

Chasing the E-Tiger

Electronic Mail Capture and Preservation tool - EMCAP

North Carolina, Kentucky, and
Pennsylvania grant initiative



Grant Web Site

<http://www.ncpress.com/ebackgroundsite/backgroundemailindex.html>



North Carolina State Archives

Preservation of Electronic Mail Collaboration Initiative

[About / Team Members / Project Documentation / Agendas and Minutes / Reports / Online Training / Timeline / Other Email Preservation Sites](#)

| Links |
|--|
| About |
| Team Members |
| Project Documentation |
| Meeting Agendas and Minutes |
| Reports |
| Online Training |
| Timeline |
| Other Email Preservation Sites |

| Other Sites |
|---|
| MARS Online Catalog |
| Friends of the Archives |
| North Carolina Office of Archives and History |
| Site Map |
| State Historical Records Advisory Board (SHRAB) |

Summary of the NHPRC Funded Preservation of Electronic Mail Collaboration Initiative



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 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.

In 2004-2005, a NHPRC fellowship awarded to Druscilla Simpson of the North Carolina State Archives provided the genesis for an electronic mail preservation solution. Fellowship funds were used to develop a software application to capture electronic mail and transform it from its native format into an XML "preservation copy," and to push the XML out to HTML for access and viewing purposes. The software is built on IMailServer, an open-source, free software registered on SourceForge. It supports open e-mail standards POP3, SMTP, and IMAP4 and supports both MySQL and Microsoft SQL Server.

The North Carolina State Archives, the Kentucky Department of Library and Archives, and the Pennsylvania State Archives are working in collaboration to test the e-mail preservation software in real time on a larger scale. For this grant, we will work with identified high level offices that produce archival correspondence and have high public interest within each participating state e.g. a constitutional office such as the Governor or the Secretary of State within each state. Additionally, we will conduct a select number of test cases within our own departments. This partnership involves working with content producers, information technology support staff, records analysts, and employees in the agencies.

There are five main goals of the project:

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.
2. Test the software to determine its scalability, efficiency, and employee compliance, using records retention schedules.
3. Produce different types of training methodologies, including web-based training, implementing work under development by Pennsylvania.
4. Test ways of providing access to these XML files and the feasibility of doing so, whether through existing online catalogs, a third party vendor, or web interfaces.
5. Extract and save attachments in their original, as received, format as a native stand-alone file, which will be wrapped in XML. The association between message and attachment will be kept and will allow navigation from message to attachment back to message.

NHPRC
National Historical Publications and Records Commission

Last Modified: 02/18/2008

[Return to top]

Home | Site Map | Department of Cultural Resources | © 2005 North Carolina Office of Archives and History.

Email Technical and Record Scheduling Environment

- Participants—NC, PA, KY
- Diverse systems—centralized versus decentralized
- All have retention schedules and training
- Management of e-mail studied and much debated

Transfer Current State of Affairs

● Manual Collections of Email

- Sneakernet, CD-ROM

● Labor Intensive

- No way to stop duplicates, have to run hashing algorithms to determine exact copies
- Duplicate messages
- Not receiving email from offices
- Different file formats

● KY & PA—MS Outlook pst files

Proposed Methodology

Build a tool to allow people to “archive” their mail on a systematic basis.

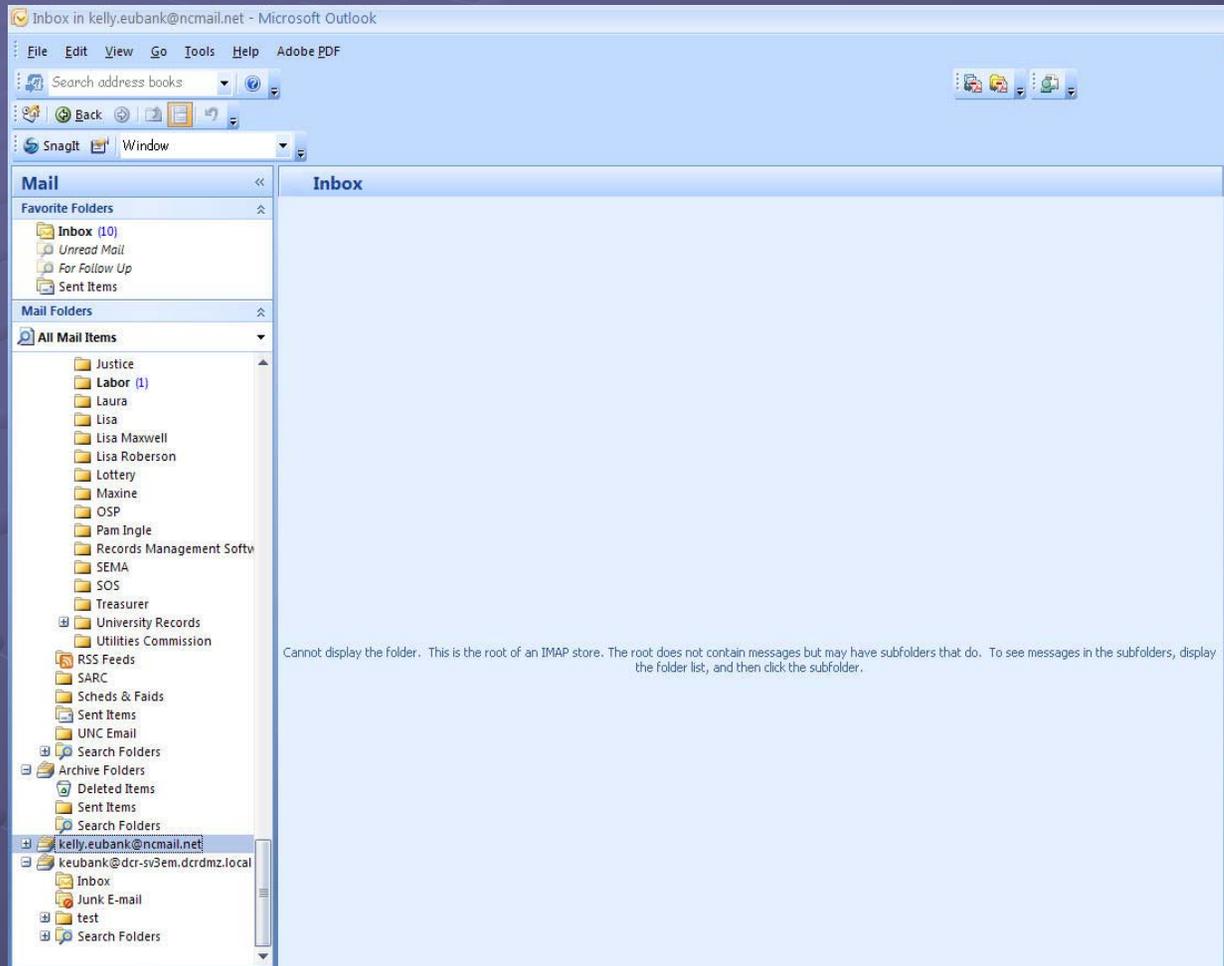
Minimize IT support

Be able to support multiple file formats

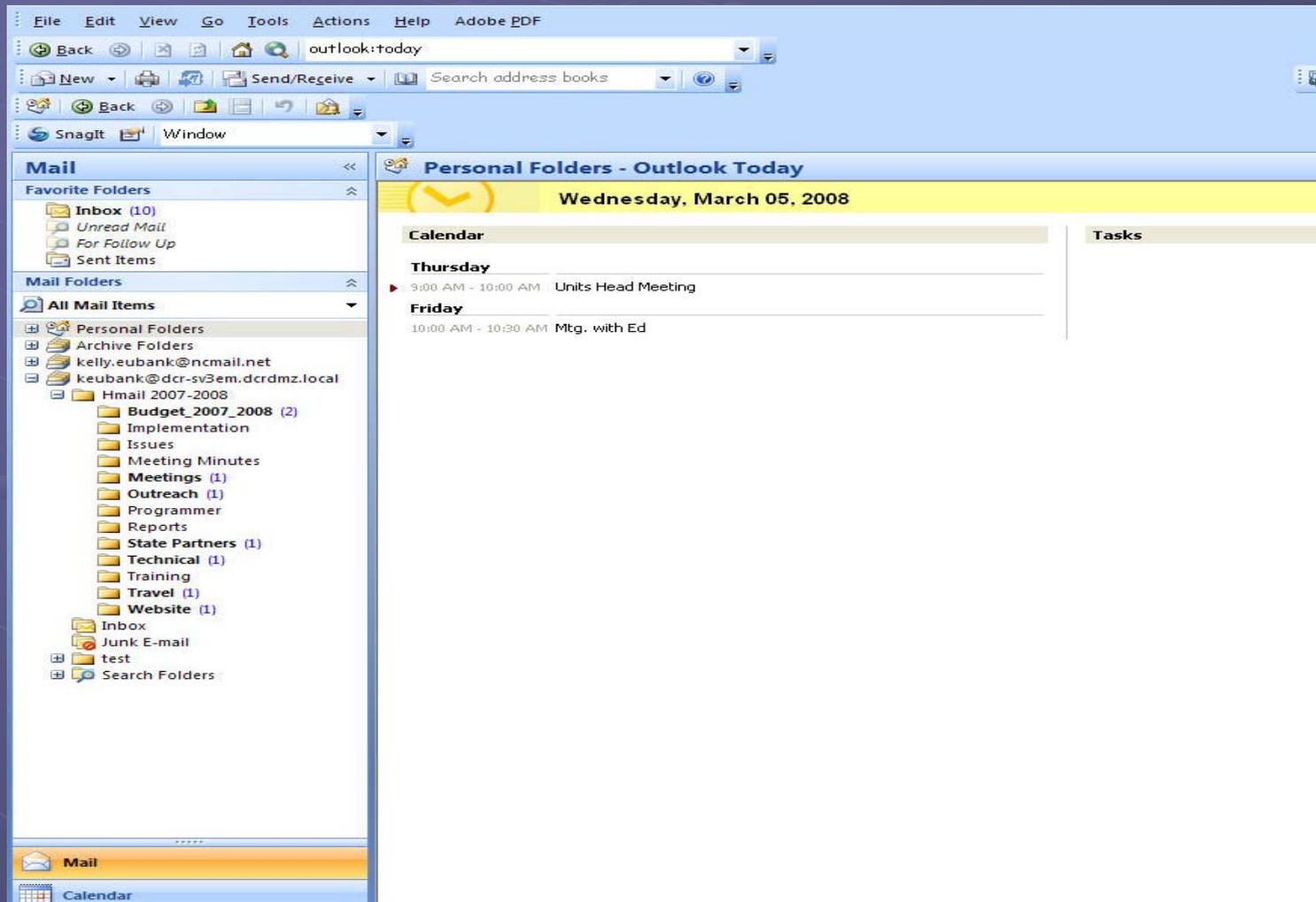
EMCAP Tool

- Open Source
- Client is configured to have Archives Folder that is mapped to our server.
- User can replicate file folder structure
- Mimics current behaviors of drag and drop
- Drops email into our collection server
- Internet message parser makes an XML copy

EMCAP User Client View



E-mail moved to hMailServer



Administrative Console

- Allows for creation of account
- Administrator can change password
- Has fields to add free text

Administrative Console Display with two test accounts

The screenshot displays the EMCAP Admin Console interface. At the top, there is a menu bar with 'File', 'Edit', 'Tools', and 'Security'. Below the menu bar, a 'Domain' dropdown is set to 'Internal Test', and a 'Status' section has three checked options: 'Read/Write', 'Read Only', and 'Off Line'. The main area contains a table with the following columns: Person Name, Address, Unit Name, Contact Email, Status, WillProcess, Creation Date, Message Count, Last Processed, Processed Last Run, and Earliest Message. Three rows are visible, with the second row highlighted. At the bottom of the console, there are buttons for 'Add', 'Edit', 'Delete', 'Take Off-Line', 'Close', and 'Start Processor'. A status bar at the very bottom shows 'User: DCRDMZ\keubank' and 'Last Visited: 2/20/2008 3:42:04 PM'.

| Person Name | Address | Unit Name | Contact Email | Status | WillProcess | Creation Date | Message Count | Last Processed | Processed Last Run | Earliest Message |
|--------------|--------------|-----------|-------------------------|------------|-------------------------------------|-------------------|---------------|----------------|--------------------|------------------|
| David Minor | david.minor | DCR-IT | david.minor@ncmail.net | Read/Write | <input checked="" type="checkbox"/> | 2/19/2008 1:51 PM | 0 | | 0 | |
| Kelly Eubank | kelly.eubank | Records | kelly.eubank@ncmail.net | Read/Write | <input checked="" type="checkbox"/> | 2/20/2008 3:29 PM | 0 | | 0 | |
| Megan Durden | megan.durden | GRB | megan.durden@ncmail.net | Read/Write | <input checked="" type="checkbox"/> | 2/20/2008 3:52 PM | 0 | | 0 | |

Account Properties Page

Account Properties Domain: GRB and Friends

Account Address: @dcr-sv3em.dcrdmz.local

Person's Name:

Unit Name:

Contact Email:

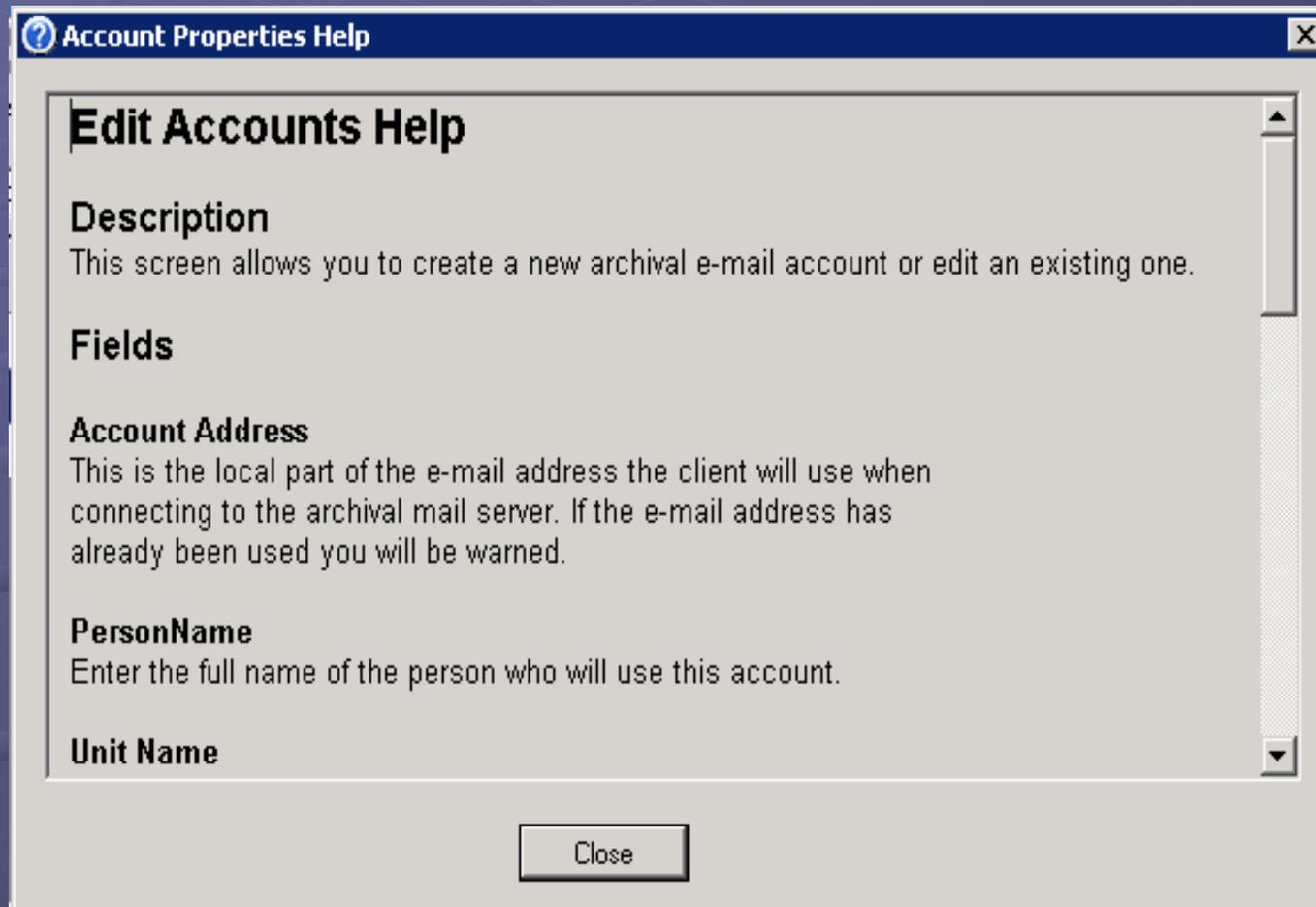
Account Status: Will Process Active

Notes:

Account Information

| | |
|---|--|
| Account Created: <input type="text" value="4/8/2008 7:59 PM"/> | Last processed on: <input type="text" value="Not yet processed."/> |
| Total Messages Archived: <input type="text" value="0"/> | Messages Archived Last Run: <input type="text" value="n/a"/> |
| Earliest Message: <input type="text"/> | Online Disk Space: <input type="text" value="Not available."/> |
| Latest Message: <input type="text"/> | Offline Disk Space: <input type="text" value="0"/> |
| Store Folder: <input type="text" value="kelly_eubank@dcr-sv3em_dcrdmz_local(1)"/> | |

Help Page for Account Management



Batch Creation of Accounts

Batch Account Creation

File Edit Options

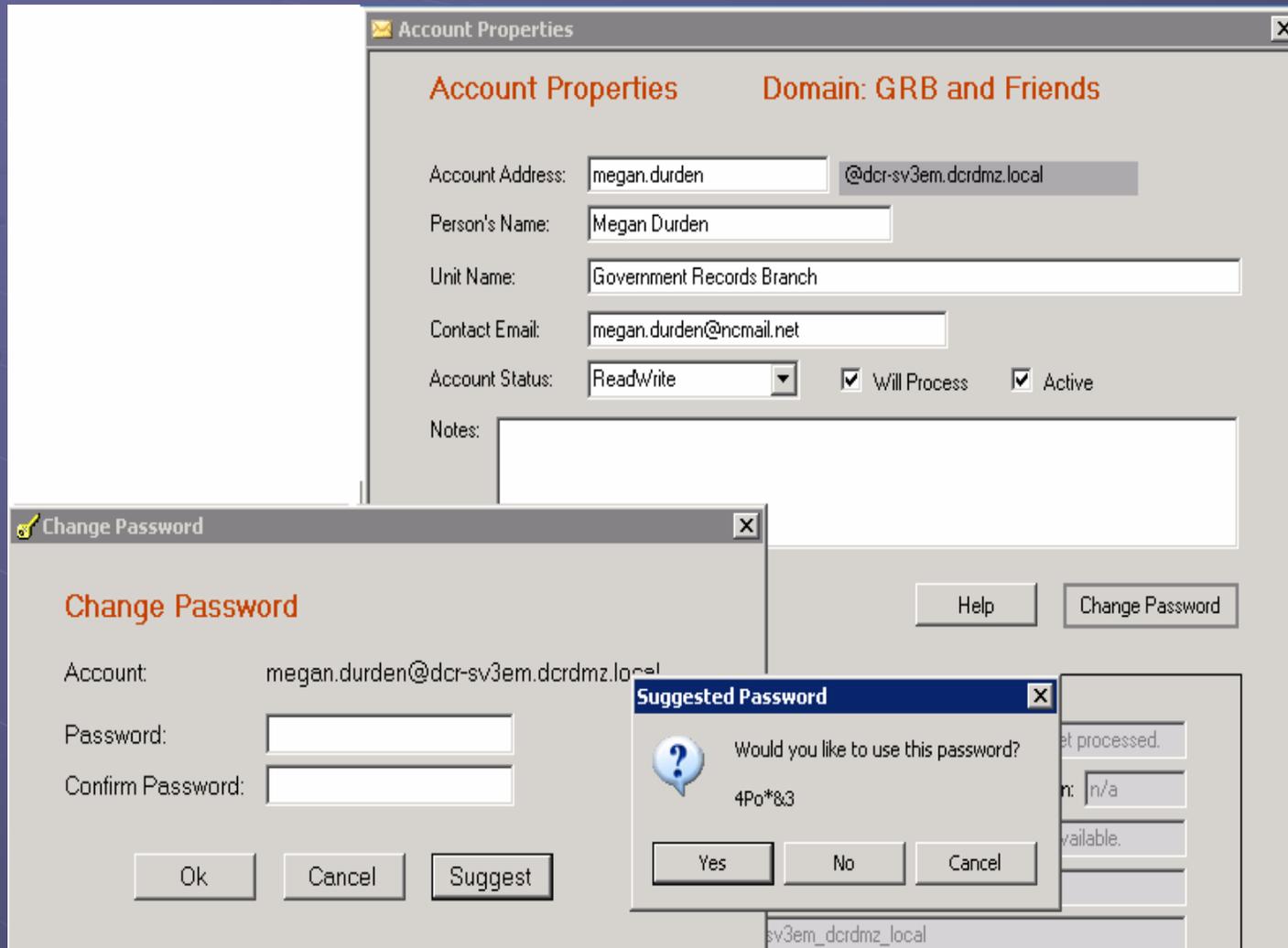
Create Accounts for: GRB and Friends (dcr-sv3em.dcrdmz.local)

Import File: Load Browse...

| | Person Name | Contact Email | Unit Name | Address |
|---|-------------|---------------|-----------|---------|
| * | | | | |

Instructions Create Accounts Send Mail Save Report Close

Tool will suggest passwords to use



Parser

- XML schema represents all email in an account
- Parses the header information in email
- Text attachments converted to Unicode
- Leaves tag in schema to point to the attachment
- Retain all original bit streams
- When we save external file, create a message digest—unique identifier

Example of the XML copy that is created by the Parser

```
<Name>X-Mailer</Name>
<Value>Unknown (No Version)</Value>
</Header>
- <Header>
  <Name>Message-ID</Name>
  <Value><18a.1823ab0.296c8bde@aol.com></Value>
</Header>
- <Header>
  <Name>X-Filter-Version</Name>
  <Value>1.7 (mailldmoz)</Value>
</Header>
- <SingleBody>
  - <BodyContent>
    <Content>The Southern Seed Legacy Project based at the University of Georgia has
    completed its web updates. We thank you for allowing us to link to your website. Each
    group has a short paragraph next to their link. Please feel free to comment on these
    paragraphs, if you feel that you would like your group described in another fashion. The
    address for the links page is: http://www.uga.edu/~ebl/links/ If you are interested in
    viewing our home page and seeing what the Southern Seed Legacy Project is all about,
    please check out our main page http://www.uga.edu/~ebl/southernheirloom/ Once
    again thank you. Katie O'Connell University of Georgia Department of
    Anthropology</Content>
  </BodyContent>
</SingleBody>
<Eol>CR</Eol>
</Message>
- <Message>
  <LocalId>0</LocalId>
  <From>"Jennifer Bean" <JBean@oldsalem.org></From>
  <To><ahweb@ncmail.net></To>
  <Subject>Historical Markers</Subject>
  - <Header>
    <Name>Return-Path</Name>
    <Value><JBean@oldsalem.org></Value>
  </Header>
  - <Header>
    <Name>Content-Transfer-Encoding</Name>
    <Value>7bit</Value>
  </Header>
  - <Header>
    <Name>Original-Recipient</Name>
    <Value>rfc822;ahweb@ncmail.net</Value>
  </Header>
  - <Header>
    <Name>X-Priority</Name>
    <Value>3</Value>
  </Header>
  - <Header>
    <Name>X-MSMail-Priority</Name>
    <Value>Normal</Value>
```

Next Steps

Development of additional .pst
file import capability

Finish training documents

Roll out of tool to state partners