



North Carolina Department of Cultural Resources

Michael F. Easley, Governor  
Lisbeth C. Evans, Secretary

Office of Archives and History  
Jeffrey J. Crow, Deputy Secretary

September 14, 2006

Lucy Barber  
National Historical Publications and Records Commission  
National Archives and Records Administration  
700 Pennsylvania Avenue, NW  
Washington, D.C. 20408-001

Dear Lucy:

Thank you so much for peer review comments you sent to me on August 17, 2006. We are very excited to receive such constructive feedback and to have the opportunity to address the concerns regarding the Electronic Mail Preservation Collaboration Initiative grant submitted by North Carolina, Pennsylvania, and Kentucky on June 1, 2006.

Enclosed you will find several pieces of supporting documentation, including an overview of electronic mail, how it works, and how we propose this project to work within these existing protocols. Also included is the final report from Druscilla Simpson, a 2004-2005 NHPRC Electronic Records Fellow, regarding her work on this project, and a position description for the programmer we have requested in the grant application. Finally, we submit a revised budget proposal, a budget summary, and a narrative summary for the questions you put forth.

Based on the peer review comments and after discussion among the partners, the revised budget is different in four significant areas: (1) we have extended the time period of the grant from twelve months to eighteen months, (2) we have requested funds for a programmer for 12 months rather than six months, (3) we have dropped our funding request for an Archivist III from the grant, and (4) we dropped the request for funding of the indirect costs at 15.5% to a cap of \$5,000. Regarding point 3, the partners in Pennsylvania have offered to take the lead on preparing the training materials with North Carolina's our collaboration. Although we are requesting a longer time period and funding for a programmer for a year, the overall funding request to NHPRC did not change significantly.

We see the development and testing of this software as a tool that we can offer to the constituents of our respective states to more fully automate the transfer of electronic mail and the records it contains. Moreover, at the conclusion of this project, we hope to offer this tool to other entities who are struggling with capturing these vitally important records.

We thank the Commission for considering our request and we are happy to answer further questions. Again, thank you for the opportunity to discuss this exciting project with you.

Sincerely,

Electronic Records Archivist  
North Carolina State Archives

Jesse R. Lankford,  
State Archivist and Records Administration

MAILING ADDRESS  
4615 Mail Service Center  
Raleigh, NC 27699-4615

PHONE NUMBERS  
Telephone 919-807-7355  
Fax 919-715-3627

LOCATION  
109 East Jones  
Raleigh, NC 27601

## Response to Reviewers' Project Summary

1. Can you provide a quick summary about the extent to which this project reflects existing practices or guidelines in the three states and in other regulatory environments? If they do not reflect those standards, is one possible outcome of the project a proposal to revise standards?

Over the past five years, North Carolina, Kentucky, and Pennsylvania have worked to address the challenges posed by electronic mail. Each state has worked with a variety of groups to establish best practices and to develop guidelines. Additionally, they have reached out to different constituencies to reinforce the need for ongoing efforts to manage and preserve electronic mail. However, they have used different software clients to manage e-mail in the absence of a federally compliant DoD/NARA 5015.2 system. At this time, none of the states has deployed an Electronic Content Management system or a Records Management software solution.

Each state has written standards and guidelines regarding the management and retention of electronic mail. At this writing, North Carolina and Pennsylvania are in the process of revisiting and updating those guidelines. Current Enterprise Content Management (ECM) systems provide tools to declare e-mail a "record" but none provide a neutral file format for its long-term retention. IT administrators are still responsible for the migration of the messages at the appropriate time to a new file format in order to make them accessible in the future. With regard to the capture of e-mails, the methodology for each state is different. North Carolina accepts e-mail in its native format but must do so via transfer on CD-ROM. With our last transfer from the Office of the Governor, Office of Civic and Community Affairs, staff was unable to open the files on the CD because of incompatibility. The creating agency had upgraded to Windows Office 2003, while the North Carolina State Archives (NCSA) is still using Office 2000. Even after upgrading our system, NCSA was unable to open the files because of denied permissions. Relationships with IT personnel in these offices are crucial, as these relationships aid in the transfer of material. As IT personnel change, the process must start over. Employees who create or receive the e-mail do not have the skills to download the files for transfer to the NCSA and must rely on their IT personnel to do this operation.

The protocol in Pennsylvania is for records in e-mail to be transformed or migrated out of their native format into a .pdf file. At this point in time, .pdf documents are a generally accepted file format for long-term access. However, they are difficult to search and require the employee to go through a relatively complicated series of steps. Additionally, they must then be transferred, requiring the help of IT.

In Kentucky, the Office of the Governor transformed the e-mail into ACSII text and scanned the incoming correspondence before transferring it to the archives. These text and scanned files were imbedded into a database containing correspondence metadata, like a customized document management system. Thus, Kentucky has these transformed copies but none of the native e-mails. As with North Carolina and Pennsylvania, these processes rely on the aid of IT personnel to do "something" to the e-mail before it can be transferred or effectively preserved in an archival setting.

This proposed project seeks to automate, as much as possible, a process of capturing and normalizing e-mail file formats into a more stable XML file, the preferred preservation format for information of enduring value. Through this proposed partnership, North Carolina, Kentucky, and Pennsylvania can test the robustness and feasibility of the system in a variety of settings. Additionally, we can also use this opportunity to educate employees about records management practices, so that pertinent information contained in e-mail does survive.

While not denying the importance of contact and collaboration with the target agency, North Carolina, Kentucky, and Pennsylvania would like to create a tool that will lessen somewhat the burden on the IT staff and make the collection of e-mail easier for state employees to comply with and easier for state

archives to collect. The software mimics their current client software with a file on the desktop, into which the user can drop and drag the e-mail. While this process will require some IT support, it basically requires setting up the machine rather than transforming files.

2. Have any of the states developed estimates about the number of emails that are likely to be considered records? Do they have estimates about the amount of storage space that will be needed over time?

North Carolina, Pennsylvania, and Kentucky use retention schedules established with agencies to determine when the designated records are transferred to the archives. In some cases, the disposition schedule calls for them to transfer after a specific number of years. In other cases, the event that triggers the transfer is “when administrative value ends.” In 2001, the North Carolina State Archives received six gigabytes of e-mail from the former governor’s second administration. The material contained both official and unofficial business; some correspondence contained official and unofficial matter in the same e-mail. The governor’s archivist, with the help of the electronic records archivist, appraised the mail message by message and reduced the amount from six to two gigabytes. We used this transfer as a training opportunity and presented our records management workshops to staff in the governor’s office regarding what is a record and what is not a record. We have continued to work closely with this office to transfer email and do so every six months. To date, we have accessioned 6.7 gigabytes email correspondence from the Office of Citizen and Community Services in the Governor’s Office. This tool will help with that transfer process, and we hope that it will remedy the issues we have had with the inability to read CD-ROMs. This office uses the Microsoft suite of products that NCSA also uses.

The North Carolina State Archives has also received e-mail from the former superintendent of public instruction at the end of his term. This office uses Groupwise software, which does not have an archiving feature. Recently, the Wake County public school system, which also uses Groupwise, spent \$30,000 to hire a programmer to write code to access e-mail in trying to answer a public request for e-mail pertaining to redistricting. We have tested the EMCAP (an acronym for E-mail Collection and Preservation) software with this Groupwise e-mail and found that it is compatible.

One of the reviewers mentioned the work being done by the Digital Preservation Testbed ([www.ditiodurrzaamheid.nl](http://www.ditiodurrzaamheid.nl)) by the Netherlands and how this may differ. The Netherlands primarily developed a tool that works with Microsoft Outlook. While it is true that Microsoft is the industry leader and all three states are Microsoft Standard states, we do receive e-mail produced by different e-mail systems. We hope this proposed tool will offer greater flexibility regarding e-mail systems.

With their last transfer of materials, Kentucky received 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). This index served as a mail log to all correspondence. The correspondence was then scanned and both the scanned images 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.

3. Can you provide some additional information about the types of institutions using the hMailServer program and whether their experiences indicates that this solution will be scalable to fit the expected number of emails that will need to be stored as records?

hMailServer is a popular, open-source product that is well tested with corporate support. hMailServer is extensible horizontally. In other words, when you need more storage, you simply add an additional server. There is a large and active user community that is continually exercising the product, reporting bugs, documenting solutions, and making feature requests. hMailServer is a favorite among web-hosting providers, and commercial support is available. The source code is available and can be modified. If we find some otherwise insurmountable problem, we can always modify the existing code. Additionally, hMailServer is free. Anyone wanting to test our archiving solution will not be required to purchase expensive e-mail server software. Some potential users of this e-mail archiving solution may be cost-conscious and want to run the solution as configured. Others may want to purchase commercial support for hMailServer, and still others may want to take our solution and adapt it to an enterprise-ready platform, such as MS Exchange Server or SendMail. Starting with a free solution leaves the widest range of choices open.

4. Can you elaborate on the records held by Kentucky to explain whether this will be a test of a different email format than Microsoft Outlook? Do you have any information on the formats of email records generated in your state government agencies?

When Kentucky transferred the e-mail records from the previous governor's administration, the records, which were transferred via thirty CDs, were created from scans of incoming correspondence and extracted text from e-mails created on a Microsoft Exchange server. We should add that the scanned files are in Tagged Image File Format (TIFF) and the text files are in ASCII format. The header metadata for both of these files is stored in fields within the Access database. During the grant period, Kentucky will work to transform these files, as much as possible, to the XML format used by hMailServer in order to integrate the previously accessioned correspondence with subsequent accessions from Microsoft Outlook, the state e-mail standard.

Both Kentucky and Pennsylvania have existing protocols that require e-mail messages to be transformed out of their existing formats into an XML file structure (Kentucky) or into .pdf files (Pennsylvania). Depending on the outcome of this grant, this tool may give these states options as to how they can accept e-mails and whether to revise their practices. At the very least, this project will help us determine the feasibility of retaining the native files, thus being compliant with existing industry and archival best practices.

Regarding file formats, in North Carolina we have little knowledge of the software systems being used in state agencies, but this grant will provide an opportunity to survey state government offices for feedback about these systems. We will explore this during the course of the grant.

In constructing its data archive, Kentucky is developing a grid that ranks formats based on long-term preservation characteristics. Attachments received in the course of the grant will be evaluated in terms of this format matrix.

Thank you, again, for your thoughtful and provoking questions and comments. We welcome the opportunity to discuss this further and are excited by the possibility of building on the work already begun in all three states regarding e-mail. We truly feel this project, if funded, will offer tremendous opportunities and learning experiences to the archival community. We look forward to hearing from you.