Advisory Committee on the Electronic Records Archives
Meeting
National Archives Building
Draft Minutes
November 2, 2011

Welcome Remarks – David S. Ferriero, Archivist of the United States

Mr. Ferriero opened the meeting at 8:35 a.m. Committee members present: Dr. Robert
Martin, Chair, Charley Barth, Laura E. Campbell, David Carmicheal, Sharon Dawes,
Dennis Day, Dr. Richard Fennell, Jerry Handfield, Dr. Christopher Greer, Andy Maltz,
John Phillips, Daniel Pitti, Jonathan Redgrave, David Rencher, Dr. Kelly Woestman.

Mr. Ferriero reviewed developments since the last meeting:
- Roll out of ERA to 16 Council-level agencies
- Roll out of new OPA features including user tagging
- About to launch new citizen archivist dashboard
- Have accepted and are processing 2010 Census records.
- The Lockheed Martin contract is completed, and final closeout activities are under way
- Restructure and formalization of business requirements group and development of detailed list of initiatives
- Improvements to public's ability to access records from ERA and staff's ability to search them
- Submission process has been made more streamlined, scalable, reliable, and flexible
- Improved processes for capturing, storing, and updating metadata
- Initiated activities to work through protests on contract award.

Comments from the Chair – Dr. Robert Martin

It will be a full day with a lot of ground to cover. He thanked the staff for getting
materials out well in advance of the meeting for member review. He asked the members
to introduce themselves.

Adoption of the Minutes (April 6, 2011)

Dr. Woestman moved approval of the minutes of the April 6, 2011 Committee meeting.
Ms. Dawes seconded, and the minutes were approved unanimously.

ERA Development and Transition Update – Mike Wash and Scott Stovall

Mr. Wash reminded the Committee that in late 2010 commitments were made regarding
the conclusion of development activities by the end of fiscal 2011 and the move into
operational status. Expectations were identified for the multiple instances of ERA. The
goal was achieved, and NARA began transitioning to the operation and maintenance

phase of the contract. Mr. Stovall emphasized the restructuring necessary to have NARA employees be in control of more of the programmatic activities and tasks. A program manager will be the primary liaison to the business requirements group and will be responsible for ensuring customer satisfaction. There will also be a project person working from the technical side, and that group will be backed up by systems engineering staff. Mr. Wash clarified that NARA will be in complete control and talked about how the agency assumed more control in the final 6 months of the Lockheed contract. He noted the importance going forward of a strong advisory board.

Committee Comments:

Mr. Pitti asked whether the architecture is sufficiently sound and open or whether additional adjustments may be necessary. Mr. Ferriero noted that it is not open. Mr. Wash said there are changes that need to be made but not substantial ones. Mr. Barth said changes, both architecturally and functionally, need to be made but there is a solid foundation. Mr. Phillips said the architectural framework is open, stable, migratable, and modular, but there was some loss of openness at the implementation level.

2012 Plan - Meg Phillips, David Lake, and Quyen Nguyen

Ms. Phillips outlined the status at the conclusion of development and the plan going forward. Sixty-eight percent of original requirements were satisfied. Some things have been removed as requirements for ERA because they will be met in a different way. Other needs were identified in the years since the requirement were originally written, including social media opportunities. The original requirements will no longer be tracked. The business requirements group includes stakeholder representatives from all over the agency and will work on where investments need to be made.

Committee Comments:

Mr. Barth asked for a breakdown of the 32% of original requirements unsatisfied, and Ms. Dawes asked for examples. Mr. Lake said whole sets of requirements totaling perhaps 10% were allocated to other systems, including description and redaction capabilities. Ms. Phillips added that some requirements are met in certain instances but not in others. Mr. Lake said templates were removed because they never really took hold. Secured storage and archival management were put off for cost reasons but are still needed.

Dr. Greer asked how prioritization happens for additional needs. Ms. Phillips described how stakeholders and users submit ideas for changes and are asked to systematically rate their idea based on benefit to NARA or the community as well as potential penalties for implementing or not implementing the idea. Ms. Phillips said in the short-term they had focused on things that had been committed to for fiscal 2011. Medium-term the focus is mainly on things that are causing a problem right now. The business requirements group collected about 90 candidate tasks and was provided the following top three priorities by the lifecycle guidance team:

- Support roll-out of ERA to all federal agencies so they can effectively accession and transfer electronic records to NARA using ERA
- Make high profile/high public demand records available online via OPA
- NARA can effectively search, review, and redact federal and presidential electronic records in our holdings for FOIA and other purposes
-

Mr. Lake and Mr. Wash emphasized that the agency is still developing the prioritization process. Mr. Wash said the core focus is on effective collection and preservation of electronic records. Mr. Pitti asked about digital preservation. Mr. Wash said there is currently a framework associated with preservation but that it still needs to be incorporated into a standard process as new file formats and preservation needs arise. Ms. Phillips said the original requirement statements for digital preservation are obsolete so the agency doesn't intend to meet them. She said what was shared with the committee in 2009 or '10 is still the current thinking. Mr. Phillips discussed socializing the prioritization process and ensuring it is open.

Footnote.com/Fold3.com - Impact on NARA Digitization Efforts - Rebecca Warlow

Ms. Warlow noted that NARA entered into partnerships with several companies for assistance in digitizing records. One was with a company named Footnote, which recently changed its name to Fold3 and its focus from general records of all sorts, primarily federal, to military records. This change has not really had any impact on the partnership. Military records are a substantial portion of the holdings. Either NARA or its partner can terminate the agreement at any time, and then the partner will no longer host the materials and will immediately provide NARA with the content they are currently working on and have not already provided.

Committee Comments:

Mr. Barth asked for clarification of how the partnering agreements work. Ms. Warlow said the records are digitized for free, and in return the partners get to host the records for five years and may charge for access. NARA and users of NARA's research rooms can access the records for free. Mr. Carmichael asked if there had been discussions with any of the partners about long-term preservation. Mr. Barth said there has been some discussion with Family Search but also noted that NARA already has redundant storage in Virginia and Utah.

Agency Adoption Report/Plan - Mike Carlson

Mr. Carlson updated his report by saying 80% of CIO Council agencies have taken training and started using ERA. Ten percent of the rest of the government has started using ERA. In a survey conducted in September, 44% reported being satisfied with ERA, 44% as neutral, and 12% as not satisfied. However, 84% are satisfied with the support they have been getting from NARA. System performance received only 20 to 30 % satisfaction. By September 2012 the entire government is due to be using ERA.

Committee Comments:

Mr. Handfield asked if the survey results would be any different if it only looked at agencies that had scheduled or transferred records. Mr. Carlson said for the most part those who completed the survey had actually created schedules or transferred records. Dr. Greer noted that many of those responsible are not records specialists. Mr. Ferriero said they are working closely with OPM to create job descriptions dealing with records management. Ms. Dawes suggested adding some questions for individual users about their job title and experience and training in records management.

<u>Demo of ERA and the Agency Experience - Mike Carlson</u>

Mr. Carlson introduced the participants for the demonstration. Julie Reaves, Office of the Chief Records Officer, will do scheduling; Ted Hull from Researcher Services will do the three modules of permanent records. Patti Stockman, NASA Records Officer, will provide user feedback since NASA has been a heavy user and participated in the second phase of the ERA pilot. She commended NARA for what they have done so far but expressed concern regarding the vision of ERA as a persistent archive.

Ms. Reaves began demonstrating the scheduling process. She said one mandate for her office is overseeing the scheduling and appraisal process of federal records, which involves determining which records have historical value and are called permanent and those which don't which are called temporary records. For permanent records there is schedule for how long the agency will hold them before they come to NARA; for temporary records, how long the agency will hold them before they are disposed of. A records schedule is the "form" which agencies propose whether records are temporary or permanent. There is also a process covering the authorities to enter for decisions originally made in paper. The disposition section is where the agency proposes the record as permanent or temporary, and the disposition instructions indicate the time for retention and transfer to NARA. There are structured options to help guide agencies and also to allow NARA to run reports on when records should be coming in. There is a section for additional information such as how large and how fast the records accumulate. There are also sections for contact information and signatory information to show the approval process. The executive summary is completed by NARA staff to provide an overview of what the schedule covers and basic information about the Federal Register Notice since the public is generally given the opportunity to comment.

Mr. Barth passed around a paper copy of SF-115 so the committee could compare the ERA process with the paper process. Ms. Stockman expressed frustration that ERA doesn't allow for sub-items and doesn't work well with the schedules NASA uses internally. Mr. Barth emphasized they want to work on ensuring NARA gets the information it needs without imposing massive additional burden on the agencies submitting schedules. Ms. Phillips talked about balancing what is requested from agencies and how tightly restricted the format. With tight restrictions less data may be provided, but it will allow for more automation.

Committee Comments for Ms. Reaves:

Mr. Rencher asked if there is an authority system behind the search that would group names, departments, etc. for variant spellings. There was discussion of executables and the extent to which the transfer process is automated. Ms. Reaves emphasized that agencies own their own records and decide when to transfer them. Mr. Rencher asked if they can anticipate how much data will come in every year. Ms. Reaves said the fields for that data are not mandatory, and any information that is provided will change over time. Mr. Phillips talked about the intrinsic value of certain records and how more time may be invested in providing metadata about more valuable records.

Ted Hull said there are currently four paths for electronic records to get into ERA:
- Legacy migration process
- The fifth partner process
- Agency-initiated transfer request (TR) with NARA doing the ingest
- Agency-initiated TR with the agency doing the ingest

Traditionally the electronic records unit received records through the direct offer process or the annual move process. Electronic records will typically continue to come to NARA through the direct offer process.

The post-transfer verification task list allows records processors to verify and review records that have been ingested to the system. When a user wants to create a TR, the first step is to find and attach the records schedule item. Ms. Stockman said that is the most challenging part when NASA creates a TR because the system generates its own identifier, but NASA knows its schedules by the NASA schedule item number and by the disposition authority, but neither of those can be used to search for a schedule. She said she had made the suggestion during the pilot, but not all suggestions were able to be implemented in the latest revision.

Once a disposition item is selected, some fields are automatically populated, but there are a lot that need to be filled out by the creator. One of the more complex parts of creating a TR is identifying access restrictions. Once all the records are ingested, the TR is set to no more shipments, meaning no more records can be accessioned against it. After processing and review of a TR is complete, it is sent to a physical custody acceptance state, and a legal transfer instrument (LTI) is generated. The Transfer Shipment Aggregate Report is used to verify that everything was received. The Transfer Manifest provides basic information at the file level.

Once a shipment is available through the post-transfer verification task list, archivists can access that shipment to do certain things. If the transfer included structured data files, the contents can be verified using the data type template verification process. If it includes non-structured data file format records transferred in accordance with the guidance from the records management unit, there is an automatic verification to compare against the essential features outlined in the NARA transfer guidance. For formats not covered by the guidance, information about the format is included in the transfer processing results

report. Mr. Hull said for large volume records they can choose to verify only a sample of the records.

Mr. Hull added that once NARA takes legal custody there is an obligation to provide the records to the public. He described the multi-step process of unzipping individual files to make them available that is used currently since OPA is still under development. Mr. Wash said the goal is ensuring that the records which should be accessible are accessible and that it doesn't involve a multi-step process to do so. Improvements to the interface between the repository and OPA are needed. Ms. Stockman also said that the email notifications of actions should say what the action is.

Committee Comments:

Mr. Phillips asked whether the operations and maintenance contract with IBM included development. Mr. Wash said part of the contract deals with corrective and adaptive activities. Development activities will be assigned through task orders. Mr. Handfield asked if you have to look at the LTI to find out when legal custody was transferred to the Archives. There was discussion on hash codes. Mr. Carmichael said generating one should be tied to NARA's ability to say what they received is what was sent. Ms. Dawes asked about the use and significance of color as it relates to helping users who are not well versed in records or how ERA works. Ms. Stockman said a bigger problem for her users is the chronology and what is really being requested in each field. Mr. Rencher cited a breakdown in the human factors department at Lockheed Martin and said it is dismal.

Mr. Redgrave noted the risk that important records could later be found to be corrupt if every record in large volume transfers is not verified, particularly for file formats not covered by the guidance. He also asked how the sampling would be done. Mr. Carmichael asked about the thought process that identifies the acceptable level of risk that files are corrupt.

Mr. Maltz asked about the ability to meet the goals for ingesting data to the system. Mr. Carlson said for some large volume records hard drives have been transported to Rocket Center for ingestion, and he has encouraged agencies to transport large amounts of electronic records on media. Mr. Wash said it is a huge issue across the federal government in terms of the available bandwidth agencies can use to transport data. Within the CIO Council discussion is starting on how to best utilize cloud services. Mr. Pitti said using the cloud to manage data in place rather than transmitting it raises issues of control, and he suggested getting a much bigger budget and managing a cloud for the federal government. Ms. Phillips added that there are areas where the process in NARA is not as fast or reliable as it needs to be to keep up with what agencies are sending. Mr. Phillips said ERA-compatible interim repositories could be cooperatively hosted in the agencies.

Mr. Pitti proposed redesigning the interface and adding some things to expedite the process. NARA may need to think seriously about what data is essential in order to

lower the burden for agencies. Mr. Carmichael said simply explaining why information is needed may also help. Ms. Phillips said the process, regardless of the system used, is genuinely complicated and difficult to manage. Access restriction information may be difficult to provide, particularly for records that may be transferred 30 years after they were created, but without that information NARA will have to do a page by page review in order to release it.

Dr. Martin asked if the NASA user manual Ms. Stockman developed would be available and whether it would be effective for other agencies. Mr. Barth said he was going to use the modules and spin them for Navy Marine Corps. NARA has also produced some good tools that might be modified by individual agencies. Ms. Dawes emphasized that nothing is more important in the near-term than making the system understandable to and usable by agencies.

Mr. Redgrave wondered if the fields that are not mandatory should be included if almost no one will fill them out. He was primarily concerned about actually getting records from the agencies who have to identify permanent records as well as what happens when a petabyte of data a day is coming in. Mr. Pitti said another issue is data that we would like that was never captured when the record was created. Ms. Campbell raised the idea of agencies with high technical competence running their own archive managed by NARA. Mr. Redgrave said the system will be overwhelmed with data without radical change in the entire structure. Mr. Maltz wondered if the same rules should apply given how much more data is created now than when everything was paper. Mr. Barth said the Navy is doing a review of all its permanent disposition schedules because the current requirements for maintaining permanent records are too burdensome.

Mr. Handfield said maintaining the records without processing them in detail is one way to triage the work. Records dealing with revenue are likely to be of value. He also emphasized that confidentiality is a moving target. There was discussion of the revenue that could be earned if users of digital archives paid a small fee. Mr. Redgrave returned to the idea of the federated cloud and the importance of common platforms across the federal government.

ERA Challenges Discussion - Meg Phillips

Dr. Martin thanked Ms. Phillips for providing the white paper to the Committee well in advance. Ms. Phillips said the paper is a compilation meant to expose the Committee at an early stage so it could identify areas in which ACERA is well suited to advise the agency. She prepared some discussion questions:
- Are you aware of things we should be worrying about but aren't?
- Are any of the challenges identified really false problems?
- Were any of the challenges a surprise?
- Are we worrying in directions which are either inappropriate or unexpected?
- Is ACERA, as a whole or in subcommittees, interested in pursuing any of the challenges?

Mr. Maltz asked whether they will be keeping the architecture and system they have. Ms. Phillips said given that they are holding permanent records, the system will undoubtedly change completely in time. Given limited funds, the agency has limited ability to completely rework the architecture right now. She noted that a lot of the challenges are policy and procedural rather than technical but agreed that the architecture is worth discussing, too. Mr. Rencher raised the issue of use cases built around the legal challenge points of the system. Ms. Dawes asked where human effort is best applied in the process. Most of the human effort is in the accessioning process, but perhaps we should think about how much is reasonable to expect of the end users who are motivated to get the record. Dr. Greer added that the system can be indexed through watching how it's used rather than creating metadata objects for every page. He also suggested looking at where we want the system to be in five years, for example, given the expectation that the system will change.

Dr. Woestman said the Archives is only responsible for providing the raw data, not for manipulating it and wondered about the committee making a statement. She also raised the idea of crowd sourcing. Ms. Campbell agreed there could be cost savings in doing less processing of records.

Mr. Phillips suggested the Committee members could send in comments on the items that could then be coalesced as sort of a general discussion that might point to areas of discussion for the next meeting. Mr. Barth suggested going through the document and categorizing each heading according to one of the five subcommittees. Ms. Dawes suggested some of the research items might be looked at by people at the National Science Foundation (NSF). Mr. Wash said they have relationships with NSF and also with the University of Maryland on research initiatives. Dr. Martin suggested they discuss assignment of items in small groups rather than as the full Committee. The Committee recessed for subcommittee break out sessions.

Subcommittee Reports

Mr. Redgrave reported for the Legal and Policy Construct Subcommittee. He said some items are micro-issues on which the subcommittee might provide a view on what sort of legal issue should be considered. The bigger thing is a vision for 2020 or 2025 that involves looking at all of the information statutes. The subcommittee will try to pull together a framework for a white paper. Mr. Handfield said it will be based on cost savings and efficiencies.

Mr. Maltz represented the Architecture Challenges Subcommittee. He said a large number of the items should be discussed by the subcommittee. A smaller number are not relevant. An even smaller number will need input from the other committees in order for the Architecture Subcommittee to work on them. The items the subcommittee will look at:
- General scalability
- ERA interoperability
- Records management services, after input from Legal and Development

- Avoiding physical transfer
- Crawling for records capture
- Finding people to write data transfer scripts, after input from Development
- Automatically identifying boundaries of a record, after input from Development
- Managing compound records, after input
- Format and software obsolescence
- Format identification
- Efficient and robust metadata extraction
- Automating generation of descriptive metadata
- Normalization of vocabulary in metadata

Mr. Phillips noted the items identified by the Development Strategies Subcommittee:
- Need for scaled-down ERA
- General scalability
- ERA interoperability
- Avoiding physical transfer

Their plan is to define the issues in each of the areas and identify potential solutions.

Dr. Woestman reported on the potential items identified by the User Interface Subcommittee:
- Researcher use of e-records collections
- How can e-records increase citizen engagement with Archives
- Automating generation of descriptive metadata

She said there should be a one page beginner's guide on OPA. She also talked about the potential of citizen archivists.

Mr. Rencher reported for the Communications Subcommittee. There is an opportunity for messaging around release of the 1940 Census and ingest of 2010. He suggested distributing something to the Committee next April on key messages.

Ms. Phillips said she hadn't expected every item to be assigned and that any input would be tremendously valuable. She also liked the idea of everyone posting some comments in the short term. Mr. Phillips emphasized that focusing on what the issues are is more helpful at this stage than trying to identify solutions.

Review of Action Items and Concluding Remarks - David Ferriero, Robert Martin

Dr. Martin said they had just reviewed the action items for the Committee with some specificity. Mr. Redgrave proposed interim subcommittee reports. The Committee settled on the end of February.

Mr. Ferriero said they had two suggested dates for the next meeting, April 24 or May 16. There was discussion with some members preferring one or the other. Mr. Ferriero suggested that an email be sent out with a couple more potential dates.

Dr. Martin said this will be his last meeting as Chair of the Committee. He was grateful for having had the opportunity. Mr. Ferriero presented him with a certificate in recognition of his service. He said that he had asked Sharon Dawes and Charley Barth to serve as co-chairs. Mr. Handfield moved adjournment. It was seconded and the meeting adjourned at 3:52 p.m.