EMAIL PRESERVATION TOOLS

SERI Educational Webinar
Tuesday, May 13, 2014
2:00 pm Eastern
QUICK WEBEX TUTORIAL

▪ Use the Chat box to interact with the host, panelists, and attendees
  • Select “All Participants” from the drop-down box to chat with EVERYONE

▪ Use the Q&A box to ask questions

▪ Click the “View all attendees…” link below your name to see a complete list of attendees

▪ If you’d like to speak, raise your hand and the host will unmute you
ACKNOWLEDGEMENTS

This project is made possible by a grant from:

INSTITUTE of Museum and Library SERVICES
PRESENTERS

Susan Gray Page
Digital Archives Coordinator
Library of Virginia
SusanGray.Page@lva.virginia.gov

Elizabeth Perkes
Electronic Records Archivist
Utah State Archives
eperkes@utah.gov
KAINE EMAIL PROJECT @ LVA

Providing Access to Born-Electronic Government Records

Susan Gray E. Page
Digital Archives Coordinator
110,956 AND COUNTING

Visit www.virginiamemory.com/collections/kaine for project background and helpful search tips!

All Emails from Kaine Administration (110956)
TRANSPARENCY ABOUT TRANSPARENCY

Kaine Email Project @ LVA

Welcome to the Library of Virginia’s Kaine Email Project, where we make accessible the email records from the administration of Governor Timothy M. Kaine, Virginia’s 70th governor (2006–2010). Users can search and view email records from the Governor’s Office and his Cabinet Secretaries; learn about other public records from the Kaine Administration; go behind the scenes to see how the Library of Virginia made the email records available; and read what others are saying about the collection. The Library of Virginia received approximately 1.5 million email messages from the Kaine Administration. We are processing and releasing these records in batches, so please check back often for new content.
5 MAJOR STEPS TO PUTTING 110,956 EMAILS ONLINE

1. Records management
2. Archival processing
3. Technical processing
4. Public access
5. Public launch
1. RECORDS MANAGEMENT

It’s all about building relationships.
2. ARCHIVAL PROCESSING

Yes, we are item-level processing 1.3 million email records.

Yes, we might be crazy.
While great care has been taken during the processing of this collection to locate, identify, and restrict access to privacy-protected information within this collection, some relevant materials may have been missed. The Virginia Freedom of Information Act (Code of Virginia, §2.2-3705.5) and the Virginia Public Records Act (Code of Virginia, §42.1-78) establish guidelines for restricting access to privacy-protected information. Privacy-protected records include, but are not limited to, certain educational, medical, financial, criminal, attorney-client, and personnel records. Privacy-protected information is considered confidential and is restricted from public access for 75 years following the date of record creation. In accordance with this legislation, the Library of Virginia can and will restrict, in whole or in part, access to records containing privacy-protected information.

By logging in and accessing this collection, the user agrees:

- That if privacy-protected information is discovered during use of this material to make no notes or other recordation of the confidential information.
- Not to publish, publicize, or disclose any confidential material to any other party for any purpose.
- That no direct or indirect contact will be made with the individuals to whom the confidential information relates.
- To contact the Library of Virginia at archdesk@lva.virginia.gov to report any confidential information found.

Improper disclosure of privacy-protected information is a breach of confidentiality that could result in the loss of access to the archival collections housed and maintained by Library of Virginia, and could result in legal penalties (Code of Virginia, §18.2-186.3).

Name: LVA
Password: LVA

Login is required only once during each session.
3. TECHNICAL PROCESSING

Cheap tools for converting PSTs to PDFs:

- PST Viewer Pro ($69.99)
- Total Outlook Converter ($49.90)
WE TURNED PST FILES INTO LOTS OF THESE:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2008-02-15_12.49.06.pdf</td>
<td>2008-03-08_11.27.16.pdf</td>
<td>2008-03-13_15.06.06.pdf</td>
<td>2008-03-04_15.45.03.pdf</td>
</tr>
<tr>
<td>2008-03-06_15.48.33.pdf</td>
<td>2008-03-27_08.44.06.pdf</td>
<td>2008-03-27_15.50.49.pdf</td>
<td>2008-03-20_13.42.11.pdf</td>
</tr>
</tbody>
</table>
THAT LOOK LIKE THIS TO THE END USER:
4. PUBLIC ACCESS

It’s not Gmail, but it’s the best we can do right now.
CLOSEST WE COULD GET TO AN “INBOX” VIEW:

<table>
<thead>
<tr>
<th>Title</th>
<th>Creator</th>
<th>Date</th>
<th>pid</th>
<th>Subject</th>
<th>Parent</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABBA Private Projects</td>
<td>Kaine, Tim</td>
<td>2010-01-15</td>
<td>336104</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[no title]</td>
<td>Kaine, Tim</td>
<td>2010-01-13</td>
<td>336103</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[no title]</td>
<td>Kaine, Tim</td>
<td>2010-01-13</td>
<td>336102</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[no title]</td>
<td>Kaine, Tim</td>
<td>2010-01-13</td>
<td>336101</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latest draft</td>
<td>Kaine, Tim</td>
<td>2010-01-13</td>
<td>336100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Letter from Hancock Timber Resource Group</td>
<td>Kaine, Tim</td>
<td>2010-01-13</td>
<td>335099</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[no title]</td>
<td>Kaine, Tim</td>
<td>2010-01-13</td>
<td>335098</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thank You</td>
<td>Kaine, Tim</td>
<td>2010-01-12</td>
<td>335097</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>land conservation statistics</td>
<td>Kaine, Tim</td>
<td>2010-01-12</td>
<td>335096</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[no title]</td>
<td>Kaine, Tim</td>
<td>2010-01-12</td>
<td>335095</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>land conservation statistics</td>
<td>Kaine, Tim</td>
<td>2010-01-12</td>
<td>335094</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>State of the Commonwealth - land conservation</td>
<td>Kaine, Tim</td>
<td>2010-01-08</td>
<td>335093</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Danville coverage</td>
<td>Kaine, Tim</td>
<td>2010-01-08</td>
<td>335092</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>400,000-acre goal — SUCCESS</td>
<td>Kaine, Tim</td>
<td>2010-01-08</td>
<td>335091</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restoration of Rights Response Letter</td>
<td>Kaine, Tim</td>
<td>2010-01-05</td>
<td>335089</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alternative Student Loan Proposal</td>
<td>Kaine, Tim</td>
<td>2010-01-05</td>
<td>335088</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Invitation</td>
<td>Kaine, Tim</td>
<td>2010-01-05</td>
<td>335087</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peter Thompson</td>
<td>Kaine, Tim</td>
<td>2010-01-04</td>
<td>335086</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preston Bryant</td>
<td>Kaine, Tim</td>
<td>2010-01-04</td>
<td>335085</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
EDUCATING OUR USERS (AND OURSELVES) ON HOW TO USE OUR SEARCH SYSTEM:

Digital Collections Search Tips

SEARCH, BROWSE, REFINE

There are many, many ways users can access the Library of Virginia’s digital collections. The Kaine Email Collection is especially large. It is also full-text searchable, which can greatly enhance access but can also overwhelm users with thousands of results.

This Tip Sheet explains at a high-level several access points to the Kaine emails.

From the main search page, users can search across all of the Library’s collections or narrow searching to high-level collections, as seen here:

- Simple Search
- Advanced Search

Users can search across all of the Library’s digital collections, including Prints & Photos, Maps, County and City Records, Civil War materials, Government Docs...you get the idea.
5. PUBLIC LAUNCH

Stress tests and blog posts and tweets, oh my!
OOPS.

System is inaccessible. Please contact your system administrator.
BUILDING ON A SUCCESSFUL PLATFORM

Out of the Box
NOTES FROM THE ARCHIVES @ THE LIBRARY OF VIRGINIA

Kaine Email Project @ LVA – BUDGET EDITION
27 January 2014

This is the first in a series of posts spotlighting recently released email from Governor Tim Kaine’s administration (2006-2010). These posts are not meant to be comprehensive but to encourage further exploration in the Kaine administration records (electronic and paper).

Governor Kaine took office shortly before the beginning of the “Great Recession,” the worst economic downturn since the Great Depression of the 1930s. From 2007 to the end of his administration in January 2010, Kaine cut nearly $5 billion in state spending in order to balance the budget without raising taxes. This week’s post focuses on the budget development process, budget cuts, and what might have happened if Virginia did not have a budget by 1 July 2006.
...AND TRYING OUT A NEW ONE AS WELL
IN THE NEWS!

WHAT'S THE BUZZ

As the first state government archives in the United States to make the e-mails of a previous administration freely available to the public online, we hope our project will generate some buzz about electronic records, archiving in the digital world, open government, and all that jazz.

Looking to join the conversation? Check below to see what's been making our ears burn recently, and then add your voice to the fray! #kaineemail

To get a taste of the collection content, be sure to check out the Library’s series of Kaine email posts at the Out of the Box blog.

Daily Press
Kaine emails detail 2006 plan to avoid budget shutdown
By Dave Ress
When the General Assembly last locked itself into a budget impasse, coming within days of missing its July 1 budget deadline, then Gov. Tim Kaine worked quietly to line up emergency loans and order emergency spending to keep all the key functions of state government running, email records in the Library of Virginia show...

Style Weekly
Kaine’s Terabyte of E-Mail Data Reveals Little Dirt and Office Poltics
By Tom Nash
The Library of Virginia long has collected the records left behind by governors. Its first batch, from Patrick Henry, came on parchment. Gov. Tim Kaine, in office from 2002-2006, left behind a terabyte — including 229 email inboxes containing 1.3 million emails. Want a peek? The library recently released 66,000 of the emails online...
www.virginiamemory.com/collections/kaine

Susan Gray E. Page
Digital Archives Coordinator
SusanGray.Page@lva.virginia.gov
GMAIL: HARVESTING & INGESTING EXECUTIVE DIRECTOR DATA

Elizabeth Perkes
Utah State Archives
GroupWise

- For 20 years, Utah used GroupWise as its enterprise email system

- GroupWise export options were:
  - Individual emails, saved as text or .eml
  - Whole accounts, saved as XML, accessible via Nexic client

- Limited searching options, bulk exports done on the backend by IT
  - Public records requests very time-intensive, and expensive
GroupWise

- Archives has email from two accounts for people who left prior to Gmail conversion:
  - Budget officer of Archives
    - A few dozen emails
  - Former state CIO who left after a data breach
    - Tens of thousands of emails

- Nexic data not very well self-described
  - Relies on local executable that isn’t being updated with OS changes
  - Folders not named in meaningful way, unknown XML structure
  - Client is user-friendly
# NEXIC CLIENT

![NEXIC Archive Client](image)

<table>
<thead>
<tr>
<th>Name</th>
<th>Subject</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rosemary Cundiff</td>
<td>developing work plan for coordinating with Family Sea...</td>
<td>6/14/2010 3:30 PM</td>
</tr>
<tr>
<td>Patricia Smith-Mansfield</td>
<td>Ogden City review</td>
<td>6/8/2010 9:00 AM</td>
</tr>
<tr>
<td>Brian Carpenter</td>
<td>Crowley Micrographics Part</td>
<td>6/1/2010 4:07 PM</td>
</tr>
<tr>
<td>Email Weecare</td>
<td>(Statewide Announcement) Reminder - a State of U...</td>
<td>6/1/2010 3:38 PM</td>
</tr>
<tr>
<td>Patricia Smith-Mansfield</td>
<td>Utah Public Notice Website training</td>
<td>6/1/2010 2:00 PM</td>
</tr>
<tr>
<td>Mark Dalton</td>
<td>Refrigeration Contact Information</td>
<td>6/7/2010 12:08 PM</td>
</tr>
<tr>
<td>Brenda J. Murphy</td>
<td>John's access</td>
<td>6/7/2010 9:58 AM</td>
</tr>
<tr>
<td>Data Warehouse</td>
<td>Cognos Query Studio Training</td>
<td>6/7/2010 9:03 AM</td>
</tr>
<tr>
<td>Patricia Smith-Mansfield</td>
<td>Management Team</td>
<td>6/7/2010 9:00 AM</td>
</tr>
<tr>
<td>Marci Soper</td>
<td>May AM 31 &amp; 65 Reports</td>
<td>6/7/2010 8:44 AM</td>
</tr>
<tr>
<td>Patricia Smith-Mansfield</td>
<td>Staff Meeting</td>
<td>6/7/2010 8:00 AM</td>
</tr>
<tr>
<td>Patricia Smith-Mansfield</td>
<td>Fed: Budget Impact Form (BIF) info</td>
<td>6/7/2010 7:47 AM</td>
</tr>
<tr>
<td>Data Warehouse</td>
<td>Data Warehouse User Group Meeting Topic</td>
<td>6/3/2010 4:52 PM</td>
</tr>
<tr>
<td>Data Warehouse</td>
<td>June Data Warehouse User Group</td>
<td>6/3/2010 2:56 PM</td>
</tr>
<tr>
<td>Patricia Smith-Mansfield</td>
<td>Fed: RE: NHPRC reauthorization -- letter from all state...</td>
<td>6/3/2010 8:32 AM</td>
</tr>
<tr>
<td>Rosemary Cundiff</td>
<td>custodial</td>
<td>6/2/2010 12:12 PM</td>
</tr>
<tr>
<td>Patricia Smith-Mansfield</td>
<td>Re: Please Review -- Non-Federal Grant Applications R...</td>
<td>6/2/2010 8:31 AM</td>
</tr>
<tr>
<td>Patricia Smith-Mansfield</td>
<td>Re: Please Review -- Federal Assistance Applications R...</td>
<td>6/2/2010 7:52 AM</td>
</tr>
<tr>
<td>Patricia Smith-Mansfield</td>
<td>Gifts and donations</td>
<td>6/1/2010 6:05 PM</td>
</tr>
<tr>
<td>Susan Mumford</td>
<td>Re: UTA Bus Pass Program</td>
<td>6/1/2010 3:26 PM</td>
</tr>
<tr>
<td>Patricia Smith-Mansfield</td>
<td>UTA Bus Pass Program</td>
<td>6/1/2010 3:22 PM</td>
</tr>
<tr>
<td>Patricia Smith-Mansfield</td>
<td>Links to reading material</td>
<td>6/1/2010 3:03 PM</td>
</tr>
</tbody>
</table>

**Selected:** 0  **Total:** 134
NEXIC BACK END
NEXIC BACK END
NEXIC BACK END
NEXIC BACK END
NEXIC BACK END

Mark:

Here you go. I hope this helps in getting the ball rolling for you on this relocation project.

Billy,

I am just following up to see if you will have the layout, pricing, and brochure information available tomorrow. Wayne Christensen from facilities management is requesting the information so that he can submit his request to the Federal Government for the building. Please let me know.

Thanks for all of your help.

Mark

Mark Dalton
Manager, State Records Center
(801) 975-4016 Fax: (801) 974-0336
mdalton@utah.gov
The State Archives hours of operation are
Monday-Thursday, 7:00 a.m. to 6:00 p.m. and closed on Friday.
Please make a note of these hours.
GMAIL

In 2012, Utah transitioned to Gmail

- Funding for this change was available as it impacted databases integrated with email
- Archives was able to connect to the Gmail API with its AXAEM system
  - Used this feature to send emails from an existing Gmail account via the AXAEM interface, impacting:
    - Records officer online training/certification
    - Patron requests for records, ordering boxes from storage
GMAIL

Apple Valley, UT

- Used Gmail
- ISP sold their domain name, no access to email
- Called Archives for help
- Archives asked APPX how to download this email
- APPX created simple interface using existing Gmail API
- Interface now used regularly:
  - By Archives, to harvest executive director data
  - By DTS, to respond to litigation and security investigations; or agencies answering public records requests
  - By agencies, because they want an easy way to move data offline, especially those leaving state employment, or share data with third parties
GMAIL

- Multiple labels can be assigned to the same email, different from concept of “folders”
  - Search email in Gmail using advanced search
  - Select hits
  - Apply a label to the hits

- Log into AXAEM
  - Provide Gmail account name and password
  - Click “Extract Contents”
  - Indicate location where email is to be saved
  - Select labels whose contents you want to export
  - Click “OK”
GMAIL
GMAIL
GMAIL
EML AS PRESERVATION COPY

- Stored as plain text

- Metadata easy to extract

- Desktop email clients know how to render it, make attachments viewable

- Attachments encoded as base64, which can be transformed to a binary and stored separately if desired, or migrated forward

- Easy to de-accession if content not preservation-worthy
VALUABLE EMAIL SAVED

- Public Safety

- Transportation

- Facilities Construction & Management
  - Non-director, 30-year employee asked for copies of his email
  - Found conversations with Capitol Preservation architect, who left long ago without email being saved
  - Found minutes to lots of meetings, plenty of value in his email account, though he wasn’t director
PROBLEMS WITH DIRECTORS’ EMAIL

- Used Google Docs instead of attachments
  - Have to read each email to know if a link is there
  - Have to have account/password still active in Gmail to access
  - Once you download the file, how do you associate it with the email, stored in context?

- Outgoing directors wiped email accounts, some forgot to do so with sent mail 😊

- Inbox keeps filling up with messages even after they left, hard to know termination date

- Sent mail filled with auto-replies 😞
APPRAISAL & NON-PUBLIC DATA

- With tens of thousands of emails to sift through, how can we weed accounts?

- Agencies could apply labels indicating retention and access restrictions

- Need appraisal interface for exported email

- To create a redacted copy, need a way to transform to PDF and use Acrobat’s features

- Need way to associate redacted copy with original during ingest into preservation system
HOW TO INGEST EMAIL

Our ingest procedure is this:

- Use BagIt to capture files with manifest and checksums, write to M-disc.
- Upload bag to AXAEM, where checksum is verified valid
- Metadata from records extracted and written to database
INGESTED EMAIL
INGESTED EMAIL
SEARCH ENGINE
ACCESS TO EMAIL

Solr search engine already indexes metadata of ingested records and makes records available for download

- Item must be marked as publishable first
- Access restrictions set at series level prevent auto-publishing records
- No staff time to read email one-by-one
- Conclusion: preserved, but not accessible
CONTACTS

Susan Gray Page
Digital Archives Coordinator
Library of Virginia
SusanGray.Page@lva.virginia.gov

Elizabeth Perkes
Electronic Records Archivist
Utah State Archives
eperkes@utah.gov
QUESTIONS & COMMENTS

It's QUESTION TIME!!
WRAP-UP

- Post-webinar evaluation will automatically open in your web browser when you exit the session.

- Next SERI Educational Webinar is Tuesday, June 10 @ 2:00 pm Eastern
  - Topic: Electronic Records Inventory

- Complete webinar schedule is available on CoSA’s website:
  http://www.statearchivists.org/CoSA_Webinars.htm

- All webinar slides available from the SERI webinar page:
  http://www.statearchivists.org/seri/STEP/SERI_Educational_Webinars.htm