United States Government Accountability Office

GAO

Testimony

Before the Subcommittee on Information Policy, Census, and National Archives, Committee on Oversight and Government Reform, House of Representatives

For Release on Delivery
Expected at 2 p.m. EST
November 5, 2009

NATIONAL ARCHIVES

Progress and Risks in Implementing its Electronic Records Archive Initiative

Statement of David A. Powner, Director
Information Technology Management Issues

November 5, 2009

NATIONAL ARCHIVES

Progress and Risks in Implementing its Electronic Records Archive Initiative

GAO Accountability • Integrity • Reliability

Highlights

Highlights of GAO-10-222T, a testimony before the Subcommittee on Information Policy, Census, and National Archives, Committee on Oversight and Government Reform, House of Representatives

Why GAO Did This Study

Since 2001, the National Archives and Records Administration (NARA) has been working to develop a modern Electronic Records Archive (ERA) system, a major information system that is intended to preserve and provide access to massive volumes of all types and formats of electronic records. The system is being developed incrementally over several years, with the first two pieces providing an initial set of functions and additional capabilities to be added in future increments. NARA plans to deploy full system functionality by 2012 at an estimated life-cycle cost of about $550 million.

NARA originally planned to complete the first segment of ERA in September 2007. However, software and contracting problems led the agency and its contractor Lockheed Martin to revise the development approach. The revised plan called for parallel development of two different increments: a "base" ERA system with limited functionality and an Executive Office of the President (EOP) system to support the ingestion and search of records from the outgoing Bush Administration.

GAO was asked to summarize NARA's progress in developing the ERA system and the ongoing risks the agency faces in completing it. In preparing this testimony, GAO relied on its prior work and conducted a preliminary review of NARA's fiscal year 2010 ERA expenditure plan.

View GAO-10-222T or key components. For more information, contact David A. Powner at (202) 512-9286 or pownerd@gao.gov.

What GAO Found

NARA has completed two of five planned increments of ERA, but has experienced schedule delays and cost overruns, and several functions planned for the system's initial release were deferred. Although NARA initially planned for the system to be capable of ingesting federal and presidential records in September 2007, the two system increments to support those records did not achieve initial operating capability until June 2008 and December 2008, respectively. In addition, NARA reportedly spent about $80 million on the base increment, compared to its planned cost of about $60 million. Finally, a number of functions originally planned for the base increment were deferred to later increments, including the ability to delete records and to ingest redacted records. In fiscal year 2010, NARA plans to complete the third increment, which is to include new systems for Congressional records and public access, and begin work on the fourth.

GAO's previous work on ERA identified significant risks to the program and recommended actions to mitigate them. Specifically, GAO reported that NARA's plans for ERA lacked sufficient detail to, for example, clearly show what functions had been delivered to date or were to be included in future increments and at what cost. Second, NARA had been inconsistent in its use of earned value management (EVM), a project management approach that can provide objective reports of project status and early warning signs of cost and schedule overruns. Specifically, GAO found that NARA fully employed only 5 of 13 best practices for cost estimation that address EVM. Further, NARA lacked a contingency plan for ERA to ensure system continuity in the event that normal operations were disrupted. For example, NARA did not have a fully functional backup and restore process for the ERA system, a key component of contingency planning for system availability.

To help mitigate these risks, GAO recommended that NARA:

- include details in future ERA expenditure plans on the functions and costs of completed and planned increments;
- strengthen its earned value management process following best practices; and
- develop and implement a system contingency plan for ERA.

NARA reported in its most recent expenditure plan that it had taken actions to address these recommendations.

Mr. Chairman and Members of the Subcommittee:

I appreciate the opportunity to participate in today's hearing on the National Archives' (NARA) Electronic Records Archive system (ERA). Since 2001, NARA has been working to develop this system which is intended to preserve and provide access to massive volumes of all types and formats of electronic records by automating NARA's records management and archiving life cycle. The system is to consist of

- infrastructure elements, such as hardware and operating systems;
- business applications that will support the transfer, preservation, dissemination, and management of all types of records and the preservation of and online access to electronic records; and
- a means for public access via the Internet.

In view of its complexity, the system is being developed incrementally over several years; the first two pieces (or increments) of the ERA system provided an initial set of functions for managing federal and presidential records. NARA plans to add additional capabilities in future increments.

As agreed, my testimony today will summarize NARA's progress in developing the ERA system and the ongoing risks NARA faces in successfully completing it. My comments today are based on our prior work in this area,[1] as well as a preliminary review of NARA's fiscal year 2010 ERA expenditure plan. Our work was conducted 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

[1]See GAO, *Electronic Records Archives: The National Archives and Records Administration's Fiscal Year 2009 Expenditure Plan*,GAO-09-733 (Washington, D.C.: July 24, 2009); *Information Management: Challenges in Implementing an Electronic Records Archive*, GAO-08-738T (Washington, D.C.: May 14, 2008); *Information Management: The National Archives and Records Administration's Fiscal Year 2007 Expenditure Plan,* GAO-07-987 (Washington, D.C.: July 27, 2007); and *Electronic Records Archives: The National Archives and Records Administration's Fiscal Year 2006 Expenditure Plan,* GAO-06-906 (Washington, D.C.: Aug. 18, 2006).

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.

Background

The ability to find, organize, use, share, appropriately dispose of, and save records—the essence of records management—is vital for the effective functioning of the federal government. In the wake of the transition from paper-based to electronic processes, records are increasingly electronic, and the volumes of electronic records produced by federal agencies are vast and rapidly growing, providing challenges to NARA as the nation's recordkeeper and archivist.

Besides sheer volume, other factors contributing to the challenge of electronic records include their complexity and their dependence on software and hardware. Electronic records come in many forms: text documents, e-mails, Web pages, digital images, videotapes, maps, spreadsheets, presentations, audio files, charts, drawings, databases, satellite imagery, geographic information systems, and more. They may be complex digital objects that contain embedded images (still and moving), drawings, sounds, hyperlinks, or spreadsheets with computational formulas. Some portions of electronic records, such as the content of dynamic Web pages, are created on the fly from databases and exist only during the viewing session. Others, such as e-mail, may contain multiple attachments, and they may be threaded (that is, related e-mail messages are linked into send–reply chains).

In addition, the computer operating systems and the hardware and software that are used to create electronic documents can become obsolete. If they do, they may leave behind records that cannot be read without the original hardware and software. Further, the storage media for these records are affected by both obsolescence and decay. Media may be fragile, have limited shelf life, and become obsolete in a few years. For example, few computers today have disk drives that can read information stored on 8- or 5¼-inch diskettes, even if the diskettes themselves remain readable.

Another challenge is the growth in electronic presidential records. The Presidential Records Act[2] gives the Archivist of the United States responsibility for the custody, control, and preservation of presidential records upon the conclusion of a President's term of office. The act states that the Archivist has an affirmative duty to make such records available to the public as rapidly and completely as possible consistent with the provisions of the act.

In response to these widely recognized challenges, the Archives began a research and development program to develop a modern archive for electronic records. In 2001, NARA hired a contractor to develop policies and plans to guide the overall acquisition of an electronic records system. In December 2003, the agency released a request for proposals for the design of ERA. In August 2004, NARA awarded two firm-fixed-price[3] contracts for the design phase totaling about $20 million—one to Harris Corporation and the other to Lockheed Martin Corporation. On September 8, 2005, NARA announced the selection of Lockheed Martin Corporation to build the ERA system. The contract with Lockheed is a cost-plus-award-fee contract[4] with a total value through 2012 of about $317 million. As of April 2009, the life-cycle cost for ERA through March 2012 was estimated at $551.4 million; the total life-cycle cost includes not only the development contract costs, but also program management, research and development, and program office support, among other things. Through fiscal year 2008, NARA had spent about $237 million on ERA, including about $112 million in payments to Lockheed Martin.

The purpose of ERA is to ensure that the records of the federal government are preserved for as long as needed, independent of the

[2]44 U.S.C. 2203(f)(1).

[3]According to the Federal Acquisition Regulation, a firm-fixed-price contract provides for a price that is not subject to any adjustment on the basis of the contractor's cost experience in performing the contract. This type of contract places on the contractor maximum risk and full responsibility for costs and resulting profit or loss.

[4]A cost-plus-award-fee contract is a cost reimbursement contract that provides for a fee consisting of a base amount fixed at the inception of the contract plus an award amount that may be given based upon a judgmental evaluation by the government of contract performance.

original hardware or software that created them. ERA is to provide the technology to ensure that NARA's electronic records holdings can be widely accessed with the technology currently in use.

The system is to enable the general public, federal agencies, and NARA staff to search and access information about all types of federal records, whether in NARA custody or not, as well as to search for and access electronic records stored in the system. Using various search engines, the system is to provide the ability to create and execute searches, view search results, and select assets for output or presentation.

NARA currently plans to deliver ERA in five separate increments:

- Increment 1, also known as the ERA base, included functions focused on the transfer of electronic records into the system.
- Increment 2 includes the Executive Office of the President (EOP) system, which was designed to handle electronic records from the White House at the end of the previous administration. The EOP system uses an architecture based on a commercial off-the-shelf product that supplies basic requirements, including rapid ingest of records and immediate and flexible search of content. Increment 2 also includes basic case management for special access requests.[5]
- According to NARA's 2010 ERA expenditure plan, Increment 3 is to include new Congressional and Public Access systems. It is also to augment the base system with commercial off-the-shelf technology to increase flexibility and scalability. NARA plans to complete this increment by June 2010.
- Increments 4 and 5 are to provide additional ERA functionality, such as backup and restore functions and wider search capabilities, and provide full system functionality by 2012.

[5]These are requests NARA receives from the current and former administrations, Congress, and the courts for access to presidential records.

NARA Has Completed Two of Five ERA Increments, but Also Experienced Schedule Delays and Cost Overruns While Deferring Functionality

NARA's progress in developing ERA includes achieving initial operating capability for the first two of its five planned increments. However this progress came after NARA had experienced significant project delays and increased costs. NARA also deferred functions planned for Increment 1 to later increments.

As we reported in 2007,[6] the initial operating capability for Increment 1 was originally scheduled to be achieved by September 2007. However, the project experienced delays due to factors such as low productivity of contractor software programmers, difficulties in securing an acceptable contract to prepare the site that was to house the system, and problems with software integration. These delays put NARA's initial plan to use ERA to receive the electronic presidential records of the Bush Administration in January 2009 at risk.

In response, NARA and Lockheed Martin agreed to a revised schedule and strategy that called for the concurrent development of two separate systems, which could later be reintegrated into a single system:

- First, they agreed to continue development of the original system but focused the first increment on the transfer of electronic records into the system. Other initially planned capabilities were deferred to later increments, including deleting records from storage, searching item descriptions, and ingesting records redacted outside of the system. NARA now refers to this as the "base" ERA system. Initial operating capability for this increment was delayed to June 2008.
- Second, NARA conducted parallel development of a separate increment-dedicated initially to receiving electronic records from the outgoing Bush Administration in January 2009. This system,

[6] GAO-07-987.

referred to as the Executive Office of the President (EOP) system, uses a different architecture from that of the ERA base: it was built on a commercial product that was to provide the basic requirements for processing presidential electronic records, such as rapid ingestion of records and the ability to search content. NARA believed that if it could not ingest the Bush records in a way that supported search and retrieval immediately after the transition, it risked not being able to effectively respond to requests from Congress, the new administration, and the courts for these records—a critical agency mission.

As we reported earlier this year,[7] NARA certified that it achieved initial operating capability for Increment 1 in June 2008, following its revised plan. According to NARA's 2010 expenditure plan, this increment cost $80.45 million to deliver, compared to a planned cost of $60.62 million.

NARA also reported that it completed Increment 2 on time in December 2008 at a cost of $10.4 million (compared to a planned cost of $11.1 million). However, it was not functioning as intended because of delays in ingesting records into the system. Specifically, before the transition, NARA had estimated that the Bush electronic records would be fully ingested into EOP, where they would be available for search and retrieval, by May 2009. However, as of April 27, only 2.3 terabytes of data were fully ingested into the EOP system. This constituted about 3 percent of all Bush Administration unclassified electronic records.[8] NARA later estimated that ingest of all 78.4 terabytes of unclassified records would not be complete until October 2009. In its recently released 2010 expenditure plan, NARA reported that the Bush records were fully ingested into EOP by September 2009.

[7]GAO-09-733.

[8]NARA's original EOP plans included a National Security System. NARA subsequently deferred the capability to ingest classified national security data, stating that the volume to be transferred from the Bush Administration did not support the establishment of a full scale classified EOP system as planned. Instead, NARA migrated the classified data from the Bush Administration to an existing classified NARA presidential library system.

NARA officials attributed EOP ingest delays, in part, to unexpected difficulties. For example, according to NARA officials, once they started using the EOP system, they discovered that records from certain White House systems were not being extracted in the expected format. As a result, the agency had to develop additional software tools to facilitate the full extraction of data from White House systems prior to ingest into EOP. In addition, in April 2009, NARA discovered that 31 terabytes of priority data that had been partially ingested between December 2008 and January 2009 were neither complete nor accurate because they were taken from an incomplete copy of the source system.

Because the records had not been ingested into the EOP system, NARA had to use other systems to respond to requests for presidential records early in 2009. As of April 24, 2009, NARA had received 43 special access requests for information on the Bush Administration. Only one of these requests used EOP for search, and no responsive records were found. To respond to 24 of these requests, NARA used replicated systems based on the software and related hardware used by the White House for records and image management. NARA's current expenditure plan reports that after completing ingest of the Bush electronic records in September 2009, it retired the replicated systems.

In fiscal 2010, NARA plans to complete Increment 3 and begin work on Increment 4. According to its 2010 expenditure plan, Increment 3 will cost $42.2 million and be completed in the fourth quarter of fiscal year 2010. It is to provide new systems for congressional records and public access, as well as improvements to the existing base system and the incorporation of several deferred functions, such as the ability to delete records and search and view their descriptions. Fiscal year 2010 work on Increment 4 is to consist primarily of early planning, analysis, and design.

NARA Faces Several Significant Risks to the Successful Completion of ERA

Despite the recent completion of the first two ERA increments, NARA faces several risks that could limit its ability to successfully

complete the remaining three increments by 2012. These risks include the lack of specific plans describing the functions to be delivered in future increments, inconsistent application of earned value management (a key management technique), and the lack of a tested contingency plan for the ERA system.

First, NARA's plans for ERA have lacked sufficient detail. For several years, NARA's appropriations statute has required it to submit an expenditure plan to congressional appropriations committees before obligating multi-year funds for the ERA program, and to, among other conditions, have the plan reviewed by GAO. These plans are to include a sufficient level and scope of information for Congress to understand what system capabilities and benefits are to be delivered, by when and at what costs, and what progress is being made against the commitments that were made in prior expenditure plans. However, several of our reviews have found that NARA's plans lacked sufficient detail.[9] Most recently, we reported in July that NARA's 2009 plan did not clearly show what functions had been delivered to date or what functions were to be included in future increments and at what cost.

For example, the fiscal year 2009 plan did not specifically identify the functions provided in the two completed increments. In addition, while the plan discussed the functions deferred to later increments, it did not specify the cost of adding those functions at a later time. Additionally, NARA's 2009 plan lacked specifics about the scope of improvements planned for Increment 3. For example, it described one of the improvements as extend storage capacity but did not specify the amount of extended storage to be provided. Also, NARA's plan did not specify when these functions will be completed or how much they would cost. NARA officials attributed the plan's lack of specificity to ongoing negotiations with Lockheed Martin.

[9]See GAO-06-906 and GAO-09-733.

Another risk is NARA's inconsistent use of earned value management (EVM).[10] NARA's 2009 expenditure plan stated that, in managing ERA, the agency used EVM tools and required the same of its contractors. EVM, if implemented appropriately, can provide objective reports of project status, produce early warning signs of impending schedule delays and cost overruns, and provide unbiased estimates of a program's total costs. We recently published a set of best practices on cost estimation that addresses the use of EVM.[11] Comparing NARA's EVM data to those practices, we determined that NARA fully addressed only 5 of the 13 practices. For example, we found weaknesses within the EVM performance reports, including contractor reports of funds spent without work scheduled or completed, and work completed and funds spent where no work was planned. In addition, the program had not recently performed an integrated cost-schedule risk analysis. This type of analysis provides an estimate of the how much the program will cost upon completion and can be compared to the estimate derived from EVM data to determine if it is likely to be sound. NARA officials attributed these weaknesses, in part, to documentation that did not accurately reflect the program's current status.

Another significant risk is the lack of a contingency plan for ERA. Contingency planning is a critical component of information protection. If normal operations are interrupted, network managers must be able to detect, mitigate, and recover from service disruptions while preserving access to vital information. Therefore, a contingency plan details emergency response, backup operations, and disaster recovery for information systems. Federal guidance recommends 10 security control activities related to contingency planning, including developing a formal contingency plan, training

[10]EVM is a project management tool that integrates the technical scope of work with schedule and cost elements for investment planning and control. It compares the value of work accomplished in a given period with the value of the work expected in that period. Differences in expectations are measured in both cost and schedule variances. The Office of Management and Budget requires agencies to use EVM in their performance-based management systems for the parts of an investment in which development effort is required or system improvements are under way.

[11]GAO, *GAO Cost Estimating and Assessment Guide: Best Practices for Developing and Managing Capital Program Costs*, GAO-09- 3SP (Washington, D.C.: March, 2009).

employees on their contingency roles and responsibilities, and identifying a geographically separate alternative processing site to support critical business functions in the event of a system failure or disruption.[12]

An internal NARA review found weaknesses in all 10 of the required contingency planning control activities for ERA. As of April 2009, NARA had plans to address each weakness, but had not yet addressed 10 of the 11 weaknesses. In addition, NARA reported that the backup and restore functions for the commercial off-the-shelf archiving product used at the ERA facility in West Virginia tested successfully, but there were concerns about the amount of time required to execute the process. In lab tests, the restore process took about 56 hours for 11 million files.[13] This is significant because, while the backup is being performed, the replication of data must be stopped; otherwise it could bring the system to a halt. Subsequently, NARA officials stated that they have conducted two successful backups, but the restore process had not been fully tested to ensure that the combined backup and restore capability can be successfully implemented.

Implementation of GAO's Recommendations Could Reduce Risks

To help mitigate the risks facing the ERA program, we previously recommended that NARA, among other things:

- include more details in future ERA expenditure plans on the functions and costs of completed and planned increments;
- strengthen its earned value management process following best practices; and
- develop and implement a system contingency plan for ERA.

[12]National Institute of Standards and Technology, Recommended Security Controls for Federal Information Systems, Special Publication 800-53 Revision 1 (Gaithersburg, MD: December 2006).

[13]NARA estimates that it has received more than 300 million files from the Bush Administration.

In its 2010 expenditure plan, NARA reported that it had taken action to address our recommendations. For example, NARA reported that a test of the ERA contingency plan was completed on August 5, 2009, and the plan itself finalized on September 16, 2009. We have not yet fully-reviewed this plan or the results of the reported test. However, if NARA fully implements our recommendations, we believe the risks can be significantly reduced.

In summary, despite earlier delays, NARA has made progress in developing the ERA system, including the transfer of Bush administration electronic records. However, future progress could be at risk without more specific plans describing the functions to be delivered and the cost of developing those functions, which is critical for the effective monitoring of the cost, schedule, and performance of the ERA system. Similarly, inconsistent use of key project management disciplines like earned value management would limit NARA's ability to effectively manage this project and accurately report on its progress.

Mr. Chairman, this concludes my testimony today. I would be happy to answer any questions you or other members of the subcommittee may have.

Contact and Staff Acknowledgments

If you or your staff have any questions about matters discussed in this testimony, please contact David A. Powner at (202) 512-9286 or pownerd@gao.gov. The other key contributor to this testimony was James R. Sweetman, Jr., Assistant Director.