
Progress Report: March 1, 2008-May 30, 2008

NHPRC EMCAP Tool

Kelly Eubank



NORTH CAROLINA DEPARTMENT OF CULTURAL RESOURCES

MICHAEL F. EASLEY, GOVERNOR
LISBETH C. EVANS, SECRETARY

OFFICE OF ARCHIVES AND HISTORY
JEFFREY J. CROW, DEPUTY SECRETARY

TO: Lucy Barber, National Historical Publications and Records Commission
FROM: Kelly Eubank, Project Director, Electronic Records Archivist

SUBJECT: North Carolina, Kentucky, and Pennsylvania's Electronic Mail Preservation Collaboration Initiative, Grant Number NAR RE-05701-07

The Electronic Mail Preservation Collaboration Initiative aims to build a prototype tool for the collection and preservation of archival electronic mail. Staff from the states of North Carolina, Kentucky, and Pennsylvania have teamed together to leverage their knowledge and skill sets to help design and test a tool that will collect electronic mail, maintain a copy of the original email and the attachment, as well as transform the email itself from its native format to an XML format based on an XML schema. The team is also developing on-line training materials. In addition, once the tool has been designed and tested, we will deploy it to participating pilot partners for phase two testing and feedback. The project team includes: Kelly Eubank (NC), project director, Megan Durden (NC), Druscilla Simpson (NC), head, IT Branch; David Minor (NC), programmer; Ed Southern (NC), head, Government Records Branch; Mark Myers (KY), electronic records archivist; Glen McAninch (KY), head Technology Analysis and Support Branch; Linda Avetta (PA), information technology generalist administrator; Cynthia Bendroth (PA), head, Division of Records Administration and Image Services.

Objective 1: Information Technology Environments

1. Continue the development of the e-mail collection and preservation software to convert e-mail from its native format into the more stable XML format and complete written documentation for the program.

MAILING ADDRESS
4610 MAIL SERVICE CENTER
RALEIGH, NC 27699-4610

TELEPHONE 919-807-7280
FAX 919-733-8807

LOCATION
109 EAST JONES STREET
RALEIGH, NC 27601

Information Technology Environment:

Kelly, Megan, and Pam Ingle met with David Gilmore (IT) of the Secretary of State's office to discuss particulars of the software tool. The group took screen shots of how the system is set up on our end and how GroupWise might be configured. Pam talked with David Gilmore about the people who might participate in the Secretary of State's office. David agreed to provide the public IP addresses and run a test once the firewall issues had been resolved.

David Minor sent requests to DCR-IT to have firewall ports open for the Governor's Office and the Secretary of State. However, the access was not given until the end of April. When Kelly spoke with the Security Officer about this, he said that he was waiting to submit all of the firewall requests for DCR in one ticket. DCR is consolidating all of its areas behind the firewall. Once submitted, ITS had to approve the requests. ITS, in turn, had to have permission from the other state agencies to open the ports. Once Larry White and David Gilmore gave those permissions, there was a technical issue with the firewalls. Eventually, Arnez Taylor of DCR-IT took down the duplicate firewall down since everyone was now behind the ESAP. Megan worked this month to complete the training manuals for using EMCAP with Outlook 2007 and 2003. We could not get started working on GroupWise because access to the SOS VPN was not permitted. The Secretary of State's office's schedule also had not been signed. Kelly worked with Pam to move the schedule up on the priority list.

As a result of this meeting, the NC team submitted requests that several ports be open between the Archives and our partners. During this time, consolidation with central IT was still occurring. As a result, the requests for access to the firewall were delayed so that it could be submitted with a larger request of access needed throughout the Department. At the end of April, access and permissions were finally granted.

Training

During this time, Megan Durden started working on the training materials for EMCAP for users using MS Outlook, versions 2003 and 2007, as well as GroupWise. These materials were circulated amongst the group for feedback as well as users who were unfamiliar with the project. Once the review was complete, the documents were posted to the web. Additionally, the administrative console has a bulk add feature. Incorporated into this feature was the ability to send an email to the user explaining to them how to set up the password and activate the account. The URL of the webpage for training was also included so that users could access this information.

Programmer Update:

At the beginning of March, the project was without a programmer. The previous programmer, while suited for the initial stage, did not have the skills to build the second part of the tool. Kelly and David Minor cancelled the contract with the prior company and resubmitted the supplemental staffing contract. We received a number of highly qualified applicants. Fortunately, one applicant was both qualified and interested. The programmer began work on March 19 with the assignment to run through April. David Minor spent a large amount of time working on the administrative console and getting it ready to deploy to the partners. The second programmer worked on completing development work on the core processing engine and XML transformation. The XML transformation requires a parser to parse each element within the email and transform each element to XML. The central processing engine should be able to import and export .pst files (large ingest), record individual actions to the account, and compare messages in order to process only the changes made.

The administrative console was released to Kentucky and Pennsylvania in early March. Kentucky installed it but found a couple of bugs in the software. The programmer in NC provided support as well as a patch.

Pennsylvania and Kentucky:

Linda Avetta submitted the Technical Architecture Review (TAR) document to Pennsylvania's central OIT (Office of Information Technology) group. This document came up for discussion in a February meeting. As a result of that meeting, a conference call with Pennsylvania's OIT group was scheduled for February 29. Initially, this call was to determine the configuration used in KY since they are also on a MS Exchange mail system. However, the call resulted in discussion about KY's environment and revealed that the key player was the OIT/Exchange team member who was satisfied with the TAR documentation and that the IMAP connection was appropriate (he was not an original member of the reviewing committee). He suggested a new diagram be created and submitted with the TAR documentation and that he would first endorse it before being submitted to the TAR board. After little movement in the past several months, it appeared PA may finally be able to make progress. However, after the OIT/TAR Board approval, the OIT Legal staff got involved and had numerous issues with open source licensing. This was finally resolved and approval was given to proceed. The server was ordered by NC and received by PA. The PHMC IT person assigned to the server is not available at the moment and Linda and Cindy are waiting for him to supply a date that he will be available to work on it. Meanwhile Linda prepared the PA portion of the "There are More Tigers in the Ring than just the e-Tiger" for the Best Practices Exchange. Cindy Bendroth and Linda Avetta also reviewed and provided comments on the NC proposed training documentation.

Kentucky installed and tested the Administrative Console on its Hmailserver application, providing feedback to David Minor.

Objective 2: Potential Partners

2. Test the software to determine its scalability, efficiency, and employee compliance, using records retention schedules.

Potential Partners

Pennsylvania:

Pennsylvania would very much like to work with two entities for this project—the Governor's Office / Office of Correspondence and the Pennsylvania Historical and Museum Commission (PHMC) Executive Office, of which Barbara Franco is the executive director. The internal executives of PMHC have agreed to participate. At this time, the question to be resolved is the issue of using beta software that touches the production environment. A determination will need to be made later as to whether other agencies will be able to participate because of this issue.

Kentucky:

The staff in Kentucky has the email from the outgoing commissioner of the Kentucky Department of Libraries and Archives (KDLA) as well as the agreement of the current commissioner of the KDLA to participate. A gubernatorial election has been held and a new administration took office in December 2007. The outgoing Secretary of the Education Cabinet determined that the e-mail of the previous Cabinet Secretary contained no archival records. The Archives plan to work with the new administration to establish an archiving procedure for e-mail with Cabinet secretaries and perhaps

the new governor's constituent e-mail. As of the end of February, only preliminary meetings had occurred with new administration officials.

North Carolina: Secretary of State

The project was unable to proceed on the original timeline for deployment to the SOS office because of issues revolving around consolidation of IT in NC and delay in granting firewall requests.

Emergency Management

No activity to report.

Governor's Office— Community and Citizen Services

At the end of April, the NC staff worked with staff in the governor's Office of Community and Citizen Affairs to configure an EMCAP account. The initial installation was not a smooth one. Because consolidation was still occurring, the server had been moved and was down. The IT support here was on vacation and we had to wait for him to return to make sure the server was reset. The Governor's office removed the EMCAP account from their client software because they getting too many pop ups. Once the programmer returned from vacation, we were able to restore the account.

Outreach

Kelly and Megan attended the conference of the Society of North Carolina Archivists. Along with Ricc Ferrante of the Smithsonian Institution Archives and the group from the Rockefeller Archives Center, Kelly presented on the EMCAP tool.

Glen McAninch presented at the Midwest Archives Conference, April 17-19, 2008 in Louisville, KY, along with Ricc Ferrante of the Smithsonian Institution Archives and Mike Smorul of the University of Maryland.

Megan Durden, Kelly Eubank, Glen McAninch, and Linda Avetta all presented on the EMCAP tool and their experiences with the project at the Best Practices Exchange in Helena, MT, May 21-23.

Other News:

Kelly Eubank and David Minor spoke with Richard Marciano regarding the EMCAP tool in development. Richard worked on a much smaller project for the National Archives in 2000 and shared his experiences with us.

As a result of the presentation at the Midwest Archives Conference, Ricc Ferrante and Glen McAninch were approached by a number of people about the possibility of taking the XML schema for the email parser to the SAA standards board. Ricc and Kelly talked further about this and Ricc agreed to contact the SAA standards committee and determine the next steps to submit the schema for review.

KY—Update: Kentucky has placed links to drafts of e-mail training modules on the wiki site. The focus of two of the draft training modules was e-mail management from a records management perspective. These drafts used two separate types of training software. One employed Moodle, an open source course management software that contains course registration, web page presentations, glossaries, quizzes, a structured forum for discussion, and event calendaring. The other draft was created using SoftChalk, a commercial course editing tool that facilitates development of web based

presentations with hyperlinks, quizzes, and glossaries. SoftChalk can produce formatted course materials as a standalone set of web pages or import into course management systems such as Moodle, Blackboard, or Angel. These training tools could also be used to create companion training for H-mail server archival administration, as well as e-mail system specific training for end users to implement data exchange with the archival Hmail server system that the grant is building.

PA—Update: In the last update we noted “in addition to the TAR approval process moving quite slowly, another unfortunate turn of events for Pennsylvania includes a hold on the e-mail training that was to have been disseminated to all PA Commonwealth supervisor and management employees in December”, the training was released this spring to supervisors ONLY but not to management staff who are not supervisors. To ensure that our targeted h-Mail testing users would have the PA e-mail related training, we identified the executive staff members in PHMC who did not receive the training and requested that the Office of Administration add them to the training list. All staff identified received the training notice in May.

NC—In March, in response to newspaper stories, Governor Mike Easley appointed an Email Records Review Panel to review the current policies regarding email management and state employees’ responsibilities under the law published by the Department of Cultural Resources. Additionally, the NC Press Association filed a lawsuit declaring that the destruction/deletion of email constituted a violation of the NC GS 132, the Public Records Law. See <http://www.ncpress.com/ebackgroundsite/backgroundemailindex.html>

Staff from the Department of Cultural Resources attended the meetings of the panel and was asked to address the panel about email/records management, explanation of the EMCAP tool and email archiving. The Panel, in its final recommendations, called for the completion of the EMCAP tool and its roll-out on an enterprise wide basis. (See appendix A) As a result of this activity the team received additional financial aid to supplement the money provided by NHPRC to retain the programmer on staff for an additional 178 hours.

E-mail Archiving Solutions

Background

Governor Mike Easley charged the E-mail Records Review Panel with studying the issue of managing e-mail as a public record. The Panel recognizes the importance of a transparent government and the public's right to access public records, including e-mail used in the course of state business. State employees, through training offered by the North Carolina Department of Cultural Resources (DCR) and the publication *E-mail as a Public Record in North Carolina*, available on the DCR Web site, are directed to treat e-mail as they do paper documents or other records handled in the course of employment by the State and to do so in compliance with existing public records laws. The content of the information in the e-mail, rather than the e-mail format, determines how an employee will manage e-mail. Employees are directed to the *General Schedule for State Agency Records* as well as their records retention and disposition schedule, specific to each agency, to determine how to manage e-mail and how long to save it.

The Panel's goal is to ensure that all employees are sufficiently trained and have the tools to comply with existing public records law and other applicable laws, using the General Schedule and the program records retention and disposition schedule for their agency's records.

Recommendations

Technology can be used to augment the above guidelines through enhanced training resources; management of records; as well as collection and preservation of e-mail records of enduring, permanent value. The Panel proposes the following solution as the most advantageous response to the challenges of e-mail retention and preservation:

- State employees are directed to treat e-mail as they do paper documents or other records handled in the course of employment by the State and to do so in compliance with existing public records laws.
- Information Technology Services (ITS) will save back-up tapes of all Executive Branch e-mail a minimum of five years [projected costs for storage are \$75,000/year]. This will provide a longer window of opportunity to recover e-mails that may not previously have been saved.
- All executive branch agencies shall collaborate with the State Chief Information Officer (CIO) and DCR toward the goal of employing a software platform that complies with the following recommendations, including saving backup tapes for at least five years.
- DCR will make online training on how to handle e-mail as a public record available to all state, local, and municipal government employees, and any other interested parties. Whenever substantive changes in public records law occur, agency heads will require recertification of employees. Additionally, agency heads may require recertification as they deem appropriate. The DCR tutorial will be updated as necessary based upon changes in public records law.

- Through the competitive bidding process, the Office of the State CIO will procure an archive system and will work jointly and in collaboration with DCR to provide the archives/records management software package to be used by state agencies in the executive branch to archive and manage e-mail and electronic records made or received in the course of business.
- DCR shall conduct random audits of state agencies in the executive branch to ensure agencies are in compliance with their records retention and disposition schedule. In addition, DCR will conduct an annual review of and report on the number of requests made for e-mails on backup tapes and report its findings to the Governor and the State CIO. The reports should include information regarding the request, who made it and the cost incurred as a result of responding to each request. The annual reports will be used to evaluate the efficacy of the retention of backup tapes and may result in changes to the required five-year retention of backup tapes.

Additionally, the Panel recognizes the need to provide an e-mail archiving solution to agencies for e-mail records of enduring, permanent value. These e-mails document the activities of state government and are a historical record. We recommend the following:

- DCR will continue to manage and provide access to e-mail records that have been archived.
- ITS, working in close collaboration with DCR, will expedite the full development of an e-mail collection and preservation tool (EMCAP) that will allow employees with e-mail of enduring, permanent value to archive their agency's e-mail using their client software and in compliance with existing public records laws.

The Panel also recognizes the need to address evolving portable data devices and the use of personal computers and personal e-mail accounts to conduct state business. We recommend the following:

- State employees who use technology including, but not limited to, home computers, personal digital assistants, or other recording technologies in the course of state business, are required to manage such records in accordance with the *General Schedule for State Agency Records* and their agency's records retention and disposition schedule. Employees are directed to ensure that all such records are synchronized or otherwise copied to the appropriate state-assigned e-mail accounts or state computers, or that a paper copy is created. All such records must be properly archived.
- DCR will incorporate into the training recommended by this Panel information regarding compliance with public records laws and records retention and disposition schedules for records created, sent or received on mobile devices, in personal e-mail accounts or on personal computers.

Implementation

- The Panel recommends that all executive branch state agencies use an e-mail system employing a platform consistent with the archiving platform offered by ITS and which will save backup tapes of e-mail for five years.
- The State CIO has the authority and will work in collaboration with agencies toward the goal of consolidating e-mail services and utilization of one e-mail system for state agencies.
- ITS will require additional funding to begin immediately implementing an archives/records management software solution for executive agencies. DCR will require additional funding to support this initiative including the audit function.
- DCR and ITS will work cooperatively to make an online tutorial on e-mail management available to all state and local employees as well as interested parties. The training results will be tracked through a content management system (CMS), possibly linked to the North Carolina Identification Management (NCID) system currently in place at ITS. This will allow agency heads to audit employees' compliance regarding e-mail records. This training will instruct employees on their responsibilities with regard to public records laws and their records retention and disposition schedule.
- Employees will be able to access their retention schedules via the DCR Web site.
- Employees will still be required to manage their e-mail and determine its value in accordance with existing public records laws and using the *General Schedule for State Agency Records*, as well as their program-specific records retention and disposition schedule.
- Agency heads will require all state employees in their respective agencies who handle public records to take either the online training or the workshop offered by DCR. Whenever substantive changes in public records law occur, agency heads will require recertification of employees. Additionally, agency heads may require recertification as they deem appropriate.
- Employees will manage their records utilizing a records management software procured through the competitive bidding process. ITS will purchase and house the software and hardware. The State CIO's office will work jointly with DCR to implement an archives/records management software program.
- Funding for additional staff in DCR to support this initiative will be allocated.
- From their desktop computers, employees with e-mail records of enduring value will have the ability to archive their e-mail with DCR. ITS and DCR will

work together expeditiously to seamlessly integrate and expedite the DCR E-mail Collection and Preservation (EMCAP) tool with Microsoft Exchange to ensure that e-mail of enduring, permanent value can be collected and preserved in accordance with public records laws and the mission of the North Carolina State Archives.

Some Potential Advantages:

This recommendation:

- Is in compliance with the provisions of General Statutes Chapters 121 and 132;
- Provides state employees with additional training to ensure understanding of public records law, records retention and disposition schedules, and the responsibilities of state employees;
- Provides state employees in the executive branch with the tools to manage their records;
- Promotes good records management practices and efficient use of state government resources;
- Makes the employees responsible for managing their e-mail records in the same manner that they manage their records in all other formats;
- Allows a more manageable quantity of e-mail to be saved than a save-it-all scenario;
- Addresses concerns regarding the availability of records for public inspection.

Some Potential Advantages of emcap:

Technology changes rapidly and both media and file formats quickly become obsolete. As a result, great expense is incurred to access e-mails in formats that are no longer supported. The proposed e-mail preservation tool, EMCAP, moves e-mail messages out of proprietary software formats into a more stable, software independent format – Extensible Markup Language or XML – an industry standard that can be read by almost any software platform.

EMCAP:

- Allows users to choose which e-mails are to be archived in accordance with public records law and the agency's records retention and disposition schedule;
- Does not require special software to be purchased by agencies;
- Is a familiar tool because it mimics current practices of "dragging and dropping" e-mails into folders;
- Does not require special training;
- Can be incorporated into the existing training vehicles;

- Allows preserved e-mails to be searched and accessed;
- Captures all the information about an e-mail, including not only the typical data of “to” and “from” but also transmission data that details how and when e-mail was sent, valuable evidential and historical information.

Some Potential Disadvantages:

- ITS may be required to expeditiously migrate all state agencies to a new e-mail platform which requires adequate funding and staff;
 - This recommendation will require funding for additional technology and staff for ITS and DCR.
-

E-mail Retention Training Solution

Background

Current resources available through the North Carolina Department of Cultural Resources (DCR) to train state employees on how to handle public records include face-to-face workshops presented monthly at DCR; these workshops are also provided for agencies on-site by request by DCR's Government Records staff. In addition, the 2002 publication *E-mail as a Public Record in North Carolina: Guidelines for Its Retention and Disposition* is available on the DCR Web site. While these resources are helpful tools for employees seeking to strengthen their understanding of e-mail records retention and public records law, the Panel recognizes that training resources must be expanded and the obligations of state employees concerning e-mail (and all other public records) more clearly defined.

Recommendation

The Panel recommends that training for managing e-mail as a public record be made available online to all state, local, and municipal government employees and be required for state employees who handle public records. State agency heads will require training for all employees in their respective agencies who handle public records. Completion of the training, in the form of certification, will be noted in the employee's personnel file. Whenever substantive changes in public records law occur, agency heads will require recertification of employees. Additionally, agency heads may require recertification as they deem appropriate.

The tutorial will be posted on the DCR Web site and linked from the "Other Resources" section of Information Technology Services' (ITS) www.ncmail.net. Agency Chief Records Officers will be notified of the tutorial's publication and communicate its availability to their agencies. This training, based on current workshop materials and in the form of an online tutorial, will cover public records law, public records found in e-mail and how to retain and archive e-mails based on DCR's *General Schedule for State Agency Records*, as well as the program records retention and disposition schedule specific to each agency. The tutorial will be updated as necessary based upon changes in public records law.

Implementation

The tutorial will be created using an open-source content management system (CMS) that allows comprehensive reporting and tracking of users and completion rates. DCR will work closely with ITS to launch a state-wide training initiative. ITS will implement and manage the CMS that will host content developed by DCR. The appearance of the tutorial will closely resemble a Web site. Brief quizzes will be included throughout the tutorial to assist the individual and the results can be recorded as described above.

Some Potential Advantages

- The CMS allows for a variety of reports to identify employees who have and have not completed the tutorial.
- The CMS allows the log-in to be linked to a pre-existing log-in (such as the North Carolina Identity Management NCID), and groups of users can be defined, allowing existing groups of employees different levels of access to the content and reporting tools.

- Confirmation of completion can be easily disseminated to each employee's personnel file.
- The tutorial can be completed in multiple sessions based on the employee's schedule and preference, rather than one thirty-minute session.
- Should future training on other topics be recommended for government employees, the content could easily be incorporated in the CMS.

Some Potential Disadvantages

- The CMS requires additional financial and human resources to provide the necessary set-up and additional programming.
-