

*Final ACERA Meeting Minutes, 11/29/07*

**ADVISORY COMMITTEE ON THE ELECTRONIC RECORDS ARCHIVES  
MEETING NO. 5  
NATIONAL ARCHIVES BUILDING**

**MINUTES  
DAY 2 OF 2  
NOVEMBER 29, 2007**

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

**COMMITTEE MEMBERS**

<u>Name</u>	<u>Organization</u>
Lewis Bellardo	National Archives and Records Administration
Laura E. Campbell	Library of Congress
David Carmichael	Georgia Archives
Sharon Dawes	Center for Technology in Government
Dr. Richard Fennell	Administrative Office of the U.S. Courts
Daniel Greenstein – <b>Not Present</b>	University of California
Dr. Christopher Greer	National Science Foundation
Jerry Handfield	Washington State Archives
Robert Horton – <b>Not present</b>	Minnesota Historical Society
Dr. Robert E. Kahn	Corp. for National Research Initiatives
Andy Maltz	Academy of Motion Picture Arts and Sciences
Richard Pearce-Moses – <b>Not Present</b>	Digital Government Information
John T. Phillips	Information Technology Decisions
Dr. Dan Reed - <b>Not Present</b>	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 Neighbors	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

## *Final ACERA Meeting Minutes, 11/29/07*

The second day of the bi-annual ACERA Advisory Committee was brought to order by Dr. Robert Kahn, the Advisory Committee chairman, at 9:12 a.m.

Mr. Andy Maltz began the day with an overview of the economic picture of the Motion Picture Industry. He mentioned that the economic power in that industry still centers around six studios. Globalization represents the biggest growth area. With the market currently distributed with 40 % in the US and 60 % international.

Hollywood still maintains the industry standard. If Hollywood should lose its' power, then the industry would lose its' standard.

Dr. Kahn – Cited the phenomenon that the music industry encountered with the Internet. Is motion picture experience likely to mirror that of music?

Mr. Maltz – No. Cinema is a different experience than music. People want the cinema experience that can only be felt communally. There is a theatre experience and a theatre product.

Dr. Kahn asked if there were any new insights since yesterday's meeting.

Dr. Thibodeau – No new insights.

Dr. Kahn – NARA cannot deal with metadata all by itself. Who is to provide the leadership regarding a metadata standard?

Dr. Thibodeau – The designated user community for ERA as an archival system is NARA staff, other agencies and the public. All three segments of this community will interact directly with the system as end users. Most of the metadata will be transparent to end users. However, NARA recognizes it needs to interact with other communities and to address the different interests of subsets of its users. In this, metadata standards are very important. NARA has participated in the development of a number of relevant standards, such as the ISO standards for Open Archival Information Systems and Records Management, and the Federal Geographic Data Committee's standards for spatial data. We participate in collaborative groups, such as CENDI and the Digital Library Federation. NARA is also open to alliances with other federations.

Dr. Bellardo – The next meeting is to address federations. We will discuss canned searches for a broad research community.

Dr. Bellardo – What role can we play with the big agencies regarding what we can do for them? Is federation the right way to go?

Dr. Kahn – Can we structure the next meeting around the ideas for federation?

## *Final ACERA Meeting Minutes, 11/29/07*

Ms. Laura Campbell – There were three scenarios of Library Organization proposed at Berkley. They are:

- 1) Library Triage
- 2) Congress of Libraries
- 3) Universal Libraries

LOC's NDIIPP program is focusing on 2) Congress of Libraries.

Mr. Phillips – What is the Product of the Committee?

Dr. Kahn - We did not have a hypothesis going in. Should we have a statement on the agenda for the next meeting?

Mr. Phillips – We have had a lot of discussion regarding implementation, recommendation, architecture, etc. One recommendation is to produce a white paper featuring “Issues for the future.”

Dr. Kahn – There is a basic difficulty of discussing technical issues in any detail.

Mr. Phillips – We could include moderating factors (e.g., what will make this work). So the product would be an issues paper with a Moderating Factors section.

Mr. Maltz – What are the scenarios to steer the ship over the next few years (e.g., limited funds, functionality)?

Dr. Ballardaro – Several years ago we had a “role flexibility concept” and a model for federal records centers. We have certain responsibilities for certain records, in real time for the agencies. Preserve our media, preserve our records into ERA. There should not be walls.

1. Dr. Kahn suggested we discuss the role of the Advisory Committee and get suggestions from the members on how best to fill that role. As an Advisory Committee, our product is advice to NARA. How best to achieve that?

### **Suggestions from the Committee?**

Dr. Woestman recommended the use of white papers.

Mr. Maltz – Feels committee should make recommendations.

Mr. Carmichael felt that the committee should make recommendations.

Dr. Dawes felt that we should present NARA with questions – not necessarily recommendations.

## *Final ACERA Meeting Minutes, 11/29/07*

Mr. Carmichael liked the idea of posing questions to NARA.

Mr. Neighbors also liked the idea of posing questions.

Dr. Kahn – we should say something about both the short and long term with the emphasis on the short term at the moment. We should not give recommendations unless they are very clear and the committee is specifically in agreement. Need to know who we are writing for before we put pen to paper.

Dr. Kahn – How is the Committee performing with respect to its charter?

Dr. Thibodeau – Read ACERA charter aloud to the Committee.

Committee - Would like to know if they are in line with NARA's expectations.

Dr Kahn – [To Dr. Bellardo and Dr. Thibodeau] “What do you want this committee to do going forward? And how can it be most helpful to NARA now?”

Action Item for Dr. Bellardo and Dr. Thibodeau – Answer above question and provide to committee.

Mr. Phillips will also provide a white paper.

Dr. Thibodeau – The charter was broadly written to give the committee certain amount of discretion.

Dr. Kahn – At this point wanted to go around the table and get an informal sense of how the committee felt about the role of the Committee and its contributions, off the record..

<Requested that Scribes Stop Taking Notes>

All input concluded for a few minutes while each member had a chance to speak.

<Scribed Started Taking Notes>

Dr. Kahn – The Committee can have an outreach function – but it would naturally be limited. We have made recommendations in the past – but we are not sure if they are actionable or not.

Mr. Carmichael – I do not receive any “heads-up” about ERA news. Committee members are contacts for ERA but they do not have any information. We have created a knowledge base.

We need to get a statement from NARA as to whether or not we are meeting NARA's needs.

## *Final ACERA Meeting Minutes, 11/29/07*

Dr. Weinstein – No attempt was made to blindside anyone with respect to Lockheed Martin.. We did discuss problems. The situation with Lockheed was news to all of us. We need this committee more now than ever.

Dr. Kahn to Dr. Weinstein – What do you view as the main product of the advisory committee?

Dr. Weinstein – The ideas around the table.

Dr. Thibodeau – Some Observations

1. 1989 – Dr. Thibodeau’s first year at NARA - A great Congressional appropriation hearing was considered one with no questions. Questions could be avoided by not making request

1998 – ERA – 300K to 2008 \$60 Million annual. I was always told not to ask for more money. Always did and got it. We have leadership willing and able to obtain increases.

2. Communication & Project Management – (Last April) Following best practices we have always relied on empirical data gained through Earned Value Management. However, EVM data is inevitably a month or more behind. My judgment, as early as January was that Lockheed would not make the September milestone for deployment, but that was based on instinct rather than data. At that time, the EVM data was positive. Only after the last ACERA meeting did Lockheed agree that they would not make the deadline. NARA decided to focus on solving the problem rather than airing it in public.
3. What the committee can do for us? It can give us important perspective that we cannot achieve ourselves because, necessarily, we are very close to the problem. One of the reasons I do extensive public speaking is that it provides me with valuable insights, even from people who have never heard of ERA or thought about any of the challenges we face. The insights we gain from the experts on ACERA, who are engaged in an extended dialogue with us are even more valuable.

NARA has subject matter experts working on plans for rollout of ERA to the entire government and to the public. The committee can give us insight on how to position ourselves to our customers. What do we need for other Government agencies to “buy-into” ERA? For the public?.

Dr. Bellardo – Earned Valued Reports are Rear Window Reports

- Concrete value – They are important for meetings to hear how we are doing and then hear from you about how we are not performing well. This is valuable.

## *Final ACERA Meeting Minutes, 11/29/07*

- Report – would not make committee more valuable.
- Budget letter from last year – Grateful for that.

Dr. Kahn – I think the committee is very disappointed that Lockheed was not able to meet the deadline. This does not seem to be NARA's fault, and the quick reaction upon determining the problem was appropriate. Please bring us up to date on progress at the next meeting. It is unfortunate, however, that the committee never had an opportunity to hear about this matter earlier.

### **Demonstration of the Hitachi Content Archive Platform**

Dr. Kahn – How did you base your search for an alternative software system to manage the information in ERA?

Mr. Dyung Le – We based the search for systems that can manage the number of objects NARA needs to preserve based on case studies. We did not find anything but HCAP. None of the other systems could handle scalability. The product had to be backwards compatible.

Reaction to the ENRON affair created a widespread need for archiving electronic information. The HCAP product was developed in this environment, recognizing from the start the need to be able to search and recall vast quantities of email.

Mr. John Redgrave – Is NARA in the mode of acquiring systems such as this? Are there other options?

Dr. Thibodeau – HCAP is the basis of the EOP system.

Mr. Maltz – These systems are not intended for federation.

### **HCAP Demonstration**

Mr. Andres Rodriguez – (Chief Technology Officer of HCAP) – HCAP addresses the problem of managing objects within a single data center. No single system can handle the mass of data available on the Internet.

- ERA – Wants to be in control of its name space. There are name space management issues.
- Email can be managed using standards such as .EML and SMTP.
- There is a simple web interface to HCAP. Just a drag and drop. There are SQL like queries as well.

Mr. Rodriguez described his experience at the NY Times where he was faced with a classic intelligence problem, namely how to provide all of the information reporters needed online, searchable and preserved virtually forever. The COTS environment consisted of expensive Network Attached storage devices at one end, and offline tapes at

## *Final ACERA Meeting Minutes, 11/29/07*

the other. After finding no existing commercial solutions and feeling that many organizations might need this capability, he formed Archivas. Since then Archivas has been used in major programs and developed major partnership. Archivas was subsequently acquired by Hitachi Data Systems and the Archivas software is the basis for HCAP

The dilemma facing the NY Times and most intelligence systems is how to address data usage over time. Unfortunately, the data usage curve almost guarantees that some time in the future what you thought was no longer important will become important. The question that has to be answered is what data you can afford to lose either through media problems, search inefficiencies or even timeliness.

There are 3 ways to lose data:

- Physically – it must be protected from disk corruption, media rot, and alteration.
- Logical – a file without metadata is effectively lost since you won't be able to easily find it years from now, so metadata and data must be preserved together and must be searchable.
- Timely – offline storage is too slow, it must be accessible quickly when it's needed.

The bad news is this data is growing exponentially; the good news is that fixed content can be treated differently.

HCAP appears to the outside world as a standard file system, one global name space. It consists of  $N$  number of independent servers connected to backend storage. The servers can be any Intel based machines; the storage can make use of the diverse architectures and standards that may exist in the marketplace. Each server can actually have a different configuration, so newer hardware can be constantly added without having to pre-process what's already in the cluster.

HCAP is near-linearly scalable, tested at NSA and each is responsible for carrying out the archives policies.

The front-end and back-end networks consist of GigE, the back-end is a private network (with multicast addresses) that defines the cluster and the front-end is usually exposed via DNS. HCAP provides search access via metadata and data. Data is fully indexed, providing full-text keyword searches.

With fixed content media exploding (some say up to 90% at 60% a year), it was realized that the solution had to be designed from the ground up. So research into archiving was being done by groups like the Open Archival Information System (OAIS), which provided the basis for the data model; and peer-to-peer systems like Napster, which provided the basis for handling Petabytes of searchable storage using cheap disk storage. Finally this approach needed to be completely open. So to handle the usage curve, HCAP had to provide: managed preservation, transparent access and scalability. HCAP allows one to deal with volatility in hardware, network, data and expansion

## *Final ACERA Meeting Minutes, 11/29/07*

The product delivers a common administrative interface to manage all tiers of storage in a single domain as well as a common interface to manage content of various types created by multiple applications. This simplifies the effort needed to manage the complete storage environment including the archive storage. It does not operate as an independent island of storage in the data center.

Dr. Kahn: Thank you for an excellent description of the HCAP system and how it deals with the storage and retrieval problem.

*Adjourn at 1:31 p.m.*