CoSA & Preservica Practical Digital Preservation 2016

Automating Email Archiving & Preservation

Webinar
23 Feb 2016
Welcome!

• **PDP Online Workshops** - with focus on records and email  
  (Tuesdays 2-4pm Eastern)
  - Nov 10 2015: Part 1
  - Nov 17 2015: Part 2
  - Mar 08 2016: Part 1
  - Mar 15 2016: Part 2
  - May 10 2016: Part 1
  - May 17 2016: Part 2

• **PDP “Hot Topic” Webinars**  
  (Tuesdays 2-3pm Eastern)
  - Ingesting records from multiple sources and systems: Dec 08 2015
  - Automating email archiving and preservation: Feb 23 2016
  - New ways of providing public access to your archive: Apr 26 2016
  - Real-world digital preservation and program/resources round-up: Jun 21 2016

Automating Email Archiving and Preservation

• The Email challenge!

• Email preservation requirements

• Email at TSLAC

• Automating ingest of Gmail example

• Summary and Questions

Mark Myers
Electronic Records Specialist
Texas State Library and Archives Commission

Michael Hope
Preservica
Why do we need Email Preservation?

Email is increasingly the default medium of correspondence.
The Email Challenge

• Email is still growing exponentially in many organizations

• It costs money to store and archive
  – so what to keep, what not to keep, how long to keep it for?

• It’s increasingly a requirement
  – Mandate
  – Compliance or Regulation
  – Litigation and Legal Defense
  – Transparency
  – Heritage/Legacy

• It’s difficult to collect and appraise
  – tools and automation can help
What makes preserving Email hard?
Archiving or Preservation?

• For less than 3 years (Email Archiving)
  – reduce pressure/costs on operational systems
  – record of organizational decisions

• For more than 10 years (Preservation)
  – Long-term records
  – Mandate/Compliance/Legal/Heritage
  – Just another type of content or record in the archive
Solutions today

• Email Archiving solutions
  – focus on near-term archiving
  – “bit-level” protection only
  – use of PST – makes search and discovery difficult

• Email Preservation tools
  – based on email to PDF or from one mailbox to another

• What’s missing?
  – email and attachments and context
  – active preservation for the long-term
  – control over access
  – easy search and render
  – categorisation for future disposition
  – be able to demonstrate “trustworthiness”
Email Preservation Requirements

- **Provenance**
  - Audit Trails and Metadata

- **Authenticity**
  - Fixity

- **Durability**
  - Self-healing Storage

- **Preservation**
  - Active File Format Migration

- **Disposition**
  - Retention Policies & Actions

- **Privacy & Confidentiality**
  - User Roles & Permissions

- **Access**
  - Search, Browse & Render

......and do it at scale in an automated & seamless way
Correspondence in the Texas Digital Archive
Correspondence Received

- Governor’s Office
  - CTS (next slide)
  - PST Files
    - Several Calendars from Gov Office
    - Individual MSG files from Outlook
- LT. Gov and Legislators
  - Another homemade CTS system
- One large email PST from director of TX Youth Commission
Correspondence Tracking System

- Correspondence Tracking System (CTS)
  - Custom built by Governor’s Office
    - MS SQL 2012 database (1GB)
    - Image Repository – (170GB)
    - GUI developed in VB 6
  - Contains all incoming and outgoing correspondence
    - Correspondence received scanned into system
    - 2008 – create a web form to load directly into database
  - Governor’s Office provided copy of the entire system and installed the system on servers setup for TSLAC
“Because of the possibility that portions of these records fall under Public Information Act exceptions. . . an archivist must review these records before they can be accessed for research.”

(Excerpted from the restriction statement in the finding aid.)
Saturday 8/29/99 at approximately 10:45 P.M., four youths escaped from Beto House. They broke out a window in the bedroom and left the facility.

Two of the youths were apprehended in Pahr at approximately 3:00 A.M. Sunday. According to the police report, this is what occurred:

A police officer observed a car parked on the side of the road. He pulled up behind it and turned on his lights. Four males ran from the car. Two Yamahaos, Jennifer Monios and Nadia Pena who were in the vehicle were questioned. These two females are EROIC employees who had been volunteering at Beto House. Ms. Moreno at Ms. Pena had a visit with two of the youths... The officer that Saturday.

When the officer questioned the women, they stated that the youth had forced their car off the road and told them to sit in it and drive away. Without specifying if they were the youth who were apprehend, the woman stated they could not be sure either of them were the four males involved in the car jacking and refused to press charges of car jacking or robbery.

When the youths were questioned, allegations of a sexual relationship between one of the female staff and a youth were made. The youth stated that they had all four escaped together. They picked Ms. Pena and she picked them in Mission, Texas. They drove them to the location on the road where the officer had pulled up behind the car. They dropped them off on the road side and said they would return in 45 minutes. Ms. Moreno and that is when the officer arrived. The police brought the personal items the youth had escaped with back to the House. The items were in the trunk of Ms. Pena's car.

Both Heidi and Steve Spilman have been notified and the police will be investigating further and have expressed a desire to pursue charges against the women.

Carmen Gonzalez has filed the appropriate alleged mistreatment report due to sexual contact allegations.

Please contact me if you require additional information.

Karen Tuccotti

Redaction of an Outlook calendar entry in a PST file

Redaction of an email
Automated Ingest: Gmail example
Ingest & preservation of email and attachments

Websites

Digital Libraries

Content Contributors
- Archivists/Collection Managers
- Non-expert contributors
- Background ingest (10s GB)
- BagIt

Emails & Attachments

Content & Records Management

Management  Preservation  Access  Admin  Ingest

Storage

Preservica®

CoSA
Council of State Archivists

Preservica
Digital Preservation
Hi Maite,

I have attached the final version of the 5.4 Release Notes. Can you please check and approve the last few changes we discussed in the meeting?

Thanks,

Jack O'Sullivan
Preservica
300 1st Avenue, Suite 202, Needham, MA 02494  Preservica is part of Tessella Group
T: (1) 956 550 0125  E: jack.o'sullivan@preservica.com  W: www.preservica.com  TW: @jOSullivan

This message is confidential and may be privileged. It is intended for the addressee(s) only. Access to this message by anyone else is unauthorized and strictly prohibited. If you have received this message in error, please delete it and notify the sender.

Please note that messages sent or received by the Preservica e-mail system may be monitored and stored in an information retrieval system.

Please consider the environment and do not print this e-mail unless you really need to.

Hi Jack,

Thanks a lot for the changes. All fine now, approved for distribution.

Maïté Braud
Operations Director
Preservica
26 The Quadrant, Abingdon Science Park, Abingdon, Oxfordshire, OX14 3YS  Preservica is part of Tessella Group
T: +44 (0)1235 555511  E: maite.braud@preservica.com  W: www.preservica.com  TW: @BDPreservation
Google Data tools - Account history

Storage

<table>
<thead>
<tr>
<th>Service</th>
<th>Usage</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Drive storage</td>
<td>Using 0 GB of 15 GB (0%)</td>
<td></td>
</tr>
<tr>
<td>Drive</td>
<td>0 GB</td>
<td></td>
</tr>
<tr>
<td>Gmail</td>
<td>0 GB</td>
<td></td>
</tr>
<tr>
<td>Google+ Photos</td>
<td>0 GB</td>
<td></td>
</tr>
</tbody>
</table>

Manage storage plan

Data tools

Dashboard
View account data

Account management
Delete Google+ profile and features
Delete account and data
Delete products

Me on the Web
View options

Inactive Account Manager
Control what happens to your account when you stop using it.
Set up Inactive Account Manager

Recent archives
Mail on November 27, 2014
Create new archive

Activity summary
Created 2 archives in the last 30 days
View history

Download your data
Manage archives
### Workflow Details

- **Workflow Context**: Upload Wizard Ingest
- **Workflow Definition**: Standard Ingest Workflow (Auto Selection)
- **Workflow ID**: 9524
- **Workflow State**: Active
- **Date Started**: 15.07.15 10:25:03
- **Number of Files**: 7
- **Total Size**: 1 MB
- **Collection Code**: code
- **Submission name**: All mail Including Spam and Trash.mbox
- **Top Level Record**: All mail Including Spam and Trash.mbox

### Step Progress

<table>
<thead>
<tr>
<th>State</th>
<th>Name</th>
<th>Progress</th>
<th>Started</th>
<th>Finished</th>
<th>Messages</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>Identify Packages</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>Import from Transfer Area</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>Process Package</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>Virus Check</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>Foxy Check</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>Metadata Integrity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>Content Integrity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>SIP Validation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>SIP Validation with Database Crosscheck</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>Characterise</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>Store Files</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>Store Metadata</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Hi Maite,

I have attached the final version of the 5.4 Release Notes. Can you please check and approve the last few changes we discussed in the meeting?

Thanks,

Jack O'Sullivan
Preservica
300 1st Avenue, Suite 202, Needham, MA 02494C2 Preservica is part of Tessella Group   jack.o'sullivan@preservica.com  W: www.preservica.com  TW: @dPreservica

This message is commercial in confidence and may be privileged. It is intended for the addressee(s) only. Access to this message by anyone else is unauthorized and strictly prohibited. Please note that messages sent or received by the Preservica e-mail system may be monitored and stored in an information retrieval system. Please consider the environment and do not print this e-mail unless you really need to.
Email Preservation

• Long-term preservation
  – original email format retained
  – all email is migrated to the open EML format at Ingest
  – attachments are protected from format obsolescence
  – conversational relationships are retained

• Accessible
  – emails are displayed in a familiar format for ease of access
  – access control for privacy and security

• Searchable
  – all metadata is ingested
  – filter on email specific fields

• Can be categorised/tagged for future action

• Can be shown to be “trustworthy”
Email Archiving & Preservation Summary

- Default form of correspondence
- A requirement and a challenge
- Archiving vs Long-term Preservation
- The Texas Digital Archive challenges
- Automation and tools are vital
Practical Next Steps

- Preservica webinar recordings
  [www.preservica.com/resources/webinars/#publish](http://www.preservica.com/resources/webinars/#publish)
  - Outlook
  - Gmail
  - SharePoint

- Practical Digital Preservation Online Workshop
  - Part 1 – 8th March 2016, 2-4pm Eastern
  - Part 2 – 15th March 2016, 2-4pm Eastern

- PDP “How To” Webinar
  - *New Ways of providing Public Access to your archive*
  - Tuesday, April 26, 2016; 2-3pm Eastern

Questions?
Thank You!

www.preservica.com
info@preservica.com
@preservica
@dPreservation

www.statearchivists.org/