

CoSA & Preservica Practical Digital Preservation 2015/16



Practical OAIS Digital Preservation Online Workshop Module 2

Practical Digital Preservation 2015/16

- **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



Sarah Koonts
CoSA

- **PDP “Hot Topic” Webinars**

(Tuesdays 2-3pm Eastern)

- | | |
|--|-------------|
| • Achieving ISO Standards for your digital archive | Oct 28 2015 |
| • 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 |

<http://rc.statearchivists.org/Content/Electronic-Records/Education-Training/CoSA-Preservica-Practical-Digital-Preservation.aspx>

Workshop Objectives

Michael Hope
Preservica



Module 1 (last week)

Understand the **fundamentals of Digital Preservation** - moving beyond the main acronyms and theory by illustrating topics with examples and demonstrations of practical real-world digital preservation workflows and processes

Module 2 (today)

Understand **how Digital Preservation fits into the Information Governance lifecycle** – including content ingested from other systems (e.g. long-term records and emails) - as well as how to provide greater “transparency” through controlled access to information for internal and public users

Digital Content and Records are Fragile



- ✓ Need to protect from loss and degradation
- ✓ Need to proactively manage file formats & metadata
- ✓ Digital content already 10+ years old or needs to be retained for 10+ years is at **risk**

Long-Term Records

Example non-permanent records	Keep for
Personal health records	Must not be destroyed, or deleted, for the foreseeable future
Employee Health records	Review 9 and then 25 years
Public contract records	30 years
Pension Corporate records	25 years
Pension personal record	Until 100th Birthday
Student records	Life of student' (which is taken to be 120 years from date of birth)

Digital Records – Short Term



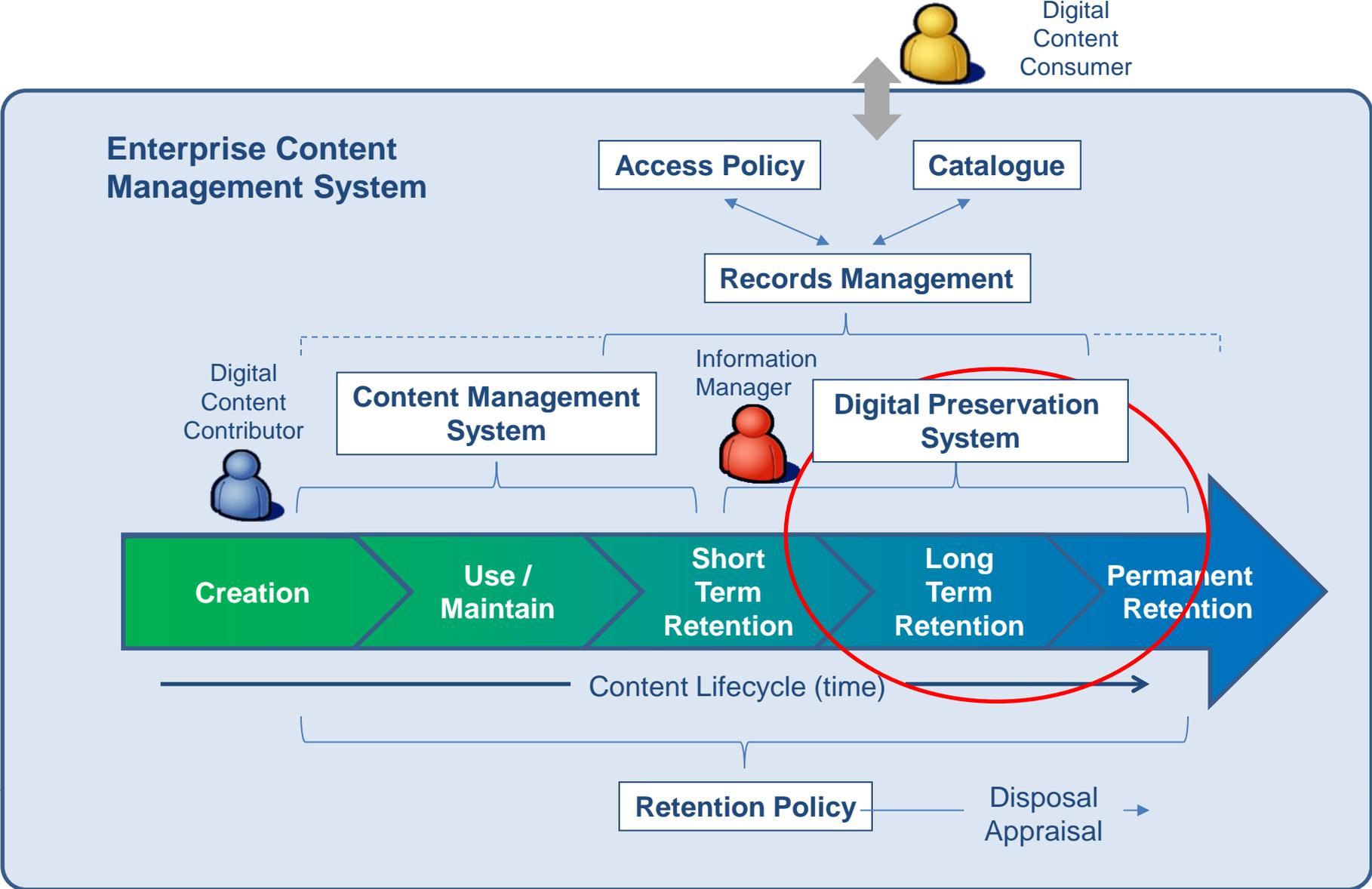
- Structured management of operational content
- Short term retention
- **Protect the file ‘bits’**

Digital Records – Long Term

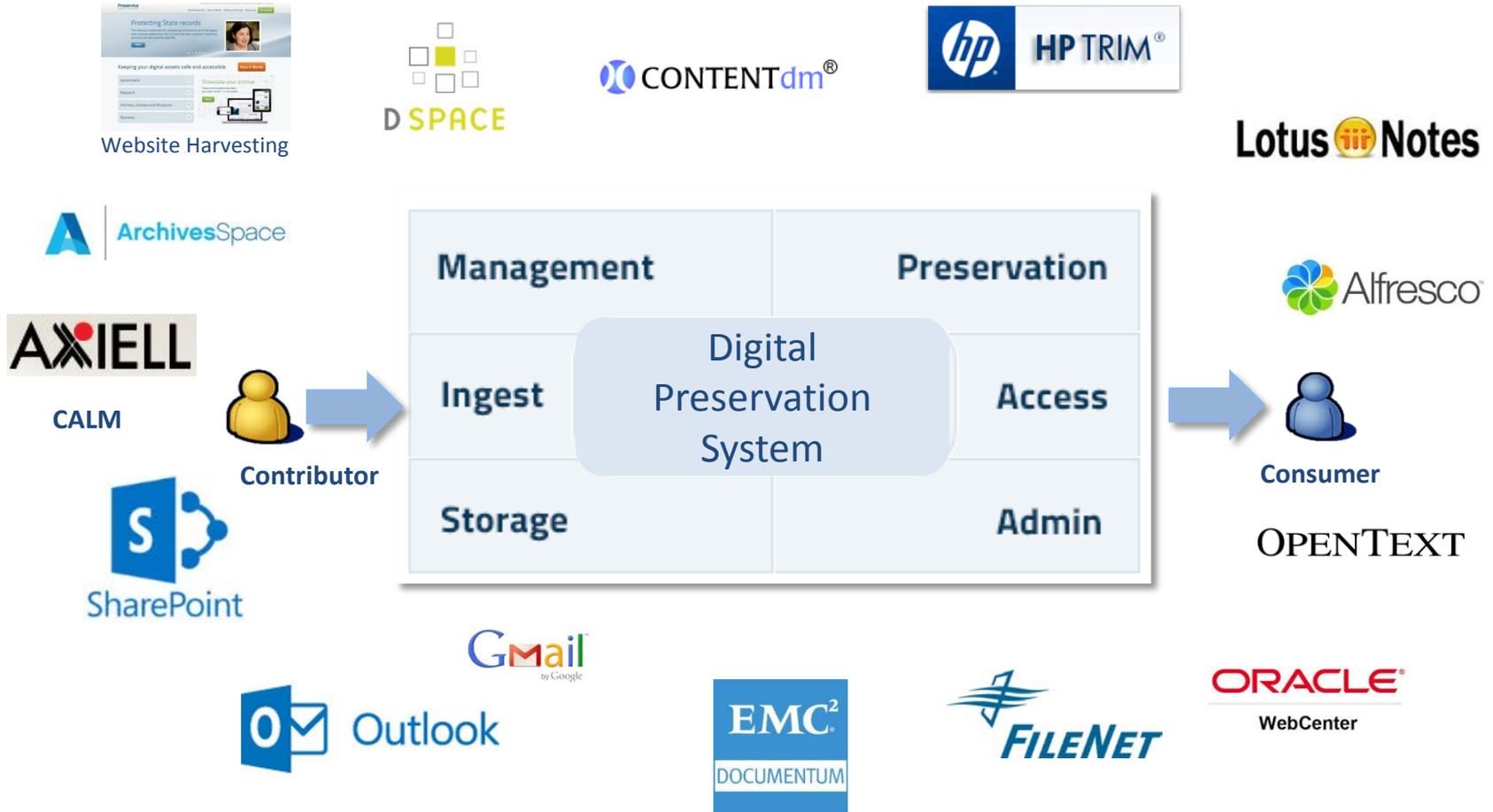
Digital Preservation
Digital Preservation System
Long-term usability and readability of digital content
Protection against file format obsolescence
Long-term and permanent retention (> 10 years)

- Long term & permanent retention of digital content
- Structured and unstructured content
- **Protect file accessibility**
- OAIS (ISO: 14721)

Making Digital Preservation part of the Information Governance Lifecycle



Making Digital Preservation Part of the Information Lifecycle



Practical OAIS Digital Preservation



Jack O'Sullivan

Technical Consultant, Preservica

Recap : Module 1

Digital Preservation Fundamentals

Module 1

Session 1

- Why do we need Digital Preservation?
- The fundamentals of preserving digital content

Session 2

- Understanding Metadata, Fixity and File Characterization
 - including example demonstration

Agenda : Module 2

DP in Information Governance Lifecycle

Module 2

Session 3

- Preservation Planning and Action
 - Ingest and Preservation long-term records from SharePoint

Questions

Session 4

- Controlling access to digital content – including practical demonstration
- Ingesting and Preserving Complex formats
 - Website harvesting (WARC files)
 - Emails and Attachments – demonstration of record classification & action

Questions & Close

- Next Steps and Close

Digital Records: What's Important?

- Authenticity
- Provenance

Module 1

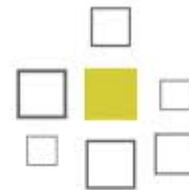
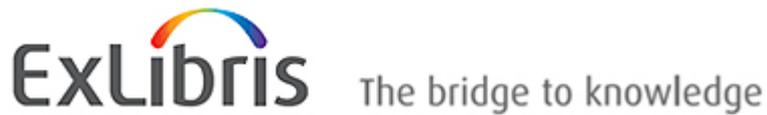
- Preservation
- Retention & Disposition
- Access, Security & Privacy

Module 2

Workshop demonstrations

The live practical examples use the Preservica Preservation system

Other systems are available:



DSPACE

Open Source
Digital Repository Application



Session 3

How do we plan and act to mitigate digital preservation issues?



Planning to Have a Plan



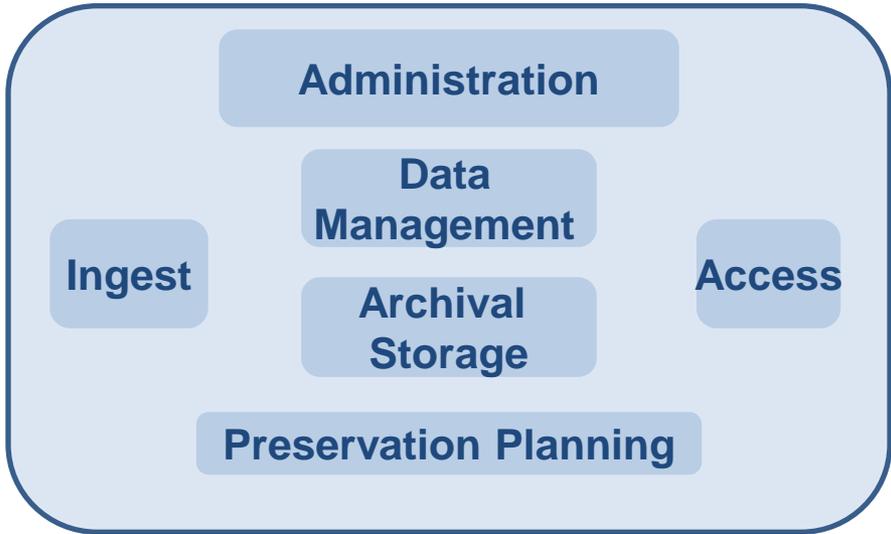
Websites



CALM



SharePoint



DEMO: MIGRATING CONTENT TO AN OAIS REPOSITORY



Passive Preservation

Passive Preservation is concerned with the secure storage of digital objects.

- Security and access control
- File integrity
- Storage management
- Media selection & refresh
- Disaster recovery



Home > [Electronic Records](#) > [SERP Framework](#) > [Strategy](#)

Strategy

> [Education-Training](#)

> [SERP Framework](#)

Policy

Strategy

Strategy Details

Governance

Collaboration

Technical Expertise

Open Standards / Neutral Formats

Designated Community

Electronic Records Survey

Ingest

Storage

Device / Media Renewal

DEFINITION: The organization charged with the preservation of permanent electronic government records must proactively mitigate the risks associated with technology obsolescence including plans related to periodic renewal of storage devices, storage media, and adoption of preferred preservation file formats.

Level 0 The Archives/RM unit does not have a plan to address technology obsolescence.

Level 1 The Archives/RM unit accepts electronic records in their native format on an ad hoc basis with the expectation that new software will become available to support these formats. *Organizations must be aware that not all formats they accept in this manner might be accessible over time. "Preservation ready" policies can address which formats are acceptable and which are not.*

Level 2
 Level 2a: The Archives/RM unit encourages records producers to retain records of long-term value in preservation-ready file formats. *Preservation formats will depend on your e capabilities.*
 Level 2b: The Archives/RM unit proactively and systematically monitors changes that may impact the digital records collections and the archival repository.

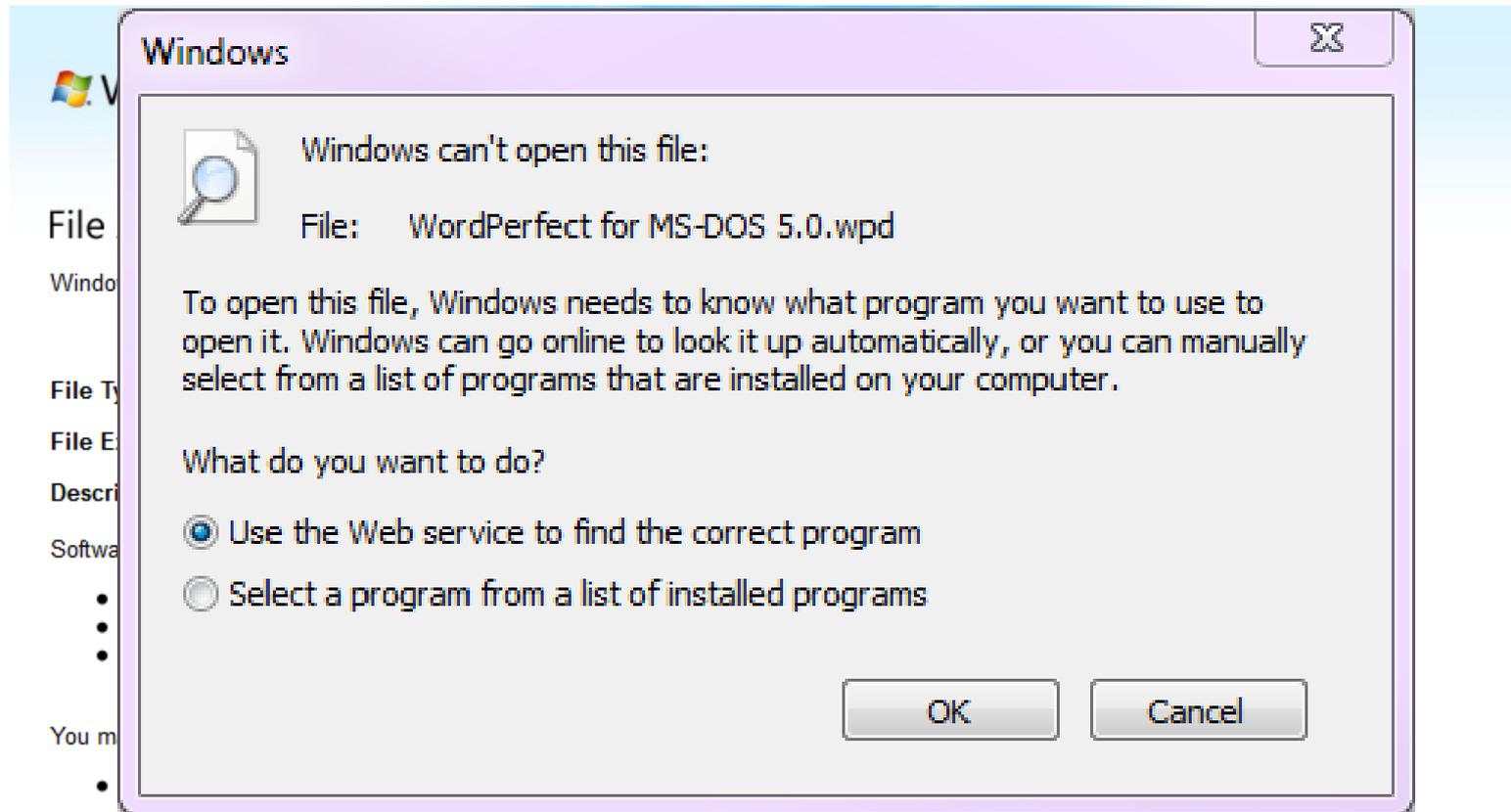
Level 3 The Archives/RM unit implements the transformation of selected native file preferred.

Level 4 The Archives/RM unit proactively and systematically monitors changes that may impact the digital records collections and the archival repository.

DEFINITION: A key capability in ISO 14721 conforming digital records in its custody, which involves two related preservation actions: cryptographic hash algorithm that normalizes any digital object regardless of length bit stream (e.g., 156 bits). This fixed length bit stream is called a digital fingerprint.

DEFINITION: ISO 14721 delineates systematic automated storage services that ensure successful transfer of Archival Information Packages (AIPs) from ingest, creation, or migration (PDI) for each AIP that confirms its "body" during any preservation action. This includes the maintenance of a separate level of metadata for PDI and facilitates transformation of preservation-ready AIPs to a format suitable for long-term storage.

The Risk: Format Endangerment



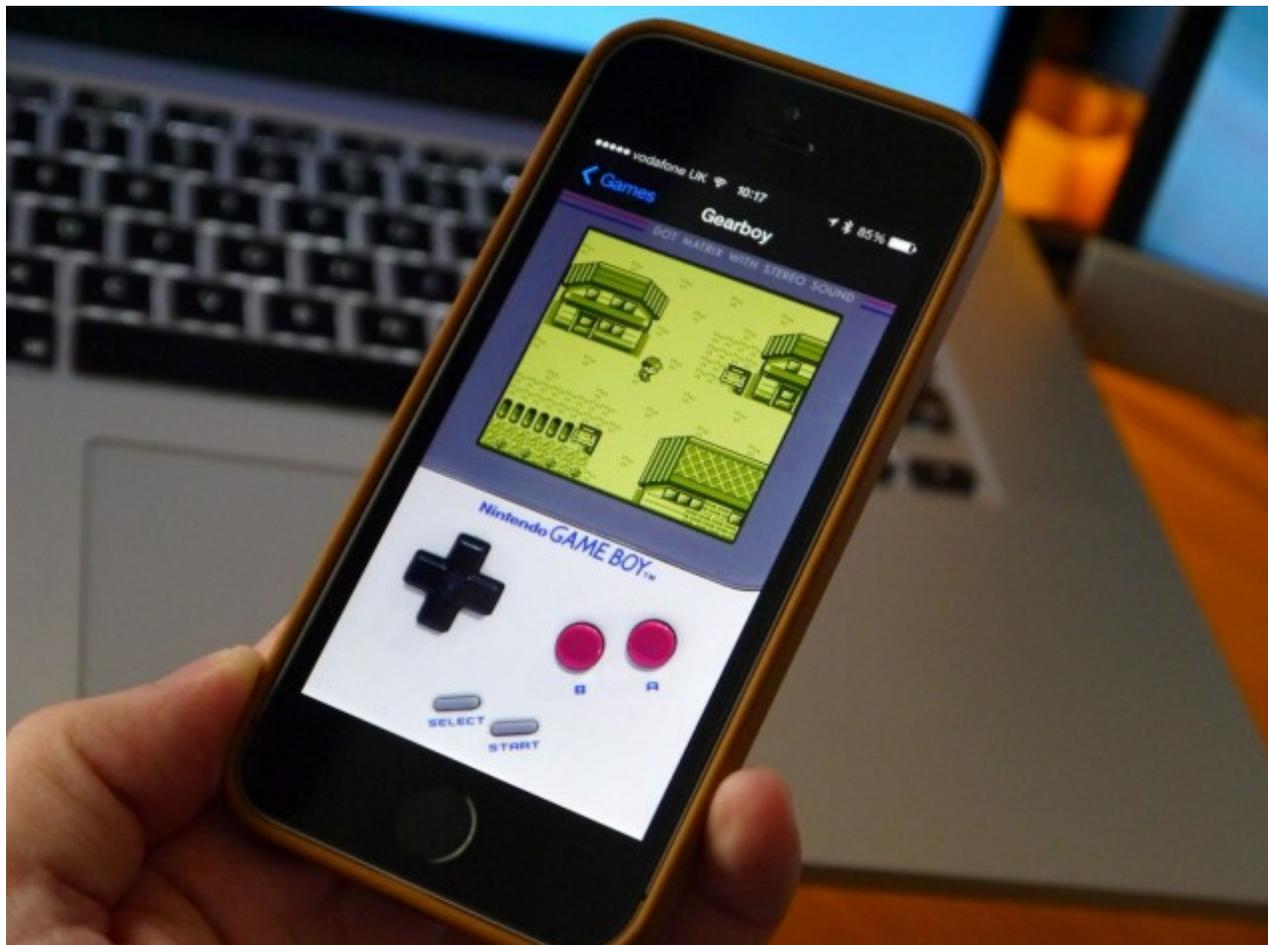
Active Preservation Planning

Active Preservation is concerned with the preservation of the underlying information.

- Institutional risk register
 - Is the information in a legacy system?
 - Can we identify it?
 - Can we validate?
 - Do we have rendering software?
 - Is software widespread in the designated community?
 - Is it a binary format?
 - Is it a proprietary format?
 - Is it well documented?
- Identify alternative formats
 - Do we have software that generate those formats?
- Identify rendering tools

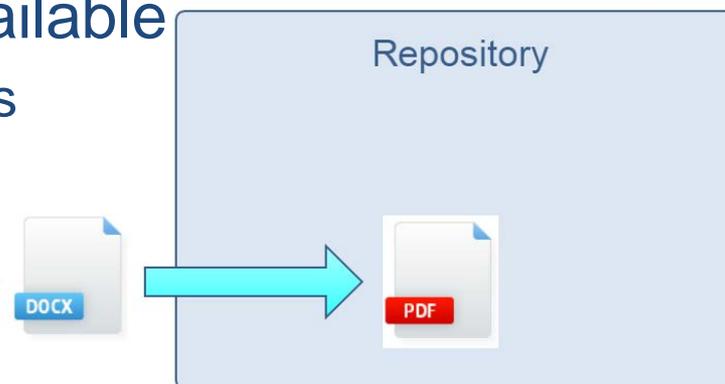


Emulation



Normalisation

- Support a small number of file formats
- Migrate on ingest to these formats
- Enhance control of your repository
- Migration pathways not always available
 - You may end up rejecting submissions
- Always likely to incur some loss

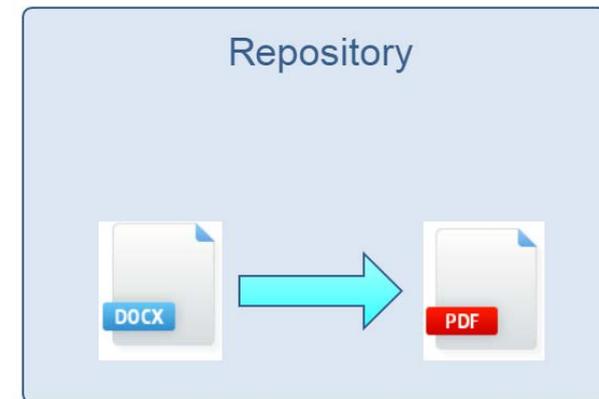


DEMO: RUNNING A NORMALISATION INGEST



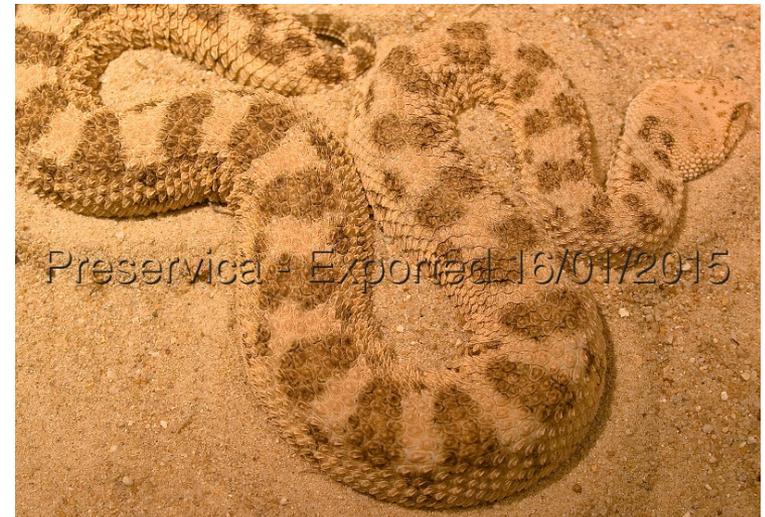
Migration

- Transform information from one file format into another
- Guard against format obsolescence
- Delays transformation until needed
- Always likely to incur some loss
- On-going process



Reasons for Migration

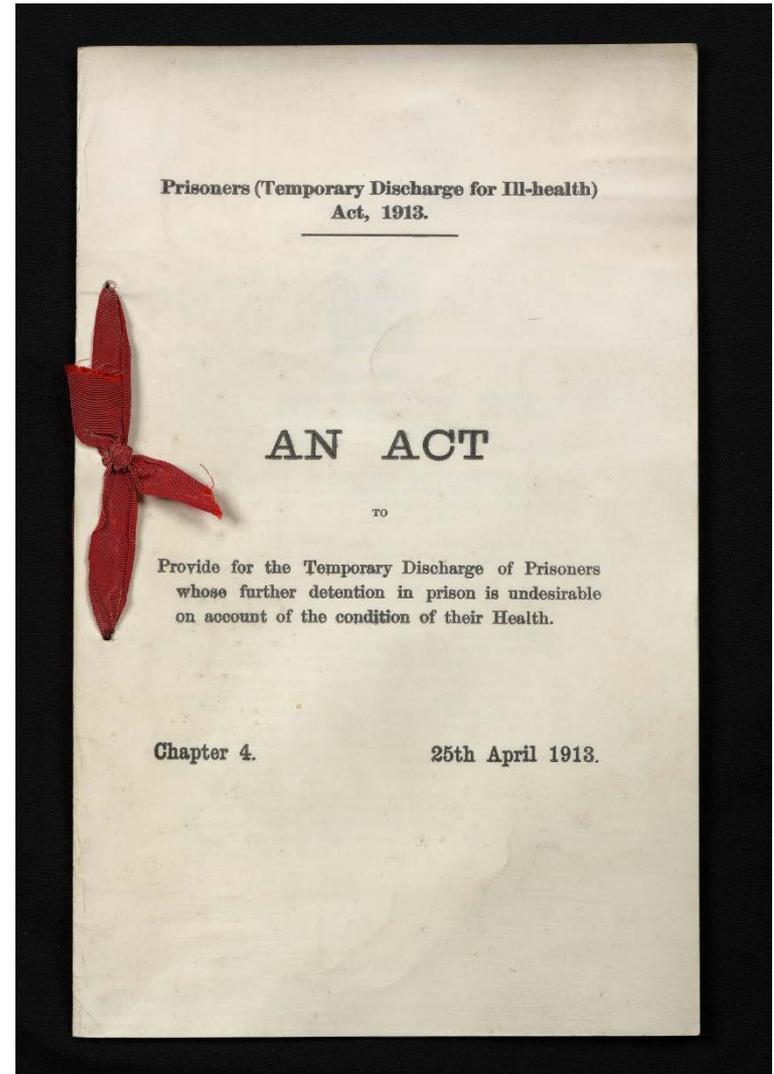
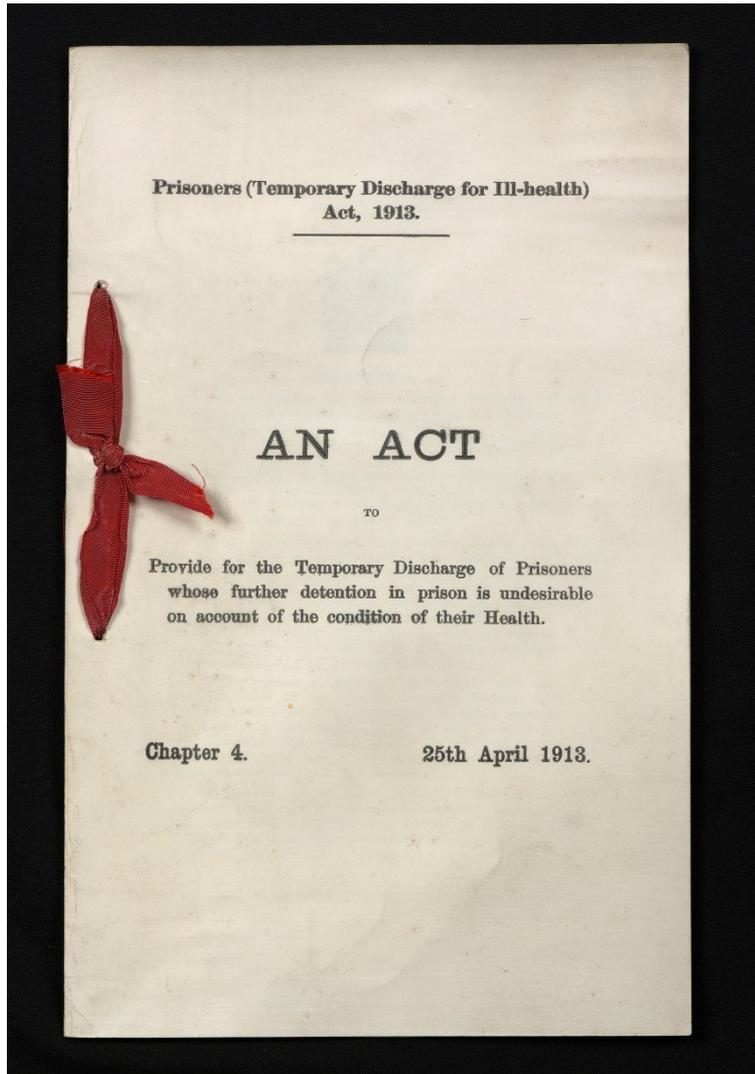
- Preservation
 - The file is unsupported in its current format
- Presentation
 - We can support the format, but it's not very useful for widespread dissemination



Mitigating Loss

- Changing format will normally lead to some loss of fidelity
- Validate – Can we accept the compromise?
 - Manually, small scale, run tests
 - Automatically, bulk migrations

Acceptable Loss



DEMO: FILE FORMAT MIGRATION



Questions ?

& short break



Project Files 1995 - All Doc... x

sdb4:8080/_layouts/15/start.aspx#/Project Files 1995/Forms/AllItems.aspx

SharePoint Newsfeed OneDrive Sites Jack O'Sullivan

BROWSE FILES LIBRARY

low EDIT LINKS

Project Files 1995

Home Documents Recent

Project Files 1995

PDP_Workshop

Sharepoint(R P test)

SingleCmp

Large SIP(100 GB)

Test

New

EDIT LINKS

+ new document or drag files here

All Documents Find a file

Name	Modified	Modified By
Geology ✱	2 hours ago	James Carr
Land_use_permit ✱	2 hours ago	James Carr
Lower Delaware River Drainage ✱	2 hours ago	James Carr
Sociology ✱	2 hours ago	James Carr
Usernet_news ✱	2 hours ago	James Carr

Search this site

Backup and Restore

sdb4:19992/backups.aspx

SharePoint Newsfeed OneDrive Sites Jack O'Sullivan

BROWSE PAGE SHARE

lowrtest

Backup and Restore

Central Administration

- Application Management
- System Settings
- Monitoring
- Backup and Restore**
- Security
- Upgrade and Migration
- General Application Settings
- Apps
- Office 365
- Configuration Wizards

 **Farm Backup and Restore**
[Perform a backup](#) | [Restore from a backup](#) |
[Configure backup settings](#) | [View backup and restore history](#) |
[Check backup and restore job status](#)

 **Granular Backup**
[Perform a site collection backup](#) | [Export a site or list](#) |
[Recover data from an unattached content database](#) |
[Check granular backup job status](#)

Site Or List Export

sdb4:19992/_admin/SiteAndListExport.aspx?WebId=d7f5c9fb-48ac-4cd3-8615-ad!

SharePoint Newsfeed OneDrive Sites Jack O'Sullivan

lowrtest

Site Or List Export

Central Administration

- Application Management
- System Settings
- Monitoring
- Backup and Restore
- Security
- Upgrade and Migration
- General Application Settings
- Apps
- Office 365
- Configuration Wizards

Readiness

- No export is in progress.
- Timer service is running.

Site Collection

Select a site or list to export. First select the site collection that contains the site, then select a site to export. To export a list, select the site collection and site that contain it, and then select the list.

Site Collection: <http://sdb4:8080>

Site: /

List: No selection

Change List

File location:

Specify the destination for the export package.

Filename:

Overwrite existing files

Example: \\backup\SharePoint\export.cmp

Export full security

Export Full Security

Export full security of the site, including author, editors, created by times, and modified by times. This also includes all users in the sites.

Export Versions

Select the version history information to

Export versions

Site Or List Export

sdb4:19992/_admin/SiteAndListExport.aspx?WebId=d7f5c9fb-48ac-4cd3-8615-ad

SharePoint Newsfeed OneDrive Sites Jack O'Sullivan

lowrtest

Site

Central Administration

- Application Management
- System Settings
- Monitoring
- Backup and Restore
- Security
- Upgrade and Migration
- General Application Settings
- Apps
- Office 365
- Configuration Wizards

Select List - Mozilla Firefox

sdb4:19992/_admin/SelectList.aspx?SiteId=0c15a83a-f737-4d17-8682-c37789a8e740&WebId=d7f5c9fb-48

Title Search Project 21-30

Title ↑	URL
Master Page Gallery	/Project Files 1995
MicroFeed	Title Project Files 1995
Mixed Files	Description
Multiple format files	
NewTestLibrary	
PDP_Workshop	
Project Files 1995	
Project Policy Item List	
Share Point Retention Policy Test	
Share Point Retention Policy Test1	

OK Cancel

Export Versions

Select the version history information to Export versions

Site Or List Export

sdb4:19992/_admin/SiteAndListExport.aspx?WebId=d7f5c9fb-48ac-4cd3-8615-ad

SharePoint Newsfeed OneDrive Sites Jack O'Sullivan

Monitoring site collection that contains the site, then select a site to export. To export a list, select the site collection and site that contain it, and then select the list.

Backup and Restore

Security

Upgrade and Migration

General Application Settings

Apps

Office 365

Configuration Wizards

File location: Specify the destination for the export package.

Filename: E:\SharepointDemo\ProjectFiles.cmp

Overwrite existing files

Example: \\backup\SharePoint\export.cmp

Export full security

Export Full Security

Export full security of the site, including author, editors, created by times, and modified by times. This also includes all users in the sites.

Export Versions

Select the version history information to include for files and list items. You can include all versions, the last major version, the current version, or the last major and last minor versions.

Export versions: All Versions

Start Export Cancel

Granular Backup Job Status

sdb4:19992/_admin/SiteBackupOrExportStatus.aspx

SharePoint Newsfeed OneDrive Sites Jack O'Sullivan

lowrtest

Granular Backup Job Status

Readiness

- No site collection backup is in progress.
- An export is currently in progress.
- Timer service is running.

Refresh | Delete Export Job

Site Collection Backup

Current Job

Status	No operation in progress.
--------	---------------------------

Previous Job

Status	No previous job.
--------	------------------

Content Export

Current Job

Status	Operation initializing.
Requested By	PRESERVICA\osuj
Site Collection URL	http://sdb4:8080
Server Relative URL	/Project Files 1995
Filename	E:\SharepointDemo\ProjectFiles.cmp
Log file generated:	E:\SharepointDemo\ProjectFiles.cmp.export.log
Overwrite	Yes
Recovery Step	To recover the data use the PowerShell import command Import-SPWeb. For more details, type Import-SPWeb -? at the PowerShell command prompt.

Previous Job

Status	Succeeded
--------	-----------

Central Administration

- Application Management
- System Settings
- Monitoring
- Backup and Restore
- Security
- Upgrade and Migration
- General Application Settings
- Apps
- Office 365
- Configuration Wizards

Session 4:

Controlling access to digital content

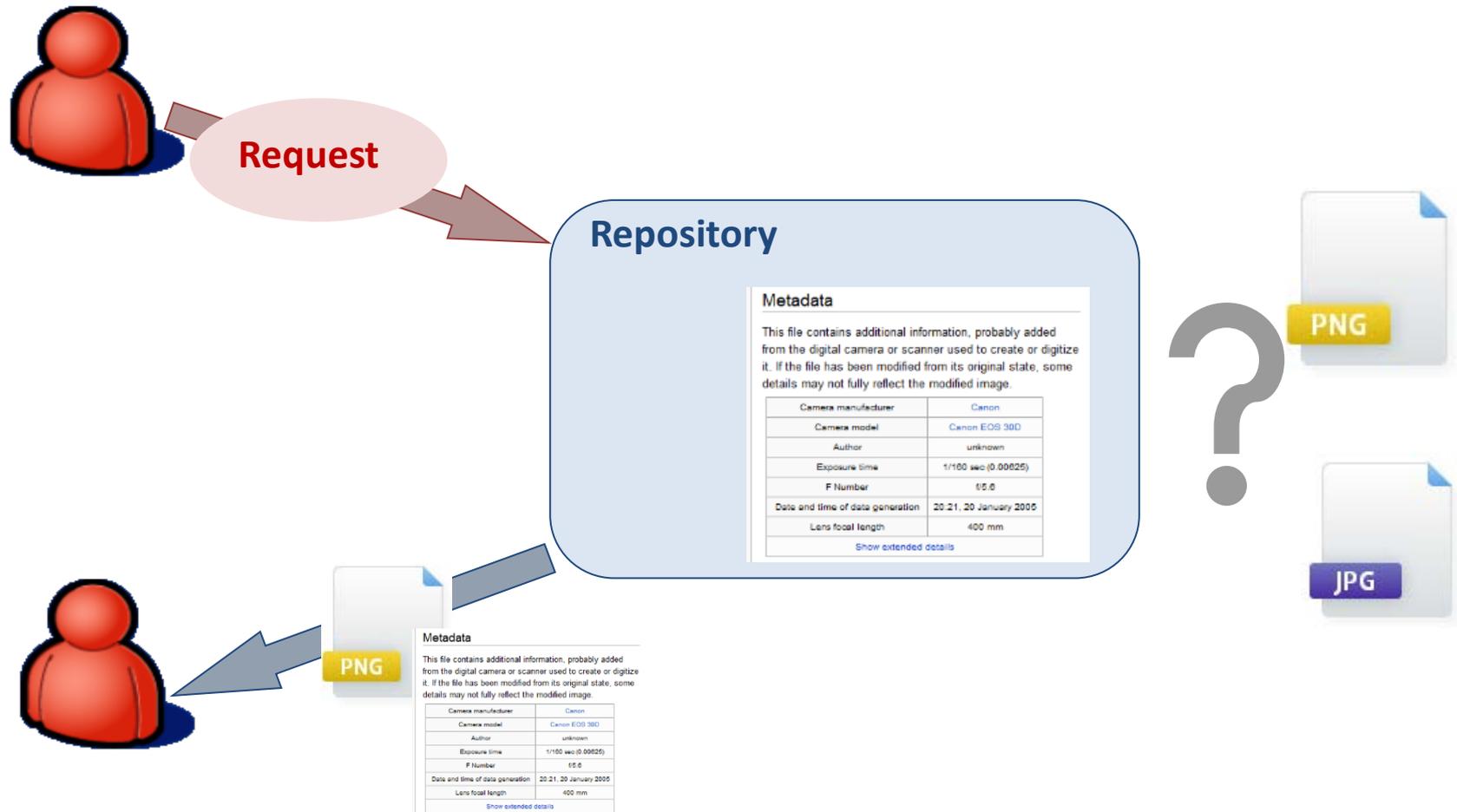


Access Control

Pulling it all together: Preservation of emails as a complex digital record



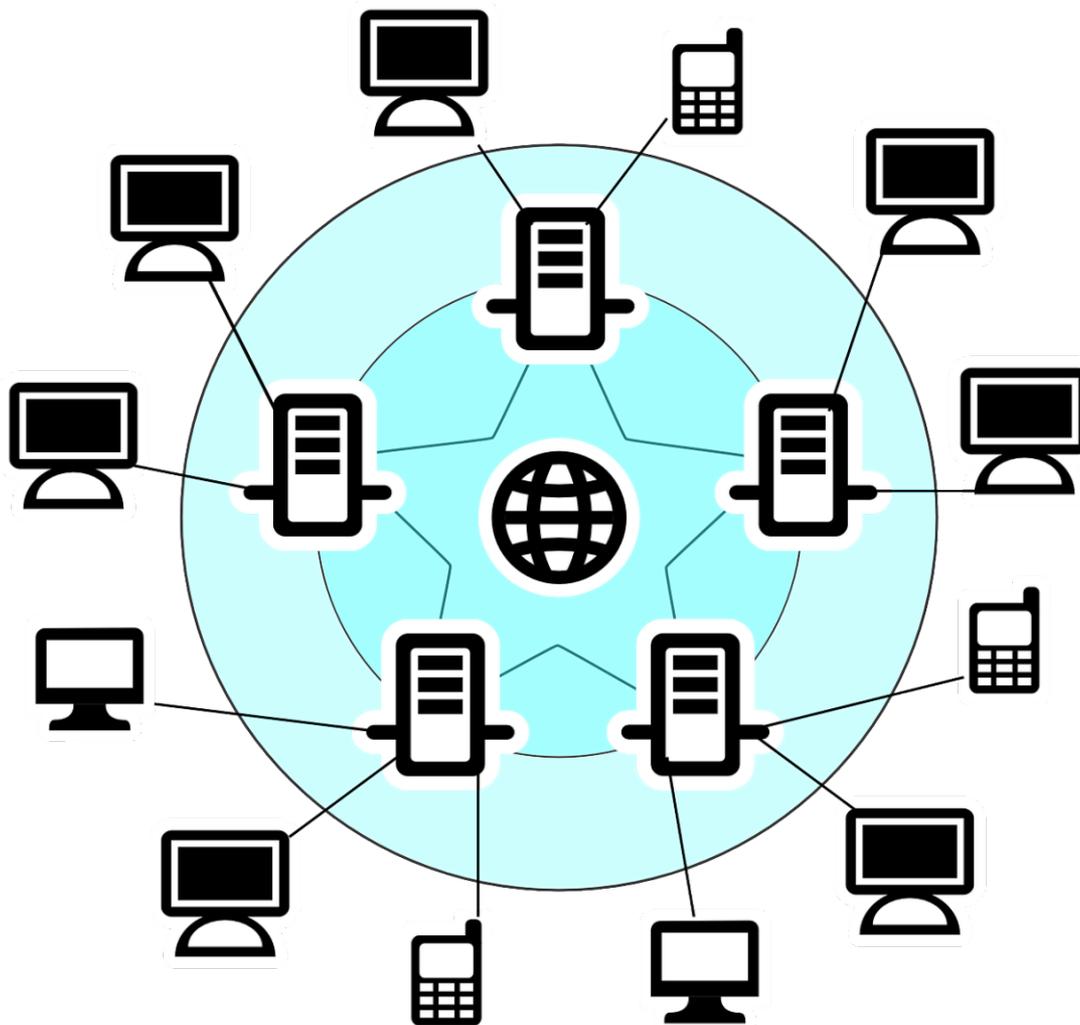
Accessing Packages



DEMO: DIP EXPORT



Widespread Access



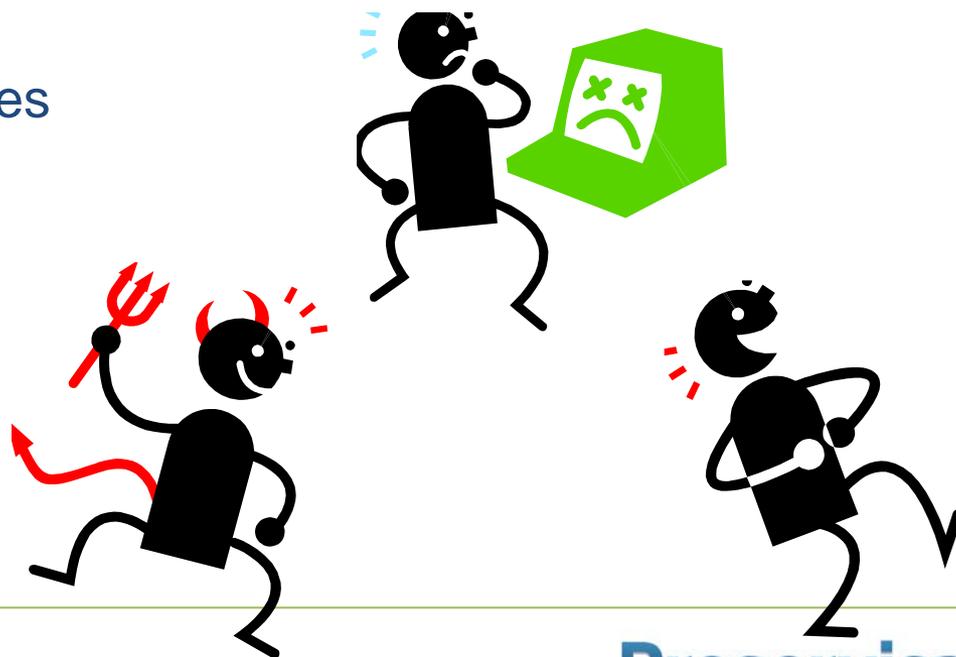
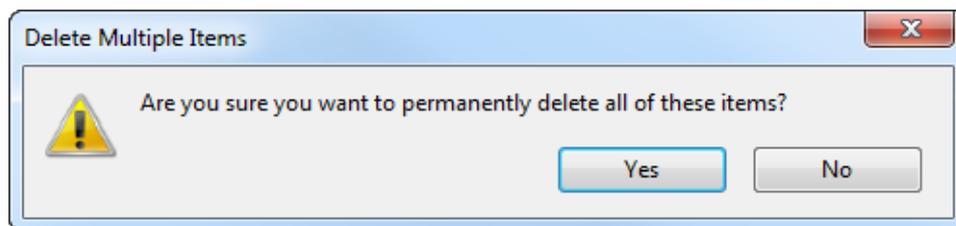
What are you protecting?

- Hardware/equipment
- Digital files
- Content of the digital files
 - Including sensitive information
- *Your investment!*
- *Your reputation!*



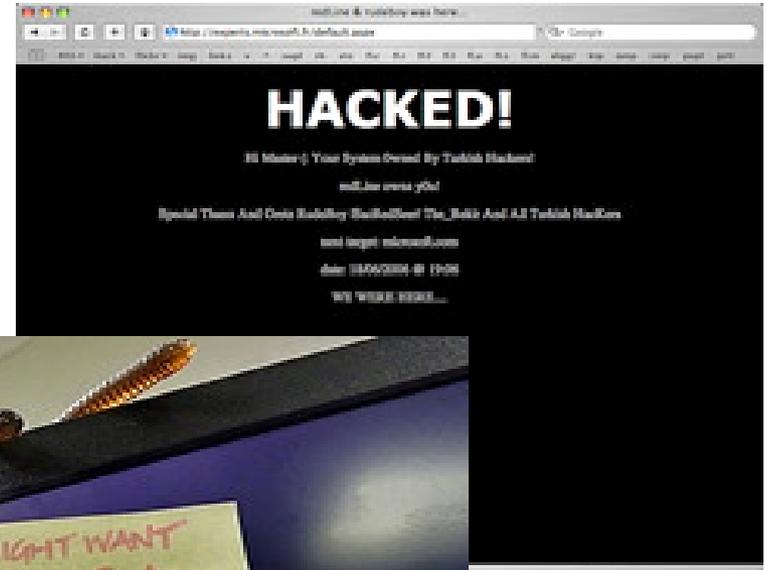
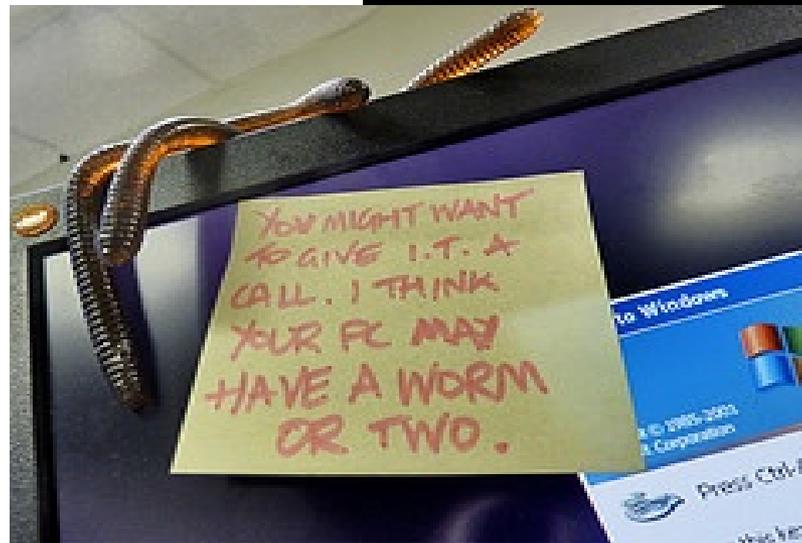
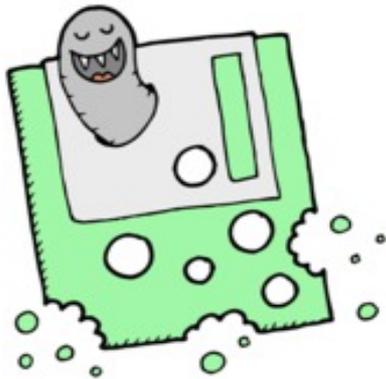
Protect Content from Ourselves

- Accidental
 - Deletion
 - Saving over original
- Intentional
 - Disgruntled employee causes harm
 - Abusing permissions to change/remove files
 - Misuse of sensitive information



Inappropriate Access

- Hackers
- Viruses
- Trojan horses
- Worms



> Education-Training

> **SERP Framework**

- Policy
- Strategy
- Governance
- Collaboration
- Technical Expertise
- Open Standards / Neutral Formats
- Designated Community
- Electronic Records Survey
- Ingest
- Storage
- Device / Media Renewal
- Integrity
- Security**

Security Details

- Preservation Metadata
- Access

> Related/Additional Framework Topics

> Resources

Security

DEFINITION: Contemporary enterprise-wide information systems typically execute a number of shared or common services that may include inter-process communication, name services, temporary storage allocation, exception handling, role based access rights, security, backup and business continuity, and directory services, among others. An ISO 14721/ISO 16363 conforming archival repository is likely to be part of an information system that may routinely provide some or perhaps all of the core security, backup, and business continuity services including firewalls, role based access rights, data transfer integrity validations, logs for all preservation activities, including failures and anomalies to demonstrate an unbroken chain of custody.

Level 0	Currently, the archival repository does not have formal disaster recovery, backups, or firewall procedures in place to protect the security of electronic records.
Level 1	The security of electronic records in the archival digital repository is protected through disaster recovery procedures.
Level 2	The security of electronic records in the archival digital repository is protected through a comprehensive firewall protection.
Level 3	The security of electronic records in the archival digital repository is protected through a comprehensive role based access rights management.
Level 4	The archival repository continuously monitors security protection processes and revises them in response to evolving technology capabilities and changing business requirements.

<http://rc.statearchivists.org/Content/Electronic-Records/SERP-Framework/Security.aspx>

Security Models

- Classes of content
- Classes of user



Public



Staff



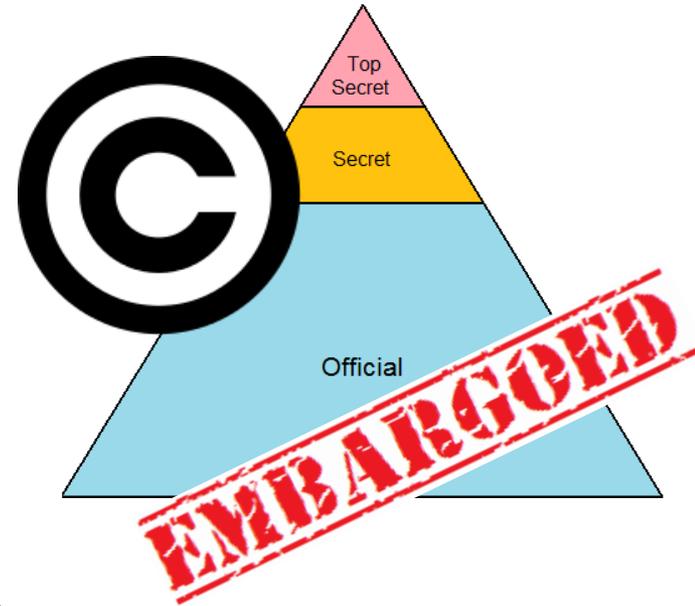
Patrons



Archivist



Researchers



- Classes of operation
 - Policy:
 - User of class X has permission to perform operation of class Y on content of class Z
- See description Edit description
Read content

DEMO: CONTROLLING ACCESS



Long-Term Non-Permanent Records

Example non-permanent records	Keep for
Personal health records	Must not be destroyed, or deleted, for the foreseeable future
Employee Health records	Review 9 and then 25 years
Public contract records	30 years
Pension Corporate records	25 years
Pension personal record	Until 100th Birthday
Student records	Life of student' (which is taken to be 120 years from date of birth)

Permanent “Long Term” Records

Example records	Long Term Issue
Copyrighted Material	Restricted access for the lifetime of the copyright
Classified Material	Declassified after 10-75 years, reviewed at 25 yearly intervals
Donor restricted material	Possibly restricted for the lifetime of the donor or other person.
Unprocessed material	Held until processed

Classification & Policy

- Automatically classify and set access permissions during ingest



- Search by classification



- Automate future actions (e.g. appraise, delete) based on policy rules

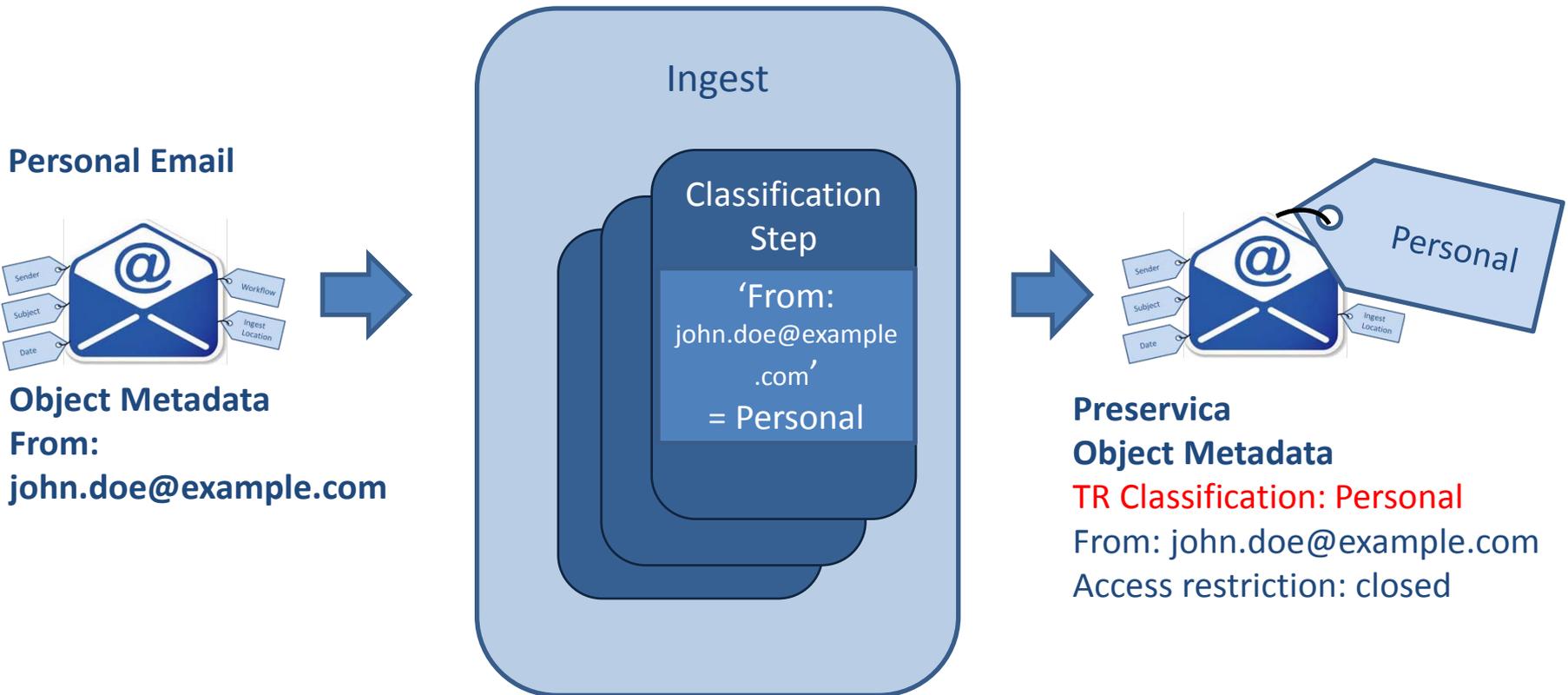
EXAMPLE: EMAIL APPRAISAL

Automated Appraisal of Email - Classifying

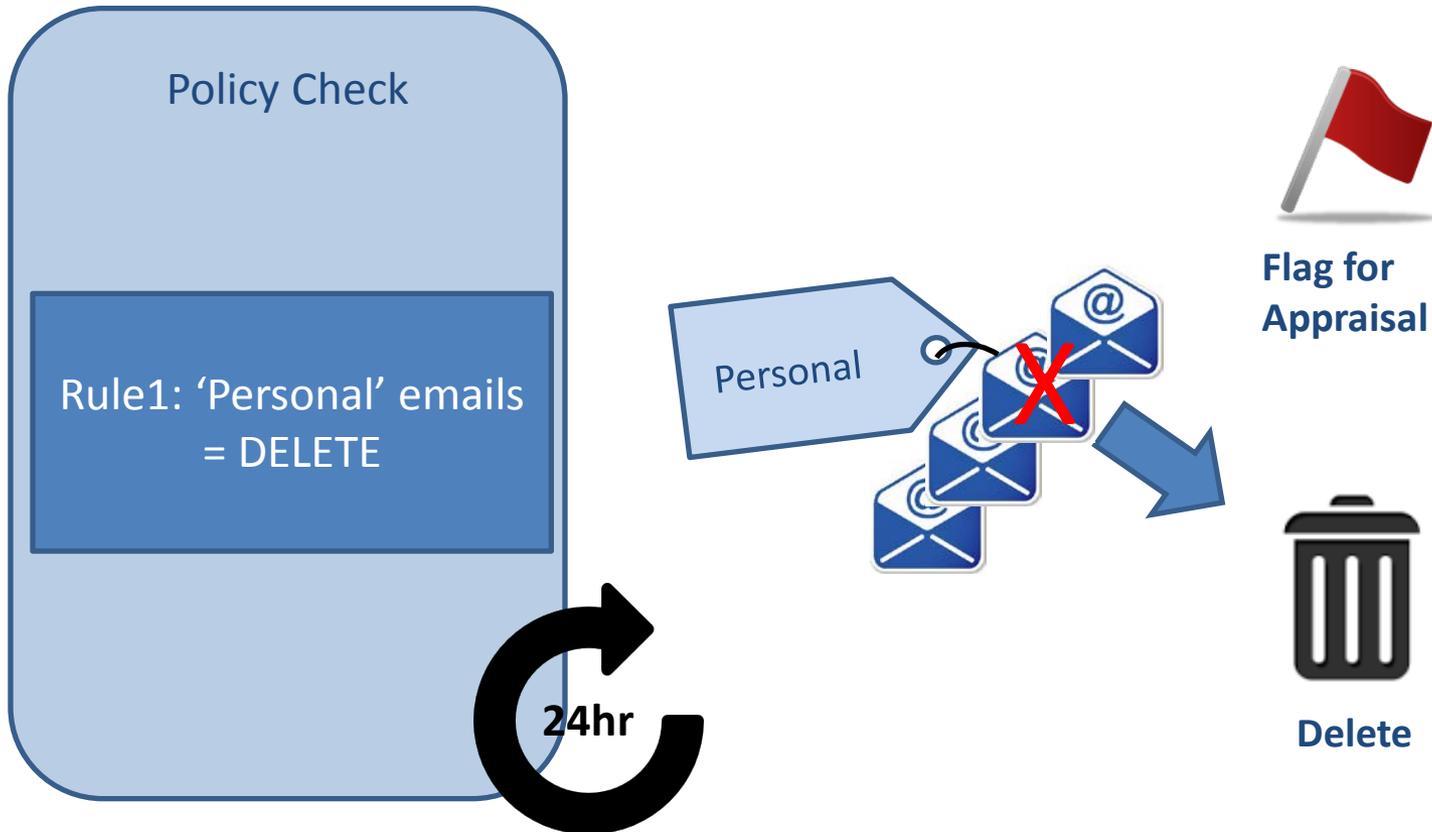


What do we know ?

Automated Appraisal of Email - Classifying



Automated Appraisal of Email - Policy



PRESERVATION USE CASE: COMPLEX RECORDS

WEB & EMAIL

Why do we need Web Preservation?

- The web is increasingly the record of our collective memory



What makes it so hard?

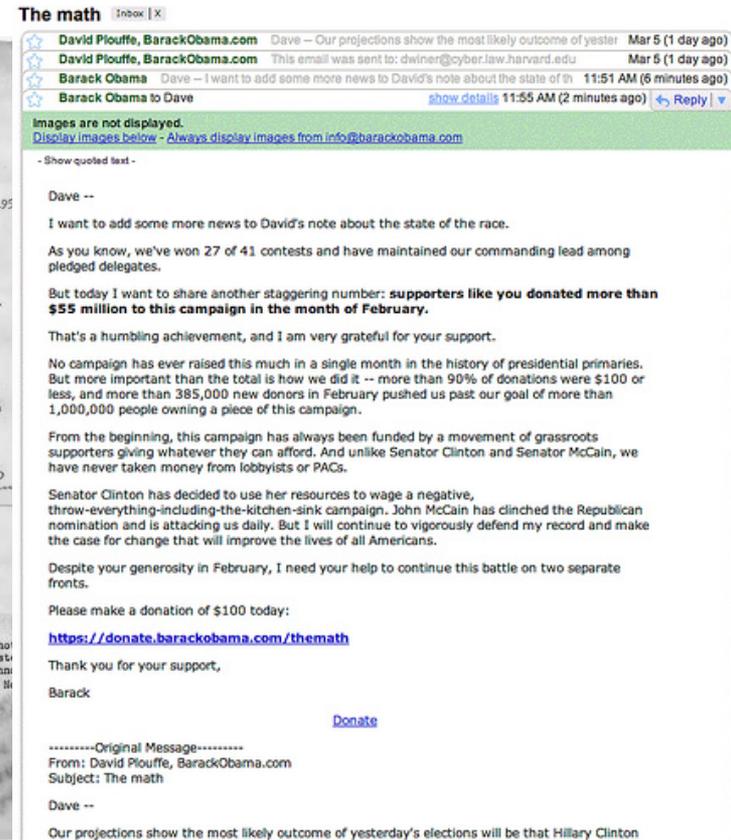
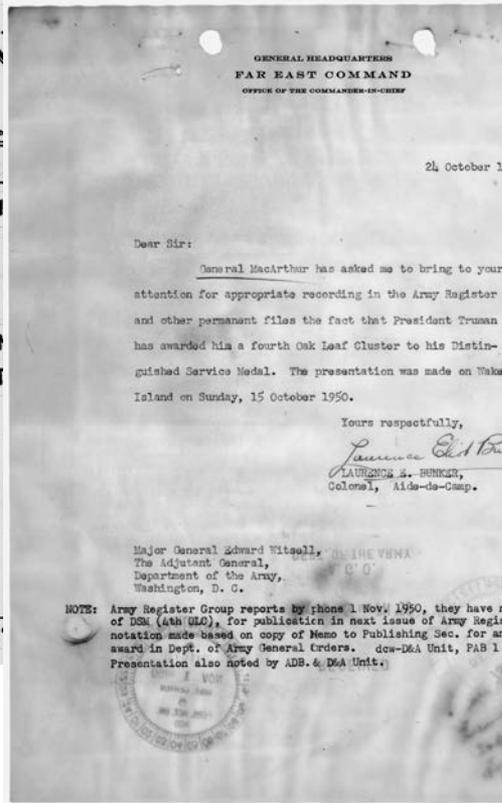
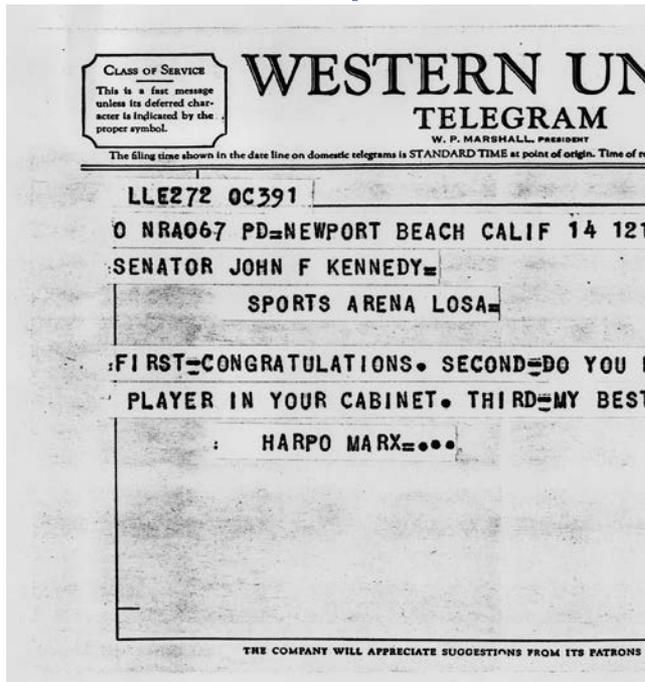
- 25 years of evolution



About Blog Help Status Jobs Terms Privacy Cookies Ads Info Brand Advertise Businesses Media Developers Directory © 2015 Twitter, Inc.

Why do we need Email Preservation?

- Email is increasingly the default medium of correspondence



What makes it so hard?



Hotmail[®]

Gmail[™]
Google



Attachments



DEMO: EMAIL PRESERVATION



Overall Recap

Module 1

- ✓ Why do we need for digital preservation
- ✓ The fundamentals of preserving digital content
- ✓ The main standards for digital preservation
- ✓ Metadata, Fixity and Characterisation

Module 2

- ✓ Digital Preservation Planning and Action
- ✓ Controlling Access to digital content
- ✓ Part of the information lifecycle (SharePoint and eMail)
- ✓ Management of long-term non-permanent records
- ✓ Handling Complex formats (like eMail and Websites)

Digital Records: What's Important?

- Authenticity
- Provenance

Module 1

- Preservation
- Retention & Disposition
- Access, Security & Privacy

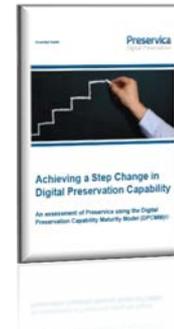
Module 2

Questions ?



Next Steps

- Next webinar: 2-3pm Eastern , Tuesday December 08 2015
 - Ingesting in multiple formats and from multiple systems
- Achieving a Step Change in Digital Preservation Capability
<http://preservica.com/resource/essential-guide-achieving-step-change-digital-preservation-capability/>
- Safeguarding your vital long-term electronic records
<http://preservica.com/resource/electronic-records-preservation/>
- www.preservica.com/resources



Next Steps

Workshop Objectives

Understand **how Digital Preservation fits into the Information Governance lifecycle** – including content ingested from other systems (e.g. long-term records and emails) - as well as how to provide greater “transparency” through controlled access to information for internal and public users

- **We value your feedback ;-)**
- **Please complete the short evaluation....**

Thank you!

www.preservica.com

info@preservica.com

@preservica

@dPreservation

www.statearchivists.org/