NATIONAL ARCHIVES *and*
RECORDS ADMINISTRATION

OFFICE *of* INSPECTOR GENERAL

Audit of NARA's Online Public Access Development Effort

May 26, 2015

OIG Audit Report No. 15-12

Table of Contents

Executive Summary

The National Archives and Records Administration's (NARA) Office of Inspector General (OIG) audited the development effort to upgrade the system for the public to access NARA's holdings online. NARA's mission is to provide public access to Federal Government records in our 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.

Part of NARA's flagship initiative in its 2010 Open Government Plan includes increasing online access and participation with the public by developing streamlined search capabilities for online holdings. This would unlock records from previously stove-piped systems. NARA reported this was accomplished with the launch of the Online Public Access (OPA) System in December 2010, also referred to as the OPA Pilot. However, this system was designed as a pilot or prototype, with improvements planned for the future.

An assessment done of the OPA System's scalability in October 2012 stated "OPA today is a collection of ad-hoc scripts and manual procedures with a poorly configured search engine. It is not scalable beyond the 18+ million documents which it currently contains." In addition, the Business Case states the system is performing below accepted industry averages for response times and, as designed this performance will decrease in direct proportion to the amount of content available in the system. The risk to NARA from this lack of scalability necessitated the need for a new OPA system.

In order to ensure NARA's current and future public search and user engagement needs for online public access to records can be met, an acquisition was initiated for a new online public access system, referred to as the Online Public Access Production (OPA Prod) System. In September 2013, NARA issued a firm-fixed price contract to Search Technologies, currently valued at over $4.5 million, for the design, development, deployment, and application support and maintenance of the OPA Prod System.

In early December, 2014 NARA launched the pre-release of OPA Prod, also referred to as the National Archives Catalog (NAC). Although the project experienced some delays, the NAC has increased functionality and is more scalable than its predecessor. This was the first phase of this multi-year project, with additional functionality planned.

Our audit found a weakness related to the lack of evidence that sufficient system testing was performed by the vendor. We made one recommendation to require evidence that sufficient system testing has been performed by the vendor for all future OPA Prod System builds containing significant changes.

Background

The National Archives and Records Administration's (NARA) mission is to provide public access to Federal Government records in our 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.

In addition, NARA's Strategic Plan Goal 1: "Make Access Happen" establishes "public access" as NARA's core purpose. It affirms that public access is the ultimate outcome of all of NARA's work. Make Access Happen also signals a significant shift in NARA's strategy and purpose. NARA has stated "We will reach beyond the traditional role of making records available for others to discover and will make access happen by providing flexible tools and accessible resources that promote public participation." The objective of this goal 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's flagship initiative in its 2010 Open Government Plan includes increasing online access and participation with the public by developing streamlined search capabilities for its online holdings that would unlock records from previously stove-piped systems. According to the 2012 Open Government Plan this was accomplished with the launch of the Online Public Access (OPA) System in December 2010, also referred to as the OPA Pilot. OPA Pilot was the online portal to NARA's records and information about its records. It provided a single, streamlined search across multiple resources at the Archives, including archival descriptions and authority records from the Archival Research Catalog, selected electronic records from the Access to Archival Database (AAD) System and the Electronic Records Archives, and web pages from Archives.gov and all of the Presidential Libraries. However, this system was designed as a pilot or prototype, with improvements planned for the future.

The Statement of Need from July 2013 indicated the amount of content in the OPA Pilot will increase to an estimated 500 million records over the next five years. This includes known datasets such as digital copies from partners in records digitization projects, George W. Bush emails, 1940 Census records, the full content of the AAD data, and data from in-house digitization projects. It is anticipated this number will grow to billions of documents thereafter, given partner work, Presidential records from prior administrations, and Federal records from the Electronic Records Archives Base System.

A contractor assessment done on the scalability of the OPA System in October 2012 stated "OPA today is a collection of ad-hoc scripts and manual procedures with a poorly configured search engine. It is not scalable beyond the 18+ million documents which it currently contains." In addition, the Business Case states, the system is performing below accepted industry averages for response times and as designed, this performance

will decrease in direct proportion to the amount of content available in the system. This lack of scalability is a risk to NARA that necessitated the need for a new OPA system. In order to ensure NARA's current and future public search and user engagement needs for online public access to records can be met, an acquisition was initiated to obtain services to establish a new online public access system, referred to as the Online Public Access Production (OPA Prod) System. NARA is developing strategic plans to re-frame the focus of the agency to public access to records. A scalable, extensible OPA system is absolutely essential to NARA fulfilling this strategic objective as stated in NARA's acquisition plan. The System's Requirement Specification document states NARA must provide a robust operational system which provides online public access to an increasing number of records, as well as social media tools and approaches for user engagement that are a key part of NARA's strategic mission.

OPA Prod Development

In September 2013, NARA issued a firm-fixed price contract to Search Technologies, currently valued at over $4.5 million, for the design, development, deployment, and application support and maintenance of the OPA Prod System.

The July 2013 Performance Work Statement (PWS) states NARA envisions OPA Prod as a cloud or hosted system to be delivered in three releases. OPA Prod Release one (OPA Prod R1) will be a rapidly developed instance supporting basic search and content management functionality plus a limited number of high priority requirements supporting transcription to be further refined through an iterative requirements process. The design of R1 will be based on the existing OPA Pilot prototype, lessons learned in the development and operation of OPA Pilot as well as this iterative requirements process. Also at the Government's option are subsequent design and development of the second release (R2) and third release (R3). R2 and R3 will be based on those requirements not deemed high priority in the previous release, but which still represent user needs/requirements. The acceptance criteria in the PWS states that 100 percent of the requirements associated with this contract are validated by the Government during testing as having been satisfied. Additionally, there can be no Severity1[1] or Severity 2[2] defects as identified by NARA.

The major system functions outlined for OPA Prod R1 are:

- Search
- Transcription
- Application Programming Interface (API)
- Ingest
- Tagging
- Account Management

See Appendix A for a description of these functions.

[1] Severity 1 (Critical): An issue where a piece of functionality or system component is broken and there is no workaround available. It prevents either use or testing of the system.

[2] Severity 2 (High): An issue where a piece of functionality or system component is broken and a workaround is available. Use or testing of the system can proceed in a degraded mode.

Deployment of OPA Prod R1 was originally scheduled for July 2014, but, deployment of a beta version of OPA Prod was rescheduled for September 2014. This was caused by delays relating to a stop gap at the early stage of the contract due to the Government shutdown, problems during implementation of the design requirements, and then again during the initiation and implementation of an infrastructure to host and store NARA records in a cloud environment.

However, the vendor was unable to meet the September 2014 milestone date due to insufficient OPA Prod System response times, security concerns, and system defects. For example, seven requirements related to search response times, and response times navigating between screens and pages were not met.

Also, a preliminary security review found two high application vulnerabilities. These vulnerabilities could lead to the disclosure of user credentials and passwords; and allow a malicious user to bypass the OPA Prod application layer to directly connect to the OPA Prod backend servers. Vulnerability scans run on the OPA Prod servers found numerous vulnerabilities due to these servers not being patched and updated. These scans revealed one critical vulnerability, five high vulnerabilities, and nine medium vulnerabilities on each of the 14 servers in the OPA Prod environment.

In addition, User Acceptance Testing (UAT) performed in mid-September 2014 identified over 150 open defects (see chart) and 120 requirements that had failed.

OPA Prod User Acceptance Test Results (09/15/14)

Defects:	Open	Closed	Total
Severity 1:	1	2	3
Severity 2:	118	124	242
Severity 3[3]:	34	37	71
Total:	**153**	**163**	**316**

119 Open Severity 1 & 2 Defects

Examples of these defects include:

- Zero results returned under different search selections;
- Missing data fields;
- Invalid data fields displayed;
- Duplicate items displayed;
- Incorrect total image numbers;
- Counts displayed incorrectly;

[3] Severity 3 (Medium): An issue that is either a partial loss of functionality or a cosmetic issue. It does not impact productivity or efficiency.

- Missing data links;
- Incorrect functionality; and
- Incorrect number of results.

These defects resulted in the failure of 120 of the 810 OPA Prod requirements. Based on the Acceptance Criteria in the PWS which requires a 100 percent pass rate of all requirements and the resolution of all Severity 1 and Severity 2 defects, a decision was made not to deploy the beta version of OPA Prod.

The vendor acknowledged not all performance requirements were met by the September deadline, and provided a remediation plan to provide NARA with another build of the system for UAT by October 31, 2014. After that UAT was completed, NARA notified Search Technologies that as of November 21, 2014 they had failed to meet the acceptance criteria for the system delivered on October 31, 2014. The open defects were reported based on two UAT phases. The first phase, UAT1, included the original defects and went through October 10, 2014. There were 28 open defects after UAT1 including 20 Severity 2 and 8 Severity 3 defects. The second phase, UAT2, was testing performed between October 31, 2014 and November 21, 2014. There were 95 open defects after UAT2, comprised of 64 Severity 2 and 31 Severity 3 defects.

NARA also stated to the vendor NARA planned on deploying the pre-release of the October 31, 2014 build on December 4, 2014 and would require a clean-up build for the final version of OPA Prod Release 1/Phase 1. The clean-up build was to fix all open UAT1 defects and a prioritized list of UAT2 defects. The UAT1 acceptance criteria would remain unchanged, as stated in the PWS. However, in order to reach resolution of Release 1/Phase 1, NARA relaxed the acceptance criteria for UAT2 through the clean-up build. There should have been no Severity 1 defects, but NARA revised the acceptance criteria to allow for no more than 10 Severity 2 defects and no more than 30 Severity 3 defects.

On December 4, 2014 NARA initiated the launch of the OPA Prod pre-release, also referred to as the National Archives Catalog (NAC). NARA's website states the catalog is powered by a completely new search engine with improved relevancy rankings and faster response times. The system had been scaled up to initially handle 100 terabytes of data[4] with a future capacity of up to 10,000 terabytes, and NARA claimed to be "more than ready to handle the millions of digital images" created through NARA's external partnerships. According to NARA's website, the NAC is the second system at the National Archives to be launched completely in the cloud, and is fully integrated with the backend system that NARA staff uses to enter descriptions and upload digital content.

A clean-up build was received from the vendor on December 15, 2014. The UAT results found 11 defects needing to be dis--positioned by the vendor. This system was deployed on January 23, 2015 with tracking tickets created for the open defects to be addressed as

[4] NARA management provided clarification that the NAC is scalable to handle 100 terabytes of data; however it is not currently scaled to that capacity.

part of the operations and maintenance process. Since this last deployment, a stemming issue was identified relating to the search results not showing all of the permeations of a word. For example, if the search included the word "fish", the search is not showing the results for "fished" and "fishing". The problem is being researched, but NARA officials think it is being caused by the way the NAC search engine is configured. NARA officials need to determine the desired query parameters and expressions that should be included in doing a search. The fine tuning of search parameters is directly related to the relevancy of search results generated.

NARA's Planned Online Public Access Effort for Fiscal Year 2015

Over 300 additional requirements are planned to be implemented in the NAC during fiscal year 2015. Sixteen requirements, many related to searches on, before, or after a specific day, month, or year are scheduled to be deployed in May 2015. The remaining requirements are planned for delivery in a build from the vendor in July 2015.

Objectives, Scope, Methodology

The overall objective of this audit was to evaluate the development and implementation of the OPA Prod System. Specifically, we sought to determine if the project was on schedule and system functionality was implemented as intended.

To accomplish our objective we performed the following:

- interviewed NARA officials responsible for the oversight, development, and quality assurance of the OPA Prod project;
- requested and reviewed contract and project management documentation; and
- reviewed applicable laws and regulations.

Our audit work was performed at Archives II in College Park, Maryland between July 2014 and March 2015. We conducted this performance audit in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objectives. We believe that the evidence obtained provides a reasonable basis for our finding and conclusion based on our audit objective.

Audit Results

1. Insufficient Evidence of Vendor Testing

We found NARA lacks assurance sufficient system testing was performed by the vendor prior to delivery of the OPA Prod builds. This weakness occurred primarily because the PWS does not specifically account for system testing for more than one build within a release. However, NARA policies, as well as generally accepted system development lifecycle practices, dictate that after any significant changes, a system will undergo adequate testing to determine if the system, and its functionality, is performing as intended. As a result of this testing issue, NARA is receiving OPA Prod System builds with a high number of defects that are not getting identified and resolved prior to UAT.

Due to the volume of defects found during UAT (see pages 6 and 7), we requested the results of the system tests that were conducted by the vendor. The only test report NARA received from the vendor was from August 2014. The results indicated some tests had failed, but there was no evidence corrections were made or that the system had been retested. In addition, the vendor provided system builds to NARA on October 31, 2014 and December 15, 2014 containing significant system changes. These builds were provided to NARA for UAT, without any test reports, and NARA has no assurance sufficient system testing (including regression testing[5]) was conducted by the vendor on these builds.

As stated in the PWS, the vendor shall be responsible for all pre-customer acceptance testing (i.e., unit testing, software integration testing, and systems integration testing) and must ensure the final system is delivered per requirements and the design. The PWS only requires three test reports as deliverables, one for each of the three releases. However, it fails to take into consideration there can be multiple system builds within a release that may contain significant changes. The PWS also states the vendor shall work within NARA's systems development lifecycle framework.

NARA Directive 805, Systems Development Life Cycle (SDLC), and its handbook, the NARA SDLC Methodology, provide guidance for project teams who are acquiring, developing, deploying, or maintaining information technology (IT) systems, IT infrastructure, or IT services at NARA. These policies, as well as generally accepted industry standards, require system testing to take place prior to UAT for systems under development or systems undergoing significant changes. The PWS is currently being revised and should include deliverables providing NARA evidence sufficient system testing has been conducted on all OPA Prod System builds containing significant changes.

[5] Regression testing is performed to confirm a recent program or code change has not adversely affected existing functionalities.

As a result of the current conditions, the vendor is delivering OPA Prod System builds with a high number of defects that are not getting detected until UAT. With a limited number of resources, the identification and resolution of a large volume of defects is an additional burden for NARA's Quality Assurance group, who estimate 70 percent of the defects they found should have been identified and corrected during system testing.

Recommendation

We recommend NARA's Chief Information Officer:

1. Work with the Contracting Officer on the Search Technologies contract to require evidence sufficient system testing (including regression testing) has been performed by the vendor prior to UAT for all OPA Prod System builds containing significant changes.

Management Response

Management concurred with the recommendation.

Appendix A - Major OPA Prod Release 1 System Functions

- **Search**
  OPA provides users with the ability to search for and access information about the National Archives' holdings. OPA provides the capability to search across multiple data sources, conduct basic searches, parametric searches, and search descriptive metadata as well as the content of electronic records and digitized records available in the catalog. The OPA system *Search* capability will enable users to view, search, browse, retrieve, export and print content based on defined user roles and permissions.

- **Transcription**
  The OPA system *Transcription* capability will allow for registered users to document the text represented in a document, or the content of the spoken language in a motion picture or audio recording. Transcriptions help users in searching for the records online as well as in reading and understanding the records.

- **Tagging**
  The OPA *Tagging* capability will allow for registered users to add a string (keyword, phrase, and identifier) to a record, which helps categorize the record in some way and allows it to be found again by searching. Tags help enhance search and retrieval.

- **Application Programming Interface (API)**
  The OPA *API* functionality will allow for individual consumers and/or external systems to login, authenticate, query and fetch records, as well as contribute data. Each API would be documented for consumers, with a description of common fields, search results presentation methods, and document presentation methods.

- **Ingest**
  OPA provides the capability to ingest and process a data file and incorporate it into OPA so that it can be stored, indexed, and searched. This includes, but is not limited to, archival description and authority data files, as well as digital objects and electronic records associated with the archival descriptions. The OPA system *Ingest* capability shall allow for: the transfer of electronic records in all supported file formats and/or media into the system; support for verification of transfer contents and record sampling; and the ability to define file folder structures.

- **Account Management**
  The OPA *Account Management* capability will allow for users to setup an account in OPA to contribute metadata such as tags and transcriptions, and to save search results. The Account Management capability will also allow for authorized users to manage user accounts and monitor user-contributed metadata.

Appendix B – Acronyms and Abbreviations

AAD	Access to Archival Database
API	Application Programming Interface
IT	Information Technology
NAC	National Archives Catalog
NARA	National Archives and Records Administration
OIG	Office of Inspector General
OPA	Online Public Access
OPA Prod	Online Public Access Production
PWS	Performance Work Statement
SDLC	Systems Development Life Cycle
UAT	User Acceptance Testing

Appendix C - Management's Response to the Report

NATIONAL ARCHIVES

Date: MAY 19 2015

To: James Springs, Inspector General

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

Subject: OIG Draft Audit Report 15-12, Audit of NARA's Online Public Access Development Effort

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

We concur with the recommendation in this audit, and we will address it further in our action plan.

DAVID S. FERRIERO
Archivist of the United States

Appendix D - Report Distribution List

Archivist of the United States
Deputy Archivist of the United States
Chief Operating Officer
Chief Information Officer