

We're not toying around:
Developing email archiving workflows
with the James W. Toy papers

Elena Colon-Marrero, Alexis Miettinen,
& Gideon Goodrich



**BENTLEY
HISTORICAL LIBRARY**
UNIVERSITY OF MICHIGAN

History of Email Archiving at The Bentley Historical Library



- 1997 - First accession that included email, President James J. Duderstadt's Macintosh computer
- 2009 - Launched "MeMail" Project
- 2021 - ePADD experimenting begins
- 2025 - Use case emerges to develop new email archiving workflows

James W. “Jim” Toy



- Toy was an activist and advocate for LGBTQ+ rights and social justice in the State of Michigan.
- He worked to establish the Human Sexuality Office in 1971, which later became the Spectrum Center at the University of Michigan, the first center of its kind at an institution of higher education in the U.S.
- Toy fought to add sexual orientation and gender identity and expression to the University of Michigan’s non-discrimination protections.
- He began transferring his papers to the Bentley in the 1990s
- Toy passed away Jan. 1, 2022 and his estate transferred his final papers later that year.

James W. Toy's Email Conundrum

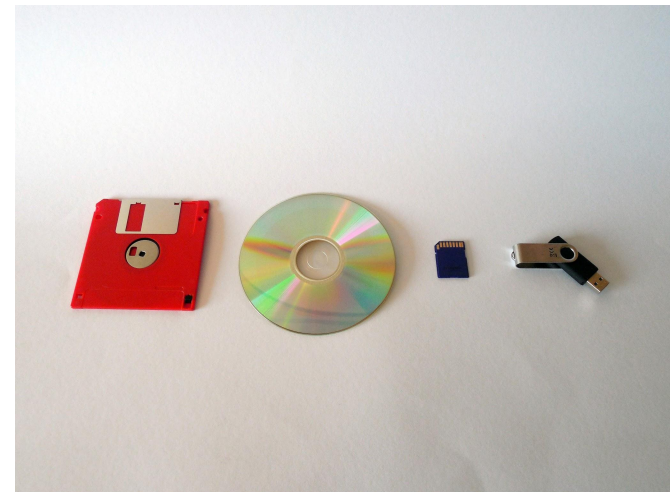


- The final transfer of materials included a hard drive containing Toy's complete university email inbox.
- Toy adopted email early on and used his university address for personal and professional correspondence.
- Emails were kept on the hard drive until they could be processed.
- Task for the Processing Archivists:
 - Export what you want to save
 - Keep it small (2GB or less)
 - Watch out for viruses

Scoping and Surveying Email



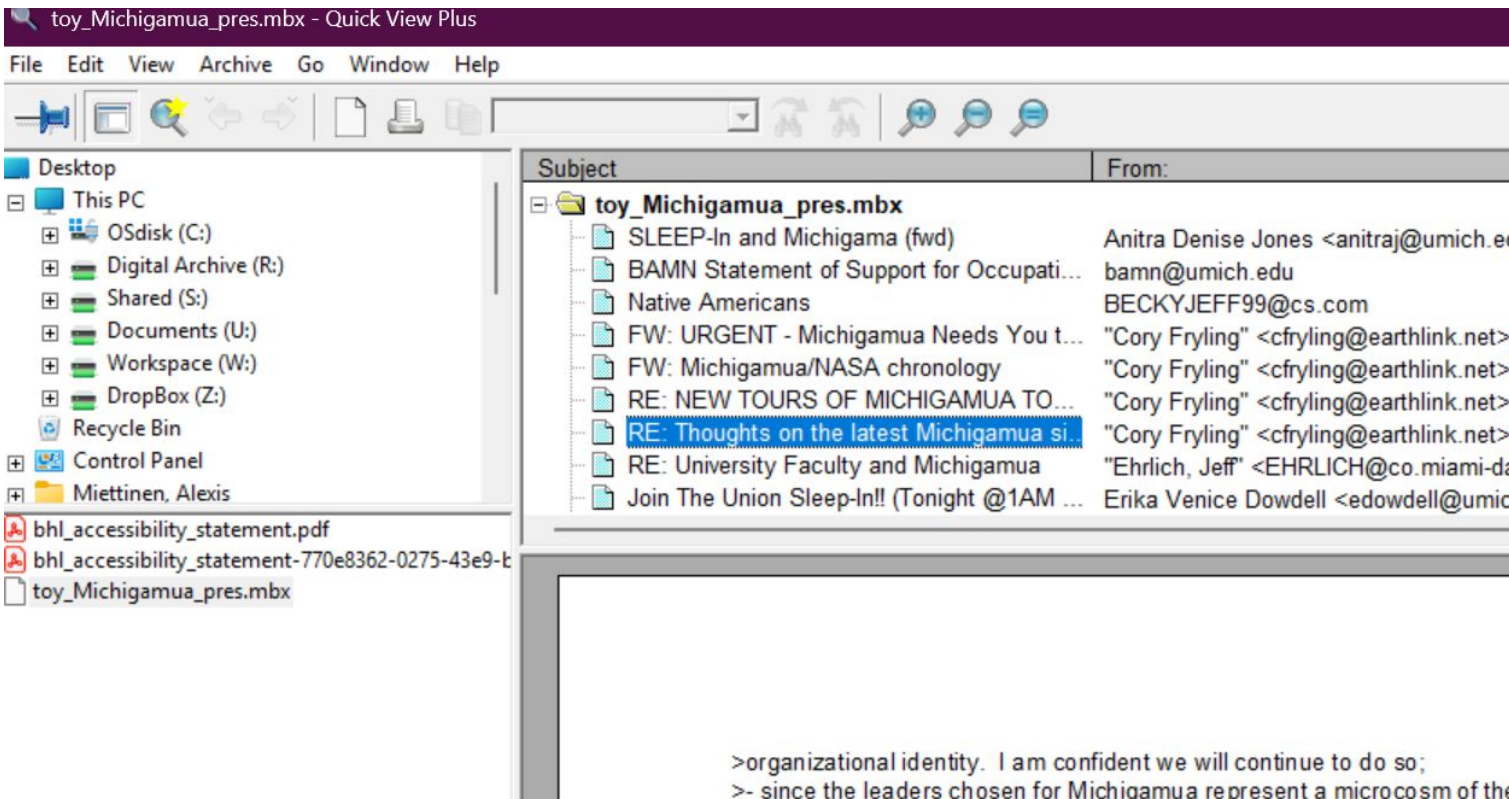
- Initial scope of the largest accession: 128 linear feet, 5.65 MB, and a note
 - “May be more removable media within the boxes”
- As we continued through boxes, more removable media continued to populate.
- There were email inboxes also on these pieces of removable media!
- After normal procedure of virus scanning, I determined it was safe to start surveying these in small batches.



Appraising and Deduping



- In organizing his inbox, Toy often labeled various topics with inbox “tags” and that is how many of the inboxes exported. This gave us a hint of what topic was in them, and helped immediately identify where there may be chance for duplicates.
- Over various pieces of removable media, there were over 80 individual “.mbox” files.
- I was unsure if my favorite tool, QuickViewPlus, would be able to open all the files and allow me to view them to appraise quickly.



>organizational identity. I am confident we will continue to do so;
 >- since the leaders chosen for Michigamua represent a microcosm of the
 >diverse campus society, I was privileged to obtain a better understanding
 of
 >organizations/people I had no contact with on a day to day basis;
 >- it is truly an HONOR to be a part of Michigamua and nobody can EVER
 >that away from me

But it was a success!



A different approach



- Determine if there was any duplication across .mbox files
 - Each inbox.mbox included every single email in Toy's gmail account
- Beyond that, was there any order we could make sense of?
 - Heavy use of foldering system, organized by topic or organization
 - Emails were only sorted if Toy intended to keep them

The screenshot displays the EPADD email client interface. The top navigation bar includes 'Import', 'Browse', 'Search', 'Export', and 'Save'. Below this, there are controls for 'LABELS', 'SORT BY', and 'ATTACHMENT VIEW', along with a '1 of 10' indicator. The left sidebar shows a list of correspondents: dtoy@att.net (8), Nancy T Young (7), Annette Homiller (5), Karen Downs (5), and david toy (2). Below this are sections for 'General Labels' (Other error w... (1)), 'Annotations' (Not annotated (10)), and 'Folders' (toy_FAMILY_pr... (10)). The main email content area shows the following details:

Folder: D:\epadd\objects\toy_FAMILY_pres.mbx
Date: Mar 13, 2000 8:38am

Labels:

From: dtoy@worldnet.att.net
To: Nancy T Young <ntyoung1@juno.com>
Cc: ayetfm@umich.edu, dtoy@att.net, homiller@earthlink.net, Ktoydowns@aol.com
Subject: **Re: eyes**

Hi,
I am also glad to hear the surgery went well, and that you had a nice weekend in NYC.

We had a quiet weekend, and ended up going to dinner with our friends the Marshes last evening at a Greek restaurant. I ate too much of course, but enjoyed it and we had a good time.

Reviewing Inboxes



- Load one inbox in EPADD at a time
- Once the ingest completes, check the reports tab to see if there are any emails that are missing metadata.
- Assess whether or not we need to keep files with:
 - Missing subject line
 - Missing date
 - Missing correspondence

Reviewing Inboxes



- There are a number of ways to tackle review of emails; our current workflow is:
 - Check the "Correspondents" tab -
 - Check the "Entities" tab
 - Use "All Messages" to complete the review
- Label messages you're keeping with "permitted for ingest"
- Export the "permitted for ingest" messages

The screenshot shows a dashboard for 'Email Archive of Jeb Bush'. The top navigation bar includes 'Import', 'Browse', 'Search', 'Export', and 'Save'. Below the navigation, there are tabs for 'Appraisal' and 'Dashboard'. The main content area displays a grid of nine summary cards:

Category	Count
Correspondents	9,838
Entities	8,120
Labels	8
Image Attachments	1,003
All attachments	2,126
Folders	4
Lexicon Search	
Reports	
More	

Corrupted Data



- Encountered corrupted emails in high priority inboxes (Order of Angell, LGBTQ+ activist groups)
- Enter: Thunderbird
- Open the offending .mbox in Thunderbird and click through messages until you find the corrupted files
- Delete those emails and save as a new .mbox

The screenshot displays the Thunderbird email client interface. On the left, the folder pane shows 'toy_FAMILY_pres.mbx' selected under 'Local Folders'. The main pane shows a list of messages in this folder, with the message from Nancy T Young (6/29/2000, 10:45 PM) selected. The right pane shows the content of this message, which is corrupted. The message header indicates it is from Nancy T Young (ntyoung1@juno.com) to dtoy@att.net. The body text is partially legible, showing a greeting and a paragraph about a family gathering. The message is marked as 'Re: Misc'.



Lessons Learned

- Use the tools that work for you
- Lots of resources out there walking through tools!
- EPADD works best as part of a multi-part assessment strategy
- Thunderbird helps with targeted removal of corrupted emails





Thank you!

Elena Colon-Marrero -
emcolonm@umich.edu

Gideon Goodrich -
gideong@umich.edu

Alexis Miettinen -
amiett@umich.edu

