# **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:







# **PRESENTERS**

#### **Susan Gray Page**

Digital Archives Coordinator Library of Virginia SusanGray.Page@Iva.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**



Collections > State Archives Collections > Timothy M. Kaine Administration (2006-2010) Emails

Timothy M. Kaine Administration (2006-2010) Emails



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

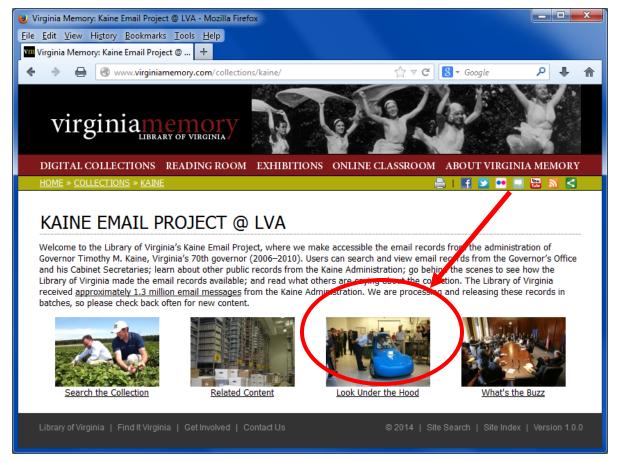
All Emails from Kaine Administration (110956)

More...





# TRANSPARENCY ABOUT TRANSPARENCY







# 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.





# PRIVACY CONSIDERATIONS =





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)





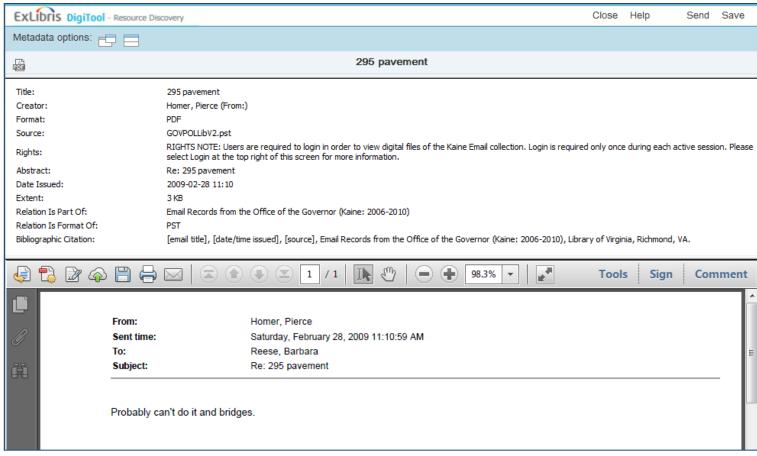
# WE TURNED PST FILES INTO LOTS OF THESE:

2007-08-07_15.39.38.pdf	☑ 2007-09-04_12.04.13.pdf	2007-09-11_17.34.20.pdf	2007-09-12_11.10.23.pdf
2007-09-27_16.50.34.pdf	2007-09-28_11.24.10.pdf	2007-10-05_09.51.34.pdf	2007-12-06_13.58.47.pdf
2007-12-06_14.38.35.pdf	2007-12-06_15.19.07.pdf	2007-12-13_13.49.33.pdf	2007-12-13_13.54.38.pdf
2007-12-20_14.28.20.pdf	2007-12-20_15.13.33.pdf	2007-12-21_14.01.55.pdf	2008-01-10_13.42.12.pdf
2008-01-11_14.20.57.pdf	2008-01-17_12.15.13.pdf	2008-01-17_14.43.38.pdf	2008-01-17_14.56.09.pdf
2008-01-17_15.04.42.pdf	2008-01-17_15.10.41.pdf	2008-01-24_15.17.58.pdf	2008-01-24_15.38.10.pdf
2008-01-28_10.31.59.pdf	2008-01-31_16.43.00.pdf	2008-02-05_20.29.28.pdf	2008-02-07_13.58.23.pdf
2008-02-07_14.19.45.pdf	2008-02-07_14.22.56.pdf	2008-02-14_15.37.07.pdf	2008-02-15_12.47.13.pdf
2008-02-15_12.49.06.pdf	2008-02-21_14.32.08.pdf	2008-02-28_15.10.29.pdf	2008-03-04_15.45.03.pdf
2008-03-06_15.48.33.pdf	2008-03-08_11.27.16.pdf	2008-03-13_15.06.06.pdf	2008-03-20_13.42.11.pdf
2008-03-25_16.02.06.pdf	2008-03-27_08.44.06.pdf	2008-03-27_15.50.49.pdf	2008-03-28_10.49.41.pdf
2008-03-31_10.53.32.pdf	2008-04-03_15.16.45.pdf	2008-04-03_16.31.37.pdf	2008-04-08_14.02.39.pdf
2008-04-08_14.17.09.pdf	2008-04-09_09.51.23.pdf	2008-04-10_12.15.53.pdf	2008-04-16_16.06.38.pdf
2008-04-17_13.56.04.pdf	2008-04-22_12.26.06.pdf	2008-04-24_13.52.05.pdf	2008-04-28_11.33.28.pdf
2008-04-28_11.37.49.pdf	2008-04-28_12.27.33.pdf	2008-04-28_16.57.47.pdf	2008-04-28_17.39.10.pdf
2008-04-29_09.24.18.pdf	2008-04-29_11.42.16.pdf	2008-04-29_11.44.14.pdf	2008-05-01_15.50.45.pdf
2008-05-01_16.39.39.pdf	2008-05-07_17.12.20.pdf	2008-05-08_12.59.00.pdf	2008-05-20_09.36.52.pdf
2008-05-23_17.00.10.pdf	2008-05-27_16.38.22.pdf	2008-05-27_17.09.23.pdf	2008-05-29_13.16.57.pdf
2008-05-29_14.10.38.pdf	2008-06-05_11.54.24.pdf	2008-06-05_13.47.36.pdf	2008-06-12_14.15.23.pdf
2008-06-12_15.09.34.pdf	2008-06-19_15.10.44.pdf	2008-06-25_22.33.55.pdf	2008-06-30_17.11.45.pdf
2008-07-07_13.40.47.pdf	2008-07-08_14.39.47.pdf	2008-07-10_14.52.11.pdf	2008-07-15_14.53.44.pdf
2008-07-17_15.24.42.pdf	2008-07-18_09.12.19.pdf	2008-07-24_10.59.28.pdf	2008-09-09_18.44.34.pdf
		·	





# 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.





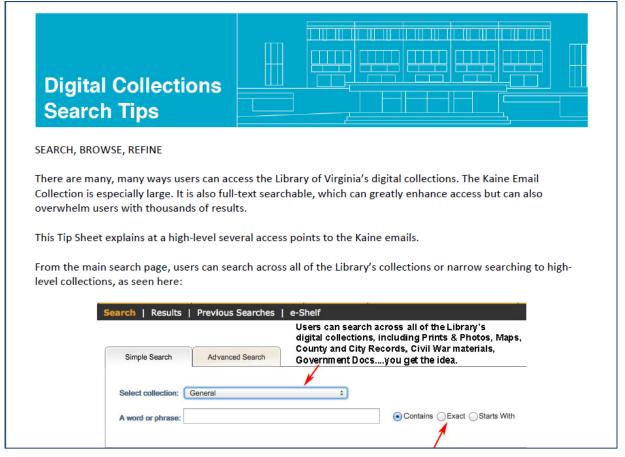
# CLOSEST WE COULD GET TO AN "INBOX" VIEW:

Brief view Table view Full view						Sort by: Date	
	cords 1- 20 of 3528					1234	
	Title	Creator	Date	pid Sut	oject Parent	Object	
ĺ	ARRA Private Projects	Kaine, Tim (From:)	2010-01-15 09:47	335104	æ e		
2	[no title]	Kaine, Tim (From:)	2010-01-13 12:20	335103	<b>2</b>		
3	[no title]	Kaine, Tim (From:)	2010-01-13 12:17	335102	<b>~</b>		
	[no title]	Kaine, Tim (From:)	2010-01-13 11:36	335101			
5	Latest draft	Kaine, Tim (From:)	2010-01-13 10:59	335100	<b>2</b>		
6	Letter from Hancock Timber Resource Group	Kaine, Tim (From:)	2010-01-13 09:54	335099			
	[no title]	Kaine, Tim (From:)	2010-01-13 09:48	335098			
3	Thank You	Kaine, Tim (From:)	2010-01-12 17:05	335097			
)	land conservation statistics	Kaine, Tim (From:)	2010-01-12 15:51	335096	<b>~</b>		
0	[no title]	Kaine, Tim (From:)	2010-01-12 15:00	335095	<b>2</b>		
1	[no title]	Kaine, Tim (From:)	2010-01-12 14:06	335094			
12	land conservation statistics	Kaine, Tim (From:)	2010-01-12 10:32	335093	<b>2</b>		
3	State of the Commonwealth - land conservation	Kaine, Tim (From:)	2010-01-08 17:08	335092	<b>2</b>		
4	Danville coverage	Kaine, Tim (From:)	2010-01-08 11:52	335091	<b>6</b>		
5	400,000-acre goal — SUCCESS	Kaine, Tim (From:)	2010-01-08 11:50	335090			
6	Restoration of Rights Response Letter	Kaine, Tim (From:)	2010-01-05 20:56	335089	<b>2</b>		
7	Alternative Student Loan Proposal	Kaine, Tim (From:)	2010-01-05 19:44	335088	<b>A</b>		
8	Invitation	Kaine, Tim (From:)	2010-01-05 13:00	335087	<b>@</b>		
9	Peter Thompson	Kaine, Tim (From:)	2010-01-04 16:42	335086	<b>6</b>		
20	Preston Bryant	Kaine, Tim (From:)	2010-01-04 15:42	335085	<u></u>		





# EDUCATING OUR USERS (AND OURSELVES) ON HOW TO USE OUR SEARCH SYSTEM:







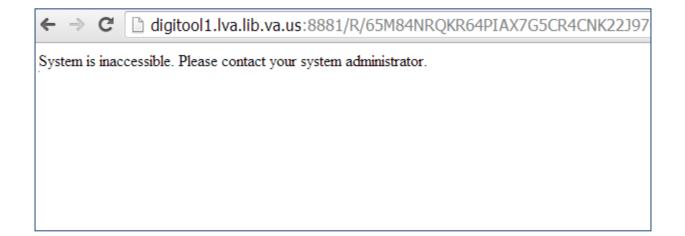
# 5. PUBLIC LAUNCH

Stress tests and blog posts and tweets, oh my!





# OOPS.





### **BUILDING ON A SUCCESSFUL PLATFORM**



Previous post:

« The Women of Smithfield

Next post:

Kaine Email Project @ LVA - Transportation Edition »

#### KAINE EMAIL PROJECT @ LVA - BUDGET EDITION

27 January 2014



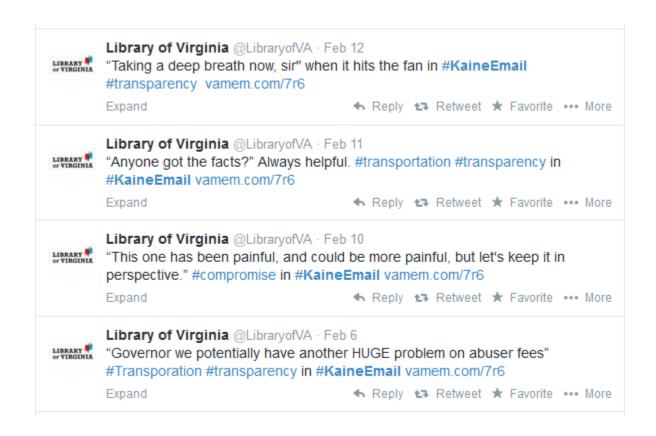
This is the first in a series of posts spotlighting recently released <u>email</u> from Governor Tim Kaine's administration (2006-2010). These posts are not meant to be comprehensive but to encourage further exploration in the Kaine administration records (electronic and paper).

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 Politics

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@Iva.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





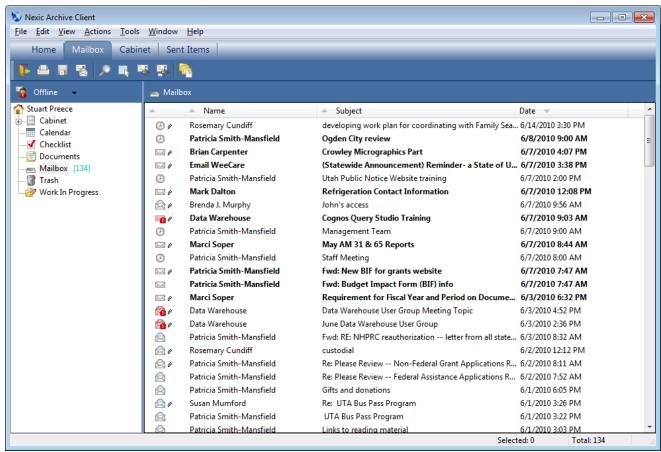
# **GroupWise**

- 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



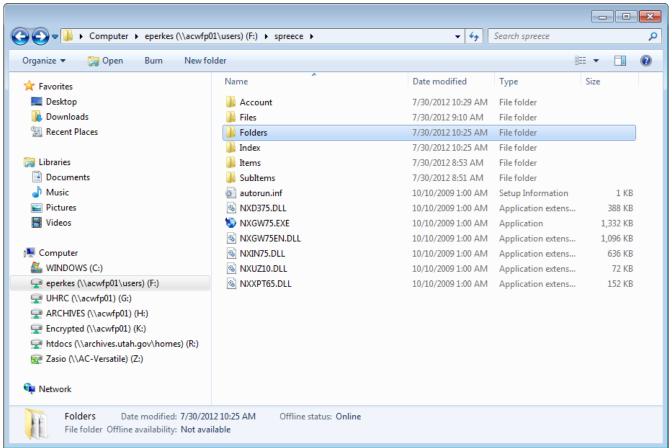


# **NEXIC CLIENT**



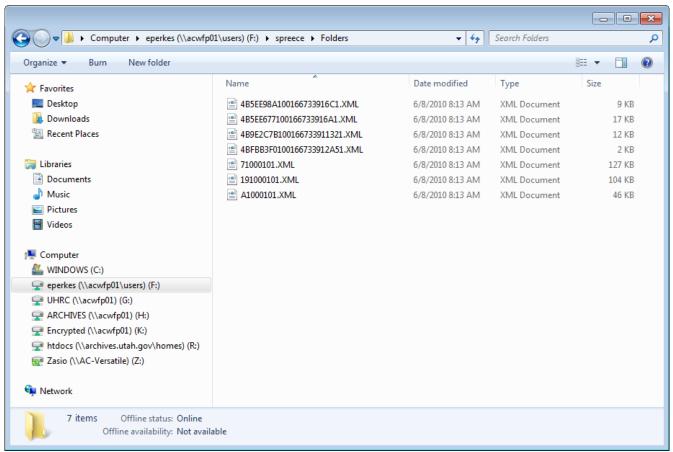






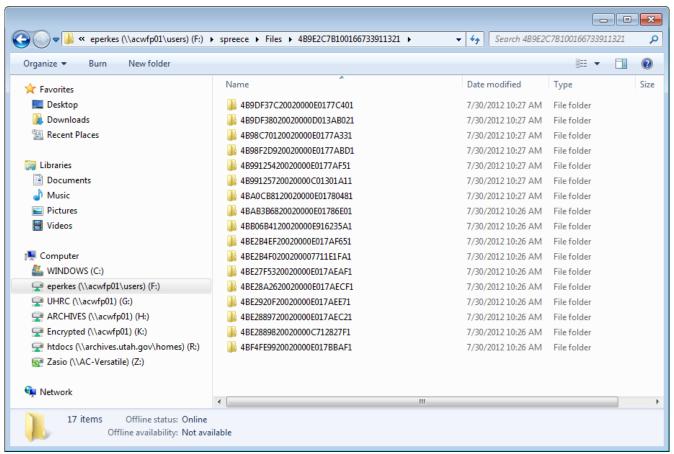






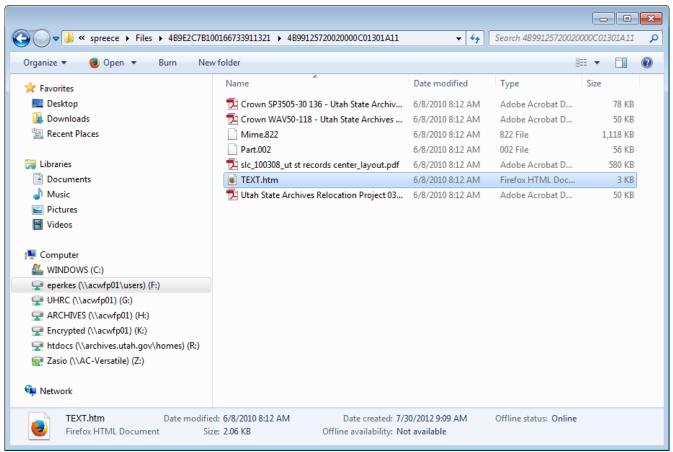






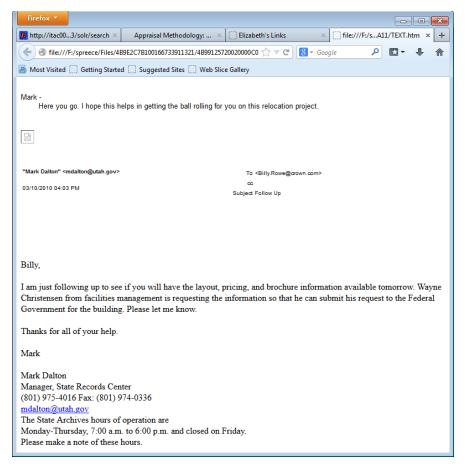
















#### In 2012, Utah transitioned to Gmail

- 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





#### 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

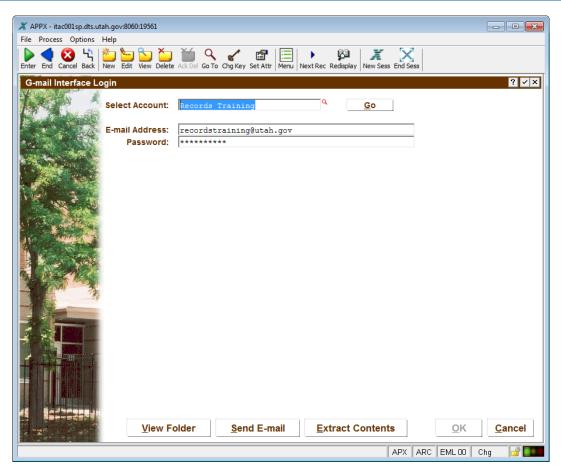




- Multiple labels can be assigned to the same email, different from concept of "folders"
  - Search email in Gmail using advanced search
  - Select hits
  - Apply a label to the hits
- Log into AXAEM
  - Provide Gmail account name and password
  - Click "Extract Contents"
  - Indicate location where email is to be saved
  - Select labels whose contents you want to export
  - Click "OK"

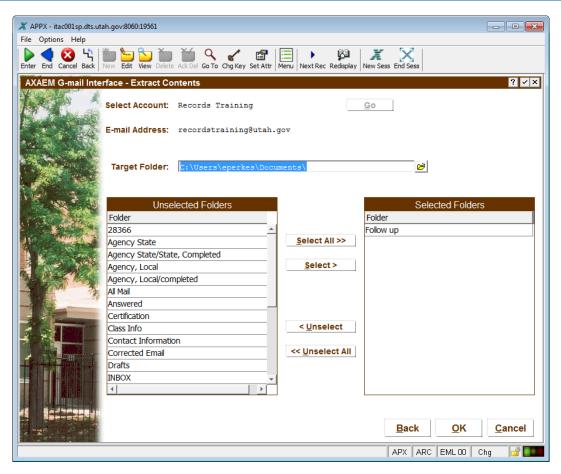






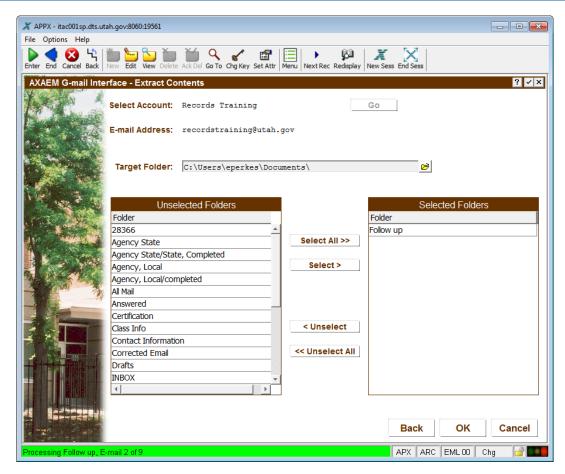






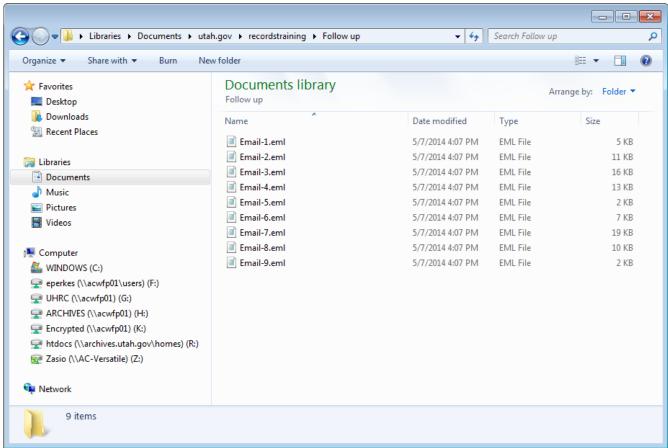
















```
Email-7.eml - Notepad
                                                                                                                                                                                                - - X
 File Edit Format View Help
Delivered—To: recordstraining@utah.gov

Received: by 10.224.170.197 with SMTP id e5csp34481qaz;

Tue, 7 May 2013 15:25:13 -0700 (PDT)

X-Received: by 10.66.251.39 with SMTP id zh7mr5083657pac.62.1367965512573;

Tue, 07 May 2013 15:25:12 -0700 (PDT)

Return-Path: <richmondcity@richmond-utah.com>
Received: from psmtp.com ([74.125.149.54])

by mx.google.com with SMTP id uhlsi20022061pab.152.2013.05.07.15.25.11
Authentication-Results: mx.google.com; spf=pass (google.com; superpass (google.com; best guess record for domain of richmondcity@richmond-utah.com designates 68.169.33.227 as permitted sender) smtp.mail=richmondcity@richmond-utah.com
Received: from richmond-utah.com ([68.169.33.227]) (using TLSv1) by na3sys009amx214.postini.com ([74.125.148.10])
with SMTP;
Tue, 07 May 2013 22:25:11 GMT

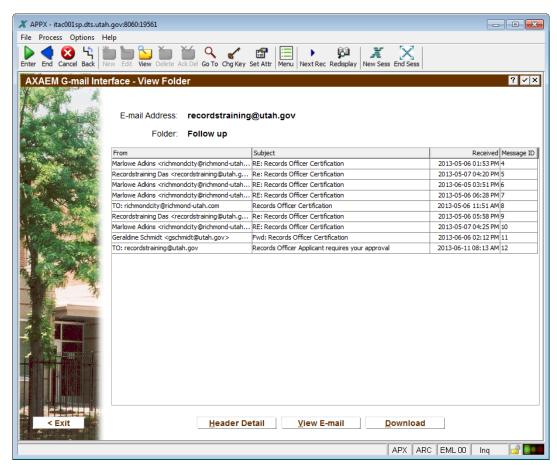
Received: from 50-79-67-102-static.hfc.comcastbusiness.net ([50.79.67.102]:55132 helo=ManagerPC)
by s1-507-13.slc.westdc.net with esmtpa (Exim 4.80)
                (envelope-from <richmondcity@richmond-utah.com>)
                id 1UZqJq-001dNQ-Do
To 102474-0010NQ-DO

for recordstraining@utah.gov; Tue, 07 May 2013 16:25:10 -0600

From: "Marlowe Adkins" <richmondcity@richmond-utah.com>
To: "Recordstraining Das" < recordstraining@utah.gov>
References: <772550156.2.1367862667427.JavaMail.aclfo@itac001sp>
<CAEy1beeR_xc+6LB9851nv=4_uD4kAvp8GwROcevy23OvWTx8xA@mail.gmail.com>
                                                                                                                                      <005c01ce4a93$5fe00d50$1fa027f0$@com>
                                                                                                                                      <008c01ce4ab9$d9befa20$8d3cee60$@com>
 <CAEy1bec1EfEgreKKJ_MTxkaJTvQ++x3VGcu7J+Qm=wNzEfsXyQ@mail.gmail.com>
In-Reply-To: <CAEy1bec1EfEgrekKJ_MTxkaJTvQ++x3VGcu7J+Qm=wNzEfsxyQ@mail.gmail.com>
Threepiy-10. CCAE/IDECIE EgreeN. MARATIVE HASSIGN
Subject: RE: Records Officer Certification
Date: Tue, 7 May 2013 16:25:06 -0600
Message-ID: <007c01ce4b71$ba76ab50$2f6401f0$@com>
MIME-Version: 1.0
Content-Type: multipart/alternative;
boundary="----=_NextPart_000_007D_01CE4B3F.6FDC3B50"
X-Mailer: Microsoft office Outlook 12.0
Thread-Index: AcSLcRZtxDcBWYmJQu0/UNCNKay4TAAAFP8g
Content-Language: en-us
X-AntiAbuse: This header was added to track abuse, please include it with any abuse report
X-AntiAbuse: Primary Hostname - s1-507-13.slc.westdc.net
X-AntiAbuse: Original Domain - utah.gov
X-AntiAbuse: Originator/Caller UID/GID - [47 12] / [47 12]
X-AntiAbuse: Sender Address Domain - richmond-utah.com
X-Get-Message-Sender-Via: sl-507-13.slc.westdc.net: authenticated_id: richmondcity@richmond-utah.com
X-Source:
X-Source-Args:
X-Source-Dir:
X-pstn-neptune: 0/0/0.00/0
x-pstn-levels: (5:99.90000/99.90000 cv:99.9000 Fc:95.5390 Lc:95.5390 R:95.9108 P:95.9108 M:96.8796 c:99.7951 )
X-pstn-dkim: 0 skipped:not-enabled
X-pstn-settings: 3 (1.0000:1.0000) s cv gt4 gt3 gt2 gt1 r p m c
X-pstn-addresses: from <richmondcity@richmond-utah.com> [db-null]
```

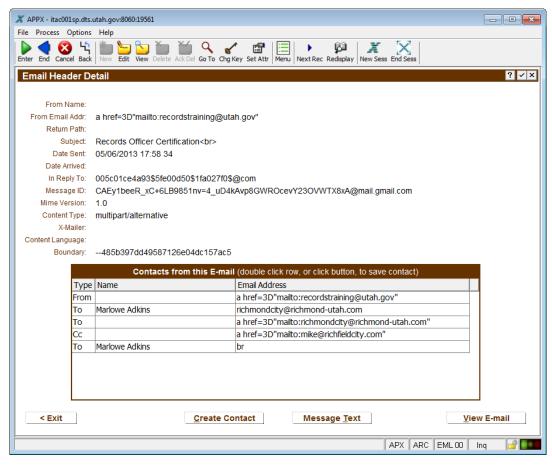
















### **EML AS PRESERVATION COPY**

- 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-PUPLIC 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**

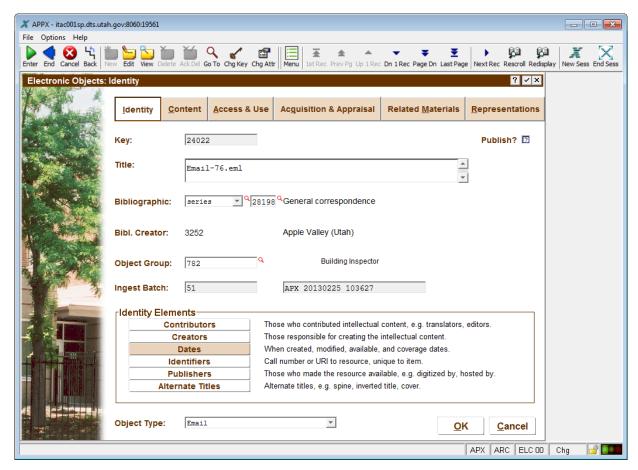
#### Our ingest procedure is this:

- 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





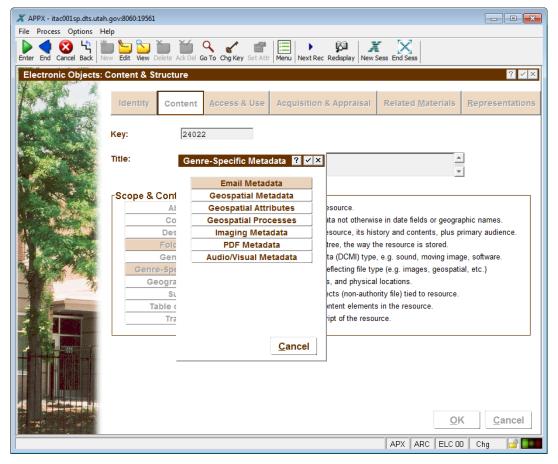
# **INGESTED EMAIL**







# **INGESTED EMAIL**







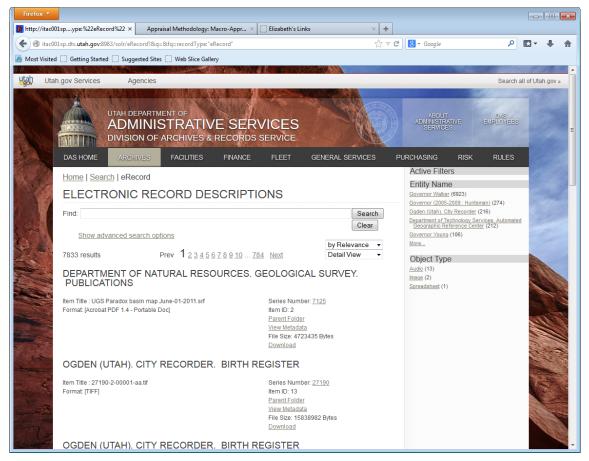
# **INGESTED EMAIL**

X APPX - itac001sp.dts.utah.gov:8060:19561     □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □	
File Options He	elp
Enter End Cancel	Back   New Edit   View   Delete   Adk Del   Go To   Chg Key   Set Attr   Menu   Next Rec   Redisplay   New Sess   End Sess
Electronic Records: Email	
Key:	24022 Date Sent: 10/29/2012 11:59 48 Date Arrived: MM/DD/CCYY hh:mm ss
From:	Nathan Bronemann
	nbronemann@applevalleyut.org
Return Path:	
Subject:	Nathan
In Reply To:	
_	CAC8tuKb5n7V87D8zWNu74bQYxKKhgvzNkHEp4GuCAP9C6EJqrA@mail.gmail.com
Mime Ver:	1.0 multipart/alternative; Content Language:
X-Mailer:	multipate, alternative,
Recipients:	References:
To: Lance Giffo	rd (Lance@cityofhurricane.com)
	<u>O</u> K <u>C</u> ancel
	APX ARC ELC 00 Chg





# **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@Iva.virginia.gov

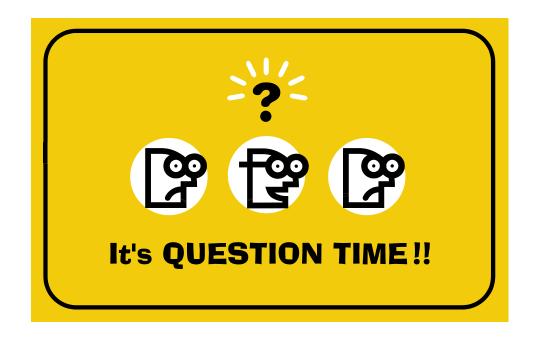
#### **Elizabeth Perkes**

Electronic Records Archivist Utah State Archives eperkes@utah.gov





# **QUESTIONS & COMMENTS**







#### **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



