Advisory Committee on the Electronic Records Archives
Meeting
National Archives Building
Draft Minutes
April 24, 2012

Advisory Committee members present: Charley Barth, Co-Chair, Sharon Dawes, Co-Chair, Laura E. Campbell, Dr. Christopher Greer, Jerry Handfield, Dr. Robert E. Kahn, Andy Maltz, Daniel Pitti, Jonathan Redgrave, David Rencher, Dr. Kelly Woestman.

<u>Welcome and Comments from the Co-Chairs - Sharon Dawes and Charley Barth</u>

The meeting convened at 8:53 a.m.   Chair Barth said this was his first meeting as co-chair with Chair Dawes, adding that he was excited about the day's business.   The Co-Chairs want to take the Advisory Committee on the Electronic Records Archives (ACERA) (ERA) in a new direction, changing the groups value proposition for the the Archivist of the United States.   ACERA will try a new agenda structure to develop this proposition.

The minutes from the November 2, 2011 meeting were unanimously adopted without amendment.

The Committee watched a promotional video on ERA in order to elicit "honest feedback" from the group.   The video can be found on the Archives YouTube channel as well as at archives.gov/era. Chair Barth said the Military Sealift Command deck logs were the first records entered into the ERA, and were rightly made permanent records. He asked the group, if they were members of the general public, would the video teach them something about ERA.   Member Kahn said the video could emphasize the electronic aspects of the program more strongly and suggested a graphic depicting the flow of information from agency to agency.   Chair Dawes said the video gave valuable information on the electronic control of paper records but should show more on the control of electronic records.   Member Woestman suggested adding information on access to the records. Member Maltz suggested adding information on what ERA does in terms of records management and why the public should care about the effort.   Chair Barth said the video is intended for federal records managers.   A second video for the general public may be necessary.

Member Kahn said the National Archives and Records Administration (NARA) should contribute to the debate on maintaining federal paper records in perpetuity. Maintaining an electronic copy should be sufficient.   Using cryptographic techniques, users could determine whether they were using the accurate record they intend to. Member Pitti said digitizing records both adds to and subtracts from the original record. Might some intrinsic value be lost in the digitization of certain documents?   Preserving digital copies is in itself a challenge.   Assurance of digital preservation is not demonstrable.

Final comments on the video: Member Handfield said it should make a greater point of the increased efficiency of the ERA program versus previous efforts.   Member

Greer said the video should say more about how federal records managers can adopt the program, as well as what are the next steps in the program.

Chair Dawes said she had had conversations with her co-chair, the Archivist and ACERA's DFO Michael Wash about how ACERA could be of the greatest value to NARA. She said the Archivist's answer to this question was: less presentation, more discussion. The agenda for this meeting was changed to reflect this principal, with brief, informative presentations in the morning and greater opportunity for discussion on pressing topics posed by ERA staff in the afternoon.

IBM Transition Status Update - Michael Wash and Scott Stovall

Mr Wash said the protests (brought by a subcontractor to Lockheed, and being a matter of public record) on the contract had been resolved favorably for the agency, having prior been subject to a 100-day stop-work order. NARA had the foresight to contract a separate O&M contract to preserve infrastructure which allowed for a smoother transition to IBM. No updates or changes were made during the 100 days. The protest questioned whether the government had in fact chosen the best value for the work.

The new contract with IBM is for O&M as well as corrective and adaptive activities, allowing for enhancements to the system as the market requires.

Mr Stovall said the new contract has been implemented in three phases: the first, running from January 27 to March 31, was for knowledge transfer and was conducted on a fixed-price basis. The second phase covers O&M, again on a fixed-price basis. The third phase covers TDLs on a fixed-price basis. Chair Barth said this arrangement posed lower risk to the government, and Member Kahn said, while this may be true, such an arrangement may incur higher costs to the same. Mr Wash said the agency would be diligent to contain costs.

Mssrs Kahn and Barth asked whether the contractor would be able to lose, delete or otherwise expunge records it had been tasked with managing. Mr Lake replied, no, by virtue of NARA's oversight of the contractors. Questions were asked pertaining to auditing capabilities and whether the contractor might be able to divert or copy records outside of the archives. Mr Lake, Member Kahn and Mr Nguyen replied that these concerns can be managed. Member Kahn suggested that NARA presume that bad actors exist who will try to harm the system, "and if something really disastrous happens, how would you go about reconstituting the nation's history here?"

Chair Barth suggested the formation of a cyber security subgroup. ACERA may add the most value with respect to the corrective and adaptive portions of the contract. Mr Stovall said concerns exist inasmuch as the contract is in its first year, especially with respect to development, testing and productivity. Continuity of operations (COOP) is a major concern and ought be addressed this year. Mr Wash said budget would be a key factor in resolving these questions.

Priorities for Fiscal Year 2012

Ms Phillips said "the ERA system that Lockheed delivered was not exactly what we had originally hoped." There is room for improvement in a number of areas, and

work is ongoing to determine what work needs to be done to better meet the needs of federal agencies.   The transition from one contract to the other, including the stop-work period, slowed progress on improvements to the system.   The new contract is being managed to allow for increased control by NARA.   NARA has outlined three priority tasks for the contractor to perform: 1) configuring online public access (OPA) storage space for public copies and 2) improving search and other capabilities in the base system.   The third task was not identified on the record.   Bodies of records received at federal records centers need to be approved as they are accessioned into the system.   Mr Lake said NARA staff wants ERA to be able to perform complicated filtering within the application itself.   The list of business requirements continues to provide guidance on where the program needs to develop in the future.   Member Rencher said ACERA could help weigh in on items in the backlog for resolution.   Member Greer asked for further comment on the role and requirements of the API system.   Mr Nguyen said the downloads for ERA need not be exposed to external systems, though they can be exposed if necessary.

Federal Agency Adoption - Michael Carlson, User Adoption Coordinator

Mr Carlson said ERA is being rolled out in two phases, first to the agencies that make up the CIO Council (the Veterans Administration being the only one of 30 not yet to have implemented it), second, to some 102 agencies.   This second phase is expected to be completed in September 2012.

A user experience survey was sent to participating agencies on March 29, 2012, closing to responses on April 13, 2012.   The final results yielded 56% of respondents reporting they were either satisfied or very satisfied with the program, with a large number being neutral.   Those reporting being unsatisfied increased from the level in the previous survey, along with those reporting to be satisfied.   The imposition of deadlines for ERA implementation seems to increase agency response.   ERA implementation is a greater priority when member agencies' upper management is aware of and involved in the process.

Chair Dawes asked whether any data were available as to which agencies were more or less satisfied with ERA, and what NARA's strategy is for accruing management support of ERA is.   Mr Carlson replied that the survey was intended to be anonymous, though larger agencies and those that have committed to the program seem to like it more. Working with records officers at agencies seems to be the best way to get agencies involved.   The survey was responded to at all levels of employment.   Most respondents have five or more years of experience in the field, and 80% spend almost all their time on records-management activities.   One must be careful to distinguish between records subject to the Federal Records Act (FRA) and those subject to the Presidential Records Act (PRA).   Adoptors of ERA will continue to be surveyed in future.

Bush Library Adoption - Alan Lowe, Director of the George W. Bush Library, and Sam McClure, ERA Transition Officer for Presidential Libraries

Mr Lowe thanked the Committee for the opportunity to speak to them.   The George W. Bush Presidential Library is currently located in a temporary facility in

Lewisville, Texas.   The permanent library is being built on the campus of Southern Methodist University in Dallas and will comprise 225,000 square feet.   Beginning later this year, staff (currently 23, including 19 archivists) and materials will be moved to the permanent facility, which hopefully will be dedicated in April 2013. Materials are processed according to the terms of the Presidential Records Act and are subject to Freedom of Information Act (FOIA) rules beginning five years after an administration ends.   Staff created a processing plan to ingest the materials.   Initial processing has been going on for over a year and 600,000 records have been processed.   There are 200 million emails to be processed.   The electronic records pose several challenges.   The collection comprises about 80 terabytes of information.   For a period of five years, the library is not open to research, though it must respond to special access requests issued by Congress, the President or courts.   The archiving and management of the electronic records will be a challenge.   The staff decided to begin by processing materials generated around certain important dates, e.g. September 11, 2001.   Electronic processing will mirror the processing of textual records.   Due to a lack of redaction tools, not one electronic record has been fully processed so far.   Development of these tools is imperative.   Consistency of archivists' review is another challenge.   The program requires close coordination with OPA staff.   Once FOIA requests begin to come in, they will dominate staff time and energy.   Classified materials may be declassified following review by the pertinent agency.   Mr McClure said classified materials are stored on a separate, air-gapped server.   NARA is developing procedures for completely electronic redaction protocols.

Member Pitti said human archivists would likely not be made redundant by software; rather, software will help archivists do their work more quickly and efficiently. Preservation of unaltered electronic documents is a concern.   Archivists are worried about the unintentional release of national security classified and personal privacy information.   Member Kahn asked how attachments to emails which may be written in proprietary software may be made manifest in the distant future.   Mr McClure said short-term solutions have been developed; long-term preservation and access may require transformation of records to new formats.   Member Kahn said NARA should think about the together-ness of, for example, emails and their attachments, whether they are contained in the same bit stream.   Mr McClure said ACERA could help by providing guidance in this and other areas.

Mr McClure said records processing began in 2008 and continued past the end of the Lockheed contract.   The priority in FY11 was to develop tools to allow the persistence of archivists' review decisions.   Redaction and OPA are continuing issues. FOIA tracking uses a different tool.   FOIA requests initiate a negotiation between the requester and NARA to provide a manageable yet comprehensive return of records. Online redaction tools should be implemented in the next several months.   When they and some other program tools come online, the library can begin processing electronic records.   Next steps include loading all the earlier presidential records into the EOP ERA system.

Online Public Access - Pamela Wright, Chief Digital Access Strategist

Ms Wright gave an update of OPA to date, its next steps and also spoke about the

Citizen-Archivist Dashboard.   NARA's open government plan was created in response to a 2009 presidential directive "that encourages agencies to be more transparent, participatory, and collaborative."   NARA is developing its second Open Government Plan, of which the Citizen-Archivist Dashboard is a part.   A vision for online public access is developing.   The public will be able to interact with the Archives' records in several new ways.   The public can also help the Archives' efforts by transcribing and engaging in other activities.   Chair Barth asked how comments from the public would be vetted for accuracy.   Ms Wright replied that, as with Wikipedia, the crowd corrects errors.   NARA is becoming more sophisticated in its understanding of how to present documents to the public.   There are 11 million records in OPA and 400,000 digitized copies.   The program continues to solicit feedback from the public.   In fact, most feedback from the public as well as that from the staff has been positive.   Chair Barth asked what efforts are made to inform interested parties about the rollout of new collections.   Ms Wright replied that staff works with the calendar to release collections associated with particular dates, for example, Labor Day yields documents on unions. Social media are used to inform stakeholders; distribution lists have not yet been used. Short-term goals include fixing the OPA ingest project; adding partner images to OPA storage, 215,000 this year and 12 million for each of the next two years; fixing OPA statistics; logging and entering George W. Bush emails.   Long term goals include implementing the automatic push from EOP to OPA; citizen description.

The transcription tool on the Citizen-Archivist Dashboard has been surprisingly popular.   800 pages were transcribed in the program's first two weeks.   "It doesn't take very many people to get a whole lot of work done."

Library of Congress' Twitter Experience - Laura Campbell and Leslie Johnston

General Counsel of Twitter contacted Member Campbell in spring 2010 to see if the Library of Congress (LOC) would be interested in obtaining a copy of the Twitter archive.   Some 200 researchers have contacted LOC about using the archive.   It will be a "gold mine" for future researchers.   LOC's goal in taking the archive was twofold: "to preserve a rich corpus of ephemera that is unique to our times for researchers and, secondly, to serve as a case study for preserving and providing access to large bodies of digital data, especially social media data."   LOC is focused on the historical archive dating from 2006 through 2010.   Tweets less than six months old are not to be released by LOC.   LOC may not provide access to the entire archive.   Researchers will not be allowed to download significant portions of the archive.   Under their terms of service, Twitter is giving us the public tweets..   The archive contains about 20 billion tweets dating from 2006 to 2010.

Ms Johnston said the Twitter archive has been the most discussed acquisition by LOC in quite some time, probably in its history.   It comprises about 21 terabytes of uncompressed data.   The tweets are received "bagged," every couple hours, essentially in real time.   LOC did not purchase any hardware on which to process or store the archive. LOC's role in maintaining the archive is to figure out how to make it available to researchers; it must be downloadable, at least for the time being, though a "data as a service" model is being explored.   LOC formed partnerships with NIST, NSF and other institutions for research in processing.   LOC's current plans allow only for on-site access

to the archive.   A very small portion of the data associated with a tweet is the tweet itself, the rest is metadata.   Member Kahn asked whether LOC might get involved with other social media formats, such as FourSquare.   Member Campbell replied, no, this is a one-time case study for now.

Discussion Topic #1:   How Could NARA Reduce Reliance on Human Processing:

Ms Phillips said the question is really about scaling up NARA's electronic records programs.   The traditional duties of archivists require a "high-touch" environment which makes processing and accession times too long.   As an example, "We have humans verifying that the electronic records the agencies send are actually the records that they were supposed to send us," how to automate this process is unknown.   How can these and other processes be made efficient, scalable and ready for the realities of the 21st century?

Member Woestman asked why so much effort is made to examine all electronic data, instead of limiting examination to the "folder level."   Concerns about the release of protected information are significant.   Chair Barth said the reality is "that we cannot keep up" with the flow of data coming in, and that there is not a human answer to the problem.   The Citizen-Archivist program may be the only solution.   Member Pitti said, perhaps data-mining software used in intelligence work could be reversed and used to ensure that protected information is not released.   NARA is looking at this possibility, as well as others in e-discovery and predictive coding models.   Commercial markets are beginning to pick up on solutions.   Legislative instruments such as immunity from liability for unintentional release will need to be developed.   Will the legal system accept the tools that are developed?   Perhaps the definition of what is an archivable, permanent record should be changed.   Member Kahn suggested a system wherein government employees would attest that an email they composed does or does not contain protected information.   Perhaps efficiencies can be found in records transfers.   Member Redgrave suggested the creation of a single federal record-creation system.   Feasible actions taken in all of the above-cited areas would have an exponential, as opposed to additive, effect on productivity management.   Member Kahn suggested that records could be tiered according to how long they must be maintained.   It is a fallacy to think of the federal government as a monolith, with matching needs and procedures across offices and agencies.   Member Greer suggested a study be performed to examine the effects of multiple interventions in the document-management stream.

Discussion Topic #2: How Could Records be Archived at Rest to Avoid Requirement to Transfer Big Data from Agencies to NARA?

Mr Wash said this discussion was closely related to that in the prior session. Many people don't realize the challenges incumbent in moving large quantities of data. Sometimes moving data still requires the use of trucks to get the storage units from one site to another.   One possible response to the question of how to store all this information is that it be kept at its place of origin.   Member Kahn said such a plan would have significant advantages in terms of survivability and continuity of operations (COOP). Archiving at rest (AAR) may lead to challenges in interoperability across sites.   The

system should be designed according to a technology-neutral framework. Mobile programming would increase the usabilty of an AAR system. Perhaps records can be archived at a third-party site, such as Amazon or in a cloud system. Data and the metadata pertaining to it could be stored in different locations. Member Greer asked, "can you construct a logical framework that allows you to effectively move control in an easy way, regardless of where the data are physically?" Giving up geographical mapping in favor of logical mapping can "set you free" in "interesting ways." AAR must be auditable. Member Rencher suggested a model wherein NARA would act more as a platform to which agencies would upload their records using their own, individually designed apps.

Discussion Topic #3: Possibilities of Challenges and Contests - Charley Barth

Challenge.gov, a federal website intended for a public audience was set up to encourage individuals to help solve some of government's problems. Department of the Navy set up a contest pertaining to helping fight pirates on the Somali coast in which user-generated ideas were placed head to head and superior ideas were advanced to the next level of competition. The response was surprising both in terms of the numbers of responses, as well as the quality of ideas. He encouraged the ACERA members to come up with challenge ideas of their own pertaining to ERA. Member Maltz suggested a challenge to develop the Citizen-Archivist program into a crowd-sourced program. Member Campbell suggested a challenge to create text-analysis archival tools, also a challenge to create a trusted network of citizen-catalogers. Member Pitti suggested challenges to develop email browsing and querying tools. Ms Phillips suggested framing the challenge within an area that is interesting or sexy to the public, for example, handling accession pertaining to incoming space shuttle records. Chair Barth suggested a challenge to let the public determine the next ACERA agenda. Member Handfield suggested a challenge to help make NARA records more accessible.

Meeting Wrap-Up

The Committee generally agreed that the structure of the meeting was an improvement over previous models. Member Woestman suggested that meetings take place over an afternoon and the subsequent morning, to take advantage of dinner conversations and overnight thinking that may occur.

The meeting was adjourned at 4:41 p.m.