



Building Email Archives in an Art Museum
Context: Policy and Appraisal Implications

Dr. Anthony Cocciolo
Archiving Email Symposium, LOC, June 2, 2015

Pratt



Background

- Consultant on an Electronic Records Start-up/Planning Grant for an Art Museum located in the United States (USAM).
- Has a paper-archives since the 1970s, and has approximately 7,000 cubic feet of paper records, including exhibition files, artist files, as well as other historical records.

Policy – BOT approves Records Schedule in 2005

Type of record	Retention Period			
	Total	Onsite	Offsite	
General (all departments)				
Correspondence	See list below			
Correspondence relating to other records	As long as the record to which it pertains			
Routine Correspondence	while useful	until discarded	n/a	
Significant Correspondence	P	3 years	Archives	Cor



Despite the policy...

- Few infrastructures in place to retain significant email correspondence.
- Migrated off of Novel Groupwise in 2002, and has used Outlook/Exchange since.
- IT maintaining the Outlook PST file after staff departure.
- Typically a supervisor has to request that it is saved upon departure of a staff member.
- Archivists are not always consulted during the departure of the staff member.
- Despite this, USAM maintained 365 PST files using 180 GB of disk space as of early 2014.



Infrastructures considered

- Creating an archive email box and attaching it to primary mailbox (like the MeMail project from University of Michigan).
- However, the training and IT resources needed to do this were through to be too great.
- Developed means to transfer email collections to archives (creating a new PST file, etc.).
 - However, we believed that the old system of seeking out PSTs on departure are the best option available in the low-resourced environment.
 - Limitation is of course that people can gut their PST file right before leaving.



How to get significant email?

- Must we go through every single message?
- Did just this for 3 mailboxes (2 curators and 1 executive).



Create a Rubric

- Scale from 1-4 (4 = Very Significant, 1 = Very Insignificant or Non-record or SPAM).
- Subject Matter – Mission Driven
- Sent Actor – Centrality to Mission
- Received Actor - Centrality to Mission
- Properties of message – read, replied to, etc.



Very Significant

- A noted designer discusses a potential commission project with a museum executive.
- An artist who is planned to have a retrospective at the museum exchanges a message with the curator about that retrospective



Very Insignificant

- An unopened newsletter from a popular magazine.
- An email from a museum executive to spouse asking him/her to pick up the groceries.



Process

- Use categorize/color code function in MS Outlook
- Set the default single-click action to “Very Insignificant”



Manual Appraisal

- Saves disk space (e.g., 36% disk space savings from one mailbox).
- 40-70% of emails deleted in 2 cases
- 5% of emails deleted in other case (seemed to have a personal assistant)
- One hour to appraise 641 messages.
- Thus, need ways to expedite this process.



Interesting Findings

- Executive: 217 significant or very significant emails in the “Deleted Items Folder.”
- Preserving sent mail is not significant to capture significant email. Nor is preserving sent mail with messages that have been acted-upon (e.g., replied to or forwarded).
- Unread messages are not necessarily insignificant (the person could have been briefed about its contents and not bothered to read it).



Speed-up solutions

- Large numbers of senders/receivers are always significant or always insignificant. E.g., curators emailing other curators are almost always significant. (40-62 significant/very significant, up to 306 insignificant/very insignificant)
- Do random sample for a given sender/receiver to see if he/she falls into one of these camps Small fraction of cases that are borderline. Figure out who those borderline individuals are, and subject those to fine-grained appraisal. For both 3 mailboxes, these were 5-7 individuals who were borderline.
- I call this taking a social-network approach.

Tools

- Many tools to-date have not helped email appraisal (e.g., sentiment analysis of email archives may be interesting for the researcher but does not help appraisal).
- Because of the sophisticated search/sorting/categorization available in MS Outlook, you can do a lot of appraisal work directly from within there.
- New tools for appraisal could incorporate web-based information (e.g., Googling peoples names, finding their Wikipedia pages).
- For developing predictive modeling or text analysis tools, I would avoid topic and content per se and instead concentrate greater efforts on uncovering the the nature of the relationship (professional, personal, or mixed), and present these inferences to the digital archivist.