Email as Cultural Heritage Resource:
Appraisal Solutions from an Art Museum Context

Anthony Cocciolo

Abstract

Purpose – This study examines strategies to appraise significant email as a cultural heritage resource, using an art museum located in the Northeast Untied States as a study site.

Design/methodology/approach – Treating this art museum as a case study site, email messages are manually appraised for retention using a rubric. Following the appraisal, strategies for expediting this appraisal process, using what is learned from the manual appraisal process, are explored.

Findings – A major finding of this study is that sent mail is almost always significant, although preserving only sent mail, or preserving sent mail in combination with inbox items that have been acted upon (replied to or forwarded), are not sufficient to capture significant correspondence. Rather, a social network approach holds the most promise to accelerate the process of email appraisal.

Originality/value – This paper provides empirically grounded strategies for appraising email for permanent retention.

Introduction

In late 2014, the hacking of emails and other digital files from Sony Pictures, Inc. made front-page news across the globe, eventually involving heads of state and resulting in a national security response (Sanger, 2014). The incident prompted commentary from a variety of angles, including the notable perspective that individuals should delete all their electronic correspondence when they are complete, and suggestions to never say anything via electronic communication that you wouldn't mind being repeated in public (Clark, Ovide and Dworkin, 2014; Okyle, 2014). Responding to this kind of advice in the editorial page of the *New York Times*, Bruni (2014) noted that "delete, delete, delete. That's a bit of your humanity being snuffed out," further arguing that:

> *It's not just creativity that's in jeopardy. It's not just candor. It's secure islands of unformed thought and sloppy talk, places where people take necessary vacations from judgment, allowances for impropriety that make propriety possible. And these aren't or shouldn't be, luxuries.*

As Sony case illustrated, email has proven to be a mechanism for communicating "unformed thought" and "sloppy talk," as well as a suitable medium for the well articulated and coherent. Because email is used by individuals most frequently as one-to-one and one-to-several style of communication, it is the natural successor to the written letter, allowing for snippets of "unformed thought" and candor that only the recipients are both privileged and sensitive enough

to receive. Personal email communication contrasts well with social media, public webpages, and press releases, where such communications may strive for candor and transparency yet do so with a heightened awareness to its public consumption. Of course, not all email exhibit refreshing candor: advertising and other messages designed for public consumption are in evidence in nearly all email boxes, as anyone who has struggled with "spam" messages can attest.

Email collections have the potential to capture the emergence of thoughts and decisions that illustrate how and why things developed the way they did. However, making email collections available in publically accessible archives faces a number of challenges, including ensuring that personal privacy is preserved and the institutions making these collections available do not inadvertently expose themselves to liability. Because of these threats, legal departments to IT departments continue to chant "delete, delete, delete," forever inhibiting the kind of historical insights that email collections can provide. In order to counteract the "delete, delete, delete" orthodoxy, archivists need email appraisal and selection methods that allay the legal concerns of lawyers, conserve valuable IT (assuaging IT executives), and contain the overall cost, which is a concern for administrators.

This project will put forward some ways to appraise emails for permanent retention, using an art museum located in the Northeast United States as a case study site. This institution, which will be referred to as USAM for brevity, has made some commitment to preserving significant email collections. The author acted as an electronic records consultant and researcher at this institution during a small one-year grant project to help it plan for a born-digital archives (2013-2014). One of the several project goals was to develop means for preserving significant email correspondence, where such correspondence would to be available to researchers on a request basis twenty-five years after creation in a reading room.

The goal at USAM was not to save every email that was sent or received; rather, it was to preserve "significant email." In 2005, the department head of the archives established a records retention schedule for the institution, which was approved by the Board of Trustees. The records schedule identified groups of records, by department, which had permanent as well as temporary retention periods, as well as highlighted vital records to the operation of the museum. Included in this records retention schedule were "Significant Email," which were flagged as having permanent value. This was defined as "correspondence that documents important activities, events, operations, policy changes," which is unlike "routine correspondence" which could be retained "while useful" and then discarded.

Most specifically, this project is motivated by the question:

> *By what method can email be appraised such that only significant email is retained?*

To address this question, significant email was selected from insignificant email (more on this in the method section), and analysis will uncover if there were strategies or shortcuts for selecting significant email from insignificant email. The potential strategies that were explored, which are not mutually exclusive, include:

1) Preserve only the sent mail
2) Preserve the sent mail, plus inbox items that have been acted on (replied To and forwarded).
3) Purge unread Inbox items
4) Purge items that are in the "deleted items" folder.
5) Preserve using a social network approach: For inbox items, consider the social role of the sender; for sent items, consider the social role of the receiver, and manually appraise messages grouped by sender (for inbox) and receiver (for sent items)).

However, before the strategies will be analyzed and discussed, relevant literature to email appraisal and its value for cultural heritage will be introduced.

Literature Review

Appraisal or selection of records for permanent or archival value is widely viewed by archivists as one of their most important tasks (Boles, 2005). Digital records, and email in particular, have prompted archivists to reconsider the need to conduct appraisal. Notable archival researchers such as Gilliland (2014) suggest that appraisal of digital records may no longer be feasible and that archivists should consider embracing records in their totality. Relatedly, as early as 1995 Bearman noted that "all selection and appraisal based on trying to create a representative record will fail simply because it is being carried out by people living in the present" (p. 383). Despite this criticism by him and others, appraisal continues to be practiced. A notable strategy includes functional analysis, where records are retained in consideration to the major social functions of the institution. For example, in higher education, Samuels describes that records in higher education should be retained in the context of its seven major functions, which are 1) confer credentials, 2) convey knowledge, 3) foster socialization, 4) conduct research, 5) sustain the institution, 6) provide public service and 7) promote culture (Samuels, 1992).

Despite criticism of archives appraisal, it needs to be continued because digital preservation is expensive, and it is wasteful to store digital material permanently with little or no historical or cultural value. Research projects such as the U.S. IMLS-funded POWRR project (Preserving [Digital] Objects with Restricted Resources) find that digital preservation is outside the reach of many small and medium-sized cultural heritage institutions because of its cost and complexity (Rinehart and Prud'homme, 2014). This is because digital infrastructure needs to be built and maintained or otherwise bought (e.g., maintaining server rooms, cooling servers, migrating files off of older hardware, paying electric bills, knowledgeable IT staff to maintain the infrastructure, software for managing the digital preservation activities, etc.). Thus, there is a cost associated with storing digital files, and saving files with limited or no value wastes scarce resources. Conversely, it may seem somewhat wasteful to have humans spend their days manually appraising individual email messages one-by-one for permanent retention. A balance must be struck where the time and cost of the digital archivist is respected (or whomever is doing the appraisal), and the storage cost savings from eliminating files or messages with little or no significance.

However, beyond the issue of cost is the ethical issue of personal privacy. Although institutions typically own the contents of staff email accounts, it is widely noted that individuals use their

work accounts for personal uses. This is especially true in the earliest uses of email, where individuals may have only had the single email account supplied by his or her employer. Thus, we are likely to find information that is not an institutional record but rather personal information that should not be archived. In some cases, this information can include medical, financial, and other personal information that should have been destroyed, but is still saved and in many cases adjacent to significant exchanges that should be permanently preserved. Given this privacy concern, the need for the appraisal of email messages is necessary.

Since email appraisal is necessary in an institutional archive context, are there automated or semi-automated technologies for doing this? Although there are a number of tools that have been developed, most of them are designed for browsing email archives rather than appraising email records for inclusion in an archive. For example, the MUSE project was explored, which allows for a number of interesting ways for viewing an email box (Hangal *et al.*, 2012). One such way is view groups of messages by sentiment (e.g., happy, angry, joyful, etc.), and then being able to zoom-in and view individual messages. Although tools such as this may make the experience of browsing email archives more interesting and enjoyable to the researcher, it is not useful for appraisal. Fortunately, some of the features of MUSE are expected to be incorporated into the ePADD project, which is a project from the Stanford University libraries (2015) funded by the National Archives that explicitly looks to include tools that can be used for email appraisal.

Prom (2011), in his comprehensive and useful report on preserving email for the Digital Preservation Coalition, explored some tools for incorporating email into archives. These tools mainly involved transforming email boxes or messages from one format to another, including Aid4Mail, CERP Mail Processor, EmailChemy, MailArchiva, Xena, among others. These tools are useful for getting messages into a format well suited for preservation (e.g., MBOX format or XML format), but will not necessarily help the archivist separate the significant email from the email with limited or no value.

There is little published research on methods for appraising email records. Exceptions include Pennock's (2006) report on curating emails, where she writes that in a "historical or cultural context, where email collections are developed without legal record-keeping requirements, there in increased merit in the 'keep everything' approach." In her report, she contrasts this with institutions that do so for legal record-keeping purposes, where most messages can be deleted. This either/or scenario does not correspond well with the situation at USAM, which is both interested in preserving significant correspondence for its cultural and historical value as well as its ability to document institutional activity, but could not employ a "keep everything" approach for ethical and resource-related concerns.

There are discussions of email appraisal in less formal settings. For example, Haws (2013)--who is the archivist for the New York Philharmonic--noted their strategy for incorporating email into the archives is to simply include all sent mail. This is a compelling strategy because it would ensure that no spam messages were included (because who replies to spam?), and since most mail clients repeat the incoming message in the sent message, it would conceivably capture inbox items that were important enough to be replied to (or forwarded). Using USAM as a case study, this project will explore the efficacy of this option, among other options.

Methods

Study Context

USAM collects contemporary and modern art, produces approximately a dozen exhibitions each year, has over 50,000 square feet of gallery space, and welcomes on the order of a million visitors annually. USAM has maintained archives since the 1970s, and currently houses approximately 7,000 cubic feet of paper records. The records include exhibition files, artist files, as well as other historical records, and is open to researchers throughout the year on a request basis. The archives is led by a head of library and archives, with an archivist with sole responsibility to the archives reporting up to this head.

In around the year 2002, USAM migrated off of Novell Groupwise as their mail server and adopted Microsoft Exchange and Outlook. Also as mentioned earlier, in 2005 a records retention schedule was developed and approved by the Board of Trustees that identified that significant email correspondence should be permanently maintained. Despite identifying the importance of email correspondence, there were few infrastructures in place to make this possible. Email collections that have been saved from deletion thus far involve the IT department maintaining the Outlook PST file for a staff member after he or she has left the organization.[1] PST files are not universally maintained: typically a supervisor has to request that it be saved upon the departure of a staff member. Although archivists are sometimes consulted during a staff departure, this is not always the case. Despite the somewhat uneven way that email collections are handled, USAM has managed to hold onto 365 PST files using 180 GB of disk space.

In USAM's electronic records project, the project team developed workflows for staff to transfer email records to the archives by creating new PST files, moving significant email messages to them, and transferring the PST files to the archives as they would do with any other type of digital file. However, members of the project team did not think that most staff would do this often or ever, because as Cox (2008) notes, "E-mail is the antithesis of the control records managers have traditionally sought to exercise over records and information systems" (p. 208). In this case, it was recognized that it could be difficult to get a two hundred fifty person staff to manage their significant email in any unified way. The project team considered the option of creating archival mailboxes that could be attached to users' primary mailbox, allowing users to transfer their significant email to these archival mailboxes. This was the practice that was piloted in the MeMail project from the Bentley Historical Library at the University of Michigan (Shallcross, 2011). However, it was believed to be too challenging from an IT perspective to setup the additional mailboxes because managing the primary mailboxes is already a heavy burden. The staff training would also be a major undertaking. Thus, the project team will rely heavily on the strategy that involves little staff involvement and minimal impact on the IT department: to seek out the PST files of staff when they depart.

In concept, the digital preservation strategy that was adopted during the electronic records grant project for USAM was to preserve the PST file as a representation of the submission to the digital archives, and to create a MBOX representation of the mailbox as the archival copy. This is consistent with notions from the Open Archival Information System (OAIS) model, which is

an international standard often used to promote digital preservation (Lavoie, 2004). In this model, a Submission Information Package (SIP) represents what was submitted to the archives from the user (in this case the PST file), and the Archival Information Package (AIP) is a representation of that submission designed to promote long-term accessibility (in this case the MBOX file). As Prom writes, MBOX is a "de facto" standard for email preservation because it widely used by many email clients, uses text to encode the email messages, uses MIME format to encode binary attachments (which is a text-based method for encoding email attachments that is used when sending binary files over the Internet), and preserves the email headers (p. 22). Thus, anyone with a text viewer can read a MBOX file. This is unlike PST files, which is a complex binary file format. Although Microsoft has provided specifications on the PST file format, and there are currently tools for reading and exporting messages contained in PST files, such as the open-source tool "readpst," the long-term ability to read these files are unknown (Microsoft, 2014). Thus, in general the strategy is not to be solely reliant on proprietary, closed-sourced software whose continuance is reliant on an ever-morphing technology sector.

Despite the advantages of MBOX files, PST files include information that is not exported to the MBOX format, and thus should be preserved as well. For example, PST files maintain information that is particular to Outlook, such as flags if a message was read, replied to, or forwarded, which would be lost if only the MBOX file were maintained. Further, the PST is more faithful to the original context in which the email was used, thus further enticing the project team to maintain the PST in addition to the MBOX file.

Email Appraisal Rubric

The goal of this project is devise methods to appraise significant email for inclusion in the digital archives of USAM. However, what are the criteria to determine a significant message from an insignificant message? The need to have a uniform method for making this determination led to the development of the following rubric (see Table 1), which includes four categories that an email message can be assigned: 4) Very Significant Record, 3) Significant Record, 2) Insignificant Record, or 1) Very Insignificant or Non-Record. Items with scores of three or four are retained in the digital archives, where items with scores one or two are not. The table includes four facets that are considered when assigning a message a score: subject matter, sent actor, received actor, and properties of message. The rubric can be applied with some flexibly. For example, to get a value of 4 or "Very Significant Record" does not require every facet to meet the "Very Significant" criteria; however, it should meet most of the criteria. For example, a personal email from a head of state may receive a "Very Significant" score because of the heightened significance of the "sent actor," even if the "subject matter" may only be "significant."

Table 1. Rubric for determining significance of email message.

	Retain		Destroy	
	4 - Very Significant Record	**3 - Significant Record**	**2 - Insignificant record**	**1 – Very Insignificant Record or Non-Record**
Subject Matter	Subject matter	Subject matter	Subject matter	Subject matter

	is very relevant to the institutional mission and mission-driven functions.	is relevant to the institutional mission and mission-driven functions.	is marginally relevant to the institutional mission or mission-driven functions.	has little or nothing to do with the institutional mission or functions. Can include information from personal life or personally sensitive information (e.g., financial) that should not be archived.
Sent Actor	Sent by individual with major salience to the institutional mission or mission-driven activities.	Sent by individual with salience to the institutional mission or mission-driven activities.	Sent by individual with minor salience to the institutional mission or activities.	Sent by an individual with no importance to the institutional mission or activities.
Received Actor	Member of the organization operating in an area very relevant to the institutional mission.	Member of the organization operating in a area relevant to the institutional mission.	Member of the organization operating within the institution, although role may not be directly related to the mission (e.g., support function).	Transitory member of the organization, low-level staff member (e.g., temporary staff), or non-staff member.
Properties of message	Ideally, message was read and replied to.	Ideally, message was read and replied to.	Message could be read or unread, replied to or not replied to.	Message could be read or unread, replied to or not replied to. Message could be marked as SPAM.
Example from an art museum context	1) A noted designer discusses a	1) A museum executive receives the	1) A message to all staff from the head of	1) An unopened newsletter from a popular

	potential commission project with a museum executive. 2) An artist who is planned to have a retrospective at the museum exchanges a message with the curator about that retrospective.	statistics for attendance of a museum exhibition at the close of the museum from the head of visitor services. 2) A museum curator sends a magazine journalist the responses to an interview that is planned to be printed in the magazine about an upcoming exhibition.	human resources with information on open enrollment of health benefits. 2) An invitation to a local gallery opening for a museum executive with no indication that it was read or responded to.	magazine. 2) An email from a museum executive to spouse asking him/her to pick up the groceries.

The rubric was developed for use in an art museum context, but can be used in other contexts as well. What is important is focusing on the mission of the organization and the major activities and functions that stem from that mission. For an art museum, those functions include producing exhibitions, acquiring objects into the permanent collection, collaborating with artists for exhibitions, commissions and other projects, and acting as an educational institution.

Thus, messages that are related to the production, display and appreciation of art, as well as the learning generated around such art, are significant; messages about support functions (e.g., computer workstations, employee benefits, routine cleaning of facilities, etc.) are insignificant.

Manual Email Appraisal Methodology

Individual mail items were manually appraised for retention from three mailboxes while the author worked as a researcher and consultant for USAM during the course of their electronic records grant project. Since this study is a test of email appraisal strategies, no emails were deleted in the course of this project, but rather copies of original mailboxes were used.

Two curators and a museum executive mailboxes were selected for appraisal because the mailbox owners occupied key roles central to the mission of the institution. The mailboxes do not encompass the entire time span that the individuals worked at the museum, but rather what had been captured and saved by the Information Technology department:

    1) Museum Executive: Mailbox date range January 2010 to May 2010
    2) Curator A: Mailbox date range August 2002 to May 2006
    3) Curator B: Mailbox date range August 2002 to April 2006

The technical steps to complete the manual email appraisal are included in the appendix. In a more general sense and from doing the appraisal, it is important to develop an awareness of the projects that the mailbox owner is working on, and the contacts that are relevant to those projects. It is also important to become aware of contacts that are not related to producing significant records, such as husbands, wives or children.

In addition, broad knowledge of the important actors within a community can help expedite appraisal and help make the distinctions between scores of "Very Significant" or "Significant." For example, in the art museum context, knowledge of artists, curators, and critics—both within the institution and within the art world more generally—is very useful. However, doing Google searches on the names of senders and receivers can provide valuable information on roles within the wider cultural context. For example, if a sender or receiver has a page in Wikipedia, this can be valuable information for situating the individual within the institutional context and indicate their wider cultural salience. However, a lack of being able to find biographical information on a sender/receiver should not cause a message to drop into insignificance. Rather, if the subject matter were relevant (e.g., creation and exhibition of artworks), the message would still be considered significant. As history has demonstrated, unknown artists do not necessarily remain unknown.

This appraisal strategy took one hour to appraise 641 messages. Thus, strategies to expedite this process would be welcome if not outright necessary.

Analysis of Email Appraisal Strategies

The goal of the analysis is to produce a data table that can be used to shed light on how email appraisal can be improved using the strategies discussed earlier (e.g., save only Sent email, etc.). Messages within the PST files were exported to a Microsoft Access database, where SQL queries were written which could count senders, receivers, and how such messages were categorized.

Results

The three email boxes analyzed illustrate how individuals treat their email differently. Curator B had a personal assistant who was clearly committed to an uncluttered mailbox because his or her mailbox was free of very insignificant records, which are often spam. Thus, the appraisal process only reduced the total messages that would be transferred to the archives by 4.6%. This is unlike Curator A, who seemed to be struggling with spam, and 41.6% of messages were removed during the appraisal process.

The executive's mailbox was an interesting one because it had 217 messages that were significant or very significant that were in the "Deleted Items" folder. When a message is deleted in Microsoft Outlook, it is moved to deleted items, where it remains until the delete items folder is emptied. This of course raises some ethical issues: six messages were "very significant;" however, the mailbox owner may have wanted them deleted. Was his or her intention to truly eliminate these mail conversation, or was he/she just finished with them and was happy to leave them in Deleted Items for future posterity? The position taken here is that a message is not deleted until it is deleted, so items in the "Deleted Items" folder can be appraised

for retention in the digital archives. This is consistent with other digital archives projects. For example, the Salman Rushdie digital archive project allows users to "peak into Rushdie's waste basket" within his emulated Macintosh environment available at the Emory University libraries (2011), which could be considered the Outlook equivalent of the deleted items folder. Measures should be taken to prevent individuals from undeleting files, which can be done in some cases since most operating systems just mark files for deletion rather than scrub them from the hard drive. However, files in an un-emptied trashcan or deleted items folder will be considered for permanent retention here as with the Rushdie project.

Tables 2 through 4 illustrate that preserving only the sent messages is insufficient to document significant correspondence. For example, in the Executive's mailbox, 162 messages were flagged as significant or very significant; however, only 21 messages were flagged as being acted upon (replied to or forwarded), meaning that only 21 messages would be included in the sent items. This same is true for the curator mailboxes: Curator A had 12 out of 132 messages that were acted upon, and Curator B has 77 out of 304 messages that were acted upon. However, nearly all sent items are significant. For example, over 80% of the executives sent mails were significant or very significant. The case is true for the curators: 94% of curator A's and 99% of Curator B's sent items were significant.

Also found were that unread messages are not necessarily insignificant. For example, in the Executive's mailbox, 15 messages that were unread were categorized as significant. Similarly, 43 messages in Curator A's mailbox were significant and unread. In the case of Curator A, some of these unread messages may have resulted from these messages being forwarded to another email account where they were ultimately read. Thus, one cannot assume that an unread message should be automatically purged.

Table 2. Results of Manual Appraisal: Executive Mailbox

	Before Appraisal	**After Appraisal**	**% Reduction**
Executive **Mailbox**			
PST File Size	377 MB	259 MB	31.3%
MBOX File Size	199 MB	168 MB	15.6%
Total Messages in Inbox	345 messages	162 messages	53.0%
4 - Very Significant Record	--	2 messages	--
3 - Significant Record	--	160 messages	--
2 - Insignificant Record	--	-50 messages	--
1- Very Insignificant Record or Non-Record	--	-137 messages	--
Significant & Very Significant email acted-upon (replied to or forwarded)	--	21 messages	--
Significant & Very Significant email unread	--	15 messages	--

in Inbox			
Total Messages in Sent Items	215 messages	172 messages	20%
4 - Very Significant Record	--	3 messages	--
3 - Significant Record	--	169 messages	--
2 - Insignificant Record	--	-40 messages	--
1- Very Insignificant Record or Non-Record	--	-3 messages	--
Total Messages in Deleted Items	1363 messages	217 messages	84.1%
4 - Very Significant Record	--	6 messages	--
3 - Significant Record	--	211 messages	--
2 - Insignificant Record	--	-50 messages	--
1- Very Insignificant Record or Non-Record	--	-1096 messages	--
Significant & Very Significant represented in Deleted Messages	--	37 messages	--
Significant & Very Significant unread in Deleted Items	--	21 messages	--
Total Messages	1923 messages	551 messages	71.35%

Table 3. Results of Manual Appraisal: Curator A Mailbox

Curator A Mailbox			
PST File Size	60.0 MB	38.7 MB	35.5%
MBOX File Size	59.0 MB	33.6 MB	43.1%
Total Messages in Inbox	756 messages	132 messages	82.5%
4 - Very Significant Record	--	0 messages	--
3 - Significant Record	--	132 messages	--
2 - Insignificant Record	--	-24 messages	--
1- Very Insignificant Record or Non-Record	--	-600 messages	--
Significant & Very Significant email acted-upon (replied to or forwarded)	--	12 messages	--

Significant & Very Significant email unread in Inbox		43 messages	--
Total Messages in Sent Items	863 messages	814 messages	5.7%
4 - Very Significant Record	--	0 messages	--
3 - Significant Record	--	834 messages	--
2 - Insignificant Record	--	0 messages	--
1- Very Insignificant Record or Non-Record	--	-49 messages	--
Total Messages	1619 messages	946 messages	41.6%

Table 4. Results of Manual Appraisal: Curator B Mailbox

Curator B Mailbox			
PST File Size	12.3 MB	11.8 MB	4.1%
MBOX File Size	8.6 MB	8.1 MB	5.8%
Total Messages in Inbox	331 messages	304 messages	8.2%
4 - Very Significant Record	--	0 messages	--
3 - Significant Record	--	304 messages	--
2 - Insignificant Record	--	-27 messages	--
1- Very Insignificant Record or Non-Record	--	0 messages	--
Significant & Very Significant email acted-upon (replied to or forwarded)	--	77 messages	--
Significant & Very Significant email unread in Inbox	--	0 messages	--
Total Messages in Sent Items	368 messages	363 messages	1.4%
4 - Very Significant Record	--	0 messages	--
3 - Significant Record	--	363 messages	--
2 - Insignificant Record	--	-1 message	--
1- Very Insignificant Record or Non-Record	--	-4 messages	--

| Total Messages | 699 messages | 667 | 4.6% |

When performing manual appraisal, the researcher observed that in most cases, significance of a message was tied to a given sender. For example, one curator emailing another curator in the same institution is almost always sending each other significant messages. This is indicated in Table 5, which shows that Curator A had 61 individuals email him or her that had messages that were always significant. Similar findings are available for Curator B and the Executive. Table 5 also indicates that there are senders who sometimes send messages that are borderline: sometimes significant and sometimes not. These people are often frequent communicators with the individual, such as a spouse or personal assistant. Thus, email appraisal can be sped up grouping by sender or receiver, and developing an understanding of the relationship between the sender/receiver and the mailbox owner, and then assigning significance in bulk. It is also important to determine who those individuals are who are borderline, such as personal assistants, who may produce relevant records and irrelevant records.

Table 5. Correspondents and their Average Message Significance

	Number of Correspondents		
	Executive	Curator A	Curator B
Between Significant to Very Significant (>3.0)	1	0	0
Significant (3.0)	40	61	55
Between Insignificant and Significant (>2.0 and < 3.0)	5	5	7
Insignificant (2.0)	26	12	10
Between Insignificant and Very Insignificant (>1.0 and < 2.0)	3	8	0
Very Insignificant (1.0)	71	294	0

Discussion and Conclusion

The study found that using easy fixes, like preserving only sent mail, or preserving sent mail in combination with Inbox items that have been acted upon, is not sufficient to capture significant correspondence. Although sent items are almost always significant, this does not mean that by only saving sent items that all significant correspondence has been preserved. This study indicates that there are large amounts of significant correspondence in Inboxes (and in one case deleted items) that are not reflected in sent messages.

This study also found that email appraisal can save valuable IT disk space. For example, in the case of Curator A, there was a nearly a 36% disk space use reduction. However, manual appraisal of messages is slow: 641 messages were appraised per hour.

Although unread messages are often insignificant, there are cases were significant correspondence was unread. For example, an executive may have not read a message because he was already briefed about it over the phone or in-person, and when he/she sees the message skips it because already knows what it is about. Thus, the unread message flag is not a sign that a message can be automatically purged. It still requires appraisal.

A social network approach holds the most promise to accelerate the process of email appraisal. This study found that message categorization was closely tied to the sender or receiver, thus meaning that in many cases all messages sent to or received from a single person were significant (or insignificant, etc.). Thus, the manual appraisal methodology described earlier in this paper could be enhanced by instead of ordering messages by oldest to newest, to order by sender or receiver, and then order by oldest to newest. This way, appraisal can be conducted on groupings of messages from or to a particular individual. After manually appraising a random sample of the messages from a sender/receiver, and all of them are significant, then the archivist may decide to flag them as all significant. Similar, if a random sample are very insignificant, there is a good chance that the rest will be as well (e.g., newsletters from popular magazine). More fine-grained appraisal is needed for individuals that send/receive both personal and professional messages. Interestingly, for all three mailboxes examined, the number of individuals that sent or received both significant and insignificant messages numbered from five to seven. Thus, figuring out who these individuals are and giving those senders/receivers the fine-grained appraisal has the potential for significantly speeding up the appraisal process. By developing a working understanding of the social role of the sender/receiver, and the relationship with the owner of the mailbox that is being appraised, decisions to retain or destroy in bulk, or give more fine-grained appraisal for borderline cases, can expedite the overall process.

Since sent messages are almost always significant, appraisal of sent items can be expedited by keeping it all, except for those five to seven borderline individuals identified earlier that should receive fine-grained appraisal.

In conclusion, in an institutional archive context, email appraisal is essential to both conserve IT resources and protect personal privacy. At the present moment, the best tools to expedite email appraisal are the tools built-in to mail clients like Microsoft Outlook, that allow sophisticated searching, sorting, and categorization. Although computers are getting more advanced at a variety of tasks, it is doubtful that they will be able to take over the job of email appraisal. For example, spam filters have become more sophisticated and effective, yet occasionally both spam gets into inboxes, and important messages get categorized as spam. Future tools developed for email appraisal should focus on a social network approach, or grouping messages by sender/receiver, and providing means for the digital archivist to quickly gain an understanding of the social role played by that individual. This could include producing biographies on individuals based on information collected from the World Wide Web (e.g., via Google searches, Wikipedia pages or other databases), and presenting that information to the digital archivist for consideration when performing categorization. More sophisticated tools could try to describe the relationship between the sender/receiver and the mailbox owner using text analysis from messages and other information available on the web (e.g., professional relationship, personal relationship, mixed relationship), and present these inferences to the digital archivist. However, the digital archivist should only use these inferences as suggestions; human judgment should be the ultimate judge of the historic and cultural value of email messages.

References

Bearman, D. (1995), "Archival Strategies", *American Archivist*, Vol. 58 No. 4, pp. 383-413.

Boles, F. (2005), *Selecting & appraising archives & manuscripts*, Society of American Archivists, Chicago, IL.

Bruni, F. (2014), "Hacking Our Humanity: Sony, Security and the End of Privacy", *New York Times,* 20 December, available at: http://www.nytimes.com/2014/12/21/opinion/sunday/frank-bruni-sony-security-and-the-end-of-privacy.html (accessed 27 January 2015).

Clark, D., Ovide, S. and Dworkin, E. (2014), "Are You Sure You Want to Use Email?", Wall Street Journal, 19 December, available at: accessed http://www.wsj.com/articles/are-you-sure-you-want-to-use-email-1419030075 (accessed 27 January 2015).

Cox, R. J. (2008), *Personal archives and a new archival calling: readings, reflections and ruminations,* Litwin Books, Duluth, MN2008).

Emory University Libraries (2011), "Rushdie Researcher Workstation Tutorial", available at: https://www.youtube.com/watch?v=oiqHv_SofNo (accessed 27 January 2015).

Gilliland, A. J. (2014), "Archival appraisal: practising on shifting sands," in Brown, C. (Ed.), *Archives and recordkeeping: theory into practice*, Facet, London, pp. 31-61.

Hangal, S., Chan, P., Lam, M. S. and Heer, J. (2012), "Processing Email Archives in Special Collections", paper presented at *DH2012: Digital Humanities Conference*, 16-22 July, Hamburg, Germany, available at: http://mobisocial.stanford.edu/papers/dh2012.pdf (accessed 27 January 2015).

Haws, B. (2013), "New York Philharmonic Digital Archives", presentation at National Digital Stewardship Alliance Regional Meeting, 14 June, New York, NY.

Lavoie, B. F. (2004), *The Open Archival Information System Reference Model: Introductory Guide*, OCLC and Digital Preservation Coalition, Dublin, OH and Heslington, UK, available at: http://www.dpconline.org/docs/lavoie_OAIS.pdf (27 January 2015).

Microsoft Inc. (2014), "[MS-PST] Outlook Perosnal Folders (.pst) File Format", available at: https://msdn.microsoft.com/en-us/library/ff385210(v=office.12).aspx (accessed 27 January 2015).

Okyle, C. (2014), "5 Lessons Leaders Can Learn From the Sony Hacking Scandal", *Entrepreneur*, 11 December, available at: http://www.entrepreneur.com/article/240843 (accessed 27 January 2015).

Pennock, M. (2006), *Curating E-Mails: A life-cycle approach to the management and preservation of e-mail messages*, University of Bath, Bath, UK, available at: http://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/curating-e-mails/curating-e-mails.pdf (accessed 28 January 28, 2015).

Prom, C. J. (2011), *Preserving Email,* Digital Preservation Coalition, Heslington, UK.: 2011), available at: http://www.dpconline.org/component/docman/doc_download/739-dpctw11-01pdf (accessed 27 January 27 2015).

Rinehart, A. K. and Prud'homme, P. (2014), "Overwhelmed to action: digital preservation challenges at the under-resourced institution", *OCLC Systems & Services,* Vol. 30 No. 1, pp. 28-42.

Robyns, M. C. (2014), *Using functional analysis in archival appraisal: a practical and effective alternative to traditional appraisal methodologies*, Rowman & Littlefield, Lanham, MD.

Samuels, H. W., (1992), *Varsity letters: Documenting Modern Colleges and Universities,* Society of American Archivists, Chicago, IL

Sanger, D. E., Schmidt, M. S. and Perlroth, N. (2014), "Obama Vows a Response to Cyberattack on Sony", *New York Times*, 19 December, available at: http://www.nytimes.com/2014/12/20/world/fbi-accuses-north-korean-government-in-cyberattack-on-sony-pictures.html (accessed 27 January 2015).

Shallcross, M. (2011), "The MeMail Project: Digital Curation at the Bentley Historical Library", *Practical E-Records, available at:* http://e-records.chrisprom.com/memail-project-guest-post-by-mike-shallcross/ (accessed 28 January 2015).

Stanford University Libraries (2015), "ePADD Project," available at: http://library.stanford.edu/spc/more-about-us/projects-and-initiatives/epadd-project (accessed 27 January 2015).

Appendix: Manual Email Appraisal Methodology

1) Make a copy of the original PST file. Original PST files should never be opened because they can be easily modified (e.g., opening a message will change it from read to unread). PST also cannot be opened as read-only files; thus, it is necessary to first make a copy. Microsoft Outlook 2010—which was used to do the appraisal—has sophisticated sorting, searching, and categorization functions. Almost no one ever uses the categories function, which allow you to color code and assign categories to messages. Thus, if the categories are not used (which they most likely won't be), create four categories in Outlook (4=Very Significant Record, 3=Significant Record, 2=Insignificant Record, 1=Very Insignificant Record) (see Figure 1). Set the auto-click option to automatically tag messages as 1=Very Insignificant Record, because most mailboxes will have a large quantity of these messages, and thus they can be easily assigned this category.
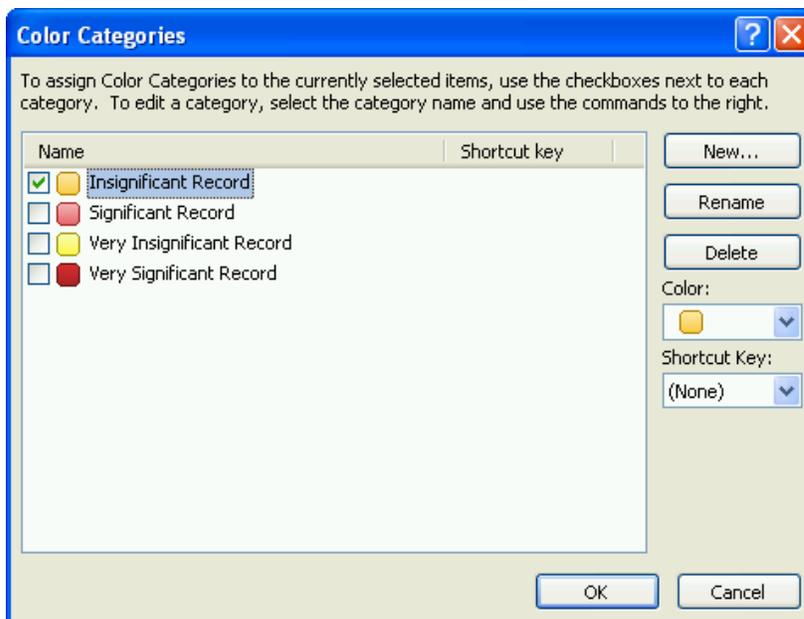

Figure 1. Create categories in MS Outlook and use these to categorize messages.

2) Change the default display in Outlook so that it will auto-preview all messages, which will provide about 3 lines of the message without the need to open the message. Also, turn off the reading pane. This will allow you to quickly ascertain the contents of the message, while not having to change the status from read to unread. If you accidentally open a message and change its status from unread to read, be sure to change it back.
3) Begin appraising by starting from the oldest message and moving to newer messages, and assigning messages to one of the four categories.
4) Be sure to mark as "Very Insignificant" messages that include correspondence related to personal finance, healthcare, and personal contacts that do not cross over into professional roles.
5) Google searches on the names of senders/receivers can shed light on their role within the institution and beyond, and help make distinctions between significant and very significant messages.

6) When complete, sort the email by category. Count the messages with Insignificant or Very Insignificant categories, and record their numbers. Delete these messages. In the descriptions produced, the number of deleted messages should be mentioned.

7) Make notes on the "Very Significant" emails, which can be used to highlight these emails in the archival descriptions of the collection.

8) Clear the categories from the remaining messages.

9) Compact the PST file (this will save disk space).

10) This PST file is ready to be submitted to the digital archives. A MBOX version should be created that will preserve the messages in a text-based format. The "readpst" open source program can accomplish this conversion.

11) Schedule the original PST file for deletion.

---

[1] PST is short for "personal store" and is the file extension for files created by Microsoft Outlook. Items that are held in PST include sent mail, inboxes, calendar items, task items, contacts, and other items available to users in Outlook.