

*DRAFT ACERA Meeting Minutes, 04/04/07*

**ADVISORY COMMITTEE ON THE ELECTRONIC RECORDS ARCHIVES  
MEETING NO. 4  
NATIONAL ARCHIVES BUILDING**

**MINUTES  
DAY 1 OF 2  
APRIL 4, 2007**

In accordance with the provisions of Public Law 92-463, the meeting was open to the public from 9:00 a.m. to 5:00 p.m.

**COMMITTEE MEMBERS**

<u>Name</u>	<u>Organization</u>
Lewis Bellardo	National Archives and Records Administration
Laura E. Campbell	Library of Congress
David Carmicheal – <b>Not Present</b>	Georgia Archives
Sharon Dawes – <b>Not Present</b>	Center for Technology in Government
Luciana Duranti – <b>Not Present</b>	University of British Columbia
Dr. Richard Fennell	Administrative Office of the U.S. Courts
Daniel Greenstein – <b>Not Present</b>	University of California
Dr. Christopher Greer – Substitute for D. Atkins	National Science Foundation
Jerry Handfield	Washington State Archives
Robert Horton	Minnesota Historical Society
Dr. Robert E. Kahn	Corp. for National Research Initiatives
Andy Maltz	Academy of Motion Picture Arts and Sciences
Richard Pearce-Moses	Digital Government Information
John T. Phillips	Information Technology Decisions
Dr. Dan Reed	University of North Carolina at Chapel Hill
Adrienne Reagins	National Archives and Records Administration
Jonathan M. Redgrave	Redgrave Daley Ragan & Wagner LLP
David Rencher	Federation of Genealogical Societies
James J. Hundley – Substitute for R. Testa	U.S. Air Force
Dr. Ken Thibodeau	National Archives and Records Administration
Allen Weinstein	National Archives and Records Administration
Dr. Kelly Woestman	Pittsburgh State University

## *DRAFT ACERA Meeting Minutes, 04/04/07*

### **1. ACERA access to ERA system**

The discussion began with the committee members discussing the possibility of having hands-on access to the ERA system.

Dr. Thibodeau stated that the system is not ready for hand-on user access. Only LMC Labs now exist to exercise system capabilities and there are already severe resource constraints associated with testing and debugging the system software in preparation for delivery. The installation of system infrastructure is not expected to be completed until May.

Dr. Bellardo inquired if the committee members could have access to the system after IOC? And the answer from Dr. Thibodeau was, "Yes, after IOC." Dr. Kahn expressed a concern that after the IOC the committee members might have to tell people that they had no impact on the system.

Dr. Thibodeau – We could allow access to the Human Factors Lab, but as it has no external network connectivity, the users must be physically in the lab.

Dr. Kahn – What content is there in the Human Factors Lab?

Dr Thibodeau – Just test data.

Dr. Kahn – Would that give one a feel of how the system works?

Dr. Weinstein proposed to table system access discussion until lunch.

### **2. Adoption of minutes and review of action items**

1. Dr. Kahn reiterated the previous request for the Human Factors documentation and the Enterprise Architecture for ERA.

There was discussion about what falls under the purview of the non-disclosure agreement. The issue here may be that LMC is protecting their technique and not the actual data. Dr. Thibodeau and Dr. Weinstein said that such documentation should be available for review.

2. Dr. Kahn inquired about the performance requirements, whether the system can handle the projected load and how it would be tested.

Dr. Thibodeau replied that requirements with respect to system performance may be examined.

3. Who were the bidders on the ERA proposal?

Dr. Thibodeau responded that there were other bidders besides Harris and LMC, but whether NARA would disclose the bidders is up to the Contracting Officer. There were two (2) levels of competition. Harris and LMC were announced at the second level. Three (3) other companies did not make it to the second level fly-off.

The question was raised as to how the committee would make use of that info?

## *DRAFT ACERA Meeting Minutes, 04/04/07*

Dr. Kahn expressed an interest in seeing what other solutions were proposed and rejected, and why those solutions were rejected. Dr. Kahn proposed that earlier in time the committee could have learned about alternative approaches, but knowing the names at this point would serve no purpose and he would be happy to close this issue. Also, the content of the proposals would be interesting, the procurement officials would probably object. Dr. Thibodeau agreed that the content of the proposals is acquisition-sensitive material.

#### 4. Resignations:

Mr. Richard Testa, US Air Force, resigned from the committee due to his retirement from the US Air Force. Daniel Greenstein resigned from the committee after taking an academic post at the University of Colorado. Lew Bellardo asked committee members to forward names of people they think might be good for the committee.

5. Dr. Kahn circulated an unsigned copy of the letter from the committee addressed to Congress. The content of the letter – Concern that if the ERA funding issues were not addressed, the system would not perform and Presidential records would be jeopardized. Dr. Weinstein remarked that the letter has been sent via a courier.

### **3. General Comments - Funding**

Every agency in the government could complain about the lack of funding. What is the effect of having sent this letter?

Dr. Bellardo - Every time a letter comes out from a group as prestigious as this, it would be considered. We are fortunate not to have to fight for the life of the program in this budget cycle.

Dr. Weinstein reported having made 50-to-60 visits to the members of Congress, mentioning this advisory committee in each of the meetings. The name made an impression.

Dr. Kahn inquired about the state of the appropriations.

Dr. Thibodeau - We were appropriated \$12.5 M above what was originally allocated.

The summary of the budget status discussion is that given the state of the budget, the appropriation is “not bad.” However, since the appropriated funds can be redirected, the letter to Congress is something of an insurance policy to make sure the increase is not diverted to other priorities. Having another group advocate for ERA is important. Dr. Weinstein reported that the senators working with ERA are very supportive. The system is at a “tipping point.” If ERA is successful, the federal agencies will get behind it.

Andy Maltz asked if the program management has more confidence in funding and is the situation significantly better?

Dr. Thibodeau – Yes. But we are also looking at alternatives. Even with the increased funding there is no room for “give.” For example, the National Security system is a stand-alone with ingest and search but no preservation capabilities. For the Presidential Libraries the most immediate need is description and access.

## *DRAFT ACERA Meeting Minutes, 04/04/07*

Andy Maltz asked what the future cost of these shortcuts would be and Dr. Thibodeau said he did not know.

Dr. Weinstein – FOIA requests are anticipated.

Dr. Kahn – Can anything be FOIAed?

Dr. Woestman – After a visit to the Clinton Library I was impressed that FOIA requests dominated the access. Probably not the best use of government funding.

Dr. Kahn – Unless everything comes in as a law suite... The requests need to be targeted. One cannot go on a “fishing expedition,” it could overwhelm the government.

Andy Maltz raises a concern about the costs of re-work and later-date compensation for the scaled down design.

#### **4. ERA Program Status – Ken Thibodeau**

In summary Dr. Thibodeau stated that the program is succeeding but the customer’s expectations are on the rise as well.

Dr. Kahn – Are you facing the prospect of running out of money in 2008?

Dr. Thibodeau – The time deadline may need to be moved but dollar limit is what it is.

Dr. Bellardo reports that the program’s relationship with GAO is “tough but supportive.” GAO has been helpful to in advising us with how to format ERA information for presentation in order to convey it more clearly. There was discussion about the permission ERA received to spend money on the

Dr. Weinstein – The good relationship with the Hill is enormously important to us. New staffers are assisting us energetically.

Dr. Kahn – What is the total planned expenditure for ERA in 2007?

Dr. Thibodeau - \$8M +\$20M + \$5M.

Dr. Kahn – What was the actual expenditure in fiscal ’06?

Dr. Thibodeau – about \$30M. Increase covers the research program and the management of the acquisition. We are now a full voting member of the White House’s Networking and Information Technology Research and Development program, along with much bigger agencies, such as DOD and NASA. We were asked if we needed additional money for research in 2008. We asked for \$1M and it was included in the President’s budget.

Jerry Handfield inquired if the money for Lockheed Martin development and the Research Program are separate.

## *DRAFT ACERA Meeting Minutes, 04/04/07*

Dr. Thibodeau – Yes. \$45M of the \$58M is going to the Lockheed Martin. The rest funds IV&V, POST, PMO, telecommunications, and payouts to the government agencies.

Dr. Kahn – Are you confident about being able to achieve your objectives?

Dr. Thibodeau – In Increments 1 and 2 – yes. Looking forward to the FOC, the agency needs a course correction. We were naïve to assume we could get full funding at the level necessary to meet NARA's needs and priorities. To date, that level has never been approved. Therefore, every year we have had to redefine a 'useful segment' according to the level of the budget passback. FY 2008 [funding] was the first time the request matched our projected need. Cumulatively, as of 2008 we will be \$108M short of what we would have needed to meet NARA's allocation of requirements to Increments 1 and 2. Obviously, a lot of adjustment will be required.

Dr. Weinstein – We have gotten adept at improvisations.

Dr. Thibodeau – The FY 2008 request includes funds for rapid ingest and rich search of the Presidential Electronic Records that will be transferred at the end of the Administration in 2009, and also improvement of the ERA infrastructure. To garner support for the 2008 request, we pointed out to Congress that, historically, Congress is the first organization wanting access to the records of an outgoing administration.

ERA is on the OMB list of major risks because of the unprecedented nature of the system. However, OMB has a positive view of ERA as a significant element in the future of e-government.

Dr. Kahn – Who decides what to keep and what to throw away, like spam?

Dr. Weinstein – The President has more authority over what gets deleted, the Congress has less, the Archivist even less.

Dr. Thibodeau – We do not anticipate finding spam, but will probably see a lot of 'trivial stuff' – we assume a very high percentage of White House records are worth preserving.

Jonathan Redgrave – You do not want to filter out the White House records. Just have to take what is given to you.

Dr. Thibodeau summarized programmatic risks.

### **Further discussion:**

Dr. Kahn inquired about the source and volume of the metadata.

Dr. Thibodeau – Some metadata comes from the legacy systems which will be replaced by ERA. Because of volume, NARA is considering hiring special contractors to type some of it in. Also, the end users, appraisers, will generate it in the course of transactions on the system, and

## *DRAFT ACERA Meeting Minutes, 04/04/07*

technical metadata about transferred records will come from the agencies. There will be directions for the agencies [to follow] on the website.

The records schedules exist on paper and are not easily reducible to structured metadata. There are tens of thousands of record schedules.

Dr. Kahn – What is an example of something that is not an easily reducible to structured metadata?

Dr. Thibodeau – A prime example is an ‘Item’ in a records schedule. An ‘item’ is some aggregate of records. There has never been a standard for defining what constitutes a schedule item. An item could be a single series located in a single, physical filing station, or it could be a single data base. But it could just as well be all of the records generated in a major program. Another example is the data model, or record layout for databases. We have automated inspection of databases to determine if the data transferred to us actually match the data model. However, with the workload pressure staff did not always input the entire layout structure, keying in the minimum to ensure we know what we received. But for full access to the database you would need full record layout in the system.

Dr. Kahn asked for an explanation of “enable”?

Dr. Thibodeau said this refers to the workflow and that the ability is there but that actual use is ad hoc.

Jerry Handfield and Dr. Kahn suggested that there may be an issue with ingest speed and inquired once again about the performance requirements, and specifically e-mail processing speed, and e-mail per minute target? For presidential records, how quickly will ERA process the data?

Dr. Thibodeau responded that 33 M of Clinton e-mails took nine months to ingest. Clinton White House stored each e-mail as a separate file. Through the first part of the Bush administration, the White House did the same thing.

Discussion of the anticipated ERA system loads, system performance and NARA’s responsibility with respect to FOIA and other requests for records. Dr. Thibodeau said that this is all part of the requirements specification. Dr. Kahn asked what the metric is. Dr. Thibodeau said that it is some multiple of terabytes per year and that we are still exploring for Increment 2.

Dr Thibodeau reminded the committee that estimation and characterization of the load is a well known problem.

Dr. Weinstein stated that there has never been a single accusation that NARA lacks objectivity in executing its functions.

Dr Kahn expressed a concern about how what NARA is doing dovetails with what the private sector is doing, the undesirability for the government to be in the standards business, the need for the industry buy-in, and the need for the committee to know what to focus on. There was also

## *DRAFT ACERA Meeting Minutes, 04/04/07*

discussion of how much authority NARA has to dictate how other agencies keep their data. Dr. Kahn said the committee should make recommendations about what restrictions NARA can enact. Weinstein asked if some of the committee members could sit down with White House staff about this issue. Dr. Kahn said that he would like to do this but that he wants to wait until ACERA knows what they want and can go into the meeting with some ideas.

Dr Thibodeau continued his presentation with discussion of schedule – Release 3 will be later than April of 2008 because people were pulled from Release 3 into testing and debugging Release 2.

**BREAK**

**RETURNING FROM BREAK – 11:42**

Dr Thibodeau continued his program status presentation. He explained how things were moved around in the site plans for at ABL and how that lowered cost. There was also a discussion of data storage with the Navy Oceanographic Office at Stennis Space Center.

Backup Discussion followed, including discussion of load sharing, back-ups, and use of other systems.

Dr Thibodeau/Dr. Weinstein discussed the need to re-issue the original plan described in the Target Release (TAR) Plan with the three (3) sites, because funding through 2008 is not sufficient for even two sites.

Dr. Kahn initiated discussion of programmatic and operational risks – What would happen if all the sites went down? What are the risks and the costs if the money is not appropriated for the project?

Dr Thibodeau – Risks:

- 1) Ingest of the Bush Materials – FY 07 budget appropriated \$3 M for analysis. Will be built in FY 08, come on line in Sept. of 2008 and tested with Clinton materials.
- 2) User Adoption.

**AI** – State Archivists will get involved with the User Adoption Group. Decide if there is anything with respect to User Adoption the rest of the committee should get involved with.

**AI** - Dr. Thibodeau to provide the Organizational Impact Assessment to the State Archives and create an abstract for the general ACERA Committee.

The question is posed if budget reduction perceived as a risk?

Dr. Kahn reiterated that he believes system performance is a risk:

- Stress on the system by users;

## *DRAFT ACERA Meeting Minutes, 04/04/07*

- Ability to switch over;
- File sizes;
- Ingest speed;
- Retrieval speed; and
- Functionality – quality of search.

Dr. Weinstein – The risks stretch over time. The system can be perfected over time.  
Dr. Kahn again requested information on the initial performance requirement numbers. Dr. Thibodeau spoke about systems constraints that are mostly due to NARA staff and not the system itself.

Break for lunch...

12:55 Meeting Adjourns

### **5. Interagency Working Group on Digital Data – Chris Greer**

There was discussion of the presentation including questions on how to divide out functions, the economies of scale, how stovepipe systems are inefficient. There was discussion about the terms “data” versus “records.” Records are a subset of data. Data is very broad and can include algorithms, software, etc. Mr. Greer also made the distinction between digital and physical. One of the committee members asked about analog records. Once a record is digitized it comes under the purview of this plan but not before. Categorized solutions may not be good because of economies of scale. Interoperability standards may not solve everything but they will make it easier.

There was also a discussion of the lifecycle concept used as a base for these assumptions to make the common framework.

Dr. Kahn – What will happen after the Working Group finishes its work?

Chris Greer – Do not know. But very focused efforts do lead to series of subgroups and interactions across the government. The vision can not be achieved in 18 months.

Dr. Kahn – What will government achieve by including OSTP.

Inclusion of OSTP in the process should lead to agreement among the agencies and OSTP. OSTP is engaged with the purpose of reaching consensus. The Committee on Science is the group to watch because they are in charge. There was a discussion of IWG membership. Dr. Kahn asked if there are minutes or reports available. Mr. Greer said they were produced but not published so the committee would need to put in a FOIA request to gain access.

PRESENTATION by Dr. Reagan Moore on the SDSC work with digital preservation and the Persistent Archive Prototype



## *DRAFT ACERA Meeting Minutes, 04/04/07*

### DISCUSSION OF THE PRESENTATION

How do you determine how many copies are enough?

Dr. Moore – Standards say seven (7) but in this design there are two (2) copies per data grid, which totals to ten (10) copies.

Does the metadata keep track of everyone and all data updates?

Dr. Moore – Yes, it is capable and necessary for things like HIPPA.

There was discussion of access controls, data redundancy, and bit error rates. There was also discussion of system administrator skills and how system administrators are expected to be archivists.

How does this work relate to ERA?

Dr. Moore – It demonstrates feasibility of ERA requirements.

How do you prevent inappropriate use of power?

Dr. Moore – Multiple environments, different administrators, and a highly distributed environment.

How do you protect files from deletion?

Dr. Moore – Any command can be ignored if necessary.

How is hardware migration envisioned?

Dr. Moore – It is managed by an administrator.

Chris Greer – Where did 174 services come from?

Dr. Moore – It took two days of going through descriptions.

Will LMC use your open source software?

Dr. Moore – No

Dr. Thibodeau – When LMC chooses to use open source software they commit to supporting it. LMC's design allows using open source for preservation. We've directed LMC to review the architecture for evolvability and scalability. Federation is especially at stake.

Are rules preserved as persistent objects?

Dr. Moore – Yes. Middleware – grid environment is focused in the middle layer.

### **6. General Discussion**

Topics for tomorrow:

- Standards and their management going forward
- Discussion of volume and performance requirements sheet provided by Dr. Thibodeau

### **8. ADJOURNMENT**

*DRAFT ACERA Meeting Minutes, 04/04/07*

The meeting adjourned at 5:00 p.m.

I hereby certify that, to the best of my knowledge, the foregoing minutes are accurate and complete.

Adrienne M. Reagins  
Secretariat  
Advisory Committee on the Electronic Records Archives

Robert Kahn, Ph.D.  
Chairman  
Advisory Committee on the Electronic Records Archives

These minutes will be formally considered by the Committee at its next meeting, and any corrections or notations will be incorporated in the minutes of that meeting.