



[Home](#) | [About](#) | [Member Meetings](#) | [Workshops and Conferences](#) | [Interest Areas](#) | [Documents](#) | [Site Map](#)

[Home](#) > [Minutes](#) > Principals and Alternates Meeting February 2004

CENDI PRINCIPALS AND ALTERNATES MEETING

National Archives and Records Administration
College Park, MD
February 3, 2004

Minutes

Digital Archiving — Technologies, Policies, and Approaches

[Technologies for Digital Archiving at CDL](#)

[ARL's Perspective and Role in Digital Archiving and Permanent Public Access](#)

[National Digital Information Infrastructure and Preservation Program \(NDIIP\)](#)

[CODATA Data Archiving Activities](#)

[Digital Preservation and Permanent Access to Scientific Information: The State of the Practice](#)

[NARA Showcase: Electronic Records Archive \(ERA\)](#)

[E-government Interagency Committee on Government Information Update](#)

WELCOME

Kent Smith, Chair of CENDI, opened the meeting at 9:30 am . He thanked NARA for hosting the meeting. He apologized for the late start but snow conditions and snow warnings caused some confusion and traffic problems (the federal government announced a two-hour delay in work schedules.)

DIGITAL ARCHIVING — TECHNOLOGIES, POLICIES, AND APPROACHES

"Technologies for Digital Archiving at CDL"

John Kunze, Preservation Technologies Architect, California Digital Library (CDL)

The California Digital Library supports the assembly and creative use of materials for libraries on 10 campuses throughout the state. It is physically located in Oakland and is not directly associated with any particular campus. The CDL has more than 60 staff on site and distributed throughout the campuses where they support co-development relationships and act as conduits for information flowing out from the libraries. CDL serves a wide range of users including the faculty, students and personnel at the 10 libraries, personnel, the laboratories at Berkley , Livermore and Los Alamos , and the public.

The CDL focuses on diverse digital library collections, including a variety of content types (text, images, and datasets). The current CDL repository is four terabytes. Development costs are estimated at 4000 man hours. The content comes from many sources including published content licensed from commercial vendors, digital content from the libraries, online research and learning materials from the faculty, and web-based content created and managed by third parties. The latter are highly distributed and particularly volatile. There are art image collections and CDL is on the verge of adding video. Some examples of materials include the UC press books, working papers from faculty, and the speeches from the Black Panther Archives. CDL is adding 50 years' worth of state census data, vital statistics for the state and immigration data. They are beginning to investigate web-based information and learning objects.

The CDL also invests in services and applications. It manages the Melville Union Catalog of over 23 million bibliographic records, and hosts the online archive of California , which includes content from 41 institutions. Through UC California Press, CDL is involved in e-scholarship efforts to develop new forms of scholarly communication, such as the Electronic Cultural Atlas Initiative. This initiative brings geospatial referencing over the whole enterprise. They also need geo-referencing as they deal with such issues as earthquake information.

CDL has invested heavily in education and assessment to further its understanding of the digital archiving domain and how to provide appropriate services. CDL's approach to technology is to build utility services that are cost effective and that can be shared with the libraries while allowing for local flexibility. This approach focuses on the commonalities rather than addressing every need of every campus. Part of this approach is to create guidelines and best practices within the system and externally.

The first utility service will be the preservation repository, a suite of utilities that make up a common preservation infrastructure that will be in part provided by CDL. The provision of framework with modular components is key because campus libraries may undertake their own efforts because they have distinct interest in a particular collection. Archiving and preservation will also take place as part of the CDL Preservation Program which was established in 2002 with the goal of providing long term access and availability of materials to the campuses. Funding has come from IMLS, Mellon and by repurposing print preservation funds.

During the development of its Preservation Program, it became clear that CDL needed to partner with other libraries and institutions because it can't do it alone and to ensure backup outside the geographic area. Current partners include the Stanford Computer Science Department, the San Diego Supercomputer Center , and Sun Microsystems.

CDL is focused on building the preservation repository rather than the access systems. (Initial access to the repository will be by record number, providing only simple addressing, rather than any true search logic.) It is important to analyze the services and the technologies needed to achieve them. Key aspects of the system include flexibility, extensibility, modularity without dependence on any particular layer, and simplicity. The Open Archival Information System (OAIS) Reference Model is used as a vocabulary. Components include digital objects and their metadata; persistent identifiers; replicated, access-controlled storage; defined submission/dissemination policies; submission and logging services, and bare-bones access.

The ingest, access and administration clients use the SOAP protocol. The data management layer uses the Archival Record Key (ARK) persistent identifier scheme (specialized URL, access to the URL, access to the metadata and access to statements of purpose). The Storage Resource Broker (SRB) from the San Diego Super Computing Center allows you to incrementally grow a grid of data bricks (a series of \$1500 computers; an array of rack-mounted machines). The SRB supports automatic replication; the write operation will not complete until the content is written to three or more other repositories to ensure redundancy. CDL is using Metadata Encoding and Transmission Standard (METS) to wrap the elements and the metadata. The PREMIS preservation metadata elements will be incorporated when that standards effort is completed. This will include the permanence ratings that were originally developed for NLM.

The CDL is also involved in capturing web-based government information. Government information is often a critical piece of the research agenda and citizens rely on it. In addition, government information is relatively unencumbered with regard to intellectual property issues.

It is conducting a 15-month study funded by the Mellon Foundation to capture web-based government information, which is considered to be a critical category of resource that is at-risk, and to provide some possible solutions for better preservation of this material (www.cdlib.org/programs/digital_preservation.html).

The methodology for the investigation of web-archiving included a comprehensive review of the literature and other initiatives, a demographic review of the '.gov' domain, and sample crawling. CDL's review found that there was little experience in web-archiving. They conducted a day-long meeting of librarians with a broad range of collection needs across disciplines to begin to identify the problems. A key finding was that people don't really know what preservation is about. They know about capture but not preservation. Web archiving practices vary widely; there needs to be sustainable support for a wide variety of experiences that include technical solutions and a common range of tools and services to help to support institutions that cannot support these initiatives themselves. This open source library might include tools for analysis, data capture, tool registration, repository management, archive auditing and curation. Analytical tools would allow the curator to determine the optimal capture strategies.

They also worked with Stanford's crawler developed by the same group from which Google emerged. The parameters of the government crawl included a depth of 10. Files of more than two megabytes were ignored. Most domains contained less than 10,000 files. Ninety-two percent returned fewer than 10,000 files, and 63 percent returned less than 1000 files. The top four file types were HTML text, gif images, jpg images and PDF. PDF made up the largest percentage of bytes (~93%).

Agreements and standards might be adopted to minimize crawling times on the sites and to improve the results; currently, there is no routinely accepted approach. For example, a "crawler hints file" is being developed. This file could contain all the URLs that the site would want to show, including content that is generally hidden from crawlers. A modification date for each URL could support re-crawling only when necessary. It could also list friendly peer institutions and announce new sites for other partners. Authenticated, digitally signed files could also be identified for the crawler. An old fashioned crawl could be performed periodically to ensure that everything has been captured. CDL will soon release some of the Apache module needed to support this hints file. The Government Printing Office has been working with CDL on this concept, and they would like to partner with others. The Registry tools will help libraries track what has been archived already and eliminate redundancy. The Archive auditing tool will enable the verification and transfer of whole collections between sites.

Minimum metadata is needed just to manage the sites that have been captured. The InfoMine Project at UC Riverside is developing a specific domain crawler that analyzes the pages and runs heuristics to extract metadata. It gives candidate metadata with varying levels of accuracy. It also does link counting. This will allow the highest priority sites to be placed at the top for review by human catalogers.

"ARL's Perspective and Role in Digital Archiving and Permanent Public Access "

Prue Adler, Associate Executive Director, Association of Research Libraries (ARL)

Ms. Adler addressed ARL archiving and preservation strategies to speak to the uniquely digital environment.

Digital archiving and preservation present new research challenges because of the changing technologies, the explosion of various digital objects, and the dependence on hardware, software, and standards that are constantly evolving.

1. Digital archiving and preservation present new research challenges because of the changing technologies, the explosion of various digital objects, and the dependence on hardware, software, and standards that are constantly evolving.
2. Digital objects require constant and perpetual maintenance due to continuous change. There is a growing consensus that digital preservation is no longer about moving data to a particular format and storing it, but providing access over time.
3. There is a growing complexity in types of data with new challenges; e.g., e-journals come with software tools.
4. IPR issues are coming with many new issues.
5. Authenticity of the record is a vexing issue.

ARL sees four key roles for libraries in digital preservation -- ensuring the information is there; ensuring access; providing selection and validation of the resources; and serving as archival and preservation agents, realizing that the economic incentives aren't usually there for the original producers.

While ARL has had several digitization projects in the past, it recently became involved in a Government Documents Project, since many of the large research libraries who are its members are also Federal Depository Library Program (FDLP) libraries. Most of the work involves reformatting from paper to digital rather than capturing born digital materials. However once the material has been digitized, it has similar problems to materials that are born digital.

Digital reformatting is a viable preservation strategy for print. ARL has been encouraging the National Endowment for the Humanities (NEH) to move money from microfilming to digitization in its Brittle Books Project. While institutional repository development is primarily focused on born digital material, there are instances of digitization as a preservation mechanism. The Massachusetts Institute of Technology (MIT), for example, includes MIT Press backfiles in its D-Space implementation. The Scholarly Publishing and Academic Resources Coalition (SPARC) members are organizing to adopt scholarly societies in order to get the backfile collections digitized. Boston College is investigating the preservation needs in the process of digitizing still images. They are also working to understand the barriers to sound and video collections.

There is a lot of activity in the area of government information preservation. This can be attributed to the increased use of electronic dissemination on the part of the agencies. Also, reforming dissemination policy has been high on the Administration's list and the new leadership at GPO is providing support for new access methods. There is no single architecture that can handle all government information. Even though some agencies are involved, it would be helpful to have more.

Preservation is high on the agenda for FDLP libraries, an indication of the investment that libraries are making in this area. ARL surveyed its members to understand a large range of issues. The results convey a very strong message about the economic effort on the part of libraries to support digitization. Ms. Adler will provide the URL for the survey.

Action Item: Ms. Adler will provide to CENDI members the URL of the ARL survey about economic issues within the federal library system.

Several key themes emerged from the last FDLP meeting in Reno , Nevada . Over the last five years, there has been a shift in how government document collections in libraries are managed as collection managers try to determine how to meet both the national and local needs. The future of the FDLP will be in service rather than in direct collection development. They will move from providing content to providing expertise, since the FDLP libraries won't necessarily house the government documents any longer. Cooperative efforts will not only enhance access but solve some space problems for the libraries, and, over the long term, libraries expect cost savings. The movement of libraries out of the FDLP puts new burdens on the remaining institutions and digitization is seen as a way to help the remaining libraries. While government documents are just one of many collections that libraries are interested in digitizing, the directions set by government documents will be key to how the overall preservation infrastructure develops.

There are significant projects in digitization of government information already underway through partnerships. Stanford University has just completed the digitization of the Nuclear Regulatory Commission publications and the University of Michigan is working on Presidential documents. A partnership project between Stanford University and Google will digitize Stanford's pre-1923 (non-copyrighted) holdings. Google has been talking to a number of libraries and agencies about digitizing content. Google claims that it can digitize 300,000 pages per day.

However, without a systematic project to coordinate the digitization of government information, there is no consistent digital conversion standard, bibliographic controls are not readily identified, and libraries are likely to lock into high price commercial

products. ARL is trying to determine how to reduce redundancy in digitization projects and to improve the quality of digitization efforts so that the results are amenable to long term preservation requirements.

ARL members asked ARL to analyze the resources required to digitize 2.2 million volumes. Raym Crowe has developed a draft Business Plan which identifies the costs, the relationships among members, etc. The result would be a distributed system that provides legacy government information in digital form without a fee.

Authenticity is a key factor when dealing with government information. Unless you can authenticate and validate the government information, you aren't going to have the robust infrastructure that is needed for e-commerce and e-government. Authenticity is particularly vexing for materials that will become government records. There is a growing complexity because of multimedia, data accompanying electronic journals and the inclusion of tools for manipulating data. Provenance and rights management are important. Permanence is required.

GPO will try to facilitate the effort. GPO will soon announce a clearinghouse for digitization efforts. This will allow libraries and other organizations to indicate what collections they are digitizing and allows others to "carve out their niche" in an effort to avoid duplication of digitization effort. Under GPO's auspices, there will be a working group to look at the standards issues so that a minimum level of quality is maintained. The Permanent Access Working Group will also be reconstituted. Four mirror sites are indicated in the Business Plan. There is a debate about whether this should be a GPO responsibility or whether there should be different sites around the country. This effort is also connected to GPO's mandate for a national bibliography. GPO won't be able to catalog everything, but a distributed search could be used to "collect" the bibliography virtually.

An ARL working group also asked Mr. Crowe to identify what it would take to digitize the top 25 government titles or series in order to gain momentum and visibility. Ms. Adler requested help in identifying the top 25 series or titles that the ADL project should digitize. Perhaps this is something that CENDI could get involved in with regard to the S&T titles. Regional interests take over pretty quickly and definitely color what people are interested in digitizing.

Action Item: CENDI should consider involvement in identifying key STI titles for digitization. The Digital Preservation Task Group will undertake review of this possible task.

There is much to do in the area of government information preservation. While all the technology issues have not been defined and the best practices are still emerging, we have to get started.

Discussion

It was noted that many CENDI agencies have digitization projects. Some of these efforts are focused specifically on government information such as technical reports, while others are more general projects to support digitization in a particular subject area. Notably, PubMed Central is supporting the digitization of backfiles for major medical journals.

The relationship between these various collections is an area for additional discussion. For example, GPO could be viewed as a "Collection of Last Resort". Two copies might be created and stored in geographically dispersed locations. A dark archive could be created that is not available for access but would be used in a disaster recovery mode to "start from scratch".

ARL is also involved in open access. Ms. Adler was asked about any "downsides" to open access. If there are any downsides, they involve the difficulties for small scholarly publishers with regard to both technology and economics. Work is underway to identify new business models that will allow these small publishers to operate in a healthy way. For instance, BioOne is helping smaller societies to move from print to electronic and to identify new models. They have hired a consultant to look at the detailed economic data for the impact of these new economic models. ARL is also concerned about value-added copyright that might be claimed by private sector organizations. For example, Google has a big digitizing initiative but claims that it will digitize and index for free and without holding any claim. However, there are concerns, particularly if Google becomes a public company. The same issues would be true in partnerships between Google and government science agencies, where Google might claim copyright in the value added to otherwise non-copyrightable government information. However, TC Evans pointed out that there are tremendous opportunities at the table and we should try to take advantage of them.

"National Digital Information Infrastructure and Preservation Program (NDIIP)"

William LeFurgy, Library of Congress

NDIIPP was created through federal legislation in 2000. It was funded up to \$175 million with cost sharing. The plan for NDIIPP, described in "Preserving Our Digital Heritage", was approved in 2002. In 2003, \$20 million was released, and \$15 million was approved for matching funds.

The basic philosophy of the program is that institutions should not wait for all the issues to be solved before trying to manage and preserve at-risk digital content. NDIIPP's specific goals are to develop a national strategy by working with Federal agencies and other institutions. NDIIPP currently has four focus areas: developing a network of partnerships, scaling up its own digital content, designing a preservation architecture (a high level design for collecting, preserving and providing access), and funding

preservation research.

The network of partnerships provides a mechanism for stakeholders to work together and learn from each other. A key question is how to build a broad network of players that goes beyond the usual digital heritage institutions. Business models are being identified in order to determine the incentives needed to ensure economic sustainability of any program. Standards and best practices are also being identified.

The first NDIIPP partnership solicitation closed in November 2003 with a focus on content. They received more than 20 proposals covering a cross-section of digital content from databases to web content to electronic journals. The proposals required one-to-one cost sharing and had to involve all aspects of life cycle management (selection, capture, etc.). The NEH will help conduct the peer review. The awards will be announced during Spring 2004.

NDIIPP is also involved in international activities. A consortium of 12 national libraries is developing technical specifications for open source tools to harvest and preserve content. An open source web crawler is under development.

The Library of Congress needs to build an advanced technical infrastructure for its own digital collections while also promoting a global approach to the problem. A new set of policies and procedures are needed to effectively manage content across the digital life cycle. In support of its focus on collection building, NDIIPP has targeted key areas of national interest. To date, more than 20 terabytes have been collected including information about the Iraq Conflict, the national elections, and the George Mason collection on 9/11 which was donated to the Library. The largest amount of digital content is in the American Memory Collection which includes images, video, audio, and text.

Work continues on the preservation technical architecture. A key principal is the separation of preservation and access. This is often a worthwhile line of discussion with rights holders. NDIIPP is working on a more detailed version of the architecture and a final draft for wider comment has been released on the project web site at <http://www.digitalpreservation.gov/index.php?nav=3&subnav=12>.

Another activity includes the testing of archive ingest and handling. The test involves about four cultural heritage institutions that already have digital preservation initiatives and a large body of material in different file formats (approximately 12 gigabytes with 57,000 objects) from the George Mason 9/11 collection. Over one year, the testers will measure the ingest processes (this includes developing metrics), document the steps taken to preserve the integrity of the bits and logical integrity of the compound documents, look at metadata that is needed, identify backup and restore procedures, and analyze the reliability of the systems. The results of each ingest will then be exported and ingested into another institution's archive.

NDIIPP has recently issued "It's About Time," which is a report outlining an agenda for digital preservation research. The National Science Foundation (NSF) will administer a grants program to implement this agenda. A call is expected in the next few months under NSF's Digital Government Program. The research is expected to span areas such as digital repository models, tools and technologies, and policy. When the NSF program is underway, NDIIPP will publicize the outcomes. This research program has an interagency Steering Committee that serves as a focused forum for the Federal government to talk about digital preservation. Mr. LeFurgy invited participation on the part of CENDI agencies.

"CODATA Data Archiving Activities"

Julie Esanu, National Research Council/CODATA

CODATA, the Committee on Data for Science and Technology, is an interdisciplinary committee of the International Council for Science (ICSU) focused on the organization, management, quality control and dissemination of scientific and technical data. CODATA has a broad portfolio of archiving activities dating from 2000 when the first working group on this topic was formed. It held its first workshop in Pretoria in May 2002.

In 2002, the CODATA Task Group on Preservation and Archiving of Scientific and Technical Data in Developing Countries was formed. The objectives of the task group are to improve understanding of S&T data management conditions in developing countries, to advance the development and adoption of archiving and records management practices, policies and tools, to provide interdisciplinary forums for discussion of issues surrounding archiving and preservation, and to build a comprehensive directory of managers, experts and archives that can provide a network of support particularly for data managers in developing countries.

On December 15-17, 2003, CODATA co-sponsored a workshop with ERPANET, an activity funded by the European Union to look at digital archiving and preservation across sectors. The focus of the workshop was on the selection, appraisal and retention of scientific data. The workshop used case studies in various scientific disciplines (physics, social science, biodiversity and astronomy) in Europe and the United States to look for commonalities and differences in practice.

The workshop was attended by 60 data managers, researchers, records managers, archivists, and librarians from 13 countries. For the archivists who attended, the workshop highlighted the fact that scientific data are not the same as records. The benefit to

the CODATA community was a better understanding of the concepts of appraisal and retention which come from the archive community. The report of the meeting will be available from ERPANET. Selected papers will be published in CODATA's online journal.

CODATA is proposing several activities as follow up to the workshop. Several years ago, CODATA developed a glossary of terms relevant to the management of scientific data, and this glossary will be updated with terms relevant to archiving and preservation. The attendees at the workshop concluded that broad statements about archiving and appraisal of scientific data would be helpful, and CODATA may work with ICSTI on this.

A follow up to the successful workshop in May 2002 in Pretoria will be held in Beijing , China , in June 2004. In addition to archiving, the program will include discussion of open access. (China is looking at liberalizing its access policies.) In particular, this workshop will emphasize biomedical and environmental sciences and related STM journals.

ICSTI and CODATA are sponsoring a "Portal to Scientific and Technical Data and Information Preservation and Archiving Resources." The U.S. National Committee for CODATA is taking the lead on this project and the prototype is being developed at the National Academies. It will provide resources and links to related work by other organizations. CODATA and ICSTI plan to launch the portal at the CODATA's biennial conference in Berlin in November.

CODATA is also involved in the issues of public domain and open access. It recently co-sponsored several meetings on public domain information in the United States and Paris . Other issues of interest to CODATA include peer review of data and the development of a core set of metadata for datasets across disciplines.

CODATA will continue to pursue opportunities to promote and advance long-term preservation, management and access to scientific and technical data. It will examine the commonalities and differences between scientific disciplines and sub-disciplines and between digital data and information. These activities will seek to learn from previous and ongoing experiences with managing the ever growing collections of digital data and information and to provide guidance to working scientists in this area.

"Digital Preservation and Permanent Access to Scientific Information: The State of the Practice"

Gail Hodge

This report was co-sponsored by ICSTI and CENDI. The goal was to update the report written in 1999 and to provide information to the sponsoring organizations and others that would allow them to decide how to focus their preservation activities with regard to scientific information. The report gathered information from over 50 specific sources and then highlighted 21 systems that are operational in the area of preservation of scientific and technical information. A list of the highlighted systems was provided.

In general, the study found that operational systems are being developed, including "off the shelf" solutions. However, there continues to be an emphasis on archiving and capturing digital information rather than preserving it. In part, this is because, other than the large data centers, there has not been a major technological change. Migration is the prevalent preservation strategy. Partnerships are increasingly important and there is increased sharing of information across sectors and scientific object types (data, text, images, etc.). Activities in the area of standards and best practices have received renewed interest.

A list of recommendations was presented to the CENDI members:

- Work with GPO, NARA , the LOCKSS-DOCS Project, the NDIIPP, and others to develop effective and sustainable preservation guidelines for government STI in the context of the R&D component of the Federal Enterprise Architecture, e-Government, the federal research process, and the environments of federal science agencies.
- Host a follow-on to the workshops sponsored by the National Science Board and NARA to continue discussions about the selection, retention, organization and on-going preservation of scientific data produced as a result of government funding.
- Support development of technical and social solutions to the federation of archives
 - Development of core metadata for different digital objects
 - -- Implementation of high level OAIS Reference Model functions
 - -- Producer-archive interaction checklists for the federal science environment
- Continue involvement with standards efforts.
 - Review the Journal Publishing DTD developed by NLM
 - Review the DTD for Technical Reports which is being developed by DTIC
 - Determine how these efforts might be addressed by the agencies and as part of the Federal Enterprise Architecture
- Support the development of technical solutions for archiving the dynamic and deep Web which is an increasing problem as scientific communication moves to the Web and as the complexity of these sites increases

"NARA Showcase: Electronic Records Archive (ERA)"

Fynnette Eaton, ERA Program, National Archives and Records Administration (NARA)

Fynnette Eaton introduced herself as being given the assignment of "Change Management Officer." She described the research initiative to frame the new responsibilities that came with dealing with rapid technological change. The Archivist of the US recognized the need to respond to problems related to electronic records and to strategically handle any type of electronic record created anywhere in the federal government. ERA is the strategic response to this environment of change. ERA's vision is to "authentically preserve and provide access to any kind of electronic record."

NARA has had over 30 years of experience with electronic records. For many years it has had a program to handle software and hardware independent files, simple files of structured databases and tables and e-mails received only as bit streams. These experiences helped NARA understand the issues related to the management of electronic records. The challenges include the diversity of formats including images, video and audio; the complexity of some records such as those from decision support systems, geographical information systems (GIS) and interactive web pages; and the volume of bytes and files. The rapidly changing environment makes the key task one of overcoming technological obsolescence. NARA is currently expanding preservation methods and formats, and the ERA Program will use these digital collections to evaluate the ERA system that will be developed.

Since 1998, NARA has been working with a group of research partners to determine a dynamic solution that expects progress in information technology, that readily accepts change and embraces the continuous improvement in customer service and performance that technology can bring. NARA has partnered with experts in industry and government, including computer scientists and engineers, to work on research initiatives through resource and fund sharing. Through these efforts NARA has been involved in key activities such as the Open Archival Information System (OAIS) Reference Model, the research at the San Diego Super Computer Center and the international InterPARES project looking at electronic systems. The first three years of ERA were spent doing research. The OAIS standard is at the center of all their follow up work.

InterPARES involves national and institutional archives from over 20 countries. The precursor to InterPARES was a collaboration between the University of British Columbia and the Department of Defense Records Management Task Force, which developed the DoD standard DoD 5015.2-STD, Design Criteria Standard for Electronic Records Management Software Applications. InterPARES I addressed a specific set of research questions focused on authenticity, appraisal, preservation, and strategy. The goal was to benchmark requirements for supporting the authenticity of electronic records; to identify baseline requirements for supporting the production of authentic copies of e-records; and to develop activity models for preservation function. Through these activities, NARA has learned a lot about the complexity of the questions that they are asking and continue to ask. Records must be authentic and persistent and the system must be scalable.

One outcome of the collaborative research has been the development of transformation as a preservation strategy. This approach does not preserve the object in its original state but focuses on precise archival requirements related to the content and architecture. This strategy actively manages records and solves some of the problems related to technical obsolescence by requiring precise specification of archival requirements related to content, context, structure and presentation of records and the collections to which they belong.

The ERA Program has three main parts – system acquisition, organizational change management, and research and exploratory development. NARA 's strategy is to attack the critical preservation problem, to define it with regard to the lifecycle management of records, to find solutions in commercially viable technologies, and to align its directions with overall IT directions with the US government.

The ERA will provide a series of services. These include the storage of records, bringing electronic records to the NARA and Presidential Libraries, preserving them, helping customers discover and access them while respecting the legal restrictions for access, providing authentic copies with property and other rights upheld, guiding and assisting federal agencies in their life cycle management, and supporting e-government. Other major system requirements include support for collection, the ability to have readily embedded records management in the workflow for each user group, security, and auditing.

NARA has identified a variety of customer groups for the ERA including the originators of records, the NARA and records center staff, and the end users of the archives. For NARA 's customers, the ERA will process electronic records of all types. For users of records, it will provide access from NARA and other locations, providing registration of users, user interfaces for users of various skills levels, with support from tutorials. The ERA will enable NARA staff to respond to user comments and requests. Record originators will be supported in the implementation of record schedules, appraisal, disposition, transfer and accessioning of records.

The timeline for ERA began with research conducted from 1998-2000. In 2001-03, a contractor team was brought in to assist with program management and NARA began to involve its staff, and through integrated product teams they developed key documents. More than 20 staff were added to the Program Management Office with an emphasis on the kinds of skills needed to deal with e-records and related technologies. The Request for Proposal (RFP) for the ERA system development was released on December 5, 2003 . The deadline for proposals is February 11, 2004 . NARA will select two contractors. From the third

quarter of 2004 to the third quarter of 2005, the two contractors will prepare a design for this system. A single developer will be selected at that time. The first increment of the system will be operational in 2007. Four additional increments will be completed by FY2011, if NARA continues to get the funds that it needs to move forward.

As part of the records management initiatives, NARA conducted a Business Re-engineering Process on the scheduling and appraisal processes. This analysis identified the need for more flexible scheduling and the ability for agencies to schedule "bigger buckets" of records at one time.

NARA has a transition team looking at the current systems and how to move them into the ERA. ERA will eventually subsume the current systems, and metadata creation for paper records will be automated. Change Management will help to ensure that NARA can successfully implement the system it acquires by managing the change within its workforce.

E-government Interagency Committee on Government Information Update

From the original E-Government Interagency Committee, three smaller working groups have emerged. Ms. Carroll presented a summary of what the Policy Group members heard from various people involved. Bev Godwin from FirstGov had provided even more information to reiterate and reinforce the involvement of CENDI members at all levels. Eliot Christian is chair of the working group on categorization of information. He has convened a group of outside experts to help to determine how to constitute the actual working group and how it should operate, and the actual working group has been defined.

Dr. Bellardo, who is on the Interagency Committee, indicated that the three working groups are now at the stage of developing formalized work plans to address the various requirements in the legislation. This effort is at various stages for each working group. Nancy Allard serves as the Secretariat for the overall committee as well as the liaison with the Web Content group. Deliverables are expected by this summer.

One of the topics being addressed by the Executive Committee is how to make the work of the group visible in the federal community and to the public. The work plans can be shared. A web site is being discussed. There must be either a comment period for agencies or involvement by the agencies during the process. The Web Content group plans to do extensive outreach in order to get comments early.

Mr. Smith suggested that CENDI be kept up to date on these activities as appropriate perhaps by having a report to the group from those involved. Ms. Carroll asked that information be provided to her and she will compile it and send it to the listserv.

Action Item : Secretariat and the Policy Working Group will determine how to provide ongoing reports about the Interagency Working Group and subgroup activities. CENDI members who are involved in these activities will report out to the group as appropriate.

[Return to Minutes Archive](#)



[Privacy/Security](#) | [Site Map](#) | [Contact Us](#) | [Members Only Site](#)

science.gov
Supported by CENDI