

The MeMail Project: Digital Curation at the Bentley Historical Library

By Michael Shalcross, Assistant Archivist, Bentley Historical Library, University of Michigan

The Bentley Historical Library was established in 1935 by the University of Michigan Regents to serve as the official archives of the university and to document the history of the state and the activities of its people, organizations, and voluntary associations. It comprises four divisions: the University Archives and Records Program (UARP), the Michigan Historical Collections (MHC), Digital Curation, and Reference Services.

Since the 1997 accession of files from university President James J. Duderstadt's Macintosh computer, UARP has managed diverse electronic records that include E-mail, desktop office files, audiovisual materials, and Web captures. Given the university's increasingly "paperless" environment, UARP launched the MeMail Project in 2009 to enhance its capacity to identify and preserve digital content of unique, essential, and enduring value. With its myriad complexities—related to creation, storage, and preservation—E-mail was selected as the initiative's primary focus.¹ A generous two-year grant from the Andrew W. Mellon Foundation in January 2010 allowed UARP to partner with the university's Information and Technology Services (ITS), bringing both archival and IT expertise to bear on digital curation. The grant also enabled UARP to hire two full-time archivists to serve as the project's functional and technical leads. The planning, development, and implementation work of the MeMail Project will provide the Bentley Library with the foundations of a program to provide enhanced archival services for an array of born-digital content.

Diverse E-mail applications and personal E-mail management practices at the University of Michigan have long posed obstacles to the archives' ability to accession, appraise, and preserve electronic mail of long-term value. The MeMail Project sought to overcome these challenges by pairing records management tools and techniques with outreach and education directed towards records creators. To gather background data on E-mail usage and potential preservation strategies, archivists interviewed staff from several campus units, analyzed business practices, and reviewed relevant policies, regulations, and laws. This research led archivists and ITS staff to conclude that an appropriately priced electronic records management system (ERMS) would permit UARP to preserve the correspondence of a target group of 1,500 administrators and prominent faculty. To streamline the identification of

significant content prior to accession, UARP planned to educate records creators in the use of file plans to self-select E-mail of value and/or have ITS develop functionality to automatically weed out spam. UARP defined functional requirements for the capture, identification, maintenance, secure storage, and dissemination of E-mail and associated metadata, and ITS drafted technical requirements for a commercial system. The partners then issued a Request for Proposal (RFP) to 18 vendors in mid-2009.

Of the two bids in response to the RFP, the most promising system cost over five hundred thousand dollars. While the partners had expected an ERMS to be costly, this amount exceeded the project budget and would have required university-wide participation and commitment. To compound matters, the University of Michigan launched an IT reorganization in late 2009, which eventually led to the consolidation of E-mail services and other IT resources on campus.² Rather than pursue a large investment in the midst of this changing environment, UARP developed an interim solution that would permit archivists to selectively capture E-mail of value and allow for the possible integration with an ERMS at a later date. As of April 2011, UARP has used this revised approach to initiate 10 pilot projects with high-level administrators from across the university.

This "mailbox method" requires ITS to establish an archival mailbox for each participant, to which UARP also has access. This mailbox is identified by the participant's University of Michigan username appended to a Bentley Historical Library prefix (i.e. "bhl-johndoe"), and it appears as a separate folder in the participant's E-mail client.³ Archivists configure the client to ensure that the archival mailbox is both visible and functional, provide an overview of UARP's collecting policies, and supply guidelines to identify correspondence of long-term value. Pilot participants are then asked to drag/drop, forward, CC, or BCC significant messages (both sent and received) to the archival mailbox for preservation. These messages and associated attachments are stored on IMAP or Exchange E-mail servers (depending on the participant's E-mail client) administered by ITS. The participants have continued access to content in the archival mailbox and may add or remove messages at their discretion. Although UARP has access to the mailboxes, archivists do not read individual messages and only check the accounts to monitor usage

(Continued on page 28)

(Continued from page 27)

for the pilot project. At an appropriate point in the project, UARP will request permission to export and accession E-mail in the RFC 2822-compliant MBOX file format.⁴

The MeMail Project is also addressing the procedures and resources needed to process and describe E-mail and other digital records. To this end, UARP has (a) established steps to securely transfer legacy E-mail and other content from a variety of storage environments, (b) created an interim repository to back up and run processes on materials, (c) identified important actions to prepare content and metadata for long-term preservation, and (d) adapted traditional arrangement and description practices to the digital environment. To establish a workflow that can accommodate diverse content types, archivists tested over 30 different software applications and identified critical pieces of infrastructure, including much-needed additions to the Bentley's collections database (BEAL) to track digital content and manage associated metadata. Policy development continues apace, as archivists work on preservation plans and policies regarding the transfer of "record" copies, digital separations, and access. A comprehensive Bentley Historical Library digital preservation policy will be formally documented in 2012.

The MeMail Project has allowed UARP to partner with additional stakeholders on the University of Michigan campus. Given the expertise needed to process, store, and administer large collections of archival materials, ITS recommended that the archives collaborate with Michigan's Library Information Technology (LIT) division to automate batch processing and enhance the functionality of the BEAL database. The Bentley Library has a long-standing relationship with LIT, and the two units signed an agreement in 2010 that allows archival records to be stored in Deep Blue, the University of Michigan Library's DSpace repository.⁵ With this choice, UARP can rely on a trusted partner for permanent storage and thereby avoid the cost of maintaining a separate infrastructure, while at the same time leveraging Deep Blue's capacity to distribute Dissemination Information Packages (DIPs). LIT also has experience developing the software and systems required to ingest and manage extensive digital collections (as exemplified by the HathiTrust Digital Library). This knowledge and expertise—in addition to LIT's established tools and procedures—will allow UARP to avoid reduplicating preservation efforts and achieve the MeMail Project's goals in a more efficient and cost-effective manner.

Over the remainder of the grant period, UARP will work with LIT to automate portions of the workflow and to develop resources to manage content. A recent draft of the processing plan calls for archivists to conduct initial virus scans and de-duplication and then arrange and describe content before placing a Submission Information Package (SIP) in a drop box. A script will then pipe the content through several micro-services, which include checksum validation, file format normalization, assignment of unique persistent identifiers, format validation, metadata extraction, Archival Information Package (AIP) generation, and deposit in Deep Blue. Procedures developed for E-mail, a relatively complex content type, will then be transferred (with necessary modifications) to other formats at a later date. Work is proceeding on specifications for SIPs and AIPs (in regard to the packaging of content and the extent/format of metadata), as well as development of the processing script and the BEAL collections database. UARP and Michigan Library programmers also will explore modifications to the interface and functionality of Deep Blue. Any improvements on these fronts will be made available to the larger DSpace community.

While much work remains to be done, UARP has made great strides towards a robust and proactive digital curation system at the Bentley Historical Library. With the full launch of the new Digital Curation Division in 2012, the lessons and advances of MeMail will be applied to the accession, preservation, and management of electronic content in the University Archives as well as the Michigan Historical Collections. Over the next year, we look forward to sharing information about the ongoing work and overall results of the MeMail Project with our MAC colleagues.

Notes

1. The name of the project underscores this focus on electronic correspondence: "MeMail" joins Michigan's iconic block "M" with "E-mail."
2. In February 2011, the University of Michigan IT Council recommended that the university adopt Gmail and Google docs as a unified collaborative platform (for more information, please see <http://www.ur.umich.edu/update/archives/110221/google>).
3. The University of Michigan currently hosts both IMAP and Exchange E-mail servers. Given the university's decentralized nature, individuals are free to

(Continued on page 31)

News From the Midwest

(Continued from page 26)

WISCONSIN

Wisconsin Veterans Museum (WVM)

To kick off the Civil War Sesquicentennial, WVM made its popular Civil War Database available on-line. Containing information about the men from Wisconsin who served in the war, this versatile database is searchable by any combination of name, residence, company, and regiment. It also contains a keyword search for a remarks field that will allow users to search for men who were killed, wounded, taken prisoner, and more. These options make the database a great tool for everyone from genealogists to local historians, and from Civil War buffs to scholarly researchers. Check it out on the WVM Web site: <http://www.wisvetmuseum.com>.

Electronic Currents

(Continued from page 28)

use any E-mail client. Archivists, therefore, developed plans to train pilot participants on a variety of clients that included Mac Mail, Mulberry, Pine, Entourage, Outlook, and the university's Web mail interface.

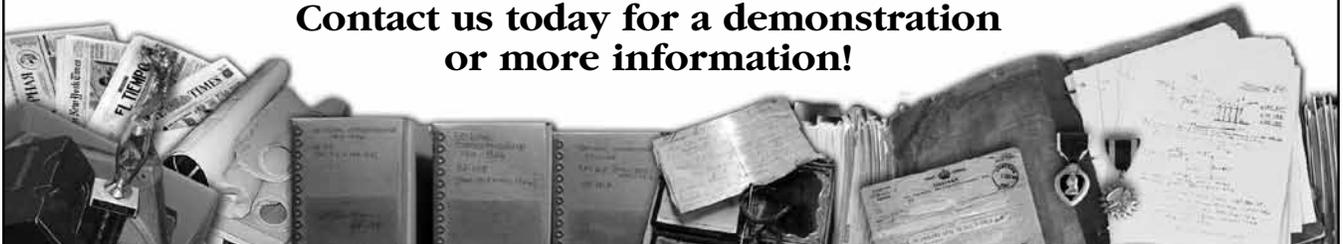
4. Request For Comments (RFC) 2822 is a standard for an Internet message format that defines the syntax and structure for electronic mail and was approved by the Internet Engineering Task Force in 2001.
5. The Bentley Historical Library's archival collection in Deep Blue is available at <http://deepblue.lib.umich.edu/handle/2027.42/65133>. Deep Blue also serves as the University of Michigan's institutional repository.



Knowledge Center for Archives

- Make your collections visible and provide for searching across and within collections
- Save time and staff resources by generating EAD finding aids, deeds of gift, MARC records, and more from your catalog records
- Get out from under your processing backlog by using software designed to help you manage your archives more efficiently

**Contact us today for a demonstration
or more information!**



Cuadra Associates | 800/366-1390 | sales@cuadra.com | www.cuadra.com/skca