Addressing the Challenge of the Governor's E-mail

Roger Christman
*Library of Virginia*

Susan Gray Page
*Library of Virginia*

Follow this and additional works at: [https://lib.dr.iastate.edu/macnewsletter](https://lib.dr.iastate.edu/macnewsletter)

Part of the [Archival Science Commons](https://lib.dr.iastate.edu/macnewsletter)

**Recommended Citation**


Available at: [https://lib.dr.iastate.edu/macnewsletter/vol43/iss1/10](https://lib.dr.iastate.edu/macnewsletter/vol43/iss1/10)
The State Library of Virginia has a long and impressive history of preserving and providing public access to the records of state and local government. But the shift from purely paper-based records to electronic records poses challenges not only to the Library of Virginia, but also to archives across the United States and around the world. In 2005, the library accessioned its first true transfer of born-digital gubernatorial records. Since then, the library has developed policies supporting the creation, transfer, and management of electronic content. As library staff, we understand and take seriously our responsibility to ensure the secure and stable management of this material and to provide open and free public access to the archival records of our government regardless of format. And yet, accessioning e-mail continues to present unique and difficult challenges.

In January 2010, the administration of outgoing governor Tim Kaine (2006–2010) transferred to the Library of Virginia approximately 1.3 million e-mail messages from more than 200 accounts. By law, gubernatorial records transferred to the library “shall be made accessible to the public, once cataloging has been completed.” The library has long had procedures in place for accessioning and processing paper records, but working with 167 gigabytes of e-mail required a new set of tools, including an additional staff member, a few key pieces of software, time, and a lot of patience to develop a process through trial and error. Four years after the original transfer, 66,422 e-mail messages from the Kaine administration were made available online. In 2014, four additional batches of e-mail were made available, bringing the total number of accessible e-mail messages to 130,644.

Setting the Stage

Budget challenges, staff vacancies, and the absence of definitive professional best practices hindered the library’s ability to move forward as quickly as we had hoped on the Kaine e-mail project. However, Tim Kaine’s April 2011 announcement of his candidacy for the United States Senate and the potential inquiries regarding his administration’s records gave us the opportunity to reassess and reconsider our priorities around these records, especially the born-digital materials.

With support from library senior leadership, a workgroup of archivists and IT staff assumed a self-imposed deadline of making the Kaine administration’s e-mail accessible to the public in time for the 2012 election cycle. The deadline proved to be unachievable due to many challenges, but the team pushed forward regardless.

We wrestled with a number of issues during this project, but first and foremost was the question of how the processed e-mail messages should be served to the public. Given a choice of limiting access to dedicated computer terminals in the library’s reading room or allowing anyone with an Internet connection to view the e-mails through the library’s online digital asset management system, we chose the latter. This seemingly simple decision had a significant impact on processing the collection.

Open Access versus Rights to Privacy

The library promotes open access to government records. However, open access must be balanced with the various laws that restrict access to certain types of records. Records with restrictions on access include those that contain personal or sensitive information (such as medical, educational, or personnel records), correspondence between attorneys and clients, and information related to clemency and restoration of rights. In addition, many inboxes contain e-mail messages of a personal nature or otherwise considered transitory and not worthy of permanent retention. To ensure that privacy-protected records were excluded from the collection, state records archivists decided that the electronic equivalent of item-level processing was warranted.

By applying classic retention scheduling concepts to the e-mail messages, archivists reviewed every e-mail in each e-mail account and segregated the e-mails that did not qualify as public records or were otherwise restricted from public access. Processed copies of the e-mail PST files were then passed on to the library’s information technology department for the technical phase of the project.

Despite our best efforts, we knew some restricted materials might slip through our manual dragnet process, so we created a virtual “reading room agreement.” To view the Kaine e-mail messages, users have to log in using a generic account that the library created for this collection. By log-
ging in, users acknowledge their researcher responsibilities regarding protected materials.

**Applying Technology**

Anyone who ever cheered on MacGyver as he repaired a radiator using only water and an egg white will appreciate what we had in our project’s toolkit: two determined archivists embedded in the library’s IT department and a few hundred dollars for processing software. When the decision was made to put the Kaine e-mail online, we began looking for a program that would export the e-mail messages from the processed PST files into full-text PDFs. Our goal was to serve static copies of the messages in a keyword searchable format. We also wanted to convert any files attached to the original messages to full-text PDF documents so users would not need additional software installed to view attachments.

Two software candidates quickly emerged, though both had limitations. Our top choice, PSTViewer Pro by Encryptomatic ($69.99 for one license), handles Microsoft Word, Excel, and PowerPoint attachments like a champion. But the most it could do with MSG files was to include them in the exported PDF as an attachment in their native format. The runner up, Total Outlook Converter Pro by CoolUtils ($99.90 for one license), easily converted all manner of Microsoft attachments, including MSG files, but the program routinely crashed when we tried to convert large PST files. In the end, we bought both programs and combined their functionalities to meet as many of our needs as possible. For all e-mail messages without MSG attachments, we used PSTViewer Pro. For the smaller number having MSG attachments, we used Total Outlook Converter Pro.

Once messages were exported to PDF format, we created CSV files for the corresponding e-mail metadata to facilitate use of the new collection. At this point we confronted head-on the inherent limitations in our current digital asset management system. In a perfect world, our resource discovery tool would let users browse, search, and retrieve gubernatorial e-mail records as effortlessly as if they were using Gmail. However, a system like ours, geared toward delivering more traditional digital content such as digitized maps, manuscripts, and photographs, presented challenges in providing access to e-mail. Although the e-mail messages are full-text searchable, the high volume of messages causes any given keyword search to return thousands of possible results. We tried a number of approaches to help users navigate the collection. For example, to better approximate the inbox environment of e-mail, we created subcollections for each individual in the Kaine administration and populated those collections with exported e-mails from the corresponding PST files. In this way, users can “step into the shoes” of specific administration officials and approximate what they saw when they logged into their e-mail accounts. We also worked within the limits of the software to optimize display options when viewing results sets. Finally, we created tip sheets to help users understand how to get the most out of the DigiTool search environment.

The final step in making the e-mail available to the public was the creation of a website. Kaine E-mail Project @ LVA is an online portal that acts as a hybrid finding aid, contextualizing the messages and integrating them with Governor Kaine’s other archival materials. The online portal contains a traditional finding aid, an interactive organizational chart for the Kaine administration, links to EAD guides for related paper collections, and links to archived websites related to the Kaine administration.

**Next Steps**

What’s next? The manual processing of the Kaine e-mail continues, but we realize that item-level processing of e-mail is not sustainable over time. That is why we are excited about a pending research project that will use computer software to try to replicate our item-level processing of the Kaine e-mail. If successful, this “Man vs. Machine” research project could help automate a large portion of our current efforts processing e-mail messages and set a path forward for working with many other collections where potentially sensitive content is commingled with public records. Stay tuned!
Notes


---

*It’s the Academy of Certified Archivists’ 25th Anniversary!*

After 25 years, the Academy is still going strong.

We are proud to celebrate our silver anniversary by introducing the new International Travel Assistance Scholarship.

For more information, visit us at [http://www.certifiedarchivists.org/](http://www.certifiedarchivists.org/)

Mark your calendar now: the 2015 Certified Archivist examination will be held August 19 in Boise, Buffalo, Cleveland, Jacksonville, Little Rock and Sacramento -- and wherever 5 or more candidates wish to take it. Applications will be available January 1 on the ACA website.