

NATIONAL ARCHIVES *and*
RECORDS ADMINISTRATION

OFFICE *of*
INSPECTOR GENERAL



Audit of NARA's Digitization Storage
and Transfer Capabilities

May 5, 2015

OIG Audit Report No. 15-11

Table of Contents

Executive Summary 3

Background 4

Objectives, Scope, Methodology 6

Audit Results..... 7

Appendix A – Acronyms and Abbreviations 23

Appendix B - Management’s Response to the Report 25

Appendix C - Report Distribution List 25

Executive Summary

The National Archives and Records Administration's (NARA) 2014 Strategic Plan is based on four strategic goals. The first—"Make Access Happen"—establishes public access as NARA's core purpose. Further, this goal includes an initiative to digitize all of NARA's analog archival records to make them available online. Similar digitization initiatives have been a part of NARA's Strategic Plans for the past eight years. In July 2014, the Office of Inspector General (OIG) issued Audit Report No. 14-12: Audit of Selected Aspects of NARA's Digitization Program. Based on the findings in that report, additional audits were planned related to NARA's Digitization Program. This audit report—Audit of NARA's Digitization Storage and Transfer Capabilities—is the second in a series of follow-on Digitization Program audits.

The overall objective of this audit was to evaluate NARA's processes for the storage and transfer of digitized records. Specifically, we evaluated NARA's efforts to ensure digitized records were adequately stored and made publicly available in a timely manner. Our audit found NARA did not maintain adequate space for storing records digitized internally by its Digitization Labs and Presidential Libraries, nor did it maintain adequate storage network transfer capabilities to efficiently move digitized records across the agency's various internal systems. As a result of inadequate storage, planned digitization projects were stopped, internal output was hindered, previously digitized records were reworked, original and digitized records were put at risk of loss, and digitization goals were adversely impacted. Further, NARA's lack of storage network transfer capabilities increased risk of process inefficiencies, delayed internal access to digitized files, and increased the potential for digitized records loss.

Additionally, NARA does not have effective processes and controls in place to manage records digitized by its partners from the time of original scan through the point the digitized records ultimately become publicly accessible. As a result, NARA's Digitization Partnership Program has increased risk of loss and file corruption of digitized records, breach of partnership agreements, and inefficiencies in making partner digitized records accessible. Finally, NARA's lack of strategic management within the agency's Digitization Partnership Program has impeded the release of over 57 million partner-digitized records currently eligible for online public access.

This report makes 13 recommendations to improve NARA's infrastructure for its digitization efforts, management over that infrastructure, performance in achieving digitization goals, and efforts to make access to its holdings happen efficiently, effectively, and timely.

Background

Over the past decade, the National Archives and Records Administration (NARA) has made the digitization of its vast record holdings a major initiative within the agency's strategic planning. NARA's 2000 Strategic Plan recognized users "increasingly expect immediate electronic access to information at no cost." Later, NARA's 2006 Strategic Plan incorporated specific digitization goals, stating NARA "will provide prompt, easy, and secure access to our holdings anytime, anywhere." Most recently, NARA's 2014 Strategic Plan included an initiative to digitize all analog archival records in order to make them available online.

In July 2014, the Office of Inspector General (OIG) issued Audit Report No. 14-12: Audit of Selected Aspects of NARA's Digitization Program. Based on the findings identified in that report, additional audits were planned related to NARA's Digitization Program. This audit report— Audit of NARA's Digitization Storage and Transfer Capabilities—is the second in a series of follow-on Digitization Program audits. The focus of this report is on NARA's processes for the storage and transfer of digitized records.

Digitizing NARA's vast holdings and making the records available online requires extensive digital storage and transfer capabilities. In May 2008, NARA issued its initial "Strategy for Digitizing Archival Materials," which identified the agency's approach to digitizing its holdings. This approach included strategies involving both in-house digitization efforts, as well as external digitization efforts performed through partnerships with outside organizations. Both in-house and external partnership digitization efforts rely upon NARA's storage infrastructure.

NARA's Office of Innovation currently manages the agency's Digitization Program. The Office of Innovation, led by NARA's Chief Innovation Officer, oversees the agency's Open Government and Digital Government Strategy efforts, the online public catalog, and Digitization Strategy. In addition, the Office of Innovation runs and coordinates NARA's internal Digitization Labs. The Office of Innovation oversees the Microfilm and Textual and Photographic Imaging Labs while Research Services oversees NARA's Audio/Video and Motion Picture Digitization Labs. Further, the Office of Innovation oversees NARA's external digitization partnerships, which to date account for 97 percent of NARA's digitization work available online. Partners digitize NARA holdings for use on the partner's web site, and after an embargo period (usually of five years), NARA receives the right to make the digitized records accessible to the public at no cost. NARA

began receiving digitized records from partners in 2007. All the digitized files created through NARA's digitization efforts must be stored and preserved so NARA can make the files accessible.

NARA's Office of Information Services is responsible, in part, for managing the infrastructure in NARA's Digitization Labs and for supporting the Office of Innovation in meeting customer needs for effective and innovative social media, open government, and digitization services, solutions, and systems. The Office of Information Services is led by NARA's Chief Information Officer, and manages NARA's nationwide information and telecommunications infrastructure, and NARA information systems.

Objectives, Scope, Methodology

The objective of this audit was to evaluate NARA's processes for the storage and transfer of digitized records. Specifically, we evaluated NARA's efforts to ensure digitized records were adequately stored and made publicly available in a timely manner. The audit evaluated whether NARA had proper controls in place to monitor and acquire storage capacity needed to house records digitized both internally by NARA and externally by partners, and the transfer capabilities needed to efficiently move digitized records across NARA's internal and public access systems.

To accomplish our objective, we reviewed the following: NARA's "Strategy for Digitizing Archival Materials for Public Access, 2007-2016," NARA Directive 816 "Digitization Activities for Enhanced Access," NARA Directive 801 "Capital Planning and Investment Control," and Office of Management and Budget Circular No. A-130 "Management of Federal Information Resources." In addition, we reviewed NARA's current and previous Strategic Plans, Open Government Plans, and Annual Performance Plans. Further, we reviewed information on NARA's digitization efforts on the agency's internal and public-facing web sites.

We interviewed NARA personnel from the Offices of Innovation, Information Services, Presidential Libraries, Business Support Services, and Research Services involved in managing, tracking, and procuring storage and transfer capabilities for NARA's digitization efforts. We obtained and analyzed internal documents relating to digitization storage and transfer capabilities such as meeting minutes, working group charters, system reports, white papers, partnership agreements and project plans, project management documents, business need summaries, purchase orders, contracts, performance work statements, and other applicable documents.

Our audit work was performed at Archives II in College Park, MD between July 2014 and November 2014. 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 used judgmental sampling in our audit work; therefore the results of this sample were not used in projecting across the entire population. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

Audit Results

1. Storage Management of Internally Digitized Records

NARA did not maintain adequate space for storing records digitized internally by its Digitization Labs and Presidential Libraries, nor did it maintain adequate storage network transfer capabilities to efficiently move digitized records across the agency's various internal systems. This lack of storage and transfer capabilities was caused by NARA's inadequate planning and management for short-term and long-term internal digitization needs. NARA Directive 816, "Digitizing Activities for Enhanced Access," requires Digitization Services—currently a component of the Office of Innovation—to provide "long-term storage and preservation of the master copies of the digital surrogates resulting from in-house NARA digitization projects and from partnership projects." As a result of inadequate storage, planned digitization projects were stopped, internal output was hindered, previously digitized records were reworked, original and digitized records were put at risk of loss, and digitization goals were adversely impacted. Further, NARA's lack of storage network transfer capabilities increased risk of process inefficiencies, delayed internal access to digitized files, and increased the potential for digitized records loss.

NARA Directive 816, issued on March 8, 2004, applies to any digitizing activity producing digital surrogates of NARA archival holdings for access purposes. Directive 816 defines "digital surrogate" as a digital image or copy of a textual or non-textual record. Further, the Directive states Digitization Services is responsible, in part, for providing guidance on the technical aspects of digitization storage for in-house digitizing activities and projects. They are also responsible for providing long-term storage and preservation of the master copies of the digital surrogates resulting from in-house digitization projects.

NARA's "Strategy for Digitizing Archival Materials for Public Access," (Digitization Strategy) dated May 2008¹, outlines the agency's approach to digitizing and making NARA's holdings available online. NARA's digitization approach includes five separate strategies:

- Strategy 1: Gather and make available archival materials that had been previously digitized in the course of performing the agency's function;

¹ NARA updated its Digitization Strategy effective December 24, 2014. Field work during this audit was conducted while the previous Digitization Strategy was in place.

- Strategy 2: Establish partnerships with organizations from a variety of sectors to digitize and make available holdings;
- Strategy 3: Conduct digitizing projects in-house;
- Strategy 4: Pursue digitization of archival materials as part of NARA's preservation reformatting approach; and
- Strategy 5: Continue to make NARA's online catalog a hub for discovering NARA's digital images.

Of the five Digitization Strategies listed above, all but Strategy 2 focus on NARA's internal digitization efforts. These internal digitization efforts take place in NARA's four Digitization Labs (Audio/Video, Motion Picture, Microfilm and Textual, and Photographic Imaging), as well as the Presidential Libraries.

Internally Digitized Records Storage Capacity

Records digitized by NARA internally require electronic storage space after converting the archival records to a digital format, as well as long-term storage for preservation and public access. NARA's Digitization Labs store their working copies of digital files on a number of different storage servers assigned to each of the four Digitization Labs. However, the storage space needed by the Digitization Labs to conduct NARA's in-house digitization projects has not always been adequate.

The issue of storage limitations in the Digitization Labs was raised to the Office of Innovation and the Preservation Programs Division in a white paper dated November 2010. The white paper stated the Digitization Labs had nearly reached their storage capacity and personnel were finding the situation to be unmanageable. As a result, digitized files were moved to tape and external hard drives to clear up needed working space on the storage network. At the time of the November 2010 white paper, only one of NARA's four Digitization Labs had sufficient storage. Of the remaining three Digitization Labs, one was already out of storage and two were less than three months away from reaching storage capacity based on normal output. The November 2010 white paper proposed the Office of Information Services and the Preservation Programs Division develop a data storage plan so the Digitization Labs could continue to provide services to NARA and its researchers. However, this proposed plan was not developed.

In 2012, NARA's former Chief Information Officer (CIO) assigned staff to work with the Digitization Labs on infrastructure issues and perform an analysis of the storage situation in the Digitization Labs. The analysis provided storage usage totals in the Digitization Labs, indicating storage would reach capacity in several weeks based on expected digitization output levels. At the time of the analysis, the Digitization Labs had recently

been allocated additional storage space on the NARA-wide Enterprise Storage Area Network (ESAN). However, the ESAN volumes allocated to the Digitization Labs only prolonged reaching storage capacity by an additional 10 months. The analysis also cited aging infrastructure, failing tape drives, server connections, and data management as other issues facing the Digitization Labs. Further, the analysis recommended the development of a project plan so NARA could plan, anticipate, and respond to the needs of the Digitization Labs over time. NARA's Chief Innovation Officer (CINO) stated it had been difficult to get resources for the Digitization Labs over the years as NARA's former CIO had not prioritized funding for the Digitization Labs and showed an unwillingness to invest in the Digitization Labs. With the issuance of a February 2012 white paper detailing the analysis above, the CINO stated she felt for the first time the needs of the Digitization Labs were sufficiently explained and understood by upper management.

Over a year and a half after the February 2012 white paper was written—despite adding additional volumes of storage on the ESAN—the Digitization Labs had only six percent storage capacity remaining. With ESAN volumes mostly full, the Digitization Labs began to rely on storage servers unused by other NARA offices. In addition to the ESAN volumes, the Digitization Labs were using eight different storage devices by June 2014. Through these temporary solutions, the Digitization Labs were able to add 125 terabytes of storage capacity between October 2013 and June 2014. However, by August 1, 2014, the Digitization Labs were down to under four percent storage capacity available across all long-term storage devices. Without efforts to find and install the temporary storage solutions mentioned above, the Digitization Labs would have been out of storage for most, if not all of 2014.

However, as a result of using various temporary solutions to add needed storage capacity, the Digitization Labs' files are distributed across multiple storage devices, making file retrieval more challenging. NARA Directive 816 places responsibility on NARA offices to “ensure that the master and duplicate copies of each digitization product maintained by NARA are properly housed and maintained.” Further, NARA's Digitization Strategy defines the act of digitizing as a series of activities, including quality control and maintenance of digital copies and metadata. The Digitization Labs are responsible for file management of digitized records, but do not have a central file repository. A central file repository would help the Digitization Labs manage their digital files by tracking where the files are stored and allowing efficient retrieval when necessary. Further, only one staff member is assigned the responsibility of storage management for the Digitization Labs. Currently, there are no contingency plans in place should this employee depart the agency or be reassigned internally. The working knowledge of file

locations and storage servers used in the Digitization Labs could be hindered without an established contingency plan.

In June 2014, the Digitization Labs received \$500,000 in funding from the Architecture Review Board for improvements to infrastructure. With this funding, the Digitization Labs were able to purchase a new tape library system. The tape library system purchased by NARA has two petabytes of storage currently, and is scalable up to over 25 petabytes at maximum capacity given continued investment. The system was installed at the end of calendar year 2014. The NARA staff member in charge of Digitization Lab infrastructure plans to move all digitized files from the existing servers in the Digitization Labs to the new tape library system (see photo below). The storage space purchased is expected to provide the Digitization Labs with enough capacity to produce at historical output rates for the next few years. However, with sufficient storage in place, it is possible the Digitization Labs will produce at a rate greater than historical output. As a result, the Digitization Labs may fill the new two petabyte storage space faster than anticipated, putting NARA back in a situation in which storage limitations negatively impact digitization efforts.



*Pictured above are storage servers for the Digitization Labs, taken 20 November 2014 at Archives II. The servers on the left are the old servers being replaced, while the servers on the right are new purchases.

In addition to the new tape library system, the Digitization Labs made two other purchases with the funding received. First, parts were ordered for a new storage server that replaces the end of life local storage servers being used in the Digitization Labs. The Digitization Labs obtained this new storage server from another NARA office, which had purchased the equipment, but did not use it. The new storage system will refresh the Digitization Labs' equipment, make the system more reliable, and offer better performance. Second, the Digitization Labs purchased a software appliance that will assist in managing files. The appliance will manage files between the local working storage servers in the Digitization Labs and the newly-purchased tape library permanent storage system. This appliance was installed at the end of calendar year 2014.

Although a long-term solution for the Digitization Labs' infrastructure issues has not been fully developed, Office of Information Services management stated the newly-purchased tape library system is part of the long-term solution. With continued investment, NARA can increase the storage amount of the Digitization Labs' new tape library system to over 12 times the current amount—potentially reaching a total capacity of over 25 petabytes of storage space. Further, Office of Information Services is considering developing contingency plans and training a second staff member as a backup knowledgeable of the storage infrastructure of the Digitization Labs. In addition, NARA management is considering the Digitization Labs' storage requirements in planning its enterprise-wide cloud storage strategy. Although these recent efforts by NARA will improve its digitization program going forward, by postponing the implementation of digital storage solutions until the end of FY 2014, NARA experienced the following adverse impacts:

- Stop of Planned Digitization Projects. Limited storage availability caused the Audio/Video Digitization Lab to stop digitizing its major planned project of U.S. House and Senate Recording tapes. This project was 78 percent of the Digitization Lab's two-year planned project storage total. Staff in the Audio/Video Lab had created an efficient process for digitizing the tapes and began producing at an increased rate. However, the increased rate of production quickly filled up available storage space, halting the project. Similarly, the Motion Picture Lab was unable to perform planned projects as it reached its storage ceiling in early September 2014. Staff in the Motion Picture Lab stated that infrastructure and storage concerns prevented the lab from digitizing all planned projects. That Digitization Lab was working on only one planned project as of September 2014, but staff in the lab stated it could digitize more planned projects given appropriate storage.

- Hindrance of Internal Digitization Output. In 2011, the Digitization Labs conducted an analysis to estimate output based upon a number of different scenarios. Inputs to those scenarios included staffing, storage, and equipment. The analysis concluded that the Digitization Labs could double their output given sufficient storage at current staffing levels. A Digitization Lab supervisor continues to agree with the analysis, and stated the assumption that sufficient storage would allow the Digitization Labs to potentially produce more than double their current output of approximately 10 terabytes per week. NARA's Annual Performance Plans from FY 2007 to FY 2013 detail four major strategies to increase the amount of archival materials available online. One of the four strategies is "exploring innovative NARA-led projects for digitizing archival material that will also allow us to develop our internal capacity in this area." Increasing storage capacity would have allowed the Digitization Labs to digitize more records internally, thereby increasing access and furthering NARA's strategic goal of "Make Access Happen²."
- Reworking of Previously Digitized Records. When the Digitization Labs create digital copies based on researcher requests, the files are saved onto temporary storage media and physically delivered to the Research Rooms. The Digitization Labs cannot save the digital files created for researcher access on the storage network due to lack of storage space. Therefore, if the same record is requested again, the Digitization Labs will have to re-digitize the archival record; thereby increasing the risk of record damage and making the Digitization Labs less productive by repeating work. One objective of NARA's Digitization Strategy is to "enhance preservation of records by reducing wear and tear on the originals for reference and reproduction." NARA personnel stated there have been instances when Digitization Labs had to re-digitize archival materials requested by researchers due to lack of storage or insufficient file management. Therefore, NARA is not fully meeting this preservation objective of its Digitization Strategy.
- Preservation Risk for Original and Digitized Records. Another objective of NARA's Digitization Strategy is to "provide access to those materials that can no longer be accessed in their original format." For example, records such as audio tapes (see photo below) may be so fragile that they are only capable of being played once. If the tape is played to create the access digital file, and after which the original tape no longer plays, the only viable record available is the newly created digital access file. Therefore, the newly created digital copy needs to be preserved to provide future access to the record. However, as noted previously,

² Work performed in the Digitization Labs had led to over 1.3 million records being made available online through NARA's public access catalog as of August 2014.

Digitization Labs have at times had to halt digitizing for preservation due to limited storage availability, which further delays needed preservation work on at-risk and deteriorating records.



*Pictured above is an at-risk audio tape from the Audio/Video Digitization Lab, taken November 20, 2014 at Archives II.

In addition, the aging local storage servers used by the Digitization Labs present a preservation risk to NARA's digitized records. These servers are used by the Digitization Labs for storage of working copies of digitized records before transferring the files to more permanent storage on the ESAN. The Digitization Labs' local storage servers were installed in 2007 and remain in use without technical refresh plans in place. Due to the age of the hardware, the Digitization Labs experienced several infrastructure issues. For example, in May 2014 one of the local storage servers used by the Digitization Labs failed. Fortunately, NARA staff was able to restore the data from backup tapes and local computers. However, if the computer files had been compromised or the backup tapes failed, NARA would have experienced record loss. With the installation of a new local storage system at the end of calendar year 2014, NARA should be at decreased risk of this type of data loss moving forward.

- Impact on Digitization Goals of Presidential Libraries. As noted previously, digitization efforts also take place at the Presidential Libraries (Libraries). Based on Library staff's survey responses, ten of thirteen Libraries stated storage limitations negatively impacted their digitization goals. The Libraries store digitized files on NARA-provided network area storage (NAS) devices. As the Libraries have filled the NAS devices to capacity, they have relied on storing

digitized files on CDs and hard drives, which creates preservation issues. Storing files on multiple forms of storage media makes it difficult for Library personnel to locate and work on the files. Further, files saved on the NAS devices are limited in backup capabilities. Sufficient storage capabilities would aid the Libraries in efficiently achieving digitization goals, thereby furthering access to NARA's holdings.

Storage System Transfer Capabilities

Another aspect of NARA's digitization storage systems involves the agency's ability to efficiently transfer internally digitized records between storage systems. The CINO stated that just adding storage space does not solve the infrastructure problems of NARA's Digitization Labs. Rather, sufficient electronic transfer capabilities are also needed to allow the Digitization Labs to move files across various storage systems, including transfer of files for public access or cloud storage.

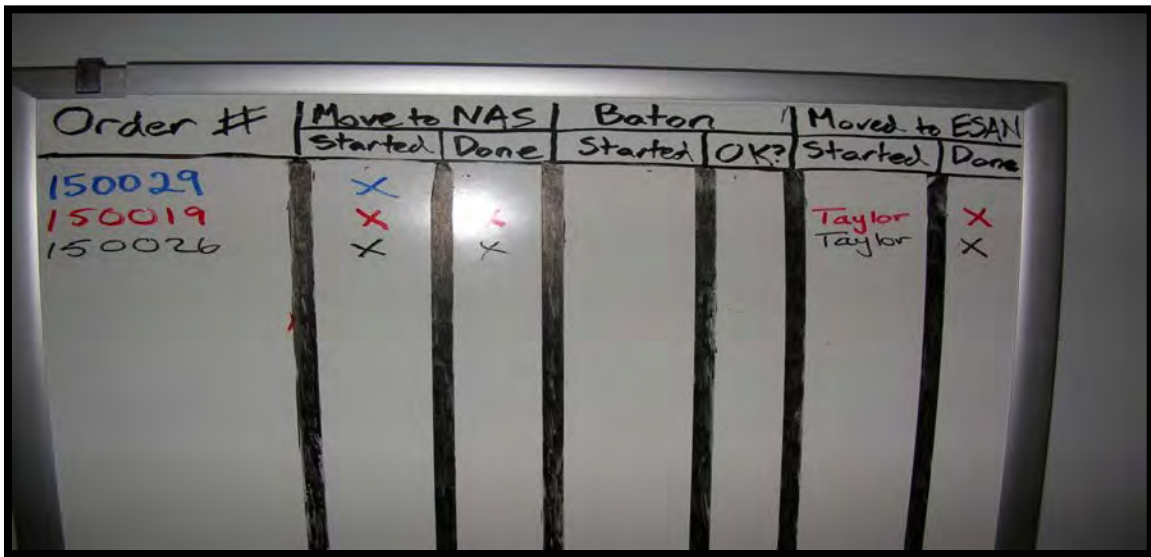
As noted previously, the Digitization Labs store their digitized files on a number of different storage servers. Local storage servers are assigned to each of the four Digitization Labs for the storage of working copies of digital files. Once Digitization Lab personnel are finished working on the digital files, the files are transferred to a more permanent storage location. However, the local storage servers are not effectively networked with NARA's more permanent storage system. Therefore, the files must be physically transferred from one storage system to another via portable storage media devices.

Similarly, the Digitization Labs' storage servers are unable to directly transfer digitized records to NARA's research rooms or public access catalog. If a researcher requests digitized records, the Digitization Labs must create a CD or store the records on a portable hard drive and physically deliver the media to the research room. Further, if digitized records from the Digitization Labs are to be made available to the public online, the completed files must be physically transferred from the Digitization Lab storage server to NARA's public access catalog via portable storage media devices. Other NARA offices, including all thirteen Presidential Libraries, store their digitized files locally and transfer the files via mailing hard drives or CDs to staff at Archives II for upload to NARA's public access catalog.

Further, with the ESAN volumes allocated to NARA's Digitization Labs reaching capacity, attempts were made to transfer files digitized by the Digitization Labs from the ESAN into a cloud storage environment. However, limited bandwidth on the network caused the uploads to fail. Therefore, the Digitization Labs can only transfer files to the

cloud via hard drive shipment to NARA’s cloud service provider. The Director of IT Operations stated bandwidth concerns were being studied and equipment such as routers and firewalls would be upgraded in FY 2015, which would increase transfer rates across NARA, including within the Digitization Labs. Due to NARA’s current lack of effective storage system transfer capabilities, the agency has experienced the following adverse impacts:

- Process Inefficiencies. Hand-delivering digitized records on portable storage media is time-consuming and inefficient. Developing capabilities to electronically transfer the files would allow staff more time to digitize NARA’s holdings and make access faster to researchers.
- Delayed Internal Access to Digitized Files. There is no central file repository where staff in the Digitization Labs, Digital Public Access Branch, and Research Services can access digitized files. Without a centralized file repository in place, NARA staff risk delays in locating previously digitized files and a greater chance of multiple digital surrogates of the same records due to rework.
- Potential Digitized Records Loss. Using portable storage media devices as a method of transferring digitized records increases the risk of record loss. These storage devices are not backed up on the network and are not part of a centralized file repository. In order to track files, Digitization Labs have developed rudimentary logging processes (see photo below). This manual file management process puts the Digitization Labs at risk of losing digital files.



*Pictured above is the Whiteboard used in the Motion Picture Lab to track internal projects, taken November 20, 2014 at Archives II.

Recommendations

To support Digitization Lab infrastructure, we recommend:

1. The Chief Innovation Officer, Chief Information Officer, and the Executive for Research Services develop a long-term strategy for internal digitization infrastructure needs and periodic technical refreshes of equipment used in the Digitization Labs.
2. The Chief Innovation Officer and Chief Information Officer develop procedures for file management of digitized records in the Digitization Labs and develop contingency plans for staff charged with managing the infrastructure of the Digitization Labs.
3. The Chief Information Officer implement a central file repository system in the Digitization Labs.
4. The Chief Innovation Officer and Chief Information Officer develop agency-wide policies and procedures for digital file preservation.
5. The Chief Information Officer implement increased storage solutions for digitization efforts at the Presidential Libraries.
6. The Chief Innovation Officer and Chief Information Officer develop a long-term strategy for increasing transfer capabilities between various internal storage systems housing digitized records.

Management Response

Management concurred with the recommendations.

2. Management of Partner-Digitized Records

NARA does not have effective processes and controls in place to manage its records digitized by partners from the time of original scan through the point in which the digitized records ultimately become publicly accessible. This condition exists due to a lack of strategic focus and management within NARA's Digitization Partnership program as it relates to storage of partner-digitized records, handling of partner-digitized records once in NARA's possession, and procedures for making the partner-digitized records publicly accessible. The Government Accountability Office Standards for Internal Control in the Federal Government state "management should design control activities to achieve objectives and respond to risk." Without sufficient processes and controls in place, NARA's Digitization Partnership program has increased risk of:

- Loss and file corruption of hard drives containing digitized records on loan from digitization partners;
- Breaching project plan agreements as they relate to the return of partnership loaned storage media;
- Loss of partner-digitized records that are now eligible to be released for online public access; and
- Prematurely releasing embargoed records in violation of partnership agreements.

Further, NARA's lack of strategic management within the agency's Digitization Partnership program has impeded the release of over 57 million partner-digitized records³ currently eligible for online public access.

NARA's FY 2014-2018 Strategic Plan includes the strategic goal "Make Access Happen." This strategic goal contains an initiative to "digitize all analog archival records to make them available online." NARA's Digitization Strategy outlines the agency's approach to digitizing and making NARA's holdings available online. As identified in the previous finding, NARA's digitization approach incorporated five separate strategies. One of these strategies—Digitization Partnerships—has delivered approximately 97 percent of the digitized record content available online.

NARA's Digitization Partnership Strategy (Partnership Strategy) involves establishing partnerships with organizations from a variety of sectors (private, public, non-profit, educational, Government) to digitize and make available NARA's holdings. NARA's Partnership Strategy states "Partnerships will enable NARA to make more digitized holdings available than we could on our own, because the partners will bear most of the expense of digitizing." Further, NARA's Partnership Strategy includes a principle to make the digital copies available as quickly as funds and capabilities allow, once

³ NARA Directive 816 defines a digitized record as "a digital image or copy of a textual or non-textual record (e.g. scanned images of documents, digitized sound recordings).

permitted to do so by the terms of the partnership. One of the terms often present in the agreements establishing NARA's Digitization Partnerships involves digitized record embargo periods. An embargo period is an established length of time in which a partner has exclusive rights to make digitized material available through their website—typically five years. Once the embargo period concludes, NARA is eligible to make the partner-digitized records freely accessible online through the agency's public access catalog.

Storage of Partner-Digitized Records

NARA first began receiving digitized records from its Digitization Partners in 2007. The partner-digitized records were delivered to NARA on hard drives, which NARA staff transferred to LTO-2⁴ tapes stored at Archives II. Once the partner-digitized records were transferred to the LTO-2 tapes, NARA would return the hard drives to the partners for reuse. NARA continued transferring the partner-digitized records onto LTO-2 tapes until 2010, at which time the volume of hard drives received from digitization partners exceeded NARA's capacity to transfer the records to tape. At that point, NARA began stockpiling the hard drives received from the partners at Archives II, until both the LTO-2 tapes and hard drives were shipped to Allegheny Ballistics Lab (ABL). At ABL, the partner-digitized records on the LTO-2 tapes were transferred to spinning disk storage, while the partner-digitized records on the hard drives remained stored on the original hard drives. Future shipments of hard drives received from the partners were stored at Archives II. However, in February 2014, the Director of NARA's Digitization Division halted all scheduled partnership shipments, in part due to inadequate storage capacity.

According to the Office of Innovation's agency-wide partnership hard drive inventory, NARA has amassed 634 hard drives. NARA tracks the hard drives with a spreadsheet that includes partner information and the location of each hard drive. The partnership hard drives in NARA's inventory date back as far as 2010. Office of Innovation personnel stated "the data on the hard drives has remained on the drives for the past six years since the initial hard drives were sent to NARA. This poses a risk to NARA because this valuable data is not in an environment in which it can be managed and backed-up." Although NARA was aware of this storage limitation since 2010, it was not until 2014 that Office of Innovation began work on implementing a solution.

On March 19, 2014, the Office of Innovation submitted a Business Need Summary outlining the need to transfer the partnership hard drives to an environment where the data can be managed and backed-up. Office of Innovation staff estimated its storage need for the partner hard drive data to be 984 terabytes. Further, the Office of Innovation

⁴ LTO stands for Linear Tape-Open, a magnetic tape data storage technology. LTO-2 tape stands for version 2 of the LTO. The current version is LTO-6.

sought a capability in which the partners could deliver digitized record data directly to a storage repository. In June 2014, funding in the amount of \$665,000 was approved for the Office of Innovation to procure the cloud storage necessary to upload data off the 634 partner hard drives. Office of Innovation staff worked with personnel from the Office of Information Services and the Acquisitions Division of the Office of Business Support Services to establish the requirements for the cloud storage contract. The CIO and the CINO agreed that cloud storage for partner data was a viable long-term solution for the business need. However, due to requirements development delays and acquisition time constraints, efforts to procure cloud storage in FY 2014 were not successful. Despite the efforts of staff involved, the need for long-term storage of partner data persists.

According to the Acting Chief Technology Officer, NARA is working to implement an enterprise-wide cloud storage solution by establishing a Cloud Governance Working Group, which will ultimately develop a Cloud Governance Board under which NARA's cloud strategy will be formulated and implemented. Office of Information Services management stated their hope is to explore a cloud storage solution for the partner hard drives again in FY 2015, if funding permits. The CINO submitted a Business Need Summary to set up space in the cloud storage environment for partners to upload digitized records, thereby eliminating the need for partners to ship their digitized data to NARA via hard drives, and for NARA to store the digitized records on hard drives. Although these planned efforts have the potential to increase the effectiveness of NARA's Digitization Partnership program, years without adequate storage for partner-digitized records has placed NARA at risk of the following:

- File Loss and Corruption. Each partner sent NARA a report documenting the information on the individual hard drives, but the quality and quantity of information varied by partner. NARA is unsure exactly what records are stored on the 634 hard drives in its possession. Until NARA is able to upload the hard drive files to a centralized storage location where staff can view all files, NARA cannot be sure what digitized records it has received and what records in a series may be missing. In the past, NARA identified corrupted files when digitized records were transferred from LTO-2 tape to spinning disk. Without a process in place to consistently inventory and verify the quality of digitized records received on an ongoing basis, NARA is at greater risk of file loss and corruption.
- Breach of Partnership Project Plan Agreement. Project Plans are described in NARA's Partnership Agreement as the document completed by the partner and NARA that provides the details about metadata, equipment, format, costs, personnel, and other required information for each selection of NARA's holdings. One of the other required information items common to the project plans is

“delivery mechanism.” The delivery mechanism requirement identified in the project plan for one of NARA’s digitization partners states “the digitized records will be provided to NARA via a partner-provided hard drive, which will be returned to the partner once NARA has retrieved the files.” Due to the lack of an adequate storage infrastructure, NARA is currently in possession of 530 of this digitization partner’s hard drives, some dating back to 2010. By amassing these hard drives with limited progress in retrieving the files off of them, NARA is at risk of breaching project plan requirements.

Handling of Partner-Digitized Records

As mentioned previously, the Office of Innovation developed an agency-wide inventory of partner hard drives. This inventory is maintained in a spreadsheet and includes details such as the creating partner and the physical location of each hard drive. However, this inventory process was not consistently implemented from the time the agency first started receiving partner hard drives in 2007. Further, NARA did not establish formal procedures for handling the hard drives upon receipt. Therefore, hard drives may have been received from partners without proper inclusion in NARA’s inventory. According to Office of Innovation staff, the inventory reflects the hard drives they currently account for in known physical storage locations. Without proper inventory and verification controls in place, NARA cannot effectively attest to the accuracy of the inventory of hard drives in its possession or ensure hard drives were not lost or misplaced over the years.

Further, during the period between 2007 and 2010 in which NARA was transferring partner-digitized records from the hard drives to LTO-2 tapes, the agency maintained a tracking sheet of files contained on each hard drive. This tracking sheet was developed based on reports provided by the partners upon delivery of the digitized records to NARA. As stated earlier, NARA transferred the digital records from the LTO-2 tapes to spinning disk storage at ABL. The spinning disk storage at ABL contained the 12.6 million records leaving embargo period on January 1, 2014. To prepare these records for public access, Office of Innovation staff created an automated program to reconcile the initial digitized record tracking sheets with the directory generated of the files stored on the spinning disks. After running the program, Office of Innovation staff was unable to locate approximately 3 million of the 12.6 million records reported on NARA’s initial tracking sheets. Office of Innovation staff stated this large discrepancy was due in part to slight changes in the file names of the records as they migrated across the different storage media over the years. However, as a result of inadequate file management and handling controls at the onset, reconciling the remaining inventory will require greater time and resources.

In addition, the tracking spreadsheet of FY 2014 unembargoed records included over 2 million records received by NARA in 2009, which—according to the partnership agreement—were subject to embargo until January 1, 2015. Office of Innovation staff stated a copy and paste error likely occurred causing the 2 million embargoed records to erroneously be included in the tracking spreadsheet of records the agency intended to make available in 2014. Without procedures and controls in place to govern the process of determining what records are embargoed versus what records can be made accessible, NARA increases its risk of releasing embargoed records prematurely and breaching its partnership agreements.

Public Access to Partner-Digitized Records

Within the scope of this audit, we assessed storage and handling aspects impacting NARA's ability to provide public access to partner-digitized records. As noted previously, NARA's Strategic Plan includes the strategic goal "Make Access Happen," which calls for the agency to make all records available to the public in digital form to ensure anyone can explore, discover, and learn from NARA holdings. However, without the necessary storage in place to facilitate the management and maintenance of partner-digitized records, NARA is impeding its efforts in achieving this goal.

In 2008, NARA received 12.6 million partner-digitized records eligible for public access, following embargo, on January 1, 2014. However, due in part to the storage and handling issues identified earlier in this report, NARA has yet to make these records freely accessible online through the agency's public access catalog. Further, despite having possession of the partner-digitized records for over five years, the Office of Innovation did not seek funding to convert the 12.6 million records into an accessible format⁵ until March 18, 2014—over two months after NARA was eligible to make the records available online. The Office of Innovation later set a goal to have all 12.6 million records converted and accessible by September 30, 2014. However, as noted above, this goal was not achieved.

Further, 61 of the 634 hard drives in NARA's possession contain records digitized by a partner whose digitized files are not subject to an embargo provision. Therefore, NARA could have made the 45 million records digitized by that partner publicly available online upon receipt of the digital files. NARA has received files from this partner since 2012, but has not made any of the files accessible to date due in part to insufficient storage

⁵ All 12.6 million records received by NARA in 2008 were not in a format compatible with NARA's public access catalog. Therefore, all 12.6 million records had to be converted to a compatible format, so that the records could be made publicly accessible. NARA spent \$30,075 in FY14 to convert the digitized records into a compatible format.

space and inadequate planning to make the records ready for timely online public access. While NARA has known it was eligible to make these partner-digitized records accessible since 2012, the Office of Innovation did not begin the process of acquiring the storage necessary to make them accessible until 2014. Therefore, between the 12.6 million records completing their embargo period and the 45 million records not subjected to an embargo provision, NARA was eligible to make 57.6 million partner-digitized records freely accessible online through the agency's public access catalog. However, to date, these records have yet to be made available by NARA.

In 2014, the same partner mentioned above—whose Partnership Agreement does not subject NARA to an embargo period—offered to provide the agency an additional 350 million digitized records created by the partner outside of its partnership with NARA. Due in part to storage limitations described previously, NARA has yet to take possession of these 350 million records or make them available online. The CINO stated Office of Innovation staff would need to perform an analysis of the 350 million records to determine what is included in the record group and whether or not they are all NARA records. However, Office of Innovation staff has not yet had the time to perform the necessary analysis. If any of these records are in fact NARA's, the records can be made accessible immediately as the records are not subject to an embargo provision—provided the storage infrastructure and other accessibility requirements are in place.

Recommendations

We recommend the Chief Innovation Officer:

7. Work with the Chief Information Officer to develop a long-term storage strategy for currently held partner data and partner data to be provided to NARA in the future.
8. Work with the partners to return partner-provided hard drives.
9. Establish procedures for the transfer of digitized records received from NARA's Digitization Partners to NARA and the handling of those records once in NARA's possession.
10. Implement procedures to ensure embargoed records are not released before the end of the partnership agreement's embargo period.
11. Implement policies outlining timeframes necessary to make records leaving embargo periods available on the date those records are first available to be released by NARA free of charge.
12. Develop a plan for converting partner-digitized records being made accessible in the future to a format compatible with NARA's online public access catalog.
13. Perform an analysis of the 350 million records offered to NARA by one partner to determine what records NARA should take possession.

Management Response

Management concurred with the recommendations.

Appendix A - Acronyms and Abbreviations

ABL	Allegheny Ballistics Lab
CD	Compact Disc
CIO	Chief Information Officer
CINO	Chief Innovation Officer
ESAN	Enterprise Storage Area Network
FY	Fiscal Year
IT	Information Technology
LTO	Limited Tape Open
NARA	National Archives and Records Administration
NAS	Network Area Storage
OIG	Office of Inspector General

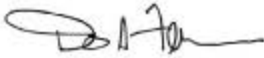
Appendix B - Management's Response to the Report



Date: 30 April 2015
To: James Springs, Inspector General
From: David S. Ferriero, Archivist of the United States
Subject: OIG Revised Draft Audit Report OIG 15-11, Audit of NARA's Digitization Storage and Transfer Capabilities

Thank you for the opportunity to provide comments on this revised draft report. We appreciate your willingness to meet and clarify language in the report.

We concur with the 13 recommendations in this audit, and we will address them further in our action plan.



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

Archivist of the United States
Deputy Archivist of the United States
Chief Innovation Officer
Chief Information Officer
Chief Operating Officer
Deputy Chief Operating Officer
Executive for Research Services