

Project Narrative

Over the past five years, North Carolina, Kentucky , and Pennsylvania have worked diligently to address the challenges posed by electronic mail. Each state has worked with a variety of groups to establish best practices and to bring to the forefront the importance of recordkeeping, standards and guidelines. Each state has developed guidelines concerning management of electronic mail and each state has reached out to different constituencies to reinforce the need for ongoing efforts to manage and preserve electronic mail.

Kentucky

Through a multi-agency Electronic Records Working Group, the Kentucky Department for Libraries and Archives (KDLA) has worked since 2002 to establish recordkeeping guidelines, procedures and standards for managing electronic records. *While KDLA has supported the federal DoD/NARA 5015.2 compliance as the standard for recordkeeping systems through the state's Enterprise IT Architecture and Standards, the absence of compliant systems in Kentucky state government necessitates the use of MS Outlook/Exchange alone as the means of managing e-mail.* KDLA's e-mail guidelines (<http://kdla.ky.gov/recmanagement/E-mailGuidelines.pdf>) describe the retention periods for various categories of e-mail, and users are urged to establish filing procedures within Microsoft Outlook for separating records by retention. With the means to migrate e-mail from Microsoft Exchange/Outlook systems to an XML open source in the hMailServer platform, KDLA will have a much better chance of maintaining records permanently and providing access to them.

In 2003, KDLA accessioned constituent e-mail from the Office of the Governor (2000-2003). The records, which were transferred via 30 CDs, were created from scans of incoming correspondence and extracted text from e-mails created on a Microsoft Exchange server. The master index to the 58,153 items of correspondence (transferred from Microsoft SQL Server to Microsoft Access as each disk was created for offline storage during the administration) served as a mail log to all correspondence. Scanned correspondence and the e-mail text were integrated into this system. The programming to create this system was in Visual Basic using LeadTools software, and it is stored in a proprietary format that is difficult to export. The tracking database, which served as the Governor's mail log, contains a fixed group of subjects or categories that were created and used by the Governor's staff. It also contains fields for routing information (from and to names, actions and dates), comments, status, log date, close date, U.S. Mail addresses, and e-mail addresses.

Pennsylvania

The Pennsylvania Historical and Museum Commission (PHMC) Bureau of Archives and History (BAH) has worked to establish recordkeeping guidelines, procedures and standards for electronic records, including records created via e-mail for the past several years.

Pennsylvania's policies and procedures are available at the Office of Administration's website at www.oa.state.pa.us in the form of Management Directives (MD) and Information Technology Bulletins (ITB). Guidelines and training specific to e-mail were recently removed from the MD arena and are in the process of being revised in the format of ITB's.

For the past three years, the BAH has offered classroom training for Commonwealth of PA employees on the management of e-mail. The course was based on the then applicable MD

210.13, “Retention and Disposition of Records Created on Electronic Mail (E-mail) Systems”, submitted separately (MD rescinded 5/2006; to be replaced by ITB). A web-based e-mail training course for supervisors, an 18 minute scaled-down version of the 2-hour classroom training, was also created and made available in 2005. The web-based e-mail training for supervisors was mandatory for all Commonwealth of PA managers and supervisors, and all were required to complete the training by February 28, 2006.

Throughout the e-mail management training, employees are instructed to be familiar with their agency specific, and general, records retention and disposition schedules. E-mail is a format, not a record type. All e-mail messages must be evaluated based on their content and those identified as a record must be retained according to the established records retention and disposition schedules. Employees are responsible for identifying e-mail messages as a record and are responsible for systematically naming and filing those messages for convenient retrieval for the duration of the retention period. For short-term messages, saving them in the native e-mail format (MS Exchange/Outlook) is sufficient. For long-term messages, they must be saved in either portable document format (.pdf, whereby they may be transferred to pdf-a in the future) or text, preferably rich text format (.rtf). Regardless of length of retention, all e-mail records must be saved on network drives (not retained on the MS Exchange Server) where they are routinely backed up according to network backup policy and guidelines.

Currently, Commonwealth employees have e-mail records stored on multiple agency servers and in multiple file formats (.msg, .rtf, .pdf, .txt). The BAH currently has no means of acquiring, storing and making accessible permanently valuable electronic records in varying media and formats, including e-mail records. With the means to migrate e-mail from Microsoft

Exchange/Outlook systems to an XML open source in the Hmailserver platform, the BAH will have a much better chance of maintaining records permanently and providing access to them.

North Carolina

Since 2002, the North Carolina State Archives (NCSA) has focused its efforts on trying to influence and educate state agency and local government employees about e-mail, its retention and its preservation. We developed a set of guidelines:

http://www.ah.dcr.state.nc.us/records/e_records/default.htm#e-mail

and have been an active participant in the Statewide Information Technology Services development of an enterprise-wide e-mail system. As a part of this involvement, we have engaged them in issues regarding management of e-mail from a recordkeeping standpoint and best practices for managing e-mail. While the system utilized at the Information Technology Services Department (ITS), which administers a majority of the state government e-mail, is ASCII text, many state employees use Microsoft Outlook for their client software.

In 2005, in an effort to address the challenges posed by this propriety software, the NCSA completed an NHPRC Electronic Records Fellowship grant administered by UNC-Chapel Hill and Duke University. Druscie Simpson, recipient of the Fellowship, researched preserving and providing access to electronic mail and its impact in terms of modern day recordkeeping and archiving practices. Often this correspondence is lost through lack of training, lack of tools or a change in technology. As a result of this project, the NCSA developed a software application to capture electronic mail and transform it from its native format into an XML "preservation copy." The program was based upon Microsoft Office Outlook technology but is also compatible with Microsoft Exchange e-mail applications.

The software is built on hMailServer, an open- source, free software registered on SourceForge. It supports open e-mail standards POP3, SMTP, and IMAP4. It supports both MySQL and Microsoft SQL Server. It works with WebMail as well, through a version called SquirrelMail. It mimics the user client software for e-mail. To the user, the "Archive" folder or directory and all files under it appear in the file structure of the client's e-mail account like any of the user's other files. When the user decides that the e-mail is a record, he or she drags and drops the e-mail to the archive folder or one of the folders underneath it. At a designated point in time (e.g.. the end of the retention period for the appropriate series for a folder or sub-folder), the Archives initiates a process to "grab" the folders and bring the mail back to a designated server for the Archives' use. Once transferred to the hMailServer, another program transforms the files from their native format into an XML format. Although the project team identified methodologies to handle the attachments, the project ended before North Carolina could address issues concerning e-mail attachments. In the current system, the attachment is saved in its native format and wrapped in an XML wrapper with the e-mail.

1. Purpose and Goals of Project

The North Carolina State Archives, the Kentucky State Library and Archives, and the Pennsylvania State Archives propose to work in collaboration to test the e-mail preservation software, built using the hMailServer software, in real time on a larger scale with identified state agencies that produce archival electronic correspondence. For this grant, we propose to work with identified high level offices that produce archival correspondence and have high public interest within each participating state. One possibility is the Office of the Governor within each state. Additionally, we propose to conduct a select number of test cases within our own

departments. This partnership will involve working with content producers, Information Technology staff and records officers in the agencies.

The first goal of this project is to provide a tool to archive e-mail easily and test the software to determine its scalability, efficiency, and employee compliance. We propose to test the compliance with records retention schedules utilizing this tool against the more conventional methodologies currently in place, e.g. transfer via CD-ROM or DVD disc. We will analyze, what, if any, material this methodology captures compared to current methodologies. We would like to develop a seamless process by which agencies can file permanently valuable archival records from their Microsoft Exchange server to each respective institution's archive server.

The second goal of the project is to produce different types of training methodologies. Pennsylvania has already developed some training resources relating to the management of government e-mail and we would like to apply these methods and develop them further for the hMailServer application.

The third goal for the grant is to continue the development of the hMailServer software to convert attachments from their native formats into the more stable XML format for preservation purposes. We propose to work in collaboration to expand the existing product to include managing build upon the existing product to expand it to include managing the attachments, so that the context of messages is preserved as well as the content. Additionally, we need to write the documentation for these programs so that other institutions that want to utilize this software are able to do so.

The fourth goal is to test importing into the hMailServer both e-mail and attachments in various standard and non-standard attachments.”

The final goal will be for each of the states to test ways of providing access to these XML files and the feasibility of doing so, whether through their existing online catalog, a third party vendor, through a web interface, or, perhaps, through the Persistent Archives Testbed (PAT), which is grid technology currently being tested with several state partners, the San Diego Supercomputer Project and the National Archives and Records Administration.

2. Significance and Relationship to NHPRC Goals and Objectives

NHPRC supports projects to research and develop means to preserve authentic electronic records, to assist archivists through a network of state partners and to improve professional education for archivists and historical documentary editors. This project speaks to those goals by automating a process of normalizing e-mail file formats into a more stable XML file. Currently, XML is the preferred format for information of enduring value. Through this partnership, North Carolina, Kentucky, and Pennsylvania can test the robustness and feasibility of the system in a variety of settings. Additionally, we can use this opportunity to educate employees about records management practices so that information does survive. Thus, another deliverable of this project will be to disseminate a portable training component (e.g. web based training). This component could be altered to fit another institution's practices and serve to enhance those efforts.

This grant would give the staff in each state the opportunity to improve their educational base of archival knowledge and test how that base can be adjusted to fit to test new technologies and methodologies. Should we be successful both in training staff and implementation of the software, we will continue to meet our statutory requirements to preserve state government information of enduring value, as well as improve the skill sets of the staff as well as e-mail users to meet the challenges that electronic records present. We can build upon existing

relationships within our state agencies and between sister states to promote the NHPRC's goals of preserving and making accessible records and archives.

Perhaps most importantly, this would offer a low cost solution to archives or memory institutions that struggle with how to continue to document American history in the digital era. For institutions with limited funding, the software would be available and give them a tool at relatively low cost (free except for storage costs) to capture important records.

3. Plan of Work for Grant period (including references to techniques)

The purpose of each institution's participation in the grant project is to establish an e-mail archive (using hMailServer) that contains "archive ready" state government e-mail records and to establish a seamless process by which agencies can file permanent records from their e-mail server to the archive server.

Timeline of Workplan:

First Quarter 2007

January-March 2007:

- Project staff will convene to coordinate work assignments, create deadlines, establish benchmarks, compile lists of needs, etc. Since each partner is at different stages regarding the preservation of e-mail, work will vary from state to state.
- Programming staff at the North Carolina State Archives will begin reviewing work done with hMailServer thus far and continue to finalize the application, documentation, and develop ways to manage attachments.

- Project staff will finalize Memorandums of Understanding with the government offices and technical support staff that they will be working with within each state.
- Project Staff in Kentucky will test exporting images and text from the Microsoft Access database and import them into hMailServer running a Microsoft SQL server. While KDLA will seek Kentucky state government expertise in extracting records from the system, the import to hMailServer may require the help of the programmer hired through the grant.
- Test export of archival e-mail from Exchange web server used by KDLA to the hMailServer installed for the grant. This work will continue throughout the rest of the year.
 - Develop and test the following OAIS archival functions within hMailServer: identification of and security for restricted records using pattern recognition or other techniques,
 - authentication of records during the acquisition process,
 - bulk loading of common descriptive elements in the hMailServer database; and
 - providing search and retrieval mechanisms through the database and/or through XML embedded structure.

Second Quarter 2007

April-June:

- NCSA's Information Technology Branch will assist project staff and archives staff in implementing solutions and enhancing the functionality of the software. They will provide services regarding set up and maintenance of the program, back ups, etc.

- Project staff will test the software prior to roll-out to the offices. This will allow the staff to identify “bugs” in the system as well as offer feedback on the documentation.

Additionally, this allows staff to become familiar with the software and to test for any outstanding issues that foster professional development as they become aware of issues concerning e-mail and provide feedback to the programmer. Often, these two groups speak different languages and this feedback is invaluable to the program's success and employee compliance.
- Staff will begin crafting training materials on the product, utilizing other staff feedback as to the clarity and simplicity of the training.
- Late in this quarter, we anticipate a rollout of the application installed at the appropriate locations (individual desktops of top administrators or the office’s e-mail server, depending upon the environment of each participating office). Preliminary training will be provided.
- NCSA IT staff will set up and activate a web-based system to track feedback from participants.
- Staff will prepare a mid-point report to NHPRC regarding the progress of the grant to this point. The report will state the progress as well as identify unanticipated issues and proposed solutions.
- Research and identify solutions to alternative collection mechanisms in addition to hmail.
- Through quarter three, collect e-mail directly from the following message stores:
 - Users’ Outlook local .pst file.
 - Users’ Netscape profile.
 - Mail stored on a MS Exchange server for a particular account.

- Mail stored on a MS Exchange server for a set of accounts.

Third Quarter 2007

July-September:

- Staff will continue to provide assistance and training to the participating agencies, utilizing the web based tracking system as well as telephone, in person and e-mail communications.
- Staff will continue to test the system internally and make recommendations regarding changes.
- Staff will build upon the training materials developed in the second quarter, develop web-based training, to build a complete complement of training materials.
- Project staff in the three participating states will communicate monthly via conference call to track progress.

Fourth Quarter 2007

October-December 2007

- Staff will finalize training materials.
- Staff in NC will analyze material collected using the application software versus material collected via DVD disc to determine if the material differs.
- Staff will begin writing articles on the project for publications such as RLG documenting the findings and analyzing the feasibility of the system.

Performance Objectives:

1. Fully develop an electronic mail preservation software application. This application will address both the text-based communications as well as the attachments of e-mail.

2. Software will be based on XML, a standard mark up language, currently identified nationally as a stable programming language for future access.
3. Fully develop documentation and training tools to be available in conjunction with the software for other memory institutions to utilize.
4. Test the software and other materials in “real-time” to determine the efficacy of use of the product versus more traditional methods.
5. At the completion of the grant, each institution will have valuable, archival electronic correspondence that will be in a stable file format and will be searchable and accessible via each institution’s on-line finding aid.
6. Promote professional development among staff regarding the complex issues that electronic mail presents. Staff will also be able to discuss at a high level identified solutions.
7. For North Carolina, depending on the feedback, we would like to explore partnering with the statewide central Information Technology Services (ITS) to put this into production on a statewide level to capture identified information that we currently receive only sporadically.
8. We would like to develop a seamless process by which agencies can file permanently valuable archival records from their Microsoft Exchange server to each respective institution’s archive server.

5. Key Personnel

Project Director: Kelly Eubank
Telephone: (919-807-7355)
Address: 4615 Mail Service Center
Raleigh, NC 27699-4615
Kelly.eubank@ncmail.net

Druscilla Simpson
(919) 807-7319
4610 Mail Service Center
Raleigh, NC 27699-4610
druscie.simpson@ncmail.net

Linda Avetta
Penn. Historical and Museum Commission
Bureau of Archives & History
State Records Center Building
1825 Stanley Drive
Harrisburg, PA 17103
(717)705-6923
lavetta@state.pa.us

Cynthia Bendroth
Penn. Historical and Museum Commission
Bureau of Archives & History
State Records Center Building
1825 Stanley Drive
Harrisburg, PA 17103
(717)705-6923
cbendroth@state.pa.us

Glen McAninch
Kentucky Dept. of Libraries and Archives
300 Coffee Tree Road
Frankfurt, KY 40601
(502) 564-8300
Glen.McAninch@ky.gov

Mark Meyers
Kentucky Dept. of Libraries and Archives
300 Coffee Tree Road
Frankfurt, KY 40601
(502) 564-8300
Mark.Meyers@ky.gov