



Preservica Email Preservation

Michael Hope
July 2017

Email Preservation Workflow

- Email selection
- Transfer
- Unpacking transfer format into archival format and structure
- Ingest
- Preservation
- Data management
- Search, view and download

Email Preservation

Explorer - (2014-10-30 09:17:12) x

Secure | https://demo.preservica.com/explorer/explorer.html#browse:7&a28e1599-a7b6-416e-887d-524f02e04a38&b6ddbc2f-c5ff-4b79-bf70-97b2c268182f/T1AC

Preservica
Digital Preservation

Jon Tilbury (0) English

DashboardIngestAccessPreservationData ManagementAdministrationHelp

Search Archive

Root > PASIG > email export 17062015.pst > Top of Outlook data file > Inbox > (2014-10-30 09:17:12) Images for Wildlife Exhibition

Sync Tree

Digital Preservation Conferences

format collection

Inbox

JISC

Old

PASIG

Digital Content

email export 17062015.pst

Search Root

SPAM Search Folder 2

Top of Outlook data file

Deleted Items

Inbox

(2014-10-30 09:17:12) Images for Wildlife Exhibition

(2014-10-30 14:03:53) inline image

(2014-10-30 14:18:44) Re: Expenses and year end accounts

(2014-11-03 09:32:55) RE: Expenses and year end accounts

(2014-11-04 09:38:25) Re: demo email

(2015-03-24 15:31:13) Fw: Images for Wildlife Exhibition

Tools View

Manifestation Preservation 2 (Active, Original)

#	Name	Description	Size	Type
	elephant-baby-285515_640.jpg	JPEG File Interchange Format	127.811 KB	File
	elephant-baby-285515_640_1.jpg	JPEG File Interchange Format	127.811 KB	File
	gavial-275959_640.jpg	JPEG File Interchange Format	79.487 KB	File
	gavial-275959_640_1.jpg	JPEG File Interchange Format	79.487 KB	File
	Images for Wildlife Exhibition.eml	Internet Message Format	56.397 KB	File

Filter Apply ☒ Show accrued files ☐ Show children ☐ Show deprecated

@

Name: Images for Wildlife Exhibition.eml
File Type: Internet Message Format
Type: File Size: 56.397 KB Reference: 61481653-7da4-47a6-896a-fb7b3d18cb86

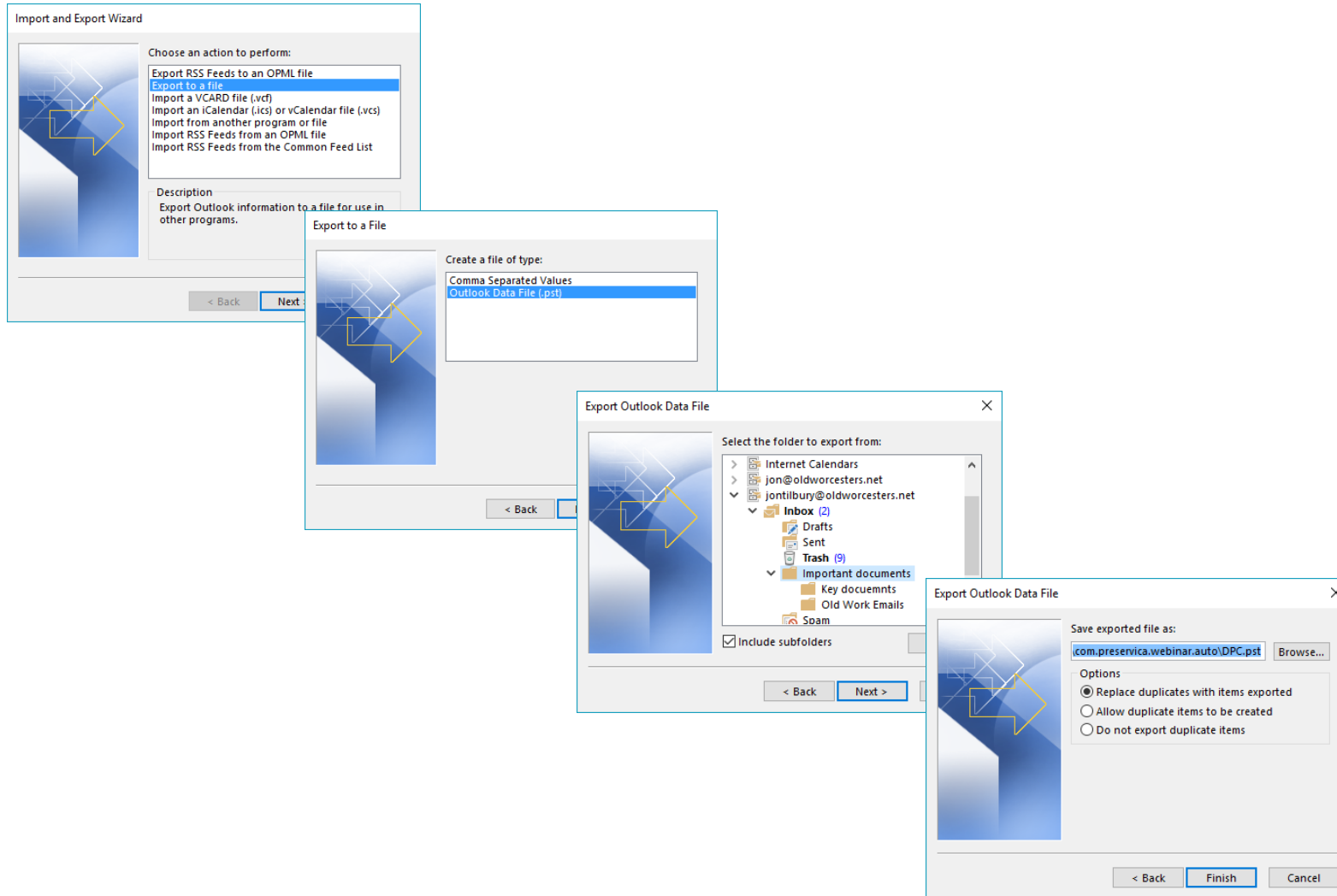
Download

preservica.com Copyright © 2016 Preservica

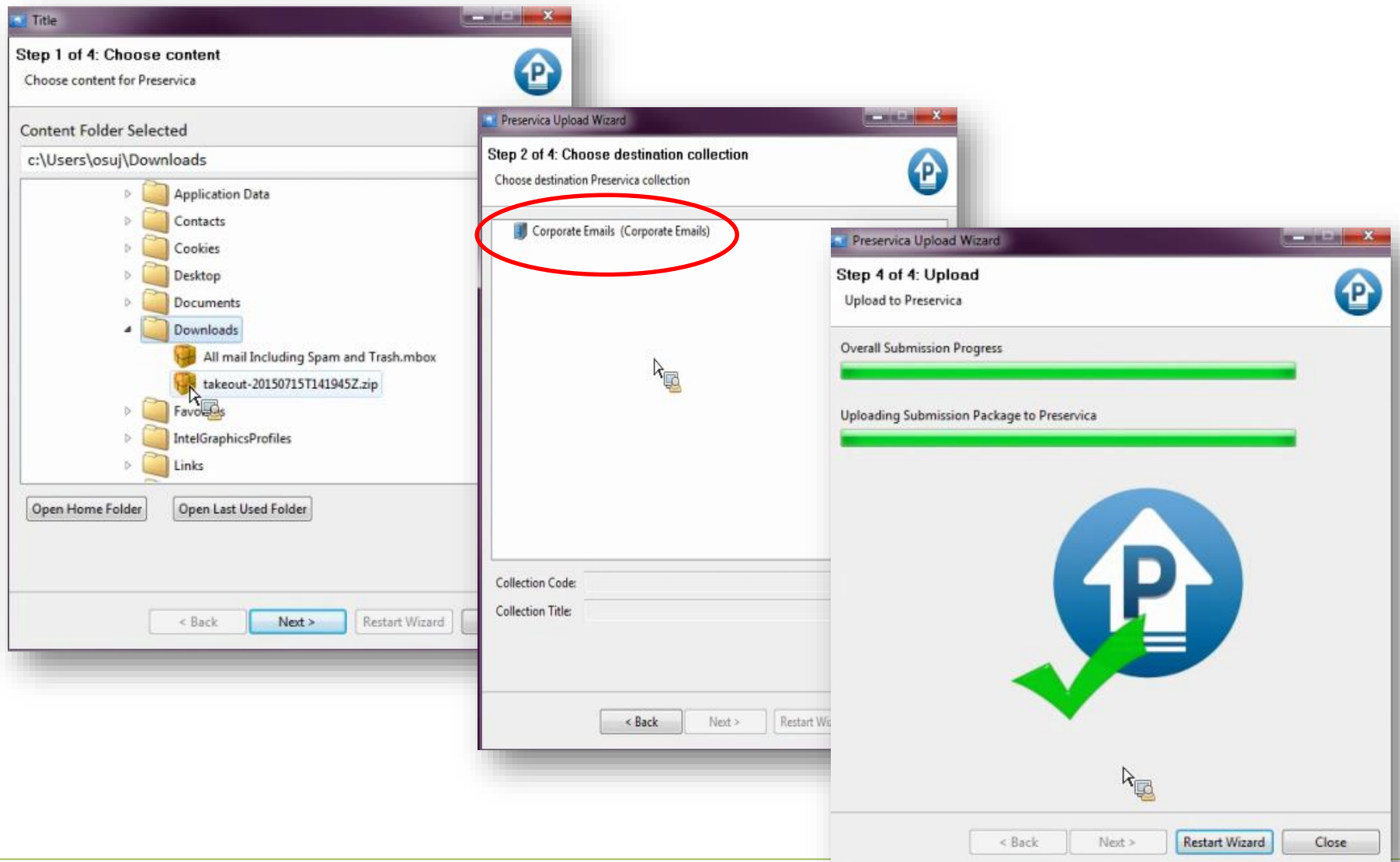
Email Preservation Issues

- Email selection
 - identify the emails of interest (by person, by action e.g. copy to folder, by keyword)
- Transfer
 - continuous via HTTP
 - continuous by file extract
 - manual extract of single mails
 - entire mailbox in PST or MBOX

Export (Outlook example)



Transfer



Email Preservation Issues

- Unpacking transfer format into archival format and structure
 - unpack PST or MBOX container into hierarchy of messages
 - handle tagging as well as folder hierarchy
 - where to put individual message file in hierarchy
 - extract message, metadata, and attachments from email file into separate objects for preservation
 - what format should the message be kept in (text, HTML)
 - Q: handling link rot : are external links / objects / images referenced by the HTML incorporated or not
 - Q: is the PST, MBOX, MSG container kept or just an artefact of transfer

Email Preservation Issues

- Ingest
 - use rules to reject unwanted emails
 - identify duplicates and ignore if email / attachments already there
 - characterise message and attachments
 - normalise attachments and message if required

Ingest


Start Waiting Running Completed Reports Manage

Workflow Details

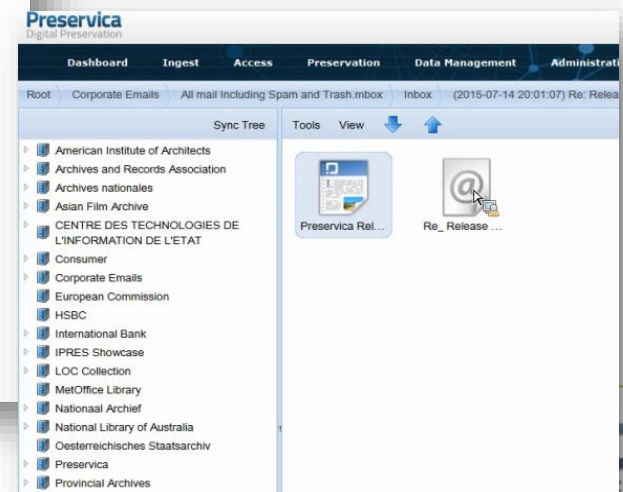
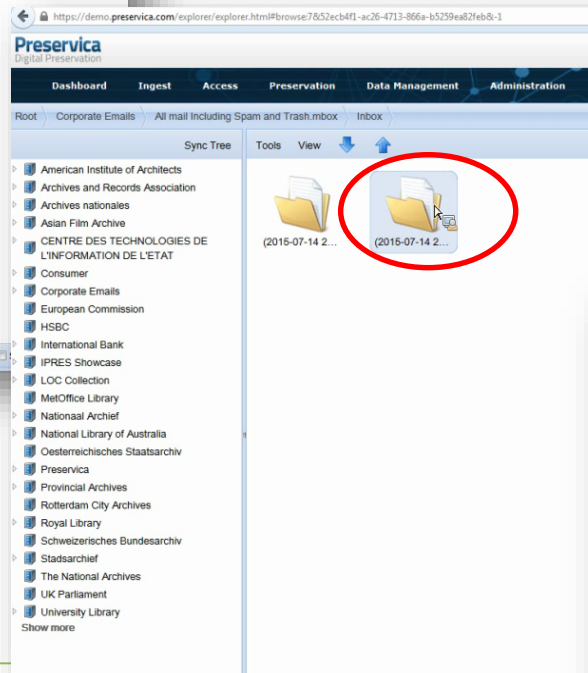
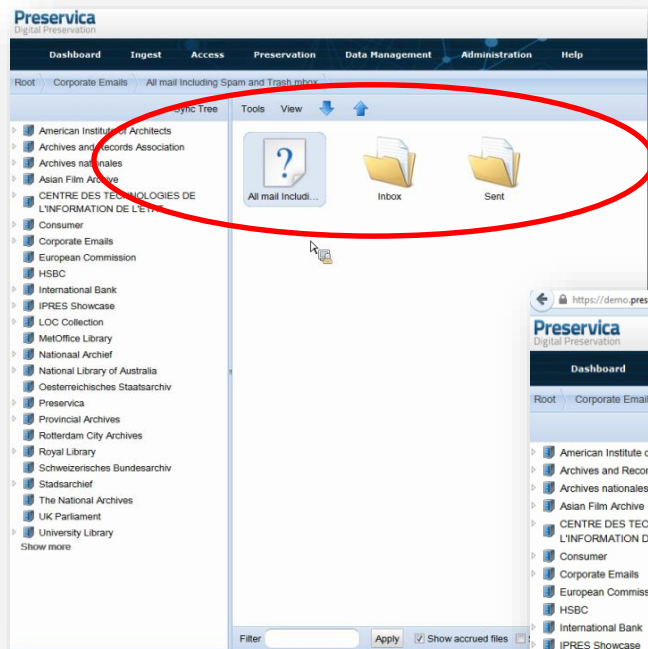
Workflow Context	Upload Wizard Ingest
Workflow Definition	Standard Ingest Workflow (Auto Selection)
Workflow ID	9524
Workflow State	Active
Date Started	15.07.15 10:25:03
Date Finished	
Number of Files	7
Total Size	1 MB
Collection Code	code
Submission name	All mail Including Spam and Trash.mbox
Top Level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

Back Terminate

Step Progress

State	Name	Progress	Started	Finished	Messages
✓	Identify Packages	<div></div>	15.07.15 10:25:03	15.07.15 10:25:07	
✓	Import from Transfer Area	<div></div>	15.07.15 10:25:07	15.07.15 10:25:10	
✓	Process Package	<div></div>	15.07.15 10:25:10	15.07.15 10:25:13	
✓	Virus Check 	<div></div>	15.07.15 10:25:13	15.07.15 10:25:31	
✓	Fixity Check	<div></div>	15.07.15 10:25:31	15.07.15 10:25:34	
✓	Metadata Integrity	<div></div>	15.07.15 10:25:34	15.07.15 10:25:37	
✓	Content Integrity	<div></div>	15.07.15 10:25:37	15.07.15 10:25:40	
✓	SIP Validation	<div></div>	15.07.15 10:25:40	15.07.15 10:25:43	
✓	SIP Validation with Database Crosscheck	<div></div>	15.07.15 10:25:43	15.07.15 10:25:46	
✓	Characterise	<div></div>	15.07.15 10:25:46	15.07.15 10:25:52	View
✓	Store Files	<div></div>	15.07.15 10:25:52	15.07.15 10:25:55	
✓	Store Metadata	<div></div>	15.07.15 10:25:55	15.07.15 10:25:58	

Unpacked folder structure




Individual Emails

Preservica
Digital Preservation

[Dashboard](#) [Ingest](#) [Access](#) [Preservation](#) [Data Management](#) [Administration](#) [Help](#)

[Root](#) [Corporate Emails](#) [All mail Including Spam and Trash.mbox](#) [Sent](#) [\(2015-07-14 19:57:00\) Release Note changes](#) [Release Note changes.eml](#) [Properties](#) [Ren](#)

Release Note changes

From	Jack O'Sullivan
To	maite.braud@preservica.com,
CC	
BCC	
Attachments	 Preservica Release Note V5.R4.M0.docx 332 KB

Hi Maite,

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

Thanks,<r>

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: @dPreservation

bf>This message is commercial in confidence and may be privileged. It is intended for the addressee(s) only. Access to this message y anyone else is unauthorized and strictly prohib
Please noe that messages sent or received by the Preservica e-mail system may be moitored and stored in an information retrieval system.
Please consider the environmet and do not print this e-mail unless you really need to.

Email Preservation Issues

- Preservation
 - conduct ongoing migration on attachments and message
- Data management
 - auto-classification of incoming emails driving security settings and retention profile
 - editing and restructuring rules
 - schema to use for extracted metadata
 - appraisal and disposal of expired messages

Preservation

Key documnts

All ▾

Older	
Jonathan Tilbury Interesting article Hi Here is an interesting article on ICT and fundamental	27/04/2017
Jon Tilbury U.S. Department of Education Parcel Services Hi Here is the paper I mentioned which looks pretty	25/04/2017
Jonathan Tilbury ROI calculator Hi This is the ROI calculator I found which may suit your	25/04/2017
Jonathan Tilbury Gartner MQ Hi Here is some information on the Gartner MQ for ECM I	25/04/2017

Thu 27/04/2017 15:53
Jonathan Tilbury <jonathan.tilbury@prese
Interesting article

To Jon Tilbury

art64-19.wpd
106 KB

Hi
Here is an interesting article on ICT and fundamental rights and the debate in ti
Jon

Jonathan Tilbury
Chief Technology Officer
Preservica

22 The Quadrant, Abingdon Science Park, Abingdon, Oxfordshire, OX14 3YS
T: +44(0)1235 428949 M: +44(0)7808 950580 E: jonathan.tilbury@preservica.com W: [le > Inbox > Important documents > Key documnts > \(2017-04-27 14:52:35\) Interesting article >

Tools View

#	Name	Description	Size	Manifestation
	art64-19.odt	OpenDocument Text	40.143 K	Preservation 3 \(Active\)
	Interesting article.eml	Internet Message Format	22.311 K	Preservation 2 \(Original\)

Preservica

Digital Preservation


Email Preservation Issues

- Search, view and download
 - faceted and fielded search via extracted metadata
 - render individual emails and attachments
 - download messages and attachments
 - viewer for a set of emails that looks like an email application
 - whole collection email analysis

Search and access

EXPLORE

test

DATASET 

FILTER

264 total results

In complete archive

Sort by Relevance ▾



FILTERS

Email subject: outlook



REFINE YOUR SELECTION

From



jontilbury@oldworcesters.n



jonathan.tilbury@preservica

Top level collection



Format collection



New Inbox



Test Area



Inbox



Video



brama



ofas

Document type



File



Record



Collection

Title ▲ ▼

Descr ▲ ▼

Date ▲ ▼



Microsoft Outlook Test Mess...

(2017-06-29 11:43:59) Microsoft OutL



Microsoft Outlook Test Mess...

(2017-06-29 11:53:19) Microsoft OutL



Microsoft Outlook Test Mess...

(2017-06-29 11:54:02) Microsoft OutL



Microsoft Outlook Test Mess...

(2017-06-29 11:55:42) Microsoft OutL



Microsoft Outlook Test Mess...

(2017-06-29 11:43:59) Microsoft OutL



Microsoft Outlook Test Mess...

(2017-06-29 11:53:19) Microsoft OutL



Microsoft Outlook Test Mess...

(2017-06-29 11:54:02) Microsoft OutL



Microsoft Outlook Test Mess...

(2017-06-29 11:55:42) Microsoft OutL

Conclusions

- Email preservation requires a full understanding of the whole life information lifecycle
- The core digital preservation problem is done
- The challenge is acquiring the correct emails and extending analytics
- The community should put its efforts into defining the framework for the email lifecycle then passing this on to vendors to code