

THE ELECTRONIC RECORDS ARCHIVES PROGRAM AT THE U.S. NATIONAL ARCHIVES

Fynnette Eaton
ERA Change Management Officer

The U. S. National Archives and Records Administration (NARA) established the Electronic Records Archives Program in order to develop the “archives of the future”. At the National Archives, we recognized the need to develop a system that can ensure access to all types of electronic files, because electronic formats are placing paper as the medium for communications and transactions. Since NARA’s mission is to preserve and provide access to essential records that are indispensable for documenting Citizens’ rights, the actions of Federal officials, and the nation’s history, we must find a method to continue to perform these functions, even as the range of technology continues to widen.

The National Archives has had a program for accessioning, preserving and providing access to electronic records that are software and hardware independent for more than thirty years. These electronic records were created on mainframe computers and most of the files that have been processed have been limited to structured data files that were fielded, fixed-length, comma-separated ASCII files. Other types of electronic records such as email have received minimal processing (bit stream preservation only). The Electronic Records Archives system is being developed to manage the records currently being created in Federal agencies today. Two examples of electronic records that have been transferred that will require the functionality that the ERA system will provide are:

The Congressional Committee that created these records has transferred the 9/11 Congressional commission records, which include streaming audio, streaming video, email, word processing and other record formats, to the National Archives. We will need to preserve these records for future researcher, so we will need to get them out of the current proprietary software and move them to a persistent preservation format, where it is possible, and, after reviewing for classified and privacy information, make these records available through the web portal that we are planning to provide. This is in the future, but it is our planned future.

The challenges facing the National Archives and all archivists interested in preserving electronic records, includes the wide diversity of record formats, such as office automation, email (in many ways, one of the largest challenges), image, video and audio formats; complexity, such as decision support systems, Geographical Information Systems, interactive World Wide Web pages; volume both in files and bytes, (pull from case for change) and the rapidly changing nature of systems used to create records.

How are we going to overcome technological obsolescence in a way that preserves demonstrably authentic records? Since 1998 the Electronic Records Archives program has been working with research partners to determine the feasibility of building a dynamic solution that incorporates the expectation of continuing change in information technology and in the records it produces; and to find ways to take advantage of continuing progress in information technology in order to maintain and improve both performance and customer service.

During these years in which NARA identified areas of research to address the problem of preserving electronic records and finding storage solutions for the increasing amount of electronic records that should be preserved, NARA has partnered with numerous world-class research institutions that are on the frontier of research into information technology, as well as other Federal agencies, state governments, non-profit organizations and private businesses. These research collaborations have provided an environment for testing and evaluating new technologies as they emerge. These institutions or organizations include:

The San Diego Supercomputer Center (SDSC), a world leader in using and providing innovative information technology. It examined the possibility of using XML as a method for ensuring long-term preservation, using examples of record collections drawn from the National Archives.

The University of Maryland Institute for Advanced Computer Studies (UMIACS), which conducts interdisciplinary research into a wide variety of computing procedures, is a partner in examining grid technology. NARA's current, fully accessioned electronic records are stored on tapes on shelves (right). The ERA Research staff is testing grid bricks (left) as a component of the ERA Research Prototype Persistent Archives. A grid brick is a disk-based storage and management device made from commercially available (COTS) components and utilizes the Storage Resource Broker middleware to manage the records it holds. The VCR-

sized grid brick on the left has the capacity to store all of the records stored on the tapes in the picture on the right.

The National Center for Supercomputing Applications at the University of Illinois has an historian Vernon Burton, as a team member looking at applying scientific data management technologies to the preservation of historically valuable collections.

The Georgia Tech Research Institute, which performs applied research to seek solutions to specific technology challenges, is working on advance decision support technologies contributing to high-confidence processing of large collections.

Other Federal agencies, such as the Army Research Lab, the National Institute for Standards and Technology and the National Aeronautic and Space administration are also helping to find a solution. One of the key research activities was the development of the Open Archival Information System or OAIS, which provides a method for developing software requirements to perform archival functions such as accessioning, preserving and providing access to electronic materials.

Another research project, InterPARES, the International Research on Permanent Authentic Records in Electronic Systems, has been developing the theoretical and methodological knowledge for long-term preservation of authentic records in digital form.

All of these activities have helped inform the development of the requirements for the ERA system.

Where are we today? We developed requirements, issued a Request for Procurement and selected two companies to compete in suggesting a design for the ERA system. In September 2005, we selected the Lockheed Martin Corporation to build the ERA system. They are currently beginning to design the system. We expect to have a first Initial Operating Capability in September 2007. As you can see from the slide, we plan to have 5 increments with greater functionality in later stages and hope to have this completed by 2011. The one sticking point that is occurring across all federal agencies is funding. So we will have to see if funding limitations have some effect on the timing of the increments.

So where are we with the Lockheed Martin Solution. It is based on the archival mission as outlined in the Open Archival Information System standard, identify, preserve and make available with common services across all three areas. The principal design considerations include evolvability, scalability, extensibility and availability. What is meant

by these terms? Evolvability requires that the system be policy neutral, so if new policies are put into place, the system will adjust. In addition, the system has to be able to change over time, as technologies change. Scalability requires the system to be able to scale up or down, depending upon volume and usage and as well to scale to other types of archives. Extensibility means that the system must be able to easily add additional features in the future without major modification. Availability means that we have no single point of failure and that the system maximizes “up time,” while balancing availability with cost.

Templates will drive preservation. A record is evidence of the activity of the Federal government. A Record Type Template is an abstraction that categorizes forms of intellectual content and captures their essential characteristics. The data file is a sequence of 0s and 1s, while the Data Type Template is an abstraction that captures essential characteristics of the data format.

One of the key elements to ERA is the preservation planning solution. By examining the records, identifying them, as for example, the record is a legal contract; the archivist can define what the objectives must be in preservation. In this case, the decision might be that pagination must be preserved, but color would also be nice. The system will offer a variety of capabilities, for example, one digital adaptation processor does a great job with pagination, but only supports black and white. A second digital adaptation processor preserves color flawlessly but does not recognize pagination, while a third one can perform both tasks, but it is still in development.

Although this sounds relatively simple, the real world is much more complex. Instead of one record, most transfers from agencies will contain series, such as correspondence, which will include possibly an HTML page, a GIF image and Outlook PST, requiring a range of digital adaptation processors.

For Lockheed, making sense of the 0s and 1s is dependent on a web of connections. Their approach is to “externalize” all the critical information and linkages and hold this in ASCII. In effect they plan to preserve all the information needed to interpret the records in ASCII (XML), preserve all life cycle data and description in ASCII (XML), preserve the linkages between information needed to interpret the records, and preserve the binding between the records and life cycle data.

One of the major benefits of this design is that provides a transition path to new generations of technology as they arise, because this plan uses the most durable format currently available for the asset catalog and future technologies will be able to use it.

Progress thus far has helped advance ERA from vision to reality. Lockheed Martin has validated that the technology for the solution exists today. There is a clear recognition of the importance of organizational change in ensuring the successful deployment of this system in the National Archives. Research supporting this work has helped clarify archival thinking about concepts like “authenticity” and “persistence.” And both Lockheed Martin and NARA have identified risk to program success and have developed mitigation strategies to control these risks.

The process already has led to some significant advances in areas of interest to archivists. These include: an asset catalog with virtually unlimited flexibility for hierarchical arrangement; a template approach that will serve as a foundation for automation strategies; an authenticity approach flexible enough to meet changing NARA policies while being rigorous enough to withstand procedural challenges; a description approach that combines preliminary automated data extraction with traditional archival descriptive practices; and a persistence approach that finally achieves the dream of a “self-instantiating archives.”

What services will ERA provide to our users. For agencies, ERA will enable agencies to use information systems for the purposes of their lines of business and when the records need to be transferred, the agencies will not have to worry about specific formats. ERA will also enable agencies to develop their records schedules electronically and work collaboratively with NARA staff in the review of these schedules.

For researchers, NARA will provide access to electronic records, from your desktop. Different types of search interfaces will be developed to reflect the interests and knowledge levels of the different types of researcher we expect to use this system. ERA will not replace reference archivists. Their knowledge will still be needed to assist with the complicated research questions that you pose to us. But for simple queries, or to get a sense of the types of materials that are available at NARA, ERA will provide web access to this information.

Clyde Relick said in his Prologue article: The ERA: Technology to Aid Archivists and Historians, “ERA is not a technology solution but rather an archival solution made possible by technology.”

NARA is taking the lead in finding a method to provide access to records that the federal government is creating digitally that need to be preserved to protect our rights and to ensure the historical record. Thank you for giving me this opportunity to talk to you about the system.