the ERA Base System only supports housing unclassified data. CERA was intended to meet the specialized requirements for ingesting, storing, preserving, and accessing classified records. However, NARA spent over $6 million on the development, software, hardware, and maintenance for a system that did not meet its intended functionality, and was never utilized.

According to a NARA official, the ERA development contractor delivered the CERA Instance, but was unable to get the system to work. The primary issue was integrating the unclassified ERA Base System, and CERA. NARA wanted to create records schedules and transfer requests in the unclassified Base System, load the data into CERA, and then have the ability to search for the data across both platforms. When a records schedule or a transfer request is created, an ID sequence number is produced identifying the record. If a user searched for a classified record in the unclassified system, it would link the ID number to the record and tell the user the record was classified or unsearchable in this system. If the user had the appropriate security clearance and access authority, they should have been able to go into CERA and search on the ID number to find the record. However, the development contractor could not get the two systems to recognize one ID number across both systems. Essentially, Base ERA and CERA were air-gapped (i.e., no interconnection or communication) by design and out of necessity. It is this requirement that prevents the two instances from keeping ID numbers synchronized. This fundamental design parameter could not be overcome at the time. Therefore, users would not be able to locate the correct records in CERA. Accordingly, CERA was never populated with records or used.

As CERA is not used, NARA continues to rely on legacy systems, external hard drives, and tapes to store and maintain classified records. According to a NARA official, the legacy systems have limitations in terms of scalability and file format handling. He further stated CERA was supposed to integrate ingesting, processing, and accessing classified electronic records. Further, the CERA equipment is reaching its end of life, and may become defunct. NARA officials have had discussions about possible using this equipment for other purposes. However, no decisions on repurposing the CERA equipment have been made.

We will continue to monitor this situation to see if the CERA equipment can be repurposed and used in some capacity. If NARA management determines the equipment cannot be repurposed, all maintenance agreements should be cancelled as soon as possible.

Scope and Methodology

In order to accomplish our objective we interviewed responsible NARA officials; and reviewed contract documentation, cost information, project management documentation, and applicable law. The contents of this report were discussed with responsible NARA officials, and those officials generally agreed with the contents. Our performance audit was performed at Archives II in College Park, Maryland between December 2013 and September 2014. We conducted this 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. We believe that the evidence obtained provides a reasonable basis for our finding and conclusion based on our audit objective. No recommendations were made as the objective was to provide an update.

Please provide a written response to these matters within two weeks of the date of this letter. If you have any questions concerning the information presented in this Advisory Audit Report, please do not hesitate to contact me at (301) 837-3000. As with all OIG products, we will determine what information is publicly posted on our website from this Advisory Audit Report. Should you or management have any redaction suggestions based on FOIA exemptions, please submit them to my counsel within two weeks from the date of this letter. Should we receive no response within this time frame, we will interpret that as confirmation NARA does not desire any redactions to the posted report.

James Springs
Acting Inspector General

cc: Swarnali Haldar, Chief Information Officer