

ELECTRONIC MAIL PRESERVATION COLLABORATION INITIATIVE

Background

2

- Builds on earlier NHPRC Electronic Records Fellows award—Druscie Simpson, 2005
- Collaboration with Pennsylvania, Kentucky and North Carolina
- All 3 institutions struggling with how to handle and collect email
- All three have strong training programs.

Current Email Systems

3

Most of the e-mail archiving solutions on the market today:

- are geared towards short or medium-term retention
- capture all e-mail indiscriminately at the point of ingest which gives no opportunity for the arrangement of those e-mail messages
- use proprietary formats
- do not have an easy migration path to other content repositories.

EMCAP -- Email Collecting and Preservation Solution

4

An e-mail archiving solution built for archivists

- ❑ Using the hMailServer—open source using the IMAP protocol
- ❑ Customized software to ingest and collect email
- ❑ Folders can be created on the archival server
- ❑ E-mail messages that are deemed to have archival value can easily be copied or moved to the archival server using any e-mail client.
- ❑ Entire existing folder structure can be copied to archive folder
- ❑ The EMCAP solution can also import mail directly from a .pst file, a mbox file, or a directory full of .msg files.

SAA Chicago 2007

Messages

5

- Original, untouched copies of each message are saved in mbox files, one for each mail box folder.
- Message are converted to XML. There is one XML file for each account.
 - ▣ XML digest will be stored in secure location.
- Back-up Storage
 - ▣ Keep duplicate sets of folders

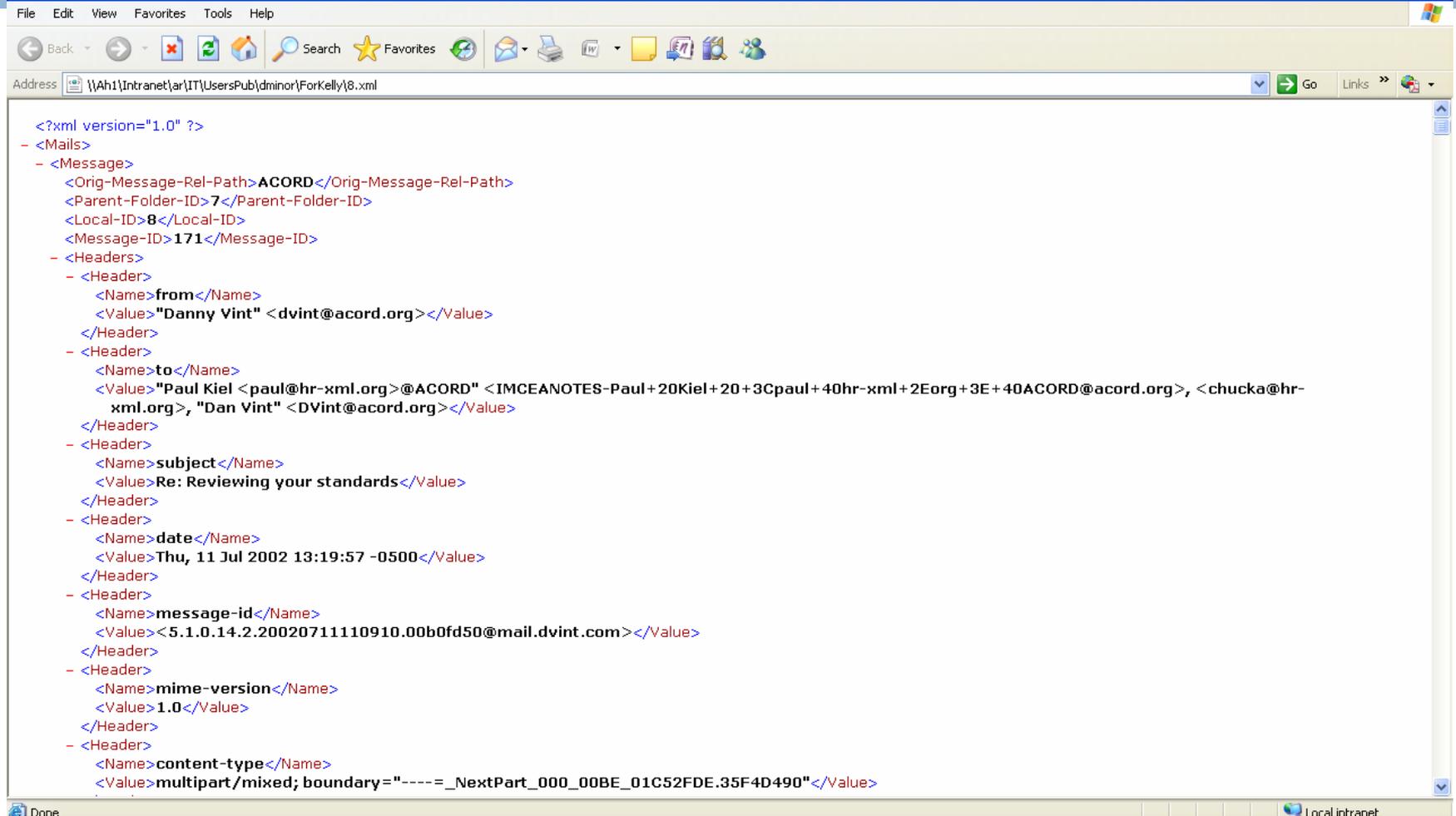
Messages-Ingest

6

- Message can be moved within the Archival server.
- Cannot be removed or duplicated.
- Collection can be set at predetermined time e.g. nightly or weekly
 - ▣ Can be removed if not yet collected.

Example XML

7



The screenshot shows a web browser window with the address bar containing the file path: \\Ah1\Intranet\ar\IT\UsersPub\dminor\Forkelly\8.xml. The browser's content area displays the following XML structure:

```
<?xml version="1.0" ?>
- <Mails>
- <Message>
  <Orig-Message-Rel-Path>ACORD</Orig-Message-Rel-Path>
  <Parent-Folder-ID>7</Parent-Folder-ID>
  <Local-ID>8</Local-ID>
  <Message-ID>171</Message-ID>
- <Headers>
- <Header>
  <Name>from</Name>
  <Value>"Danny Vint" <dvint@acord.org></Value>
</Header>
- <Header>
  <Name>to</Name>
  <Value>"Paul Kiel <paul@hr-xml.org>@ACORD" <IMCEANOTES-Paul+20Kiel+20+3Cpaul+40hr-xml+2Eorg+3E+40ACORD@acord.org>, <chucka@hr-xml.org>, "Dan Vint" <DVint@acord.org></Value>
</Header>
- <Header>
  <Name>subject</Name>
  <Value>Re: Reviewing your standards</Value>
</Header>
- <Header>
  <Name>date</Name>
  <Value>Thu, 11 Jul 2002 13:19:57 -0500</Value>
</Header>
- <Header>
  <Name>message-id</Name>
  <Value><5.1.0.14.2.20020711110910.00b0fd50@mail.dvint.com></Value>
</Header>
- <Header>
  <Name>mime-version</Name>
  <Value>1.0</Value>
</Header>
- <Header>
  <Name>content-type</Name>
  <Value>multipart/mixed; boundary="----=_NextPart_000_00BE_01C52FDE.35F4D490"</Value>
```

The browser's status bar at the bottom shows "Done" on the left and "Local intranet" on the right.

Attachments

- Automatically extracted and saved in regular files on the file system.
- Will automatically be converted to PDF format.
- The XML file will reference the original mbox message files for each folder, and the attachments for each message.
 - Utilizing relative links for message and attachments.
 - Digest stored separate and secure.

Technical

- Each account is archived separately and receives a global unique identifier.
- Each message within an account is assigned a globally unique identifier.
- Original, untouched copies of each message are saved in mbox files, one for each mail box folder.
- Messages are converted to XML. There is one XML file for each account.
- Designed to work with Digital Repository, as collection server.



Configuration of Architecture

11

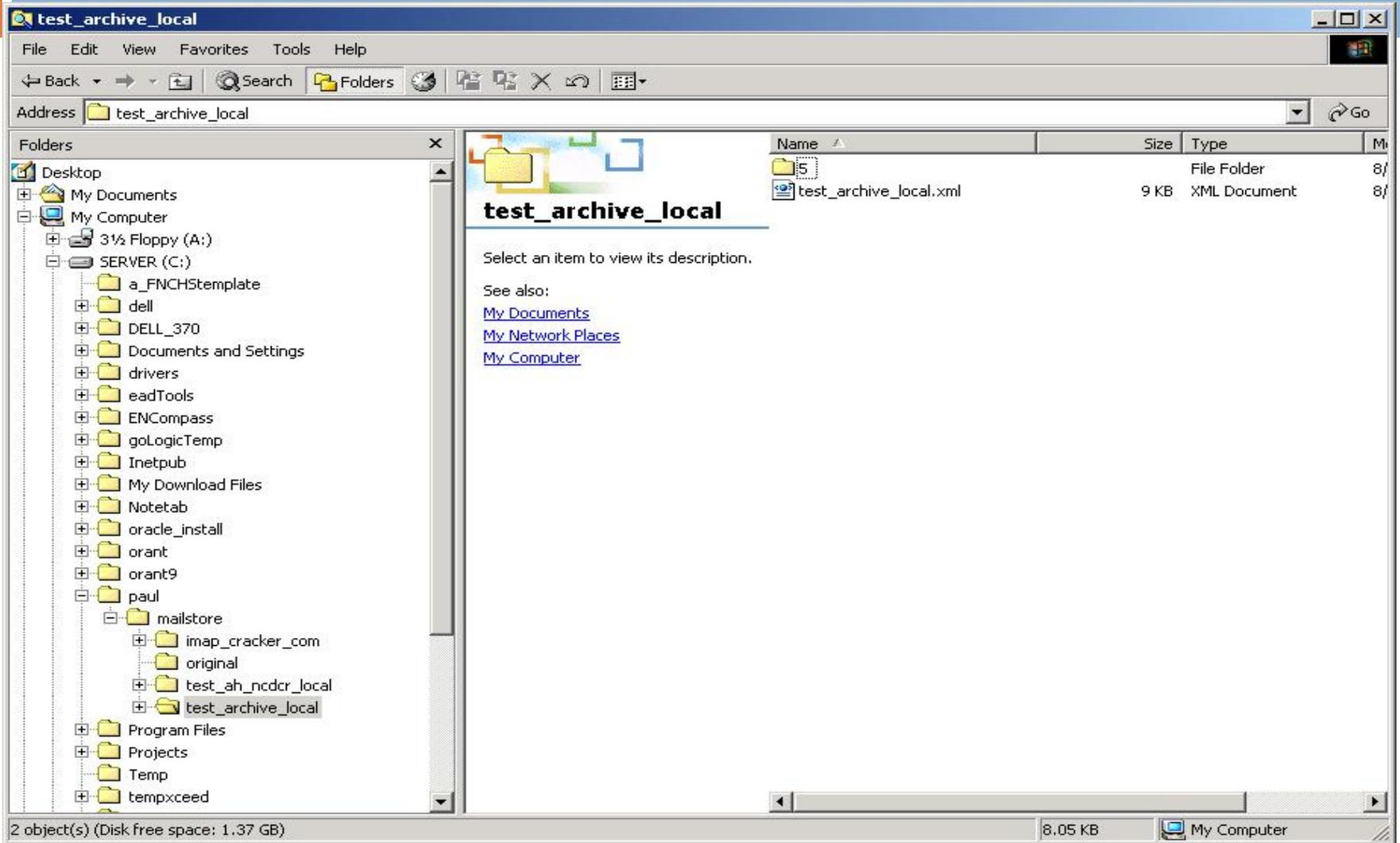
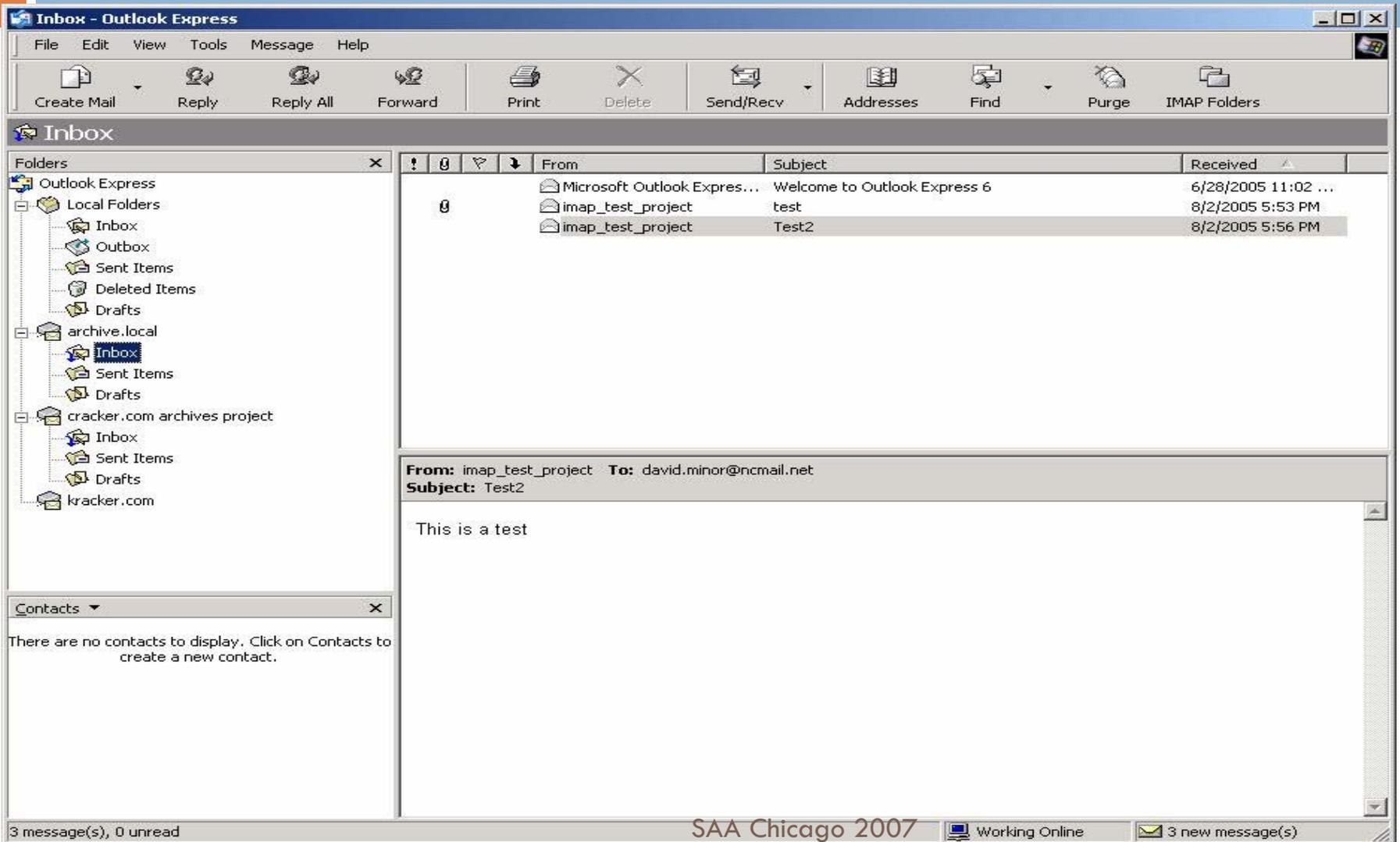


Image of Client Screen

12



Example of Report For Archives

13

NORTH CAROLINA OFFICE OF ARCHIVES & HISTORY
History for all the People

Email Report

(snapshot taken 7/29/2005 11:15:26 AM)

Overall Summary

Number of accounts:	1
Number of messages for test_archive_local:	3
Number of folders for test_archive_local:	1

Message Detail Summary

Account	Message Number	To	From	Subject	Date	Folder	Number of Headers
test_archive_local	4	"New Outlook Express User"	"Microsoft Outlook Express Team" <msoe@microsoft.com>	Welcome to Outlook Express 6	Tue, 28 Jun 2005 11:02:59 -0400	Inbox	8
test_archive_local	5	"david.minor@ncmail.net"	"imap_test_project" <imap@cracker.com>	test	Tue, 2 Aug 2005 17:53:37 -0400	Inbox	9
test_archive_local	6	<david.minor@ncmail.net>	"imap_test_project" <imap@cracker.com>	Test2	Tue, 2 Aug 2005 17:56:45 -0400	Inbox	10

Metadata—What about metadata?

14

- Extraction of metadata and searching
- Metadata components to be stored—To, From, Date, email address, transmission information
- In future, consider leveraging commercial software to search metadata for retrieval.

Partnerships within States

15

- Also going to test the tool with selected partners in the states
 - ▣ Pennsylvania—Pennsylvania State Archives, Governor's correspondence
 - ▣ Kentucky—Kentucky Department for Libraries and Archives—Commissioners office (?)
 - ▣ North Carolina—Secretary of State's office with possibility of the Corporations Division and the Notary office,
 - Other possibilities include Governor's office and State Emergency Management Office.

External Collaborations

16

- Working closely with the Smithsonian Institute and the Rockefeller Archive Center on
 - ▣ Collaborative Electronic Records Program (CERP), focusing on email
 - ▣ Collaborating on XML schema
 - ▣ www.siarchives.si.edu/cerp/cerpindex.htm

Electronic Mail Preservation Collaboration Initiative

17

An e-mail
archiving
solution built
for archivists

Questions?

<http://www.ah.dcr.state.nc.us/records/emailpreservation/>

