ADVISORY COMMITTEE ON THE ELECTRONIC RECORDS ARCHIVES
MEETING NO. 2
NATIONAL ARCHIVES BUILDING

MINUTES
DAY 1 OF 2
APRIL 05, 2006

In accordance with the provisions of Public Law 92-463, the meeting was open to the public from 9:00 a.m. to 4:55 p.m. The meeting commenced at 9:00 a.m.

COMMITTEE MEMBERS – 3 MEMBERS WERE NOT PRESENT

Name	Organization
Dr. Daniel Atkins – not present	University of Michigan
Lewis Bellardo	National Archives and Records Administration
Laura E. Campbell	Library of Congress
David Carmicheal	Georgia Archives
Sharon Dawes – not present	Center for Technology in Government
Luciana Duranti	University of British Columbia
Dr. Richard Fennell	Administrative Office of the U.S. Courts
Daniel Greenstein	University of California
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 – not present	Digital Government Information
John T. Phillips	Information Technology Decisions
Jonathan M. Redgrave	Redgrave Daley Ragan & Wagner LLP
Adrienne Reagins	National Archives and Records Administration
David Rencher	Federation of Genealogical Societies
Mr. Richard L. 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

Welcome: Dr. Allen Weinstein, Archivist of the United States

Allen Weinstein, Archivist of the United States, introduced himself to the Committee, thanked the members for their participation, and announced that Dan Adkins accepted a National Science Foundation position; consequently Mr. Adkins can not continue his ACERA Committee involvement. Allen Weinstein indicated that the current budget realities have caused NARA to ask Lockheed Martin Corporation (LMC) to reassess their scope for the first Increment of ERA. He continued to report that NARA received recognition from the White House for their work in Information Technology Research.

Ken Thibodeau, ERA Program Director, covered administration items and asked the committee members to direct reimbursement questions to his staff. Ken Thibodeau announced that Robert Chadduck, ERA Research Director, was appointed to chair a White House committee on Data Management.

Allen Weinstein updated the committee on the status of the ERA Program indicating that the emphasis now is on building the infrastructure, developing the lifecycle processes, and moving towards the goal of facilitating public access.

1. Comments by the Chair: Dr. Robert Kahn

Robert Kahn gave an overview of the meeting agenda that would include a discussion of the LMC Business Development presentations delivered the previous day followed by a discussion of open architecture and a review of the short paper Robert Kahn had written entitled "Thoughts on an Open Architecture for Electronic Archiving" that would be distributed to the committee. Dr. Kahn proceeded to go through the Action Items from the previous meeting and updates were noted by the scribe. (The minutes updated from this discussion are posted on the CORE.gov site.)

Action Item 1: Assistant to the Archivist, Deb Wall, was assigned the action to follow up on any open items from the first meeting.

2. Discussion on the LMC Presentations delivered on April 4, 2006

Robert Kahn asked the committee members for their views on the Lockheed Martin presentations.

Campbell – Excellent. They provided user perspective. ERA maintains capability at the single record level.

Horton – LMC is not working on long term preservation, NARA will determine the medium.

Bellardo – The Focus/strategies and methodologies/technologies of ERA is to liberate information from medium, software, hardware.

Thibodeau – Another focus is on durable storage.

Comment: There is some question of whether or not the states can share with NARA even though extensibility has been planned.

Dr, Thibodeau explained that most of ERA will be based on Commercial Off-the-Shelf (COTS) products and smaller entities will need to purchase infrastructure, which could be a substantial investment.

Question: Will NARA permit the ERA software to be made available to the states? Can a specification of the architecture be made available?
Answer: Security may be an issue.

Dr. Thibodeau was asked to talk about executables. He responded that the security folks have some concern about involving unsecured people in the development of code or allowing them open access to the source code. An executable version of code can be released but not source code.

Archivist – ERA is being designed to allow fee for service operation especially when dealing with the Federal Records Centers. There will be no fee for service for Presidential records and NARA internal documents. Increment I of the system is due to be delivered on 9/07. The date has been pushed back because of budget issues. Increment I will deal with Ingest and preservation followed closely by records scheduling.

There was a discussion about how storage will be handled including what goes on tape and what will be cached on disk.

In terms of fee for service, the libraries are looking for new business models. Perhaps they can leverage the ERA development for this. There may be a way to enable a more general discussion on this topic and this may be brought up at the smaller system architecture working group.

Campbell – Industry is also looking for Business models

Kahn mentioned the issue of "air gaps" for security purposes, which stimulated a conversation about air gaps and how they will play with ERA. Not having an air gap means that there must significant trust in the system and its overall security.

Comment: We need to be clear in differentiating between system development specifications and softer issues like user adoption.

Question: Will a business model be available?
Answer: Some of the higher level components will be available. This may be useful for state archives to start building .their own systems

Kahn talked about what has historically happened with access to source code. This may result in different versions of the same program if we allow access to this form of the software. The Internet is an example where open source code has been made available for various purposes.

Comment: There is strong community support for open source software. It would also be in NARA's best interest to have an open architecture since that would eliminate dependency on a single source or company.

Dr. Thibodeau talked about the costing and how it plays with open architecture. This has been a huge problem so far because it would require this small agency to give even closer scrutiny to security issues. The security folks are very concerned about the mosaic effect. How do we get these people to buy into open source?

Comment: It would be worth identifying leveraging points. Issue #1 is how the information collected by ERA will be used. Issue #2 is potential application basis. It was asked if a subcommittee should be formed to compare issues.

Question: Is the Harris proposal available for public use?
Answer (Dr. Thibodeau): They should be approached directly. We can provide the contact information for Harris.
Action Item 2: Bellardo or Thibodeau – provide Harris contact information.

Comment: There are other agencies that are working on long term preservation. We may be able to share information with them to ensure smooth interaction of systems.

Kahn's Comments on LMC Presentation

1. How to run a metadata registry is an area that Kahn's organization has experience with.
2. He described three (3) approaches, representing different stages of evolution:
   a. Monolithic approach
   b. Modular version – plug & play
   c. Real time versions (of both a and b)
3. What is NARA view of what the LMC effort?

Kahn's View – What should be the role of NARA in interacting with other agencies? NARA is currently in the box management business. NARA should not have to deal with most data typing and related maintenance tasks – this is a hopeless job. They need a standard way in which agencies can submit records. They need to simplify the process – make it simpler and much more manageable. This leads directly to the open architecture approach.

Comment: We could also tell the agencies that being able to manifest the "original" record format would be up to them, even if NARA retained the original digital object.

Bellardo – NARA has developed six (6) or seven (7) areas where there are requirements for agencies to submit records: G15, scanned, pdf, email, digital photo, and web. We'll still take other forms but only when absolutely necessary. The idea is to narrow the scope of what agencies can send to us so we can deal better with others.

Comment: It would be useful for the committee members to get URLs for the format requirements of the relevant government agencies.

Action Item 3: Bellardo – Send URLs, as appropriate, of the relevant format requirements out to the committee members.

Kahn – If an agency doesn't give NARA information in the standard format, the responsibility for manifesting the information, once accessed, should fall to the agency.

Campbell – Best manifestation version – The library has some "best edition" work on digital media (progress in videos standards) and can share it.

Archivist – This is a great discussion. We tend to get focused on day to day business.

Kahn – This is an Advisory Committee for ERA and if our guidance is helpful, we've done our job.

3. Discussion on the Open Architecture Approach

Kahn opened the discussion with a slide depicting an open architecture model for an archival system. He indicated that an open architecture makes a lot of sense when you are trying to involve the public. The purpose of the slide is to make the discussion less abstract.

At the most abstract level, he described three (3) parts:
1. Archival electronic storage with deposit & access mechanisms
2. User interfaces
3. External interfaces to other network-based systems, such as metadata registries

Kahn then discussed the diagram in the slide containing arrows and light blue boxes and stated that the metadata registry is what allows you to be able to search for records.

Kahn added that he assumes real time submission is included as part of an open architecture, but this may not be able to be part of ERA because of the security or other concerns. Kahn raised the general question of how user's and external systems should interface with the system.

Request: The group requested that the slide be posted on core.gov.

Action Item 4 – Put example PowerPoint slide on core.gov

Robert Kahn distributed a four (4) page paper entitled "Thoughts on an Open Architecture for Electronic Archiving" and gave the committee members some time to read through the content. Kahn explained why open architecture is a good thing for ERA and NARA. He talked about who has responsibility for what by using the Internet as an example where social controls were perhaps even more important than actual technical controls via standard interfaces, protocols and objects.

We needed to create the social institutions that support the archiving process.
- NARA can commission these institutions
- They can be staffed or coordinated by NARA
- Complications arise because of external contractors
- Structure is complicated, so there is motivation to move towards an open-architecture

Question: Kahn asked if NARA can buy into open architecture, or the social aspects as a need or, if not, can you restate the need.
Answer: Weinstein and Bellardo will defer to the opinion of the committee.

Dr. Thibodeau reviewed the "must" requirements in the ERA Request for Proposal (RFP). The RFP lays out three (3) requirements. The system must be:
- Evolvable,
- Scalable, and
- Extensible.

Dr. Thibodeau talked about how NARA has supported the OAIS model since 1995. He added that he likes the idea of an independent source of advice and it would be very useful to have the committee's input as he goes down the path with LMC.

Martha Morphy (Observer) – Questions generated here and yesterday are very valuable. We don't always get external input. Any guidance to move this forward would be very helpful.

Comment: It is desirable to have plug and play without rework. I support using state and local government authorities to help build standards.

Comment: There is a tension between two (2) disparate time scales – short term and long term. It might be useful to think of things in these terms during the discussion.

Testa: We talked about decentralized storage with centralized control. He envisions one (1) construct rather than several that would house multiple levels of classification addressing security without air gaps presents significant cost savings over the long run.

Kahn - Protect the objects with encryption, wherever allowed, rather than use air gaps. I take a long term look. This is a high level consideration, but could be a lower level requirement. For example: Sending email to book in the archives to ask questions – such as who wrote you. The book has a unique persistent identifier, which could be resolved into an IP address that was valid at that moment in time.

Comment: We need to address issues such as how the data comes in without attaching ourselves to one (1) component so the system can evolve.

Testa: Wants more detail in the paper.
Kahn hesitates because he doesn't want to influence the design in a negative way by making the system over specified a priori.

Kahn – How should we think about this system?

Greenstein – OAIS model – let's start there to map servers or standards to a model. Will help community to understand what NARA wants to achieve. It may be useful to map the current concept of ERA functionality to OAIS. Knowing where the services mapped could be helpful.

The Chair explained why he did not do that.

Kahn – Open Architecture work is done to leverage innovation and promote evolution in a multi-vendor environment. The web is one (1) such example of an information access system and we need the ability to support alternative architectures and multiple independent systems over time within a common meta-level framework.

Duranti – OAIS is only one (1) part of what ERA wants to be. ERA will cover more processes. ERA will also support the agency business. We can look at a number of models such as DoD. We should look at the components of open system such as record keeping and record preservation.

Campbell – We need to put the business problem in the framework we have. We need to decide what should be open and what should not be open.

Thibodeau – OAIS is intended to be technology agnostic – requires more architecture work; i.e., interface between OAIS and producer such as audit standards tell the system how to perform audits to be compliant. Standards can help committee in conversation about open architecture.

Phillips – What is open depends on perspective. Develop a strategy for openness. It is flexibility. Here's what we think in the next year or two (2) then modify as you go forward. Otherwise, we will be too rigid in system development.

Maltz –Requests to have the ERA system architecture made available and pointed out that the high level architecture was one (1) of the top ten (10) concerns from the last ACERA meeting.

Kahn resisted having the committee take responsibility for the LMC architecture, which is the bailiwick of NARA and LMC; but Kahn encouraged the committee to feel free to gather additional information.

Kahn – We can have a discussion of what NARA is doing, but NARA has to have its own statement of the ERA architecture going forward.

Phillips – We need to determine what the open architecture is and a strategy for addressing concerns.

Campbell – This is a design review. Did we raise the question of where should the architecture be open – if it can't be then why?

Thibodeau – If this committee comes up with a robust architecture for Electronic Record Archives, it will reflect well on NARA and I can use it as a gage to judge what LMC is doing going forward.

Greenstein – You've been six (6) years on this; don't you have an architecture already?

Thibodeau – We based the acquisition on the SDSC architecture, that's what we wanted. We wanted to be abstract. The OAIS model doesn't give us an architecture because it is so general.

Question: Is there an Information Architect on the NARA Staff?
Answer: Yes

Comment: So, wouldn't we start there?

Dr. Thibodeau responded that what we have is notional and does not get down to this level of detail.

Reed – There is lots of interest from outside. We need a more concrete idea of what ERA is.

Archivist – Everything you just said could have a major impact on how NARA proceeds.

Kahn – In many ways, this problem is more complicated than was the original Internet.

Archivist – That's our next brochure!

Kahn – Is this PPT slide a good model?

Campbell – What are the services and transactions that need to be open? This is even more important than the model.

Kahn – It's like Pandora's Box to start with services. I was trying to start at a lower level.

Greenstein – Take a look at the high level statement of preservation and then select a model or create one to use to move forward.

Kahn expressed concern that in trying to replicate the current world in the new system will make the system too constrained by the past, rather than open it to the ideas of the future.

Kahn –To take what exists in the physical world and automate it fundamentally restrains design. The picture today may look very different tomorrow, i.e., changes in media, mediums and even means of expression. Future: the explicit storage of records may disappear, in favor of dynamically created records, where everything is in the background until you want to see it.

Lessons Learned on developing the Internet = eight (8) bits was assumed to be enough to specify the destination network for the indefinite future. We were proven wrong within a year

Bellardo – Proposed that a sub-group be formed on this topic of dynamic records and present it finding later.

David Rencher – We all come to the table with different domains. I propose we have two (2) subcommittees; Architecture components and Service components.

Lunch Break

Committee Reconvenes

Kahn asked Dr. Thibodeau to make comments on the assumptions in Dr. Kahn's paper that was previously handed out to the committee.

Thibodeau - Assumption #1: I'm right at home with. An additional goal is that ERA is a system within a system; digital preservation system and business system moderation for NARA (managing lifecycle of record), thus, we don't expect you to look at all the regulatory and operating restraints we operate under.

Dr. Thibodeau pointed out that Assumption #12 is similar to the OAIS model. In addition to metadata about the record itself, we need metadata about the records' relationship to each other. We will store digital objects with system but we will store records with one (1) to one (1) or one (1) to many.

Kahn – How does this affect the assumption? Can you mark this up (#12) with your input? Metaobjects are another way to deal with certain aspects of relationships between objects or records.

Action Item 5: Thibodeau – Edit Assumption #12 in Dr. Kahn's paper.

Handfield – Asked Dr. Thibodeau if for Assumption #6, is there any provision for 24 X 7 access?

Thibodeau – The budget issues cut back planned access to 10 -12 hrs/day, 5 days a week at Initial Operating Capability (IOC). There will be no public access up front. Public access will come at the end of 2008.

Handfield – With 24 x 7 access, backups, and redundancy will be increasingly important.

Thibodeau – Eventually there will be separate instances and offsite backup for disaster recovery.

4. Discussion on the Subcommittees Structure

Kahn – How should we break out into subcommittees? He proposes three (3) groups:
     a. Open Architecture
     b. Services and transactions

c. Collaboration and utility

Kahn said that each group should define a set of (issues/areas) that you know how to deal with – make a mini work plan for 6 months, 9 months and 12 months. Come up with a plan and report back.

Kahn consulted with the committee members and then assigned the members to the subcommittees as follows:

COMMITTEE MEMBER SUBCOMMITTEE ASSIGNMENTS ON DAY 1

Name	Subcommittee
Laura E. Campbell	C
David Carmicheal	C
Luciana Duranti	B
Dr. Richard Fennell	B
Daniel Greenstein	B
Jerry Handfield	C
Robert Horton	C
Dr. Robert E. Kahn	A
Andy Maltz	A
John T. Phillips	B
Jonathan M. Redgrave	B
Dr. Dan Reed	A
David Rencher	A
Mr. Richard L. Testa	A
Dr. Kelly Woestman	C

Kahn went through the list of Issues identified during the first ACERA meeting and assigned each them to one (1) or more subcommittees as follows:

COMMITTEE TOP TEN ISSUES IDENTIFICATION AND PRIORITIZATION ASSIGNMENTS TO THE SUBCOMMITTEES

	Top Ten Identified Issues	**Votes**	**Assignment**
1.	High level open architecture framework	**12**	**A**
2.	Usefulness and usability for different communities	**12**	**C**
3.	Incentives for collaboration and implementation	**11**	**C**
4.	ERA applicability beyond NARA	**10**	**C**
5.	Consistency of terminology	**10**	**B, A**
6.	Centralization vs. decentralization	**9**	**A**
7.	Leverage, other stakeholders, and initiative	**8**	**C**
8.	Outreach	**8**	**C**
9.	Authenticity defined as identity and integrity of records	**7**	**B**
10.	Managing cryptographic objects	**7**	**A, B**

COMMITTEE COMPLETE LIST OF ISSUES IDENTIFICATION AND PRIORITIZATION ASSIGNMENTS TO THE SUBCOMMITTEES:

	Total List of Identified Issues	**Votes**	**Assignment**
1.	Centralization vs. decentralization	9	A
2.	Labor associated with ingest	5	C
3.	Assure complex record types are defined	3	B
4.	Integration with legal framework	6	B
5.	GOTS and proprietary vs. COTS	4	C
6.	Allow multiple standards for metadata	5	A, B
7.	Scalability of manpower to the number of documents	4	C
8.	Relation to federal enterprise architect	5	A
9.	Leverage, other stakeholders, and initiative	8	C
10.	ERA applicability beyond NARA	10	C
11.	Organizational impact within NARA	1	C
12.	Authenticity defined as identity and integrity of records	7	B
13.	Digital library and archives convergence	4	B
14.	Document chain of custody	3	B
15.	Hardware lifecycle and migration	1	A
16.	High level open architecture framework	12	A
17.	Analysis/viability of persistent preservation	4	A, B
18.	Record-type template creation	5	B
19.	Out-of-scope vs. in-scope	2	B
20.	Plan integration with existing RMAs	4	A, B, C
21.	Conceptual search alternatives	3	A, B
22.	Virtual storage using centralized librarian	1	A

	Total List of Identified Issues	Votes	Assignment
23.	Proprietary source availability	1	B
24.	Curatorial best practices	3	B
25.	Incentives for collaboration and implementation	11	C
26.	Is legal/policy regime appropriate	5	B
27.	Migration issues for encrypted records	4	B
28.	Usefulness and usability for different communities	12	C
29.	Privacy	6	B, C, A
30.	More opportunities for NSF, LC, and NARA cooperation	4	C
31.	Ability to track archivist's access	1	B
32.	Consistency of terminology	10	B, A
33.	Mechanisms that allow applications involving multiple archives	6	B, C, A
34.	Outreach	8	C
35.	Interagency work processes	2	B
36.	Threats analysis	5	A
37.	Secure the information, not the medium	2	A, B
38.	Extracting data for the new archives	2	C
39.	End-user culture change and training	6	A, B
40.	Managing cryptographic objects	7	B, C, A

5. Subcommittee Reports

Each subcommittee reported on their group's activities.

Kahn reported on Group A Subcommittee activity:

Discussion on how to think about the high level open architecture resulted in looking at three (3) levels:

    1. External systems   high
    2. NARA processing – whatever LMC produces   middle
    3. The "Stuff" plus administration of the stuff   low

Emphasis on Protocols: Electronic submission of batches – How to deal with synchronous and asynchronous imputes/outputs and all of the constraints of supporting the non-real time batch input.

Could be good to think about a module to stage large volumes of transfers without disrupting the running system.

Spent time on metadata:

Differences, where appropriate, between data and metadata
How people access metadata –search – batch jobs

Authority files – ways for external transactions to access the appropriate information, or to authorize the appropriate parties for access.

Testa – Requests architectural views from LMC or NARA.
Action Item (6) – Thibodeau to supply services architecture and working group to review it.

Greenstein reported on Group B Subcommittee activity:

Group B began by determining their role, as they understood it, in the process of developing the ERA system. Group B will attempt to identify an *idealized* service architecture reflecting the lifecycle orientation of the business process, identify key service components, and the policy environment for NARA and key stakeholders.

Specific tasks included in achieving these broader goals include but are not limited to:

- Manage expectations amongst users both inside and outside the federal government;
- Identify where the ERA is open and where it is closed;
- Help stakeholders identify where and how they can and cannot interact with ERA;
- Help stakeholder communities identify services that are being offered and those that are not, and develop strategies for filling the gap that they perceive;
- Help NARA assess demand for service amongst stakeholder communities in order to prioritize its own decision making and spending;
- Create opportunities for third-party service definition and development (without undermining the integrity of the record and or NARA's traditional role of ensuring public access to the archival record), and leverage these opportunities as revenue generating ones; and
- Emphasize and encourage an evolving trend within NARA that helps specific communities serve their own access needs.

Laura Campbell reported on Group C Subcommittee activity:

We reviewed 15 issues and group them into 2 areas:

1. Learning
2. Real Issues

Identified three (3) Product Areas:

- Commonality of Needs – such as authenticity and scalability looking at best practices.
- Inspiration for Collaborations – understanding preservation solutions and defining a framework for collaboration.
- Incentives – Looking at fees for service that can suit state archivists.

By next meeting:

- List of Desired Learning (Lessons)
- Definition of Eligible Collaborations

By Next Meeting - To Do:

- Bellardo to distribute NARA's record management standards and guidance in the six (6) areas (expanded formats)

6. Wrap Up

Kahn opened floor to suggestions

Reed – Spoke about having a stakeholder group representatives, like lawyers, to let NARA know what their needs are as users.

McMillen – Suggested we combine subcommittees B and C.

Kahn – We should document the output from each group to use as a central reference – put the comments on the shared work space.

Reed – How to make questions more concrete?

Kahn – Tomorrow, among other things, we'll try to focus in on the protocols.

Archivist – We're winded now, let's meet tomorrow. If be, we'll break after lunch.

Bellardo – Identifying the links between the three areas may help scope the work of each group.

Kahn – Need ½ page write up on each group's mission. Other topics to consider: maintenance and testing, system integrity.

Maltz – Need parameters.

Kahn – Security should be off the table for now.

7. Action Items:

Action No.	Description of Action or Request	Assigned to:	Status:
1.	Follow up on Action Items from the first ACERA Meeting	Debra Wall	
2.	Contact NGC to see if we can provide the contact information for Harris	Lew Bellardo or Ken Thibodeau	
3.	Get URL of expanded formats out to the committee.	Adrienne Reagins	
4.	Put example PowerPoint slide on core.gov	Adrienne Reagins	
5.	Edit Assumption #12 in Dr. Kahn's paper.	Ken Thibodeau	
6.	Post the Human Factors documentation and Enterprise Architecture on core.gov	Ken Thibodeau	

8. ADJOURNMENT

The meeting adjourned at 4:55 p.m.

I herby 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.