Audit of the Electronic Records Archives System's Ability to Preserve Records

OIG Audit Report No. 13-03

February 15, 2013

Table of Contents

Executive Summary

The National Archives and Records Administration (NARA) Office of Inspector General (OIG) completed an audit of the Electronic Records Archives (ERA) System's ability to preserve records. NARA developed the ERA system to enable the agency to realize its strategic vision: "ERA will authentically preserve and provide access to any kind of electronic record, free from dependence on any specific hardware or software, enabling NARA to carry out its mission into the future." During this audit, we assessed NARA's capability to preserve electronic records to ensure the continued existence, accessibility, and authenticity of electronic records over time. Further, we assessed future plans for increased functionality.

OMB Circular A-109, *Major Systems Acquisitions*, identifies a number of major system acquisition management objectives in ensuring each major system operates effectively in its intended environment and demonstrates a level of performance and reliability justifying the allocation of the Nation's limited resources for its acquisition and ownership. Similarly, NARA 805, *Systems Development Lifecycle* and its supplemental handbook, within the scope of this audit, provide policy on requirements establishment, and deployment and acceptance activity. The purpose of these activities is to install the system in the operational environment and conduct acceptance testing to ensure that the system and its associated products perform in accordance with specified technical and contractual requirements.

Although NARA's ERA System has completed its developmental phase, the preservation functionalities identified in the contract and used to promote the need for this major system acquisition were not fully achieved. This is particularly apparent in the system's inability to automate and scale the process of transforming electronic records into a format independent of specific hardware or software. This condition exists due to a number of inadequacies involving requirements management, acceptance testing, project communication, and status representation. As a result, the ERA system is currently unable to mitigate the risk of electronic format obsolescence[1]—a major objective identified by NARA in carrying out its mission into the future. Further, with the decreased funding and limited resources available among competing priorities in the Operational and Maintenance phase of the program, the likelihood of the ERA system meeting the preservation mission needs in the foreseeable future is further challenged.

Our audit identified several improvements to be made in further ERA preservation enhancements and efforts. We made five recommendations to more accurately identify the ERA's preservation functionality and to further ensure future enhancements are reflective of NARA's needs.

[1] NARA's ERA Requirements Document defines this risk in terms of inaccessibility of electronic records.

Background

According to original contract documents, the Electronic Records Archives (ERA) system was developed to enable the National Archives and Records Administration (NARA) to realize its strategic vision: "ERA will authentically preserve and provide access to any kind of electronic record, free from dependence on any specific hardware or software, enabling NARA to carry out its mission into the future." The ERA system as a whole represents a major system acquisition at NARA both in terms of mission criticality and financial resources. Further, it is the largest information technology project ever undertaken by NARA.

NARA began planning for the ERA system in the late 1990's, leading to the establishment and funding of ERA Program Management Office (PMO) in 2000. The ERA PMO was approved and funded under a separate line item of NARA's annual budget request to ensure annual funding for this mission-critical program. After five years of study and research by NARA into the possibilities, approaches, and design requirements of the ERA system, two companies—Lockheed Martin Corporation (LMC) and Harris Corporation—were selected to compete in designing a technical solution to preserve NARA's electronic information. At the time, NARA described ERA as "a revolutionary system that [would] capture electronic information, regardless of its format, save it permanently, and make it accessible on whatever hardware or software is currently in use."

In 2005, LMC was awarded the design contract to build the foundation of the ERA system that would be developed in five increments. In announcing the contract award, the former Archivist of the United States emphasized the importance of this mission-critical system, stating "the need for ERA is urgent, since there is an unprecedented number of electronic records now being created by the Government's departments and agencies. The most important of them will be preserved and will be accessible indefinitely. This simply must happen…ERA's failure is not an option."

As development continued into 2010, the ERA system became the subject of Office of Management and Budget (OMB) TechStat[2] Reviews. NARA took actions to address TechStat concerns, including accelerating ERA's development process for completion by the end of FY 2011. In June 2011, NARA's newly appointed Chief Information Officer (CIO) cited the TechStat Accountability Sessions as being instrumental in helping NARA assess and plan a successful path forward for ERA. During this timeframe, the ERA team revised requirements documentation, including those related to preservation.

NARA has described ERA as a "system of systems," with multiple components performing different archival functions. These include four essential functions the system must perform: Submission, Metadata, Repository, and Access. The Repository function involves the review and preservation of electronic records. This audit focuses on the preservation component of the Repository function.

[2] TechStat Accountability Session (TechStat) is a face-to-face, evidence-based accountability review of an IT investment; it enables the Federal Government to intervene to turn around, halt or terminate IT Projects that are failing or are not producing the results for the American people.

Objectives, Scope, Methodology

The overall objective of this audit was to evaluate and report upon NARA's capability in preserving electronic records to ensure the continued existence, accessibility, and authenticity of electronic records over time. Specifically, we assessed the ERA system's current capability of preserving electronic records and evaluated future plans for increased functionality.

To accomplish our objective, we interviewed key NARA and ERA personnel from the Offices of Research Services and Information Services. We reviewed the ERA contract, requirements, and program documentation. In addition, we gathered and reviewed historical ERA information, meeting minutes, status updates, and presentation material. Further, we examined NARA announcements and press releases regarding the ERA system and its preservation functionalities. We compared the capabilities of the production system with the latest system requirements[3], and identified plans for future enhancements. We examined applicable Federal and NARA policy and guidance including OMB Circular A-109, *Major Systems Acquisitions*; OMB Circular A-130, *Management of Federal Information Resources*; Federal Acquisition Regulation Part 39, *Acquisition of Information Technology*; NARA 801, *Capital Planning and Investment Control*; NARA 805, *Systems Development Lifecycle*.

Our audit work was performed at Archives II between February 2012 and November 2012. 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. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

[3] Electronic Records Archives Requirements Document (RD v4.1) dated 17 April 2011.

Audit Results

ERA System's Preservation Capabilities

Although NARA's ERA system has completed its developmental phase, the preservation functionalities identified in the contract and used to promote the need for this major system acquisition were not fully achieved. This is particularly apparent in the system's inability to automate and scale the process of transforming electronic records into a persistent, resilient format. This condition exists due to a number of inadequacies involving requirements management, acceptance testing, project communication, and status representation. As a result, despite the vast amount of time and resources devoted to this effort, the ERA system is currently unable to mitigate the risk of electronic format obsolescence—a major objective identified by NARA in carrying out its mission into the future. In addition, with the decreased funding and limited resources available among competing priorities in the Operational and Maintenance phase of the program, the likelihood of the ERA system meeting the preservation mission needs in the foreseeable future is further challenged.

OMB Circular A-109, *Major Systems Acquisitions*, defines "major systems" as programs that are critical to fulfilling an Agency mission, entail the allocation of relatively large resources, and warrant special management attention. OMB A-109 identifies a number of major system acquisition management objectives, which include ensuring each major system operates effectively in its intended environment and demonstrates a level of performance and reliability that justifies the allocation of the Nation's limited resources for its acquisition and ownership. In addition, OMB A-109 states each agency acquiring major systems should provide strong checks and balances by ensuring adequate system test and evaluation.

NARA 801, *Capital Planning and Investment Control*, which derives its authority, in part, from OMB A-109, establishes NARA's policy for Information Technology (IT) investment management. NARA 801 stipulates investments authorized under the directive must follow NARA 805, *Systems Development Lifecycle (SDLC)*. Further, if contractor assistance is used in the development of an information system, NARA 805 requires the contractor to be informed of NARA's SDLC procedures. NARA 805 and its supplemental handbook, within the scope of this audit, provide policy on requirements establishment, and deployment and acceptance activity. The purpose of these activities is to install the system in the operational environment and conduct acceptance testing to ensure the system and its associated products perform in accordance with specified technical and contractual requirements.

The ERA system as a whole represents a major system acquisition at NARA both in terms of mission criticality and financial resources. In terms of criticality, the former Director of ERA stated "ERA equals NARA…NARA as a paper archive is going to shrink over time, and as an electronic records archive is going to grow over time." The ERA system is the largest information technology project ever undertaken by NARA. Further, for the final year of ERA's development contract, the system represented over 18% of NARA's total requested appropriations. In total, NARA has recognized actual expenses exceeding $380 million in the development of the ERA system. To put this into perspective, NARA's annual appropriation for the entire agency in FY 2012 was $391.5 million. Therefore, a program of this magnitude and criticality is clearly subject to the developmental controls and oversight policies established at the Federal and agency levels for major system acquisitions. A common emphasis of these controls pertains to the establishment and definition of mission needs and system requirements.

Requirements Definition and Management

In order to gain an understanding of the early requirements related to ERA preservation, we reviewed the original ERA contract awarded to Lockheed Martin Corporation (LMC). The original contract was provided by ERA program office and Information Services personnel and included NARA's initial ERA Requirements Document. In the section entitled "Major System Capabilities," the initial Requirements Document states:

> "To achieve NARA's mission and support the broad range of its responsibilities, the system should eliminate or minimize records' dependence on any specific hardware or software. The system should maximize the types of electronic records and types of digital data it can handle. The system should be able to ingest electronic records from a wide variety of sources, including any entity in the Federal Government or private donors, created using any type of application on any computing platform. The system should be able to ingest electronic records currently in the holdings of NARA. The system should provide discovery and delivery of documentary material to anyone with interest and legal right of access, from now until the end of the republic. The system must accommodate unscheduled, permanent, and temporary electronic records regardless of record type, format, or physical media."

Further, the initial Requirements Document states the system must provide capabilities for automated archival processing of electronic records. Specific to preservation, these automated processes must include long term storage of electronic records and transformations of electronic records to maintain accessibility and authenticity. In

addition, within the scope of preservation, the initial Requirements Document states the following:

- ERA will authentically preserve and provide access to any kind of electronic record, free from dependence on any specific hardware or software, enabling NARA to carry out its mission into the future;
- ERA will ensure that electronic records transferred to NARA remain free from corruption and accessible regardless of changes in information technology;
- The system must provide capabilities for automated archival processing of electronic records themselves, including long term storage of electronic records and transformation of electronic records to maintain accessibility and authenticity; and
- ERA shall provide the capability to transform electronic records/data types into a hardware and software independent format.

The initial Requirements Document also describes NARA's preservation goal, which is to preserve electronic records in persistent formats enabling access to authentic electronic records indefinitely into the future. Based on these early requirements, it is clearly evident that the ability to preserve records in a persistent format is a required functionality and attribute of the ERA system. Many of these preservation requirements have remained throughout the revisions to the ERA Requirements Document. Furthermore, ERA Requirements Personnel indicated the Statement of Objectives (SOO) were the primary basis by which NARA and LMC negotiated the scope of the ERA increment requirements. In reviewing the SOO contained within the original ERA contract package, we identified a number of performance objectives developed by NARA. The following ERA performance objectives pertain specifically to preservation:

Measurement Indicator	2007	2008	2009	2010	2011	2012
Percentage of archival electronic holdings managed at the planned Preservation and Access Level	60%	80%	85%	88%	92%	95%
Percentage of archival electronic records preserved in a persistent format	5%	10%	15%	20%	25%	30%

In a memorandum issued to the GAO on 21 May 2010, the current Archivist of the United States stated "NARA has not made any changes to the original [ERA]

requirements documented in the 2003 Requirements Document." However, in the same memorandum, the Archivist states functional priorities will be mapped to the underlying contract requirements and the ERA Requirements Document will be updated as appropriate. Revisions made to the initial ERA Requirements Document in 2010 and 2011 removed some of the early requirements related to preservation. However, the last revision of the Requirements Document drafted for the ERA development contract, dated 17 April 2011, still contained many of the preservation requirement attributes identified in the original contract package. Further, additional preservation requirements were included in the latest Requirements Document, such as:

- The system shall provide the capability to select a set of ingested data files to be transformed to a different format;
- The system shall provide the capability to transform data files into ASCII[4] to ensure hardware and software independence;
- The system shall store the files output from transformations to XML format[5]; and
- ERA will enable the creation and management of preservation plans and strategies in order to drive the process of transforming electronic records to different persistent formats while maintaining their authenticity.

NARA's SDLC Handbook states Requirements Definition Activity focuses on developing detailed system specifications for the end product. As such, it begins to refine the Concept of Operations and define expected systems behavior in terms of performance, capacity, data inputs and outputs, processing, etc. From a project management perspective, the Requirements Definition activity will provide the inputs necessary to begin a detailed risk assessment and continued systems development planning. It is during this activity that the project scope should be closely examined to determine if resource allocation and project expectations have been assessed accurately. The Project Manager should use this information in discussions with Product Owners and the Guidance Team to revise project and product plans as needed to ensure success.

While interviewing and discussing preservation requirements with ERA Requirements Personnel, they acknowledged the requirements negotiation and definition process with LMC "wasn't that great." In describing this process, ERA Requirements Personnel stated NARA would start by providing LMC with a SOO which expressed the basic, top-level objectives of the acquisition for the particular increment. Next, NARA and LMC would negotiate the scope of the requirements for that increment based on the SOO. LMC

[4] American Standard Code for Information Interchange (ASCII) is a set of digital codes representing letters, numerals, and other symbols, widely used as a standard format in the transfer of text between computers.
[5] Extensible Markup Language (XML) is a flexible text format for creating structured computer documents in machine-readable form.

would then issue a proposal, and following NARA approval, LMC would decompose[6] the requirements. However, the requirements LMC decomposed did not map to those in NARA's initial Requirements Document. Therefore, when an increment was complete, NARA performed acceptance testing based upon the requirements decomposed by LMC because NARA's SOO was too high level and NARA's Requirements Document did not map to those in which LMC performed during the increment.

The ERA program office eventually decomposed NARA's requirements in July 2010. ERA Requirements personnel stated it was something they "should have done a long time ago, but hadn't." Until then, NARA's decomposed requirements were not mapped back to those developed and used by LMC in executing the contract. In terms of having LMC going back and mapping requirements for previous increments, ERA Requirements personnel stated at that point there would have been no real benefit for the effort. Despite developing their own decomposed set of requirements, NARA still tested against those developed by LMC in the final increment of the development contract.

A number of prior OIG audits and evaluations identified similar issues with ERA requirements management. Most recently, in July 2011 the OIG issued an advisory report in part highlighting ongoing concerns related to the lack of updated system requirements in accordance with SDLC criteria and policies[7]. The GAO had previously identified similar concerns. For example, in a June 2010 report[8] the GAO stated "NARA has not effectively defined or managed requirements for the ERA system…Although NARA established an initial set of high-level requirements to guide the system's development, these requirements are not traceable to work in later phases, or increments, of the system."

Preservation Framework Prototype and Demonstration

According to the ERA Business Analysis Team[9], for most of FY 2009, efforts were taken to develop and vet—on a conceptual level—what the ERA system would do in terms of preservation. Starting in FY 2009, during Increment 3 of the ERA contract, LMC began designing a preservation framework prototype based upon these concepts. According to the Increment 3 SOO, the preservation framework design was to incorporate the

[6] Requirements decomposition involves breaking down the work needed to execute the project objectives and required deliverables. Each descending level represents and increasingly detailed definition of the project work.
[7] OIG Advisory Report No. 11-16, Implementation Status of the Electronic Records Archives System Requirements.
[8] GAO Report No. 10-657, Electronic Records Archive: Status Update on the National Archives and Records Administration's Fiscal Year 2010 Expenditure Plan.
[9] The Business Analysis Team was later dissolved into the ERA Preservation Board and Working Group.

functionality necessary to plan, schedule, execute, monitor, and report preservation activities in ERA.

Beginning in December 2009, demonstrations of the preservation prototype were given to the user and stakeholder communities. During a meeting with the Advisory Committee for Electronic Records Archives (ACERA) in April 2010, the former ERA Program Director provided a status of the preservation efforts, indicating the prototype will identify the formats of ingested records and provide the ability to select tools to do preservation and preservation planning. The former ERA Program Director stated "Everything we heard back from the users, this went very well."

The final preservation prototype demonstration took place during the April 2010 ACERA meeting. During this meeting, the former ERA Program Director detailed the preservation prototype functionality as follows:

- A flexible framework to enable the deployment of various software tools for the purpose of transforming electronic records from one format to another;
- Enhanced capabilities to identify formats of ingested electronic records, and to persist their important archival and technical characteristics; and
- Initial electronic records preservation planning.

The preservation prototype was expected to represent the framework that would be put into the production system, allowing for the identification of the formats of ingested electronic records and the ability to select tools to perform preservation activities in ERA. However, during the demonstration, the ERA Business Analysis Team stated the prototype was limited in scope, specifically, it did not:

- Determine the optimal strategy for a given digital format;
- Determine the optimal tool for each transformation;
- Include ingest, archival catalog entries, search, or file extract services;
- Include preservation planning or risk assessment; or
- Allow for scalability.

In addition, the preservation prototype was limited in the format code types in which files could be transformed. The transformation tool used within the preservation framework prototype during the ACERA demonstration converted the Extended Binary Coded Decimal Interchange Code (EBCDIC)[10] to American Standard Code for Information Interchange (ASCII). The Business Analysis Team explained this tool was selected

[10] A standard eight-bit character code used in computing and data transmission.

because it represented a number of NARA's record holdings, and specifically, because the tool was available for free. The Business Analysis Team stated the next step was to get technical experts on staff to produce white papers identifying the optimal tools and formats used in electronic records preservation. In a meeting held in April 2012, the Team Leader of the ERA Preservation Working Group stated three such technical white papers had been written. However, in June 2012, the Team Leader announced only one white paper had been reviewed and approved at that point. The approved technical white paper pertained to the EBCDIC format—the same format used in the April 2010 ACERA preservation prototype demonstration.

Further, concerns were discussed during the ACERA preservation prototype demonstration related to the registry application used to identify the files in need of transformation. A member of the Business Analysis Team stated "if you don't have the right tools to recognize formats, if [the format] comes back as 'unknown,' you're kind of stuck." The ERA system uses the PRONOM[11] application to perform this function. However, early prototype testing discovered the application was unable to identify a number of file signatures of records ingested into ERA. In the minutes of an ERA Program Management meeting held on 13 August 2010, it was mentioned that there were potential issues involving the sole use of PRONOM as the file format registry and the ability to maintain file format identification if additional registries are added. These were identified as issues NARA must address. By the time ERA system was placed into production two years later, these issues had not been resolved.

As of April 2012, according to Preservation Programs (RX) personnel, PRONOM is unable to identify formats of between 20 and 70 percent of the records ingested into ERA. This further complicates efforts elsewhere in the ERA system (including the development of technical white papers mentioned earlier). The Team Leader of the ERA Preservation Working Group stated the preservation of electronic records begins with the identification of an inventory of the different formats; however, the ERA system's ability to perform this function remains insufficient.

Additionally, during the April 2010 ACERA meeting, the Business Analysis Team stated the preservation prototype was based upon Conceptual Framework version 1.1, however, they stated version 2.0 was to be completed in the summer of that year. Two years later, while interviewing the Team Leader of the ERA Preservation Working Group, he stated the conceptual framework is still in draft and efforts are still being taken to incorporate comments made during a peer review.

[11] A web-based technical registry to support digital preservation services, developed by the National Archives of the United Kingdom.

Acceptance Testing

NARA's IT Architecture Systems Development Lifecycle Handbook identifies deployment and acceptance activity as part of the systems development phase. The SDLC Handbook includes two types of testing activity: system testing and acceptance testing. Of these, system testing is conducted to validate the built system against the requirements. The purpose of the deployment and acceptance activity is to install the system in the operational environment and conduct acceptance testing ensuring the system and its associated products perform in accordance with specified technical and contractual requirements. According to the SDLC Handbook, acceptance testing involves the users and operators in understanding system usage and ensures the system is delivered and performing as intended.

While interviewing the Director of Preservation Programs, we inquired about the results of the preservation framework acceptance testing. She stated the Chief of Electronic Records Preservation (RXE) and his staff "are on the front lines in terms of acceptance testing." The Director stated that although it fell under her office's responsibility starting in the June-August 2011 timeframe, she could not really speak of the acceptance testing results. The Director stated she realizes at the end of FY 2011 there was a big push to complete ERA development, but due to NARA's ongoing reorganization, she did not believe it was her responsibility at that time.

In a separate interview, the Chief of Electronic Records Preservation indicated his staff participated in acceptance testing of the ERA preservation requirements. However, he stated it was not "full fledged acceptance testing" as compared to past projects. In addition, he stated the tests were all "canned" and did not reflect real world usage. Further, although RXE will be responsible for preservation and transformation aspects of the ERA system, the Chief of Electronic Records Preservation stated he did not sign off on any of the acceptance testing; instead he indicated the acceptance testing was approved by other senior ERA officials. NARA's CIO stated he delegated acceptance testing approval authority to the ERA Program Management Office (PMO) Test Lead. The CIO stated the process in place at the time stipulated test approval could not be granted if severity level 1 problems existed. The ERA PMO Test Lead reported 53 issues identified during acceptance testing of ERA preservation functionality; none of which reached the highest severity level (severity level 1). Even though a number of issues were identified at lower severity levels, one of the considerations cited in the Team Lead's acceptance recommendation was that there was no additional time left in the development contract for another build.

In addition, after interviewing ERA PMO personnel who conducted the preservation requirements acceptance testing and reviewing testing documentation, we found only limited preservation testing was performed. Further, despite SDLC Handbook guidelines, the preservation framework was never tested in a production representative environment. The ERA PMO Test Lead stated she performed the preservation acceptance testing in the Customer Acceptance Test (CAT) environment. She stated a test environment should be similar to the production environment; however the CAT environment used was "nothing like the production system." Further, the ERA PMO Test Lead indicated that LMC handed off products that appeared to be incomplete.

As described earlier, LMC's requirements decomposition was used instead of NARA's during acceptance testing, therefore the Test Report reviewed did not map back to NARA's Requirements Document. Although the Test Report indicated capabilities generally functioned as expected, it also highlighted concerns about limitations in the testing environment restricting the degree of testing performed. Despite an announcement at the ACERA meeting in April 2010 that the production version of ERA's preservation framework would be built in Increment 4, it was not actually delivered until the end of Increment 5—days before the conclusion of the development contract. Therefore, due to the short timeframe remaining for acceptance testing, LMC contract staff were limited in their availability to research and address issues identified by NARA's testers.

Production Environment and Current Status

According to ERA PMO personnel, at the end of the development phase, LMC had set up a "generic" preservation framework, however, only one algorithm—which allows for the conversion of EBCDIC to ASCII—was developed for the ERA production system. This is despite announcements made to stakeholders in April 2010 that additional tools would be developed in increments 4 and 5. Similarly, the Team Leader of the ERA Preservation Working Group stated the framework provided at the end of development is "bare bones." When asked if the preservation framework is currently functional, the Team Leader of the ERA Preservation Working Group stated he believes there is a good chance it works, adding he "sure as hell hopes it works."

The "ERA Status and Accomplishments" webpage on NARA's public website—which was last updated on 11 April 2012—states the electronic preservation processing capability in ERA is not yet occurring. According to the website, "although a framework has been developed into ERA, [NARA has] not yet converted records from one format to another, although the basic capability to do so has been designed into the system." The Team Leader of the ERA Preservation Working Group reiterated this by stating the preservation framework delivered at the end of FY 2011 and accepted by NARA had not

actually been used in production, further adding "so far it hasn't been needed." However, according to NARA's Performance Measurement and Reporting System (PMRS), during FY 2012, RXE was responsible for transforming 100 percent of standard EBCDIC records to ASCII records in accordance with NARA's preservation and access plan using the ERA system.

Despite the ERA preservation framework being accepted and placed into production at the end of FY 2011, it was not put into use by RXE until the end of April 2012. It was at that time RXE determined the preservation framework was non-functional. RXE contacted the ERA help desk, which led to the discovery that all preservation servers were inoperable. It took until June 2012 to resolve the issue enough to allow for limited functionality. However, RXE had already concluded it would not be possible to meet the PMRS goal of 100 percent EBCDIC conversion using ERA due to the system's inability to query a specific file format type from the Asset Catalog Database. As mentioned previously, the PRONOM application used by NARA has not consistently or adequately identified the file formats of records ingested into ERA. Further, RXE indicated the preservation framework did not include a workflow capability to assign and review preservation jobs. As a result of these functionality issues with the ERA preservation framework, the Chief of Electronic Records Preservation and the RXE staff continue to rely on legacy systems outside of ERA to perform preservation tasks. In a recent meeting, the CIO acknowledged ERA was unable to subsume legacy systems used for preservation. He stated that legacy preservation systems are now a required part of the ERA workflow. The CIO stated preservation activities will continue to utilize legacy systems for some time as opposed to replacing them with the ERA system.

In comparing NARA's ERA contract and Requirements Document to the preservation framework delivered and deployed by LMC on 24 September 2011, it is apparent the capabilities provided are not as robust as those established in NARA's requirements and advertised to stakeholder, users, and the public. This is due in part to NARA's inadequate requirements tracking and management during system development as described earlier. In addition, this negatively impacted acceptance testing, as NARA put itself in the position where it had to perform this testing based on LMC decomposed requirements. The delivery of the ERA preservation framework was further impaired due to the rushed and unrepresentative acceptance testing that took place in the days before the LMC contract ended.

In addition to those identified above, the Chief of Electronic Records Preservation and the RXE staff voiced concerns about the production version of the preservation framework in terms of the one conversion tool delivered at the end of the contract. As noted previously, part of the reason the EBCDIC to ASCII conversion tool was selected

was because it was free. RXE is hesitant about relying on this tool without a thorough understanding of its makeup. If the free tool is not reliable and sound, the Chief of Electronic Records Preservation stated it will only create unusable outputs.

Communication/Management Concerns

Based on interviews of ERA officials and personnel, as well as analysis of ERA meeting minutes, it became evident that communication channels involving the development of the ERA preservation framework were not always effective. Furthermore, important information and concerns did not consistently reach or were not fully acknowledged by NARA senior management. This disconnect was present in efforts involving both internal and external stakeholders.

NARA established ACERA as a deliberative body to advise the Archivist of the United States on technical, mission, and service issues related to ERA. This includes, but is not limited to, advising and making recommendations to the Archivist on issues related to the development, implementation, and use of the ERA system. However, comments made by the expert and stakeholder panels at such meetings have illustrated a level of frustration in the lack of information shared on a timely basis and the straightforwardness of the information provided by ERA officials. During the 7 April 2010 meeting, the former CIO stated "we didn't want to show you something that was giving bad results." Further, it was demonstrated during such meetings that user input was not always solicited. During the meeting, one user asked "when is someone going to come and ask us about user requirements?"

Conversely, input provided by ACERA expert and user panels were often not given adequate consideration as evidenced by the same recommendations being brought up meeting after meeting with no resolution**.** One such concept involved "technology chasing" which, in terms of preservation, places NARA in a position of constantly needing to retool its transformation capabilities to meet the ever growing universe of record file formats. ACERA members suggested taking a less costly policy based approach that would define common formats federal agencies use in creating records, and then place the burden on the agencies to provide conversion tools with their records if they choose to depart from the formats NARA manages.

Furthermore, an ACERA member recommended NARA give consideration to commercial preservation vendors who might be able to provide this service at less cost to NARA than developing evolving preservation capabilities in-house. Other recommendations related to placing the responsibility of preservation and storage of records on the agencies in a cloud environment. ACERA members explained that NARA

continues to take a paper archives approach by centralizing its storage of electronic records. As a result, a member of the panel voiced concern NARA's budget would increasingly go toward electronic storage costs at the expense of other NARA services. Such cloud concepts were also recommended by the Federal CIO. Following the completion of the development contract, the CIO and ERA program office have discussed moving some of the ERA system to the cloud.

As with the ACERA members, user input into requirements was also recently voiced as a concern by the Chief of Electronic Records Preservation, specifically in terms of the ERA preservation framework. During an April 2012 interview, he stated he was surprised no consideration was given to the preservation framework workflow process. Without such a process, there is no capability to assign transformation assignments or provide layers of review. The Chief of Electronic Records Preservation stated as a result "accuracy and authenticity is a big question mark." Further, he expressed concern that even though his team is responsible for preservation, they were not consulted or involved in the development of the ERA preservation process. However, Business Analysis Team meeting minutes during the FY 2010 timeframe indicate some degree of participation by Electronic Records Preservation staff in the ERA preservation prototype design and review efforts.

Additionally, during our meeting with the Chief of Electronic Records Preservation, he provided a list of needs and concerns his team identified related to ERA Preservation Transformation. The list identified issues related to the lack of a workflow process, inability to accurately query the ERA asset catalog, inexistence of processes for verification and running statistical reports, and inability to differentiate between certain file formats. The Chief of Electronic Records Preservation indicated he voiced his concerns and provided this list to the Director of Preservation Programs and other ERA officials after it was compiled in October 2011.

During a meeting with the Preservation Programs Department's ERA Transition Officer (who also serves as the Team Leader of the ERA Preservation Working Group), he also provided a list of ERA deficiencies. The list was developed by the Electronic Records Preservation Working Group. The list included concerns related to records characterization, compound records, contextual metadata, transformation tools, enhancements to preservation framework, the capability to fully verify structured data, and the incorporation of the planning portion of the ERA framework.

Despite discussing major deficiencies and concerns of the ERA preservation framework with both the Preservation Programs Department's Chief of Electronic Records Preservation and ERA Transition Officer, when we later interviewed the Director of

Preservation Programs Department, she indicated she had not received a report on things that need to be done, or the need for strengthening or enhancements. Another Preservation Programs official at the meeting stated the transformation tool works fine, but there are concerns with the record search capability. However, the Preservation Programs official stated it is "really early days to evaluate that framework," and it is more of a "confidence" issue, indicating the RXE staff does not yet have much experience in transforming records in ERA. Based on these interviews, it appeared as though there was a major disconnect between the staff with knowledge of the production system and those responsible for the management of the system. Further, when we asked about the results of the preservation framework acceptance testing, the Director of Preservation Programs stated she did not oversee the ERA preservation function at the time of the testing, although she indicated the reorganization involving her office had taken place prior to the end of the ERA's development phase.

In addition, as mentioned earlier: ACERA members, the Federal CIO, NARA's CIO, and ERA PMO personnel all discussed moving elements of the ERA system to the cloud. However, when we asked the Director of Preservation Programs if she was aware of such efforts or considerations, she stated she had not heard anything regarding ERA and the cloud. Further, an RX official at the meeting added "there is a lot more [we] would have to know before considering an ERA cloud as a viable alternative."

Promotional/Status Reporting

From the program's early inception until present, NARA has used various methods to communicate the capabilities, functionality, status, and progress of the ERA system. Some of these methods included conferences, press releases, white papers, enterprise architecture updates, web pages, and advisory committee discussions. Despite the reduction in scope of the ERA's preservation framework functionality, most of the promotional documentation reviewed remained consistent with the originally defined mission needs and requirements. Further, a number of status updates provided to stakeholders reflecting the progress of ERA preservation efforts were often inaccurate. Examples of the promotional and status documentation spanning the ERA development timeframe are summarized below:

- In terms of preservation, as early as 2003, during the 18th Annual Preservation Conference at NARA, the director of the ERA program pressed for an archival approach to preserving electronic records focused on the properties of records that must be preserved, rather than the artifacts of specific technologies used to create, store, or communicate them. Further, the director discussed preserving the data in

a persistent form, which makes the records independent of any particular hardware or software.

- In a press release dated 8 September 2005 announcing the award of the ERA contract to LMC, NARA announced "The ERA system will capture and preserve the electronic records of the federal government, regardless of format, ensure hardware and software independence, and provide access to the American public and federal officials." Further, in the press release, the former Archivist of the United States described the purpose of the system, stating "The Electronic Records Archives' goal is clear and simple: a system that accepts, preserves, and makes accessible—far into the future—any type of electronic document." The NARA press release continues, stating "ERA will be a comprehensive, systematic, and dynamic means for preserving virtually any kind of electronic record, free from dependence on any specific hardware or software. When operational, ERA will support NARA's mission by making it easy for the public and government officials to discover, use, and trust the records of our government, and to make it easy for NARA to deliver those records in formats people can use."

- Shortly after the press release mentioned above, NARA Notice 2005-292 was issued. In terms of preservation, the NARA Notice states "ERA will enable NARA to authentically preserve any type of electronic record created by any entity in the Federal Government and provide this electronic information anytime and anyplace to anyone with a legal right to access it."

- In the February 2007 issue of the ERA publication "For the e-Record," the ERA Transition Officer states "the problem is that there is a plethora of other kinds of electronic records that also must be preserved for the long term—independent of any particular hardware or software." In the same publication, an article states NARA and LMC are working hard to tell the users how ERA will work and what impact it will have on NARA. Further, it states "be assured…we will do the best we can to keep everyone informed about ERA's progress." Lastly, the publication includes a chart indicating initial appraisal and preservation plans were developed in Increment 1, search and preservation frameworks and full preservation plans will be completed in Increment 2 (FY 2008), and expanded preservation and capacity will take place in Increment 3-5 (FY 2009-2012). As described previously, these dates were not met.

- In the November 2009 ACERA meeting, NARA ERA Preservation personnel indicated a prototype based on the conceptual framework for digital preservation

was to be constructed. Further, they stated that sustainable format is so important that NARA has promulgated guidance regarding sustainable formats to the federal community. The ERA preservation personnel stated NARA contemplates building a staff of IT specialists with the task of working with archivists to identify those formats that will be most challenging, and to assess the needs of the research community for certain formats. This would be a collaborative process, determination would be recorded into a reference-driven transformation plan, and the ERA system would execute the transformation on specific groups of records. These efforts have not yet been completed.

- In the April 2010 ACERA meeting, the ERA Transition Officer indicated the ERA preservation conceptual framework would be completed by the summer of 2010. Further, he stated technical white papers for various file formats would be completed in 2010. Lastly, the ERA Transition Officer stated "first we have to ingest everything into ERA and now we have a tool that tells us what we have." However, as described previously, none of these statements were accurate. ERA's preservation conceptual framework was only recently completed, the first technical white paper for the EBCDIC format was not approved until June 2012, and up until recently the tool used to identify the records ingested into ERA was unsuccessful in identifying file formats as high as 70% of the time.

- ERA's Requirements Document version 4.0, dated 30 July 2010, was posted on NARA's public webpage and remains available for view under the header "ERA Project Information." This version of the Requirements Document for the most part contains the same preservation requirements as those contained in the original LMC contract, which reflect the capability to "eliminate or minimize records' dependence on any specific hardware or software." Further, there is little indication the scope of the preservation functionalities and capabilities has been reduced based on the project documentation available on the ERA's public website.

- During the November 2011 ACERA meeting, the CIO stated the core focus is on effective collection and preservation of electronic records. The CIO further stated there is currently a framework associated with preservation but it still needs to be incorporated into a standard process as new file formats and preservation needs arise. These comments were made despite the preservation framework not functioning when it was put into the production environment after the end of the development contract in September 2011. Further, during this meeting the Electronics Records Lifecycle Coordinator stated the original preservation

requirements statements became obsolete between 2009 and 2010; however, this was not reflected in the revised Requirements Documents.

- Further, the Application Architecture component of NARA's Enterprise Architecture specifies the business applications and service components of the agency. In the most recent version, dated 13 September 2011, the Application Architecture description of the ERA system states "ERA is a comprehensive and dynamic means for preserving virtually any kind of electronic record, free from dependence on specific hardware or software. ERA will make it easy for customers to find records and easy for NARA to deliver them."

- On 20 December 2011, the Archivist of the United States approved a charter for the Electronic Records Preservation Board and Electronic Records Preservation Working Group on Electronic Formats. The charter states "ERA has the capability to ingest, validate, and verify records from agencies, store them in secure locations, and use 'plug-in' tools to 'transform' records to new digital file formats when needed to avoid obsolesce or improve access." However, this Charter was drafted without the knowledge of whether or not the deployed version of the preservation framework was even functional.

- Up until recently, Archives.gov contained a webpage entitled "ERA Misconceptions and Facts." One of the misconceptions listed pertained to the ERA system not being able to solve the problem of long-term preservation of electronic records as hardware and software technology changes over time. The webpage refuted this, stating "ERA allowed NARA to make a quantum leap forward in the preservation of electronic records and building a flexible and adaptable framework that will let NARA evolve as electronic recordkeeping evolves."

NARA officials expressed the importance of ensuring information provided to the ERA stakeholders is accurate and representative of the program. On June 28, 2012, the Electronics Records Lifecycle Coordinator in coordination with the CIO and COO, issued a document entitled "Lessons Learned from NARA's Electronic Records Archives Project." As an introduction to the paper the Electronics Records Lifecycle Coordinator states she and the CIO "agree that NARA's future credibility depends on being very honest about what the ERA experience was like so it's clear that we understand and are prepared to move forward." In the Lessons Learned paper, the Electronics Records Lifecycle Coordinator states:

> "ERA has laid the groundwork for a sustainable preservation solution for the National Archives, but work remains to be done in several areas. NARA is continuing work on preservation policies, including a risk assessment methodology, which will determine when staff would intervene to preserve the content of a record using something other than its transfer format. ERA also faces challenges in improving the process of format identification, a necessary precondition to format migration. The existing NARA collection of electronic records includes many records in older formats or encoding schemes that are not currently recognized by tools such as DROID. NARA is actively supporting the expansion of the set of formats included in PRONOM through sharing of the work of research partnerships but more work needs to be done to automate format identification in ERA."

Although the Lessons Learned paper describes a number of ERA preservation deficiencies—including those identified previously in this report—it does not provide a clear picture of where the preservation capabilities currently stand. According to the RXE personnel responsible for performing this activity, the preservation framework is barely operable. Further, the issue involving the PRONOM format identification application was identified over two years ago with little indication of any resolution.

Future ERA Preservation Efforts and Constraints

Following the conclusion of the development phase of the ERA system at the end of FY 2011, the Archivist of the United States issued a charter establishing the Business Requirements Group (BRG). The BRG "exists to identify and express the business requirements of NARA business units for ERA and provide guidance to the ERA Program Management Office on these subjects." The BRG is responsible for providing high-level business direction for ERA's ongoing evolution, in part by prioritizing potential changes from a NARA-wide perspective and reviewing and approving changes to the business requirements, goals, and prioritization agreed on by the group.

During our interviews we inquired about the future functionality of the ERA preservation framework. A member of the BRG from the Preservation Programs Office indicated there is not enough money to address all the preservation priorities, "so it can be very frustrating." Further, the ERA Transition Officer stated he submitted proposals to the BRG with preservation requirements. However, he explained that with the new Operations and Maintenance contract award (and subsequent protest) no preservation work was planned for FY 2012. He stated the executive management team had not selected the preservation requirements as top priority. BRG members from the ERA PMO provided a similar response, indicating of a list of over 100 BRG tasks submitted, none of the preservation tasks has "risen to the top of the BRG list." Preservation

capabilities and requirements will have to continue competing for limited funds in future years against other ERA priorities.

According to the Electronic Records Preservation Board and Electronic Records Preservation Working Group on Electronic Formats charter, future work will focus on identifying—for each digital format—the action needed to prevent format obsolescence which potentially makes the records difficult to use or unusable. Part of these efforts involve preparing additional technical white papers (such as that mentioned earlier for the EBCDIC file format). In a meeting with the Director of Preservation Programs and RX official, the RX official stated NARA does not have to write every technical white paper—as a lot of the documentation is being globally developed—so NARA is not alone in these efforts. Further, she stated "although the technical white papers may sound like a big piece of the work, technically it's not." Based on the charter, future preservation efforts will also include developing preservation and access plans, identifying formats and records characteristics in NARA's stewardship universe, identifying emerging formats, and conducting an ongoing technology review.

The "Lessons Learned" paper issued by the ERA Electronic Records Lifecycle Coordinator also mentions these efforts, stating NARA is continuing work on preservation policies, including a risk assessment methodology, which will determine when staff would intervene to preserve the content of a record using something other than its transfer format. The paper also indicates NARA is actively supporting the expansion of the current set of formats included in PRONOM through sharing of the work of research partnerships, but more work needs to be done to automate format identification in ERA. Further, the ERA Electronic Records Lifecycle Coordinator states NARA anticipates someday it will request funding to begin a new development phase to create ERA 2.0. The Coordinator states since ERA's purpose is to preserve electronic records permanently, the current ERA system was designed to evolve. It will need to take advantage of better hardware and software as it becomes available so it can continuously improve to better meet the changing needs of federal agencies, researchers, and NARA staff.

In a recent meeting, the CIO emphasized NARA's early vision "to develop a revolutionary system that [would] capture electronic information, regardless of its format, save it permanently, and make it accessible on whatever hardware or software currently in use" did not represent the actual ERA system development requirements, but instead outlines what needs to be considered for future needs. He stated digital preservation is still immature and requires further invention and innovation. The CIO stated digital preservation research has not yet resulted in the tools needed to reach NARA's preservation vision, which may still be decades away. Further, he stated it may be more

cost effective to continue using legacy systems and manual processes to perform preservation activities. The CIO stated there has been a lot of effort within the Federal Government related to digital preservation, but NARA's efforts are just part of the journey and it needs to be recognized that we are still in the early stages of solving a very critical problem.

Other future concerns the ERA system must contend with relate to the need for a Technology Refresh for the continued successful operation of the ERA system. The Information Systems Business Needs Summary sponsored by the CIO indicates the hardware and software comprising the ERA system is over five years old—the majority of which has either surpassed the manufacturers' End of Life (EOL) date or will do so within the next 12 months. According to the summary, as the system components approach and surpass EOL dates, their reliability greatly decreases and poses increased risks of unreliability, data loss, or unavailability of the ERA system. Based on its research and industry best practices, the ERA Program Management Office has determined all ERA system hardware, software and components shall be planned to be refreshed during a repeating four year cycle. The summary indicates this refresh will provide the system with newer, more reliable hardware and software incorporating improvements in efficiency and security.

Conclusion

The ERA system is the largest information technology project ever undertaken by NARA. However, the development of ERA's Preservation Framework was not conducted in accordance with Federal and Agency system development and acquisition policy. This is particularly apparent in the ERA system's inability to automate and scale the process of transforming electronic records into a persistent, resilient format. The initial contract for the ERA system states "the purpose of the ERA system is to enable NARA to realize its strategic vision: 'ERA will authentically preserve and provide access to any kind of electronic record, free from dependence on any specific hardware or software, enabling NARA to carry out its mission into the future.'" After over six years of development and an evolving system purpose, the preservation framework delivered is barely functional and the scope of its capabilities has been greatly reduced.

Throughout its development, NARA announcements and status updates have described the ERA preservation achievements in terms of "a quantum leap forward." However, after the end of the development phase, NARA's ERA status and accomplishments website simply states "the electronic preservation processing capability in ERA is not yet occurring." This lack of functionality is a direct result of numerous inadequacies identified during this audit involving requirements management, acceptance testing, project communication, and status representation. Now in the operations and

maintenance phase, without efforts to correct these issues, NARA's ERA system preservation needs will remain unrealized.

Recommendations

We recommend NARA's Chief Information Officer:

1. Ensure the ERA Program Manager follows NARA 805 SDLC Handbook and System Development Guidelines for any enhancements or modifications to ERA, including the Requirements Definition Activity and Requirements Review Process.
2. Establish a test environment for ERA that is representative of the production environment and use this test environment to ensure future enhancements or modifications to the system perform in accordance with specified technical and contractual requirements.
3. Implement a process for documenting, analyzing, and tracking suggestions and recommendations made by ERA stakeholders and ACERA.
4. Conduct and document a thorough assessment of the production version of the ERA system's preservation framework capabilities.
5. Establish a quality control process for reporting ERA preservation status to internal and external stakeholders and the public.

Management Response

Management concurred with the recommendations.

Appendix A – Acronyms and Abbreviations

ACERA	Advisory Committee for Electronic Records Archives
ASCII	American Standard Code for Information Exchange
BRG	Business Requirements Group
CAT	Customer Acceptance Testing
CIO	Chief Information Officer
COO	Chief Operating Officer
DROID	Digital Record Object Identification
EBCDIC	Extended Binary Coded Decimal Interchange Code
EOL	End of Life
ERA	Electronic Records Archives
GAO	Government Accountability Office
IT	Information Technology
LMC	Lockheed Martin Corporation
NARA	National Archives and Records Administration
OIG	Office of Inspector General
OMB	Office of Management and Budget
PMO	Program Management Office
PMRS	Performance Measurement and Reporting System
SDLC	Systems Development Lifecycle
SOO	Statement of Objectives
XML	Extensible Markup Language

Appendix B - Management's Response to the Report

NATIONAL ARCHIVES

Date: FEB 11 2013

To: James Springs, Acting Inspector General

From: David S. Ferriero, Archivist of the United States

Subject: Comments on revised draft of OIG Draft Audit Report 13-03, "Audit of the Electronic Records Archives System's Ability to Preserve Records"

We thank you for your revised draft report entitled, "Audit of the Electronic Records Archives System's Ability to Preserve Records." In particular, we appreciate your efforts to work with the Information Services office to obtain feedback and make several changes to the report in preparation of the final draft.

We concur with the recommendations in the report.

DAVID S. FERRIERO
Archivist of the United States

NATIONAL ARCHIVES *and*
RECORDS ADMINISTRATION
8601 ADELPHI ROAD
COLLEGE PARK, MD 20740-6001
www.archives.gov

Appendix C - Report Distribution List

David S. Ferriero, Archivist of the United States (N)
Jay Bosanko, Chief Operating Officer (C)
Michael Wash, Executive for Information Services and Chief Information Officer (I)
Mary Drak, Performance and Accountability Staff (CP)