UNITED STATES OF AMERICA

NATIONAL ARCHIVES AND RECORDS ADMINSTRATION

ADVISORY COMMITTEE ON THE ELECTRONIC RECORDS ARCHIVES (ACERA)

APRIL 30, 2013

MINUTES

The Advisory Committee met via web/telephone conference. Some NARA staff, the committee chair, and the Archivist of the United States, dialed-in from the Archivist Board Room in the National Archives, 700 Pennsylvania Avenue, N.W., Washington, DC at 1:00pm, Sharon Dawes, Committee Chairman, presiding.

MEMBERS On-Line

SHARON DAWES, Committee Chairman
DANIEL PITTI, Assistant Director, Institute for Advanced Technology in the Humanities
University of Virginia
LESLIE JOHNSTON, Library of Congress
KELLY WOESTMAN, Pittsburg State University
DENNIS DAY, Director, Air Force Declassification Office
ANDY MALTZ, Director, Academy of Motion Picture Arts and Sciences
STEVE LEVENSON, U.S. Courts
JONATHAN REDGRAVE, Partner Redgrave LLP
BOB KAHN, CEO and President, Corporation for National Research Initiatives

NARA STAFF

David Ferriero – Archivist of the United States
Meg Philips – Electronic Records Lifecycle Coordinator
Mike Wash – Chief Information Officer
Kimberly Scates – ACERA Secretariat
Scott Stovall – Strategic Systems Director

April 30, 2013

Welcome and Comments from the Chair

Chair Sharon called the meeting to order at 1:10 p.m.

Welcome Remarks

Archivist Ferriero welcomed those present for the meeting and thanked everyone for their flexibility in testing the new way of meeting. He started by sharing the impact of the sequestration on the National Archives and expressed how Congress is really paying attention to haw agencies are handling their sequestration plans. We were required to reduce our budget by about 20 million dollars. It was a decrease from 391 million to 371 million and we were ordered to do these reductions along our 4 appropriated lines, 5% for each of those lines: Operating Expenses, Office of the Inspector General, Repairs and Restoration, and National Historical Publications and Records Commission. Operating expenses got the largest hit, 18.8 million dollars. Our strategy in planning was to preserve the agency mission and minimize disruptions in agency services to the public. He would be happy to answer any questions anyone has throughout the meeting.

Impact of Sequestration:

- We will delay or reduce investments in information technology, defer the preservation of some records, and reduce maintenance of NARA-owned facilities
- Sept 2011-reduce employee recruitment, retention, and relocation incentives
- November 2011-hiring freeze
- March 2013-reduced NARA-sponsored conferences and instituted procedures to apply increased scrutiny to new conferences

ERA Program Update

Mike Wash and Scott Stovall provided the ERA program update. Mike said, we are about to launch and move into production, the replacement for ARC our Archival Research Catalog and it's a system that's known as the Description Authority Services (DAS). It's actually hosted in a cloud environment so it's providing another example of our move to a flexible and extensible type of approach for our systems.

We continue to offer new types of records. And we've also just shared in the launch of DPLA; which just about a week ago 1.2 million objects within the Online Public Access are referenced using DPLA. The pilot form of OPA is limited. It was a pilot used for learning, and for us to go from where we are today, with about 18 million items in OPA to over 500 million objects in the next several years.

Scott Stovall spoke about IBM. From an operational status perspective, we feel that IBM is continuing to successfully operate ERA, and all of ERA Instances at ABL. They've proven to be a pretty effective partner not to say that it has been perfect, but they've been pretty effective, and I think we're overall very satisfied with what IBM continues to do for us from an operational standpoint.

Operational Challenges: So, we issued two TDLs to IBM and one we called the Cloud TDL and in this particular task, if you wanted to call it a task, rather than a TDL. We asked IBM to replicates the base code in the cloud for test purposes and they've successfully done that. They've taken the ERA Operational Code base, the application, and placed it in to a cloud environment and we've done acceptance testing.

Mike Wash spoke: We've run into a couple of cases where the user requirements weren't pinned down quite as much as we'd hoped and one case of an oversight where the names of internal reference units were misidentified. One group assumed that the reference unit names should be something, when we got to UAT, the other user said these are the wrong user names. So, we had to make a change. That cost us a couple of weeks, and some money. A couple of other cases with reports, reports that were delivered to the users and user acceptance testing, weren't as descriptive as reports that were envisioned by the users so we had to delay for a couple of weeks the activity, and redefine the requirements. We had to make sure the reports that were going to be delivered would meet the user's needs. What we've been finding as we've made changes is that some new types of failure modes are starting to emerge.

We associate it with some of these processes not being synchronized. Some, in some cases, a process may be still operating when it had been expected to be finished. So the results aren't being carried forward through the work process as it should. What we don't know is whether or not the system was designed in an asynchronous fashion, rather than some sort of synchronized fashion. Which means we had to do an assessment of the design of that workflow control. What type of problems we really need to resolve. The activity that we had of actually moving our development and testing environment into a cloud environment or into a new type of environment, is what started to reveal these things.

A few things going on as Mike said, it's probably easier, even though its OPA, it's probably easier to think of it as making EOP records available through, publicly available, or through the OPA system. We have as Mike mentioned the OPA pilot, that's currently in production. It has scalability issues and we know as the volume of records rose that we have to put into our public access system, we will start to see a further degradation of performance times within our pubic online access pilot. Unfortunately, we have a legal obligation to make presidential records available to the public in January of 2014. So, to not risk litigation we have kind of taken 2 approaches for making these records available. We're going to leverage the OPA pilot so there's a work activity we call Task 39, it would be a TDL to IBM, where we have them make some required fixes for the Presidential Library folks to the system so that we can redact content and make it publicly accessible in the OPA pilot. So there's an activity on one thread to make the content available through January 2014.

Scott Stovall also said, in parallel, there's an activity to basically refresh the OPA system, we call it OPA Prod. So OPA pilot, which is in production now, will have a TDL to IBM to fix the errors that we have in our system and make it slightly more scalable so we can add EOP records into the OPA pilot, so we can meet the January 2014 obligation. In parallel, we're running an activity that will establish a new instance that's more scalable to meet our needs into the future.

Constraints: In essence, we have two constraints, at least two constraints, that we are working through right now. One is just the infrastructure, there's a limited amount of bandwidth we're capable of supporting on access request OPA and we watch that very carefully. When the DPLA announcement was made, we watched to see just how much of our bandwidth we were using up. That was one and moving it into a cloud with content delivery type of network distribution eliminates that. So that's the strategic solution. The other constraint is really by design, when OPA pilot was created, it wasn't intended to have a high number of multiple accesses or downloads. And what we've found, with subject matter experts reviewing the system was, the more you put in, the slower the response.

So it wasn't designed in a parallel fashion to be able to support the scalability. So again from a design prospective, we will create a scalable structure so if more content goes in, we will be able to provide support for that, which also gets into our virtualized infrastructure . So if we go from using 500 million objects in OPA to billions, which we dictate, we'll still be able to provide timely access to these records.

Helpdesk: We mentioned earlier that the Help Desk continues to get good marks from the user community. I can't say that users are overall happy with the system. Their satisfaction still remains a concern. I think it's incrementally better than it has been. I've heard form users that they're happier with the system than they were before. Since IBM has taken over, the contract is significantly better, but there is still some dissatisfaction with the users with ERA as a system. We would like the system to be better than what it is. Success is the storage refresh and data migration is continuing to reduce the footprint at ABL, which is going to be a long-term huge benefit for NARA. And establishing the cloud development and test environment was a big thing, we've already learned an enormous amount from that activity.

Packaging: So on page 10, the As far as our architecture, when we starting looking at the Online Public Access and understanding where we were today with the pilot, one of the learnings that wasn't clear from the users using the OPA system, but from a development perspective and us being able to use the system to serve out records to the researchers and the public. What was clear was, we weren't managing the interface of OPA efficiently in a way that make it easy to get more records into it. One of the strategic concepts that evolved from this study of OPA was a search engine information package approach. The package structure for information that flows into OPA from the National Archives, so things we want to make publically available, information that flows through OPA will go through a standardized interface. Within that interface, it would include the records that would be available and the metadata associated with that record and other things that are critical to support how you would search for it.

So in some cases, we have records that are very database centric, maybe the way you search the database may be different from how you search for documents. The information associated with how to make it more optimally searchable, can also be accompanied in a package along with the information itself. So, adopting standard interfaces approaches to these modular types of components of ERA are going to be additional ways for us to be able to make quick improvements to the way the system will work. The SEIP Search Engine Information Package Is really one of our first reveals of the types of standard interfaces that we are going to start to bring in to the internal work flow. On a very similar front, in some of the work we've been doing with agencies, Mike Carlson, who you guys have all met from previous ACERA meetings, Agencies are starting to ask the National Archives very advanced questions. Particularly in light of records management. So if we're going to be working more into an electronic records world, then NARA what's your scheme? When you have an agency records officer asking what's your schema or records do you want to receive. That's really cool.

The Department of Defense and the Department of Interior are pushing us really hard. So, what it's starting to do is create another opportunity to establish a standard interface for a records information packet or a standard records information packet and start to help us in the early stages of ingest, we can define that schema, then it start to move further upstream into tools.

We are very excited about the SEIP. As we move into OPA Prod, as Scott referred to as Task 39, for the Bush presidential record, that will be an early application of SEIP. We are looking at METS as a aspect of the SEIP. The question mark is there, but I am a strong advocate for METS, as an approach that we will use. Also NEM. NARA is one of the lead agencies working with NEM across the Federal government for standardizing information exchange types of activities so that SEIP is registered, when we settle down with it, with the NEM standards.

Discussion Question 1 Overview(Meg Phillips):

We've talked quite a bit about the presidential directive and change we're expecting over the next 10 years. Agencies required to manage email electronically by 2016 and all of their permanent records electronically by 2019. We're really expecting quite a change in the overall dynamic of electronic records management. There are also other mandates that have changed. Things like the open government and cloud first initiative. So a lot stuff is changing around us since the time that the original ERA requirements were written up.

And we wanted to use this opportunity to get a summary of what you think the most important advice for NARA to keep in mind, as we adapt ERA in the next decade and as we think through how electronic records management in the Federal Government needs to work and work with ERA. This is really our invitation to you to give us whatever advice you've got and what you think is most important for us to keep in mind as we plan for the transition over the next decade or so. So that's the only background and it's intended to be very open-ended.

Many of the tasks organized under the auspices of the presidential directive are really designed to help agencies and they are asking for our help. And also try to influence the entire information

ecosystem. So, that things like packages which include the metadata that we want, can be easily passed among the different kinds of records systems that are used in agencies. One thing that we're already doing is reaching out to the people at the Department of Defense that run the certification program for record management operations under DOD 1515.2. Most of you probably know that standard pretty well. It has had quite a bit of influence over the market place for records management application.

We're opening a dialogue with the people that have more influence than we do and they are happy to work with us. They are asking us like agencies are, "What is your schema?" "What does NARA want to receive?" and "How do you want this to be structured." So I hope we can work backwards through the lifecycle of information, starting with the need to provide public access and figure out, what information we need and make this available through a search engine information package. Ok what does that mean and how we need to receive it from the agencies so that we can archive it and make it available. What we should ask for in the records management application and records creating application that are being used in agencies. We are really hoping to influence commercial sector to get some of the things that we want done, completed. Up to this point, we haven't been explaining that clearly, how we wanted those commercial products changed. As Mike keeps saying, defining our requirements clearly is definitely the first step. We need to be able to answer the question when people ask us "What do you want?" and "What would make your life easier?" If they are willing to make those changes, we will have an answer ready for them. We completely agree that making changes upstream will have the biggest impact on our overall results.

Advice: Advice was given from all of the members present.

Questions that arose:

Daniel Pitti: To what extent are DOD records management standards comparable to ISO standards in the same arena?

As these agencies move toward moving everything into an electronic management being fairly electronically done, how much influence does the National Archives have in assisting all of these agencies in doing this?

Dennis Day: Will ERA be adapted to ingest classified records with permanent retention?

Will ERA be modified to allow work loads for internal use of federal agencies to allow development of proposed schedules?

Kelly Woestman: When you talk about running out of space on the east coast, is there any talk of using the caves in Kansas City more with this system. About the cloud or is that off the table?

Leslie Johnston: How do you explain to people what's actually possible in terms of discovery and what's not possible?

Steve Levenson: Who's going to become the central government buyer? Who we're going to designate when we come up with these requirements to come out of this presidential directive? We say do these

6 to 100 things and we come up with a part it's going to be in. Who's going to designate the

central government buy in?

Question 2 discussion overview (Meg Phillips):

This is really a question coming out of this transition to an online meeting and the value we've gotten from talking to people outside of NARA. We've definitely got to make sure that we continue to reach out to people with expertise that we don't have, or perspective that we

don't have. Our stakeholder communities and related industries that have lessons to teach us. So we want to get your opinions about the best ways for us to engage with the outside world and continue to get connected and get even more connected as time goes on, with you and people like you.

ACERA has been our primary mechanism for getting input on electronic records issues over the last several years and we've gotten better at running ACERA meetings. There's a whole wide world out there and we would really like your input on how we can, on an ongoing basis, solicit input from people like yourselves, going forward, so we have the best possible advice.

Advice: Advice was given on this topic from all of the members present.