



June 19, 2017

TO: David S. Ferriero
Archivist of the United States

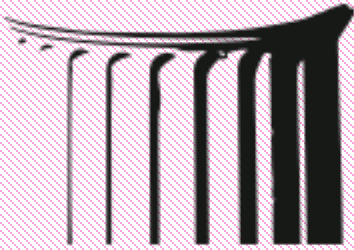
FROM: James Springs *James Springs*
Inspector General

SUBJECT: *Audit of NARA's Online Access to Digitized Holdings*

Attached for your action is our final report, *Audit of NARA's Online Access to Digitized Holdings (OIG Audit Report No. 17-AUD-12)*. Based on your June 14, 2017 response to the draft report, we consider all recommendations open and resolved. Once your office has fully implemented the recommendations, please submit evidence of completion of agreed-upon corrective actions so that recommendations may then be closed.

Consistent with our responsibility under the *Inspector General Act, as amended*, we are required to publicly post the report on our website. We also may provide copies of our report to congressional committees with oversight responsibility over the National Archives and Records Administration.

Please call me with any questions, or your staff may contact Jewel Butler, Assistant Inspector General of Audits, at (301) 837-3000.



OFFICE *of*
INSPECTOR GENERAL
NATIONAL ARCHIVES

Table of Contents

Executive Summary	3
Background	4
Objectives, Scope, Methodology	6
Audit Results	7
Finding 1. Availability of NARA’s Digitized Holdings.....	7
Recommendations.....	14
Finding 2. Reporting of Online Availability of Digitized Holdings.....	17
Recommendations.....	18
Appendix A – Acronyms	19
Appendix B – Management Response	20
Appendix C – Report Distribution List	24

Executive Summary

Audit of NARA's Online Access to Digitized Holdings

OIG Report No. 17 AUD 12

June 19, 2017

Why Did We Conduct This Audit?

We conducted this audit because the National Archives and Records Administration's (NARA) most recent Strategic Plan established access as the agency's core purpose.

Specifically, the plan contains the strategic initiative for NARA to make all records available to the public in digital form to ensure that anyone can explore, discover, and learn from NARA holdings. The audit objective was to evaluate NARA's controls in place to achieve its Strategic Goal of "Make Access Happen" by providing online access to its digitized holdings through the National Archives Catalog (NAC).

What Did We Recommend?

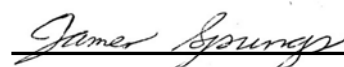
We made eight recommendations to further NARA's achievement of its strategic goal to Make Access Happen. The recommendations instruct NARA to make the necessary resource commitment to providing access to its digitized holdings via the NAC, realize current and future cost savings, and revise its internal reporting metric concerning online access to digitized holdings.

What Did We Find?

We found that although NARA has made progress in making its digitized holdings available via the NAC through improved controls, there are still millions of digitized holdings that need to be placed on the NAC which have not yet been made publicly accessible. We attribute these conditions to inadequate system performance, incompatible partner digitized holdings, inadequate planning and goal setting, and a lack of resources. As a result, the American public has been unable to access the full extent of NARA's digitized holdings online through the NAC.

Further, the agency has reported its main opportunity to provide access lies online. The online portal for researchers to discover the digitized holdings of NARA is the NAC. Public access to records allows Americans to claim their rights of citizenship, hold their government accountable, and understand their history so they can participate more effectively in their government. NARA's inability to make accessible online the full extent of its digitized holdings affected the agency's ability to serve its core purpose and meet the initiatives of its strategic plan.

Also, we found NARA's Performance Measurement and Reporting System (PMRS), NARA's formal mechanism for reporting metrics for online access to digitized holdings, was inaccurate. Specifically, it indicated more than triple the actual amount of digitized holdings available. The reporting mistake resulted from a calculation error within the metric from the source data provided. The Office of Inspector General (OIG) notified NARA of this reporting error. Had NARA relied upon the PMRS metric for reporting purposes, it would have significantly overstated actual progress toward making access happen through placing digitized holdings online through the NAC.



James Springs

Inspector General

Background

The mission of the National Archives and Records Administration (NARA) is to provide public access to Federal Government records in NARA's custody and control. Public access to government records strengthens democracy by allowing Americans to claim their rights of citizenship, hold their government accountable, and understand their history so they can participate more effectively in their government. Additionally, NARA's vision is to be known for cutting-edge access to extraordinary volumes of government information and unprecedented engagement to bring greater meaning to the American experience. To accomplish its mission and achieve its vision, NARA established the strategic goal "Make Access Happen" in its 2014-2018 Strategic Plan, which establishes public access as NARA's core purpose. Under the strategic goal, NARA's strategic objective is to make all records available to the public in digital form to ensure that anyone can explore, discover, and learn from NARA holdings. NARA intends to accomplish this objective through three initiatives: description, digitization, and processing.

NARA's 2014-2016 Open Government Plan (OGP) established the agency's flagship initiative: Innovate to Make Access Happen, which describes NARA's description, digitization, and online access efforts. The 2014-2016 OGP states to make access happen NARA needs to establish more substantial "roots" that increase the number of records NARA digitizes and strengthen the "core" systems that serve as NARA's platforms for all of its online access. NARA's 2015 Annual Performance Plan (APP) stated the agency's main opportunity for making access happen clearly lies online. Therefore, NARA operates the National Archives Catalog (NAC) to provide an online portal to NARA's records and information about the records.¹ The 2016-2018 NARA OGP modified the flagship initiative to Make Access Happen through Innovation. The 2016-2018 OGP set ten initiatives, including the NAC, to increase access to records of NARA through innovation. For the NAC, the 2016-2018 OGP states over the next two years, NARA will focus on improving the performance and scalability of the system, and using open source platforms, so that it can effectively provide access to tens of millions and eventually hundreds of millions of records.

NARA's Chief Innovation Officer plans, directs, and coordinates NARA's Open Government efforts, the online public catalog, digitization strategy and partnerships efforts, and NARA's web and social media presence. NARA's online public catalog, the NAC, is maintained by the Digital Public Access Branch (VEO), under the Office of Innovation's Digital Engagement Division. In addition to maintaining the NAC, VEO creates, manages, and owns the tools that

¹ Launched in December 2014, the NAC contained over 21 million digital objects, with more than 12 million from partners as of January 1, 2017.

locate and provide digital access to holdings, formulates and implements NARA's approach to providing digital access to electronic and digitized holdings, and provides training and quality control of metadata for digital access to NARA's holdings. VEO accomplishes its responsibilities through its staff of ten employees, plus a branch chief. Information Services supports the Office of Innovation in meeting customers' needs for effective and innovative social media, open government, and digitization services, solutions, and systems. Information Services provides support for the NAC, including requirements writing, project management services, systems engineering, acquisition strategy, and system testing. NARA's progress in completing its strategic objective of digitizing all analog archival records to make them available online, is measured and tracked via NARA's Performance Measurement and Reporting System (PMRS).

The majority of digitized holdings at NARA have been the output of NARA's digitization partnerships.² To date, partners have provided approximately 235 million images, along with the corresponding descriptive metadata. NARA custodial units (e.g. Research Services, Presidential Libraries) also provide VEO with descriptions and digital images for upload to the NAC. The process is outlined via the *Adding Descriptions and Digital Objects to the National Archives Catalog* document. VEO tracks and approves submitted requests to create access to NARA's holdings. Once approved, VEO will enter the description into NARA's Description and Authority Service (DAS) system.³ If a digital image is provided with the description, DAS will contain a link to the location of the image, usually the Amazon S3 Cloud. From DAS, weekly export files are created for ingest into the NAC. The NAC's ingest process will pull the description from DAS and the image from a Cloud location, thereby creating the digital object available to users online.

NARA's Office of Inspector General (OIG) previously conducted four audits relevant to the current audit objective. In Fiscal Year (FY) 2014, the OIG issued Audit Report No. 14-12, *Audit of NARA's Digitization Program*. This report made 11 recommendations to strengthen NARA's Digitization Program. The report also suggested six follow-on audits, three of which were conducted in FY15. OIG Audit Report No. 15-10, *Audit of NARA's Digitization Partnerships*, made 11 recommendations to strengthen management of NARA's digitization partnerships. OIG Audit Report No. 15-11, *Audit of NARA's Digitization Storage and Transfer Capabilities*, made 13 recommendations for NARA to strategize and provide long-term storage for its digitized holdings. OIG Audit Report No. 15-12, *NARA's Online Public Access Development Effort*, made one recommendation to require evidence that sufficient system testing was performed by the vendor in future system builds containing significant changes. As of the date of this report, seven recommendations related to NARA's digitization program remain open.

² NARA defines partners as external entities such as foundations, universities, libraries, individuals, professional associations, commercial entities, and Federal, state, and local agencies with which NARA collaborates to digitize and make archival holdings available in digital format.

³ DAS is the data-entry system used by staff to describe NARA's holdings and archival creators.

Objectives, Scope, Methodology

The objective of the audit was to evaluate NARA's controls in place to achieve its Strategic Goal of "Make Access Happen" by providing online access to its digitized holdings through the National Archives Catalog. To accomplish our audit objective, we identified and reviewed the following documentary evidence:

- NARA's 2014-2018 Strategic Plan;
- NARA's 2014-2016 and 2016-2018 OGPs;
- NARA 101, Part 5 – Innovation, and Part 10 – Information Services;
- Adding Descriptions and Digital Objects to the National Archives Catalog;
- NARA's Strategy for Digitizing Archival Materials for Public Access, 2015-2024;
- NARA's Principles for Partnerships to Digitize Archival Materials; and
- Government Accountability Office (GAO)'s *Standards for Internal Control in the Federal Government*.

Further, we reviewed NARA's 2014 and 2015 APPs, 2014 and 2015 Annual Performance Reports, FY14 and FY15 Summary Reports, and 2016 Agency Financial Report; meeting minutes; cost estimates; performance metrics; previous OIG audit reports; system requirements; performance work statements; internal reports; risk assessments; and other internal documents. We conducted judgmental sampling on processing times and availability of records. The results of the sample cannot be projected to the intended population. We obtained testimonial evidence from NARA personnel in the Office of Innovation, Office of Information Services, Office of Strategy and Performance, Research Services, and the Office of Legislative Archives, Presidential Libraries, and Museum Services. Testimonial evidence obtained was corroborated with documentary evidence when available, and was also corroborated through confirmation from multiple NARA personnel. We also obtained physical evidence via observation of VEO staff working in the NAC and DAS systems.

This performance audit was conducted in accordance with generally accepted government auditing standards between October 2016 and February 2017 at Archives II in College Park, MD. The generally accepted government auditing standards require 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 the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

The audit was conducted by William Brown, Senior Program Auditor.

Audit Results

Finding 1. Availability of NARA's Digitized Holdings

Although NARA has made progress in making its digitized holdings available via the NAC through improved controls, there are still millions of digitized holdings that need to be placed on the NAC which have not yet been made publicly accessible. We attribute these conditions to inadequate system performance, incompatible partner digitized holdings, inadequate planning and goal setting, and a lack of resources. As a result, the American public has been unable to access the full extent of NARA's digitized holdings online through the NAC.

NARA's mission is to provide public access to Federal Government records in its custody and control. NARA's 2014-18 Strategic Plan established access as NARA's core purpose and stated NARA would make all records available to the public in digital form to ensure that anyone can explore, discover, and learn from NARA holdings. Based on statistics for online visits versus in-person research visits, the agency's main opportunity for making access happen clearly lies online. To provide online access to its digitized holdings, NARA operates the NAC, the online portal to NARA's records and information about its records.

The largest quantity of digitized holdings has come from NARA's digitization partners. To date, partners have provided approximately 235 million images to NARA; 12.9 million of which were available on the NAC as of January 1, 2017. While the 12.9 million images represent a significant increase in access since our previous audits in the digitization area, there were still millions of digitized holdings NARA could have made accessible, but did not. The most significant sources of digitized holdings not yet accessible include the following:

- **Records No Longer in Embargo Period** – NARA digitization partnership agreements contain an embargo provision, where NARA will not have full and unrestricted rights to use the partnership digitized records until five years after the records were digitized. Once the embargo period is over, NARA has the right to publish all partner digitized images on the NAC. To date, all records digitized by partners in 2007-2011 are out of embargo and eligible to be made accessible. There are approximately 35 million images no longer in an embargo period, but not available on the NAC.
- **Records Without an Embargo Period** – One digitization partnership agreement does not contain an embargo period. Therefore, all images provided by this partner were immediately eligible to be made accessible via the NAC. This partner provided 45 million images to NARA through June 2014, and we estimate the partner has digitized approximately 20 million more since. None of the approximately 65 million images digitized by this partner are available via the NAC.

Custodial units provide digitized holdings to the NAC for access as well, but do not have the quantity of digitized holdings the partners have provided. A NARA official stated access was becoming more ingrained in the NARA workflow, so more access would be created moving forward.⁴ A limiting factor for access currently is the use of hard drives by custodial units to provide images to VEO. Custodial units must save their digitized holdings to a hard drive, then deliver the full hard drive to VEO. VEO must then upload the images off the hard drive onto the Cloud. Electronic Records Archives (ERA) 2.0⁵ is the proposed solution to cut out the hard drives from the process and allow images to flow directly into ERA 2.0, where the NAC can pull the image.

Inadequate System Performance

NARA was hindered in its ability to make access happen because of inadequate system performance of the NAC. Limiting factors for the NAC performance include its procurement and development, its ingest capabilities, and the stovepiped⁶ nature of the system.

Procurement and Development. The development of the NAC was hindered by a combination of poorly written requirements from NARA and poor performance under the contract by the vendor. NARA user acceptance testing of the system found a multitude of defects during each round of testing. OIG Audit Report No. 15-12 reported on the development efforts of the Online Public Access (OPA) system (which became the NAC) and included a finding concerning testing of the system. The production version of the OPA system was delayed several times, and NARA officials expressed much concern and criticism over vendor performance. Despite efforts to hold the vendor accountable, NARA eventually had to accept the system as is.

The system was not robust enough to be horizontally scaled to meet the increasing volume of data NARA planned to ingest into the NAC. Multiple NARA officials expressed dissatisfaction with this vendor's development efforts and the agency's inability to hold the vendor to the contractual requirements. This situation led to NARA being unable to create the significant access it could have provided given a fully functional system. Subsequently, NARA declined the option years on the contract and re-competed the operations and maintenance (O&M) and development, modernization, and enhancement (DM&E) contract for the NAC.

In March 2016, NARA selected a new vendor for both the NAC O&M and DM&E contracts.

The new vendor had developed DAS, an effort in which NARA had a positive vendor experience. The long-range plan for the new vendor is to move the NAC and DAS systems into a more tightly integrated system, attempting to eliminate the stovepipes present with the

⁴ NARA projects internal digitization output to triple from FY16 to FY18.

⁵ Since 2008, NARA has been using its ERA system to take in and store electronic records from across the Federal government. ERA 2.0 is an on-going modernization of the existing system.

⁶ A stovepiped system is a pejorative term for a system that has the potential to share data or functionality with other systems, but which does not.

separately developed systems. While the new vendor had not yet provided completed task orders when the audit began, NARA officials from multiple offices expressed optimism the vendor's work would be satisfactory for the agency and would improve the systems. This sentiment was a common theme throughout the audit. As of the end of field work, both task orders underway were ahead of schedule, and user acceptance testing of the first task order found no defects. NARA officials expressed continued satisfaction with the new vendor's performance.

While there is optimism the system problems will be alleviated moving forward, the NAC as delivered by the contractor was not the robust solution NARA needed to make significant access happen. Further, the software supporting DAS lacks scalability to store and query the volume of data necessary. Once the systems are scaled to allow for a significant quantity of digitized holdings to be ingested and hosted, the agency will see a potential \$1 million per year increase in Cloud storage costs. One NARA official expressed uncertainty as to whether NARA was aware of the need and whether the agency was prepared to handle the cost increase.

Ingest Capabilities. The current ingest process for placing digitized holdings into the NAC causes significant delays in "making access happen." Once VEO processes the description in DAS, a batch of descriptions is exported to the NAC Queue, where the contractor picks up the descriptions for ingest into the NAC. The batches sent to the NAC Queue are picked up by the contractor on a first-in, first-out basis. When ingest of the first batch is complete, the contractor will automatically pick up the next batch. As the NAC pulls the images related to the descriptions, the images undergo tiling, where zoom features are added. Current ingest times for the NAC can take anywhere from a week to two months based on size of files and system use.

During the audit, the new vendor for the NAC and DAS was developing two task orders: Task Order 3, DAS Delta Export; and Task Order 4, Image Decoupling. Both task orders were ahead of schedule as of the end of fieldwork, with both expected to be in production by March 31, 2017.

- For Task Order 3, DAS will be improved so when it exports files to the NAC, it exports only those files which have been created since the last export. Currently, DAS exports all files during its export process, a time-consuming process, instead of only the descriptions entered in the previous week. A NARA official expects significant system improvement once DAS can process only the delta, and not the total.
- For Task Order 4, images will be processed separately from, but in parallel fashion with other objects ingested into the system. This task order will improve the NAC through relieving ingest bottlenecks and allowing for better, and more readily available, access to NARA holdings.

While NARA officials were optimistic about system processing gains realized through implementation of these task orders, one NARA official expressed concern these improvements could not be used until the full technical refresh of the systems occurred. As currently planned, a

backend technical refresh necessary for the NAC and DAS to handle increased volumes would not be in production until early FY19. A backend technical refresh will also allow NARA to replace a proprietary ingest system with free, open source technology, resulting in potential annual cost savings of \$50,000. A prototype for the design of the backend refresh is planned for the end of FY17, but budgetary concerns make the accomplishment of this prototype uncertain. NARA placed a monthly limitation⁷ on the volume of data the NAC and DAS could ingest. The task order improvements could increase the limit at first, but even lower limits might have to be placed on the systems once initial gains are realized. For now, the two million images per month limit is expected to allow consistent ingest into the NAC until a technical refresh can be implemented.

Stovepiped systems. While NARA is moving toward eliminating the stovepipes between the NAC and DAS, the two systems as developed were limited in their compatibility. Due to compatibility limitations between the two systems, NARA is spending funds for Cloud storage which could potentially be put to better use. The approximately 35 million images NARA was able to move into Cloud storage are in two different Amazon Cloud buckets: the S3 Cloud and the Glacier Cloud.⁸ The DAS description of partner images will provide a link to the image in the S3 Cloud. Upon ingest, the NAC will marry the description from DAS with the image in the S3 Cloud. Once the image is copied to the NAC, NARA will have a third copy of the same image in the Cloud (one in the NAC, one in the S3 Cloud, and one in the Glacier Cloud).

The copy is kept in the S3 Cloud because DAS is not configured to recognize the image located in the NAC. Therefore, if the image in the NAC needs to be adjusted, the image will have to be re-ingested by the NAC, marrying the S3 Cloud image with the DAS description. Although VEO receives only a few requests per month to change images in the NAC, all images in the NAC are also stored in the S3 Cloud in case an image needs to be updated. The OIG conservatively estimates NARA would save at least \$1,000 per month if it deleted the duplicate copies in the S3 Cloud once the image was ingested in the NAC. One NARA official stated the estimate was very conservative, and believed NARA's cost savings would be between \$2,000 and \$3,000 per month in actuality. Further, the official stated the cost savings realized would increase as NARA increases the volume of images ingested into the NAC.

NARA has a planned task order to update the systems so DAS will recognize the image in the NAC, and the S3 Cloud copy can be automatically deleted upon image ingest into the NAC. However, there is uncertainty as to when this task order could be put into production. Already in FY17, VEO staff has managed to reduce the monthly Cloud storage bill by deleting superfluous data from the S3 Cloud. By continuing this effort, and deleting duplicate images in the S3 Cloud, NARA can realize immediate cost savings.

⁷ Maximum of two million images ingested per month.

⁸ The Glacier Cloud is cheaper, long-term storage while the S3 Cloud is more expensive working storage.

Incompatible Partner Digitized Holdings

All partner images received by NARA on hard drives were not in a format that could be made available on the NAC.⁹ Therefore, NARA executed a contract to upload some partner images into the Amazon Cloud while simultaneously converting the image into a NAC-compatible format. Prior to the images being converted, VEO was unable to make the images accessible. However, NARA only converted a portion of its overall partner images. All of the estimated 200 million partner images in NARA's possession on hard drives will need to be converted before the images can be made accessible. Moving forward, NARA is considering writing a script to convert partner images when uploaded to the Cloud from the hard drive.

Not only do the partner images require conversion, all metadata received from partners must be converted before access can happen. The partner metadata is not in a format that can be imported into DAS. Each set of partner images would often require a unique, manual conversion of metadata to format the descriptions into a DAS-compatible version. The process was manual and very time intensive. VEO staff was burdened with having to find a unique conversion method for every batch of partner digitized holdings worked. Eventually, VEO was able to develop a Python¹⁰ script to standardize the metadata conversion process. With the implementation of the Python script process in March 2016, VEO was able to more quickly convert partner metadata, thereby increasing access to partner records from that point forward. The Python script will be used for all partner metadata received, but not yet processed. One NARA official stated short of hiring a programmer, the partner metadata conversion process was working as best it could.

Inadequate Planning and Goal-setting

Planning. NARA has not developed a plan for how, or if, it will use the Glacier Cloud to promote efficiency and economy of its storage of partner images. The Glacier Cloud provides cost effective long-term Cloud storage and is another area where NARA has the opportunity to realize cost savings. A NARA official stated Glacier Cloud storage is about one-tenth the cost of S3 Cloud storage. As NARA is charged for large movements of data out of the Glacier Cloud, a NARA official stated there is a need for NARA to have someone actively manage the data flow in the Glacier Cloud to maximize its use. Another official stated there is a need for NARA to perform data analytics to identify the best use of the Glacier Cloud. Multiple officials stated they had a vision of using the Glacier Cloud economically, and only using the S3 Cloud to work on images just before the images would be made accessible.

Additionally, the agency has been unable to execute a plan to move approximately 200 million partner images off over 600 hard drives and into Cloud storage despite an open OIG

⁹ See page 11 of OIG Audit Report No. 15-10.

¹⁰ Python is a widely-used programming language.

recommendation and the Chief Innovation Officer declaring the hard drive images as a high risk area for the office. OIG Audit Report No. 15-11¹¹ detailed problems with the FY14 attempt at procuring Cloud storage for the hard drives. In FY16, the Office of Innovation included \$650,000 for Cloud storage of hard drive data in its mid-year budget request, which was not approved. The \$650,000 total includes the cost of upload to the Cloud and the annual carrying cost of the storage. Once uploaded, NARA estimates spending \$400,000 per year in Cloud storage cost for the partner images. A NARA official estimates the upload of all hard drives to take approximately two years.

There is a risk decision to be made by the agency in regard to these hard drives. As some of these hard drives are several years old, there is risk of data loss if the images are not moved off the hard drives. Conversely, there is risk of spending significant funds on Cloud storage when the images cannot be made accessible. For example, if NARA had moved forward with the acquisition in FY14, it could have spent over \$2 million in Cloud storage costs for partner images before the images could begin to be made accessible. The agency must decide whether to allot the funds to move the hard drive images into a different storage environment to decrease the risk of loss, or keep the images on the hard drives to save funds until a future date when the images on the hard drives can be quickly made accessible via the NAC.

Goal-setting. Under the overall goal for access, NARA has set goals of different image volumes and completion dates. These goals were not always accomplished, and were not always set to maximize the volume of access. The limitations noted elsewhere in this finding contributed to NARA not successfully achieving these goals. During the audit, NARA worked to create a robust access goal, and while the OIG is satisfied with NARA taking the initiative to set a long-term access goal and strategize options for achieving the goal, there are several concerns with how the goal will be reached.

In February 2013, the Chief Innovation Officer informed the Executive Leadership Team (ELT) 12.6 million images from partners would be made available by January 1, 2014 (when the embargo period ended). When this goal was not met, a new goal was set to make the 12.6 million partner images available by September 30, 2014; however, the goal was not met either.¹² In FY16, the VEO was tasked with the goal of making 10 million images available via the NAC during the year, with the majority coming from partners. This goal was accomplished, though a NARA official stated the 10 million images was the maximum VEO could have accomplished during the year. With an additional resource and improved processes, a NARA official stated an FY17 goal might be more ambitious (between 20 million and 40 million images added), but annual targets were still under discussion as of end of fieldwork.

¹¹ See page 19 of OIG Audit Report No. 15-11.

¹² 12.6 million partner images were not available via the NAC until FY17.

During initial audit meetings, we were told of a new goal of making 500 million images available, which would be accomplished by FY18. However, during the course of the audit, agency staff determined achievement of the goal by FY18 was not only unlikely, but impossible given the system development work necessary for the NAC to handle the volume of images. Therefore, the achievement date was pushed back to the end of FY20. The Office of Innovation developed several strategies to meet the goal, ranging in cost from \$16 million to \$26 million. As of the end of fieldwork, the goal had not been formally adopted by the ELT.

While the later achievement date makes accomplishment of the goal more likely than before, much work has to be done to achieve it. Almost half of the 500 million images are currently digitized: the remaining portion will be digitized over the next three and a half years. Most of the digitization work is planned to occur internally requiring a substantial increase in internal digitization output. To increase the output, the Office of Innovation planned a build out of a digitization center at Archives II. An information technology plan for the digitization center was approved by the Chief Information Officer in August 2016, but the plan is not yet funded. One NARA official expressed concern the equipment currently used for digitization would be unable to handle the increase in work.

The majority of internal digitization output would come from the digitization of NARA's existing microfilm collection. Microfilm collections are listed on NARA's Digitization Priority List, and one NARA official stated the microfilm would be of interest to researchers. However, other NARA officials voiced concerns over the digitization of microfilm. One official expressed concern of microfilm digitization being planned because microfilm is the cheapest and easiest form of NARA record to digitize, and not necessarily the most valuable to researchers. This official was critical of NARA expending resources to meet a superficial number that sounds good. Another NARA official echoed the same concerns of creating access to microfilm if researchers would have no interest, and the potential waste of resources to create and sustain access if the images were never viewed. Another concern is the quality of the microfilm scan being less than that of an original record.

Lack of Resources

Another limiting factor in the Office of Innovation creating access was lack of personnel resources. VEO had one staff member in FY16 dedicated to uploading descriptions and digital objects to the NAC. When the employee was out of the office, entries to DAS and uploads to the NAC were limited, if not completely stopped. NARA officials acknowledged the single contact at VEO. One official conveyed satisfaction with the work of the VEO contact, but stated access from their office would increase if there were two people working on uploading digital objects into the NAC. In FY17, VEO added a new position trained in DAS entry and the NAC upload. With this second staff member devoting a majority of their time to making access happen through the NAC, access will increase when both VEO employees are present, and not come to a

halt when one employee is unavailable. With the addition of the new employee, a NARA official stated the branch was now adequately staffed.

NARA has plans to improve the NAC and DAS to provide access to tens of millions and eventually hundreds of millions of records. However, according to NARA officials, continued uncertainty with the Federal budget situation casts doubt on what needed fixes the agency will be able to accomplish in a timely manner.

Recommendations

We recommend:

Recommendation 1: The Archivist of the United States prioritize and commit necessary resources to address all risks to NARA's program for providing online access to NARA's digitized holdings via the NAC.

Management Response

NARA concurs with this recommendation. NARA will develop a list of risks and estimated costs to mitigate each risk. Also, NARA will develop expenditure plans that prioritize risk mitigation activities for funding within the NAC and DAS budgets.

Target Completion Date: August 31, 2018

OIG Analysis

We consider NARA's proposed actions responsive to our recommendation. This recommendation will remain open and resolved pending completion of the corrective actions identified above.

Recommendation 2: The Chief Innovation Officer and Chief Information Officer develop long-term cost estimates for the continued storage of digitized holdings in the NAC.

Management Response

NARA concurs with this recommendation. NARA will develop requirements and an Independent Government Cost Estimate for continued storage of digitized holdings in the NAC.

Target Completion Date: November 30, 2018

OIG Analysis

We consider NARA's proposed actions responsive to our recommendation. This recommendation will remain open and resolved pending completion of the corrective actions identified above.

Recommendation 3: The Chief Innovation Officer and Chief Information Officer commit resources necessary to complete the backend technical refresh of the NAC and DAS.

Management Response

NARA concurs with this recommendation. NARA will develop a technical refresh plan and include the costs of the technical refresh in the spend plan described in response to Recommendation 1.

Target Completion Date: August 31, 2018

OIG Analysis

We consider NARA's proposed actions responsive to our recommendation. This recommendation will remain open and resolved pending completion of the corrective actions identified above.

Recommendation 4: The Chief Innovation Officer and Chief Information Officer implement solution to delete duplicate copies of digitized holdings to realize cost savings.

Management Response

NARA concurs with this recommendation. NARA will define requirements for the solution to delete duplicate copies of digitized holdings.

Target Completion Date: October 31, 2018

OIG Analysis

We consider NARA's proposed actions responsive to our recommendation. This recommendation will remain open and resolved pending completion of the corrective actions identified above.

Recommendation 5: The Chief Innovation Officer and Chief Information Officer develop a strategy to optimize the use of NARA's Cloud infrastructure as it pertains to providing online access to NARA's digitized holdings.

Management Response

NARA concurs with this recommendation. NARA will estimate long-term requirements for Cloud storage and related services as it pertains to providing online access to NARA's digitized holdings. Further, NARA will develop a strategy to optimize NARA's Cloud infrastructure.

Target Completion Date: December 31, 2018

OIG Analysis

We consider NARA's proposed actions responsive to our recommendation. This recommendation will remain open and resolved pending completion of the corrective actions identified above.

Recommendation 6: The Chief Innovation Officer make a documented risk decision on the hard drives.

Management Response

NARA concurs with this recommendation. NARA will document the risk decision on the hard drives.

Target Completion Date: April 30, 2018

OIG Analysis

We consider NARA's proposed actions responsive to our recommendation. This recommendation will remain open and resolved pending completion of the corrective actions identified above.

Recommendation 7: The Chief Innovation Officer create a formal timeline to move partner data off the hard drives and make it accessible.

Management Response

NARA concurs with this recommendation. NARA will create a formal timeline for moving the partner data off the hard drives and making it accessible.

Target Completion Date: June 29, 2018

OIG Analysis

We consider NARA's proposed actions responsive to our recommendation. This recommendation will remain open and resolved pending completion of the corrective actions identified above.

Finding 2. Reporting of Online Availability of Digitized Holdings

We found NARA's Performance Measurement and Reporting System, NARA's formal mechanism for reporting access metrics for online access to digitized holdings, was inaccurate. Specifically, it indicated more than triple the actual amount of digitized holdings available. The reporting mistake resulted from a calculation error within the metric from the source data provided. The OIG notified NARA of this reporting error. Had NARA relied upon the PMRS metric for reporting purposes, it would have significantly overstated actual progress toward making access happen through placing digitized holdings online in the NAC.

GAO's *Standards for Internal Control in the Federal Government* states management should use quality information¹³ to achieve the entity's objectives. Management uses the quality information to make informed decisions and evaluate the entity's performance in achieving key objectives and addressing risks. PMRS is the official source for statistical management information at NARA. In particular, it reports how NARA is doing relative to the goals in NARA's Strategic Plan and the various APPs. Those results form the basis for reports to Congress each year.

PMRS Metric 3101 tracks the percentage of traditional records digitized and available online through the NAC. This metric supports the strategic goal of Make Access Happen. The metric was revised in March 2016 to report only those records digitized and available on the NAC.¹⁴ A user can view the metric based on pages available, images available, and then the number of cubic feet available. The cubic feet availability is determined by converting the number images to cubic feet, then dividing by another PMRS metric on NARA's total cubic feet of traditional records. Data support for the pages and images online is provided monthly to the PMRS contractor by VEO in the form of VEO's tracking spreadsheet of uploads to the NAC.

We downloaded the PMRS metric in January 2017, where the metric reported over 69 million images available online via the NAC through December 2016. However, VEO's tracking spreadsheet, where the number of images is updated based on system reports from the NAC, reported 21 million images online via the NAC as of the first week of January 2017. Additionally, an update posted to NARA's internal website by the Chief Innovation Officer reported 21 million images were available via the NAC as of January 5, 2017. PMRS officials were confident in the system's number, as the data came directly from VEO. We showed the PMRS metric to an Innovation official, who stated the metric was incorrect. The official assumed PMRS might have been taking a cumulative total instead of adding the monthly increases, but was unsure of the cause. The OIG attempted a number of calculations to

¹³ Quality information is appropriate, current, complete, accurate, accessible, and provided on a timely basis.

¹⁴ The revision occurred based on the implementation of Recommendations 9 through 11 from OIG Audit Report No. 14-12.

determine how the PMRS metric more than tripled the actual number, but was unable to reconcile the difference based on data available to us. As a result of the OIG notifying VEO of the discrepancy, an internal discussion took place to revise the PMRS metric to accurately reflect the quantity of images available via the NAC. However, no changes had been made to the metric as of the end audit fieldwork.

Recommendations

We recommend:

Recommendation 8: The Chief Innovation Officer ensure the revision of PMRS metric 3101 to accurately measure the quantity of digitized holdings available via the NAC, and the on-going monitoring of the accuracy of the metric.

Management Response

NARA concurs with this recommendation. NARA will establish a plan to ensure the revision of PMRS metric 3101 to accurately measure the quantity of digitized holdings available via the NAC and for monitoring its accuracy on an ongoing basis.

Target Completion Date: December 29, 2017

OIG Analysis

We consider NARA's proposed actions responsive to our recommendation. This recommendation will remain open and resolved pending completion of the corrective actions identified above.

Appendix A – Acronyms

APP	Annual Performance Plan
DAS	Description and Authority Service
DM&E	Development, Modernization, and Enhancement
ELT	Executive Leadership Team
ERA	Electronic Records Archives
FY	Fiscal Year
GAO	Government Accountability Office
NAC	National Archives Catalog
NARA	National Archives and Records Administration
OGP	Open Government Plan
OIG	Office of Inspector General
O&M	Operations and Maintenance
OPA	Online Public Access
PMRS	Performance Measurement and Reporting System
VEO	Digital Public Access Branch

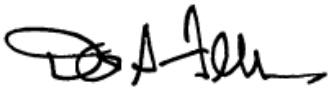
Appendix B – Management Response



Date: JUN 14 2017
To: James Springs, Inspector General
From: David S. Ferriero, Archivist of the United States
Subject: Management's Response to OIG Report 17-AUD-12, *Audit of NARA's Online Access to Digitized Holdings*

Thank you for the opportunity to provide comments on this final report. We appreciate your willingness to meet and clarify language in the report. We understand that, as with all audits, this Report, including its Executive Summary, Background, Audit Results and Recommendations, only addresses matter up to the conclusion of the OIG's audit field work, which ended in February 2017.

We concur with the eight recommendations in this audit, and in response, the attachment provides a summary of our proposed actions. As each recommendation is satisfied, we will provide documentation to your office. If you have questions about this action plan, please contact Kimm Richards at kimm.richards@nara.gov or by phone at 301-837-1668.



DAVID S. FERRIERO
Archivist of the United States

Attachment

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

**Action Plan Response to OIG Report:
17-AUD-12, Audit of NARA's Online Access to Digitized Holdings**

Recommendation 1: The Archivist of the United States prioritize and commit necessary resources to address all risks to NARA's program for providing online access to NARA's digitized holdings via the NAC.

Planned Action: The Offices of Innovation and Information Services will jointly develop a list of risks and estimated costs to mitigate each risk. Information Services will develop expenditure plans that prioritize risk mitigation activities for funding within the DAS and NAC budgets.

Target Completion Date: August 31, 2018

Recommendation 2: The Chief Innovation Officer and Chief Information Officer develop long-term cost estimates for the continued storage of digitized holdings in the NAC.

Planned Action: The Chief Innovation Officer will develop requirements for the continued storage of digitized holdings in the National Archives Catalog. Pending receipt of requirements the Chief Information Officer will develop an Independent Government Cost Estimate.

Target Completion Date: November 30, 2018

Recommendation 3: The Chief Innovation Officer and Chief Information Officer commit resources necessary to complete the backend technical refresh of the NAC and DAS.

Planned Action: The Offices of Innovation and Information Services will mutually develop a technical refresh plan and Information Services will include those costs in the spend plan included in Management's response to recommendation 1.

Target Completion Date: August 31, 2018

Recommendation 4: The Chief Innovation Officer and Chief Information Officer implement solution to delete duplicate copies of digitized holdings to realize cost savings.

Planned Action: The Chief Innovation Officer will define requirements for the solution to delete duplicate copies of digitized holdings to realize cost. The Chief Information

Officer will provide input for the implementation of the solution submitted by Chief Innovation Officer.

Target Completion Date: October 31, 2018

Recommendation 5: The Chief Innovation Officer and Chief Information Officer develop a strategy to optimize the use of NARA's Cloud infrastructure as it pertains to providing online access to NARA's digitized holdings.

Planned Action: The Chief Innovation Officer will estimate the long-term requirements for cloud storage and related services as it pertains to providing online access to NARA's digitized holdings. The Chief Information Officer and Chief Acquisition Officer will develop a strategy to optimize NARA's Cloud Infrastructure.

Target Completion Date: December 31, 2018

Recommendation 6: The Chief Innovation Officer make a documented risk decision on the hard drives.

Planned Action: The Chief Innovation Officer will document the risk decision on the hard drives with input from Information Services and Research Services.

Target Completion Date: April 30, 2018

Recommendation 7: The Chief Innovation Officer create a formal timeline to move partner data off the hard drives and make it accessible.

Planned Action: The Chief Innovation Officer will create a formal timeline for moving the partner data off the hard drives and to make it accessible.

Target Completion Date: June 29, 2018

Recommendation 8: The Chief Innovation Officer ensure the revision of PMRS metric 3101 to accurately measure the quantity of digitized holdings available via the NAC, and the on-going monitoring of the accuracy of the metric.

Planned Action: The Chief Innovation Officer will establish a plan to ensure the revision of the Performance Measurement and Reporting System metric 3101 to

accurately measure the quantity of digitized holdings available via the National Archives Catalog and for monitoring its accuracy on an ongoing basis.

Target Completion Date: December 29, 2017

Appendix C – Report Distribution List

Archivist of the United States
Deputy Archivist of the United States
Chief Operating Officer
Deputy Chief Operating Officer
Chief of Management and Administration
Chief Innovation Officer
Chief Information Officer
Accountability
United States House Committee on Oversight and Government Reform
Senate Homeland Security and Governmental Affairs Committee

OIG Hotline

To report fraud, waste, or abuse, please contact us:

Electronically: <https://www.archives.gov/oig/referral-form/index.html>

Telephone: 301-837-3500 (Washington, D.C. Metro Area)
1-800-786-2551 (toll-free and outside the Washington, D.C. metro area)

Mail: IG Hotline
NARA
P.O. Box 1821
Hyattsville, MD 20788-0821