

SOLOMON R. GUGGENHEIM FOUNDATION

Electronic Records Management Start-Up Project Generously funded by the NHPRC

Preserving Significant E-Mail Correspondence

August 25, 2014 – Prepared by Anthony Cocciolo, Electronic Records Consultant

Objective: To develop guidance and workflows, and test software to preserve email.

Problem Statement:

SRGM records retention schedule identifies significant correspondence as being a permanent record group, and most of such correspondence has been conducted over email for the last fifteen years. However, there are not well-established procedures for preserving email.

Analysis of Environment for Solution:

This pilot requires has two facets. The first is providing a way for staff to transfer email to the archives. The second is once that email has been transferred to the archives, how it will be stored to promote preservation and access.

Staff Transfer of Email to the Archives

The consultant explored a number of options. One option was to have some kind of option in MS Outlook to flag an email to be sent to the archives. However, no software could be located to achieve this option. Also, this option—if such a solution exists—would likely be platform dependent and staff access email on a variety of platforms (e.g., MS Web Outlook, iPhones, Blackberries, etc.). The other option that was considered was the creation of separate Archive mailboxes for every user. In this scenario, this mailbox could be attached to a user's mailbox and he/she could copy in messages, or even forward messages, to this mailbox. This is the strategy that was used in the University of Michigan's MeMail Project.¹ This option would have the advantage of being device independent; however, would require significant IT resources to accomplish (e.g., establishing the accounts, making sure that users can access the mailboxes from their workstation, ensuring that users have awareness of these mailboxes, etc.). Thus, this option is less than ideal.

The final option is that the archives provide guidance for staff to transfer email files to the archives manually (e.g., creating a PST file and submitting it to the archives). This option can be coupled with the practice in place for the last several years, which is attempting to get ahold of PST files from departing staff.

Preserving and Providing Access to Transferred Email

Once email has been appraised for inclusion in the electronic records repository, it is necessary that the archive has means to preserve it and make it accessible.

Solution:

Staff Transfer of Email to the Archives

The solution selected is that archives staff will continue to pursue PST files when staff leaves the Foundation. Additionally, the archives should be integrated into staff departures procedures so that significant email can be transferred upon departure.

Also, provide staff with the option to transfer significant folders of email before departure. See Appendix A: "Transferring Significant Email Correspondence to the SRGF Archives."

¹ <http://e-records.chrisprom.com/memail-project-guest-post-by-mike-shallcross/>

Preserving and Providing Access to Transferred Email

Microsoft Outlook Personal Folders Files (.PST)

Microsoft Outlook is heavily used at SRGF and there are over 500 PST files on network shares, which do not include active PST files used by staff.

Although the PST format has been openly specified by Microsoft, it is a complex binary format that requires sophisticated software to open.² The recommendation is that it be converted into MBOX format, where folders (Inbox, Sent Items, etc.) are created in simple text-based formats well suited for long-term preservation.³ This is automatically done by Archivemata using the readpst tool. The original PST file is also maintained.

For accessing the converted PST files, MBoxViewer is available on the obsolete media laptop.⁴ Other mail programs, such as Mozilla Thunderbird, can be used for reading MBox files.⁵

Outlook Message Format (.MSG)

There are 3,357 files in this binary format that has been openly specified by Microsoft.⁶ Like the PST, this format requires sophisticated software to read it (specifically MS Outlook). It is recommended that these files be converted to EML files, which is a text-based format where attachments are encoded with the MIME format.⁷ Please see instructions on normalizing this format for preservation and access available in the appendix of the ER Processing Manual.

Novell Groupwise E-Mail (.MLM)

There are 4,307 .mlm files on the network. This is a binary email format used by Novell Groupwise and is actually encoded as WordPerfect 5.X files (thus, there are no email attachments). The script pres_and_access.vbs created by the consultant can be used for normalization these files into preservation and access formats (DOCX and PDF).

Microsoft Outlook Express Message Format (.EML)

There are 1,552 files of this type on the network. This is a text-based format with MIME encoding for attachments. This format is suitable for long-term preservation. Instructions on normalizing this format for access are available in the appendix of the ER Processing Manual.

Technical Notes

The following Perl Library was installed on the Archivemata server to read MSG files and covert them to EML files:
<http://www.matijs.net/software/msgconv/>

Commands for reading converted MSG to EML and PST to MBOX.zip are included in the text file, "Email commands for Archivemata."

Note that Archivemata uses readPst (<http://www.five-ten-sg.com/libpst/rn01re01.html>) to convert PST to MBOX.

² <http://msdn.microsoft.com/en-us/library/ff385210%28v=office.12%29.aspx>

³ From Christopher Prom, Assistant University Archivist, Univ. of Illinois Urbana-Champaign: "in general, if an institution can get email into one of the MBOX or EML formats, it has taken a very big step on the road toward preserving email." Via <http://blogs.loc.gov/digitalpreservation/2013/07/the-what-of-email-archiving/>

⁴ <http://sourceforge.net/projects/mbox-viewer/>

⁵ <http://www.mozilla.org/en-US/thunderbird/>

⁶ <http://msdn.microsoft.com/en-us/library/cc463912.aspx>

⁷ <http://blogs.loc.gov/digitalpreservation/2013/07/the-what-of-email-archiving/>

Appendix A: Transferring Significant Email Correspondence to the SRGF Archives

Updated 24 June 2014

Overview

SRGF Archives is interested in preserving significant email correspondence (e.g., correspondence with artists, emails with important attachments like major project reports, policy changes, etc.). Email can be transferred to the Archives at time of departure from the museum, or earlier using the process articulated below. This process assumes the use of Microsoft Outlook (screen shots appear in Windows version 2010, but other versions are fairly similar). Note that you would want to consider submitting not simply messages that you receive (messages in your “Inbox”), but also messages that you send (messages in your “Sent Mail”).

Steps for Sending E-Mail messages to the Archives

To send email to the Archives, create a new Outlook PST file, and transfer the PST file to the archives. To create a new PST file, select New Items -> More Items -> Outlook Data File (see Figure 1).

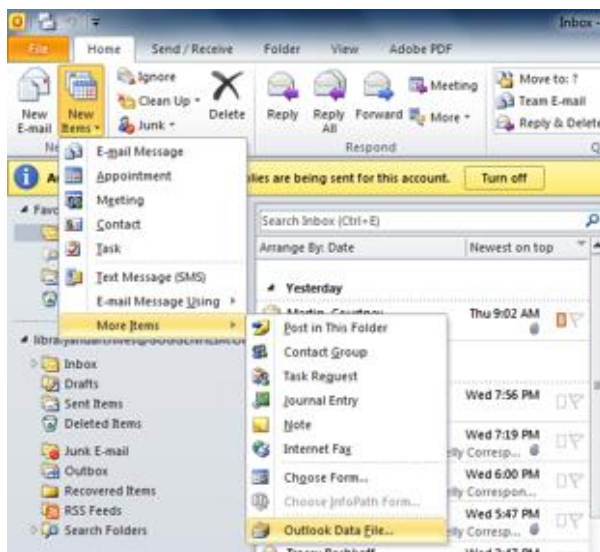


Figure 1. Create new PST file.

Click “OK” to save to save the file to somewhere on your local computer. Make sure you remember where the PST is saved to. Then, a new option will appear in your leftmost folder list label “My Outlook Data File.”

Right click the newly created Outlook data file, and select “New Folder.” Enter a folder name (e.g., “Still Spotting Exhibition Sent Messages”). Click “OK” (see Figure 2).

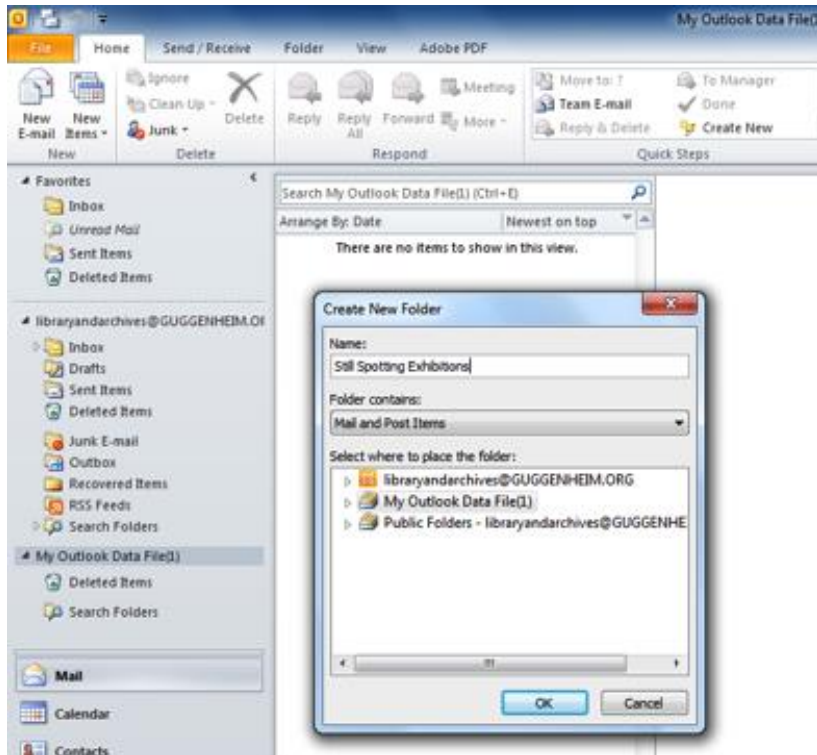


Figure 2. Create new folder within PST file.

Select messages from your mailbox which you'd like to move to the archives. You can select multiple contiguous messages by holding down the SHIFT key and selecting the messages, or holding down the CTRL key to select single messages to add to the selection. Drag messages to the new folder you created in the Outlook Data File.

Note that this procedure will move messages—rather than copy them—to your new folder. Do not copy messages—move them. If you need to reference your archived email at a future date, contact libraryandarchives@guggenheim.org and they will make it available to you.

When you are done moving messages into the new PST, you can close it by right clicking it, and selecting “Close.”

You can now submit the PST file to the Archives using the standard transfer process for electronic records.