

Tami Morse McGill

Gmail as institutional memory

Archiving correspondence in the cloud

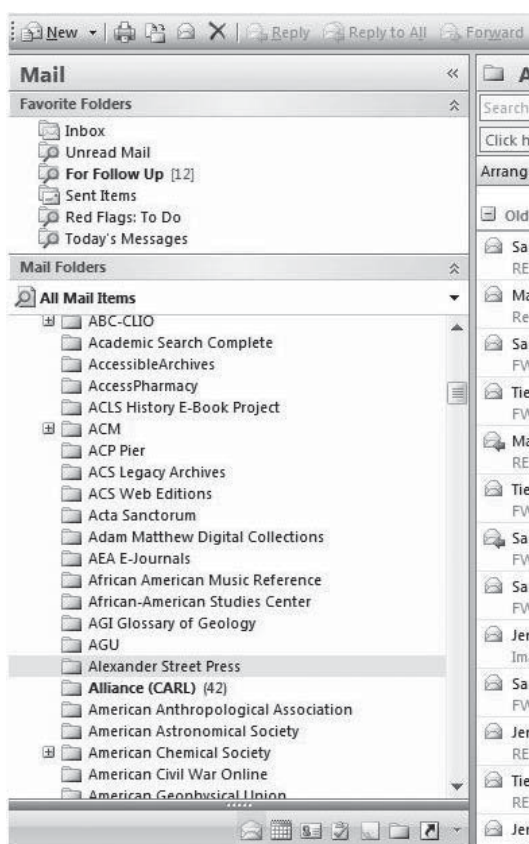
The University of Wyoming (UW) Libraries' Collection Development Office (CDO) has collected a large archive of e-mail correspondence over the course of four years. This archive contains records of communication with vendors, and many of the e-mails have attachments containing price quotes, license agreements, and other important documents. It has become an invaluable resource for CDO.

UW uses Microsoft Outlook and Exchange Server as its e-mail management system. While this works well for day-to-day communications, limitations in both Outlook and the university's e-mail policies made it inadequate for managing CDO's archive. This article discusses the workaround we devised for these limitations.

The initial state of the e-mail archive and problems with Outlook

The archive was created by CDO's first electronic resources librarian. Over time, the archive had grown to more than 450 MB and was still growing. It was stored as an Outlook archive (.PST) file. E-mails were sorted in folders based on either the product discussed or the vendor

with whom the correspondence took place. In some cases, subfolders were used to file e-mails first by vendor, then by specific product.



Gmail account access in the Outlook client.

Both the ERL and the head of collection development needed shared access to the archive. When we tried to share information about a particular vendor contact by referring to messages in the archive, though, we discovered we had been filing e-mails in two separate copies of the archive, and couldn't see the messages the other had filed. Our library systems department was able to resynchronize the two copies, but we were informed that Outlook .PST files could not be shared.

Our first idea was to create a new, active, and shareable e-mail account to contain the archive.

This didn't turn out to be feasible. The university limited the size of active e-mail accounts to 300 MB; our archive was already well over that limit, and (understandably) University IT would not make an exception for our case.

Tami Morse McGill is catalog librarian at the University of Wyoming Libraries, e-mail: tmorsemc@uwyo.edu
© 2009 Tami Morse McGill

Gmail as an interim solution

At around this same time, Google implemented an Internet Message Access Protocol (IMAP) interface for its Google Mail (Gmail) service. This implementation received some coverage in the technical press and blogs, and inspired the idea to use Gmail as an alternative to Outlook for the CDO archive. The IMAP protocol allows two-way communication between e-mail accounts, which would enable us to access the e-mail archive from the Outlook interface, and move e-mails from Outlook to Gmail fairly easily. The protocol also supports access to one mail account from multiple devices, and we could link to the account from more than one Outlook client simultaneously (see Google's "Getting Started with IMAP for

subject would either need to be duplicated in more than one folder, or stored in a new folder whose name described all the pertinent subjects. Gmail uses labels instead of folders, and more than one label can be assigned to a given message. This would provide more flexibility in categorizing the messages in the archive. Gmail also offers superior search capability, so e-mails in the archive could be retrieved efficiently.

Gmail would also have disadvantages as a storage service. The first and most concerning is the obvious one: we would be placing our valuable archive in the hands of a third party. Google's future intentions for Gmail cannot necessarily be predicted. It could become a for-fee service. Google could decide to dis-



The same set of e-mails in the Gmail interface.

Gmail" Web page,¹ or for a more technical explanation, the IMAP base specification.²)

Storing the archive in Gmail would solve our two primary problems: space and multi-user access. At that time, Google allotted 6.5 gigabytes of space to each Gmail account, and had been steadily increasing this storage allotment ever since Gmail was first made available. Even if the archive were to grow beyond the free space allotment, additional space was available at a very reasonable cost.

In addition to solving our immediate problems, Gmail offered other attractive features. The folder arrangement in Outlook meant that messages covering more than one

continue its Gmail service, and we could lose our archive or need to transfer it again at short notice. It's also difficult to assess how strong a commitment Google has to the privacy of e-mails in Gmail, now and in the future. (The Electronic Privacy Information Center's Gmail Privacy Page³ covers the issues; Tim O'Reilly refutes many of them.⁴)

We would also be giving up our control over the systems aspects of the archive. Our in-house IT department has a commitment to providing reliable service to the campus community, and informs us ahead of time of any expected downtime for our critical systems. The same can't necessarily be said

of Google, though experience with personal accounts in Gmail has shown it to be a very reliable service.

The transfer process

The first step in the transfer process was to create a new account in Gmail and enable its IMAP interface. This was accomplished easily through Google's Gmail signup page and the preferences page for the new account.

Connecting Outlook via IMAP to the new Google account was much less straightforward. It involved creating a new account in Outlook and configuring port numbers, server names, and security protocols to match Gmail's requirements. This needed to be done for each e-mail account

needing access to the archive. Google provides detailed instructions on their "Getting Started" page, as do several technology sites on the Web (see, for example, the How-to Geek blog.⁵⁾

The transfer was done by one person, and only one IMAP connection to the Gmail account was implemented until the archive was completely transferred. We opted to copy e-mails rather than move them, so the original e-mails were retained in the Outlook archive file as a safeguard, and the accuracy of the transfer process could be checked by comparing the messages in Gmail with those in the Outlook archive file.

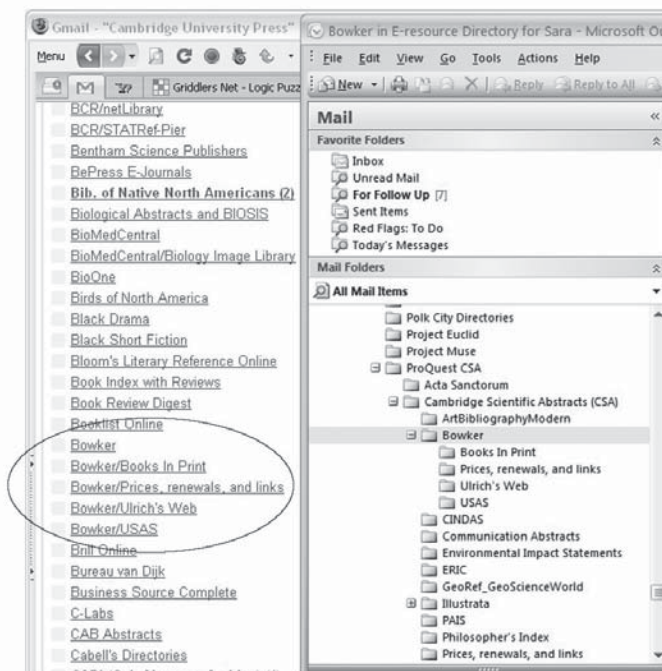
The IMAP interface allowed e-mails from the archive to be transferred by dragging and

dropping them in the Outlook client interface. Folders from Outlook could be dropped in the Gmail account, and all the e-mail contained in them would be copied, as well. The folder names automatically became labels in Gmail, and the e-mails contained in the folders were given that label automatically. Folders with subfolders could be copied all at once, and

their subfolders would be copied, too. Subfolders became single labels with the names of containing folders separated by slashes; for example, the subfolder "Correspondence" under the main folder "Gale" would become "Gale/Correspondence" in Gmail. On the Outlook side, the labels would still appear as folders and subfolders.

The folder-to-label conversion gave us our first problem. There is

a limit on the length of labels in Gmail, and long folder names or deep subfolder trees would sometimes exceed that limit. If the problem resulted from too deep a folder tree, and the subfolder name was unique enough, the subfolder could be copied to Gmail separately, instead of as part of the tree. Name uniqueness was a problem in some cases, though. Some folders in the Outlook archive were named for companies, with subfolders having generic names like "Trials," "Renewals," or "Correspondence," as in the Gale example above. Without the company folder name included, these subfolder names would become meaningless labels in Gmail. A generic label like "Correspondence" would end up grouping e-mails from multiple companies under one



An illustration of the folder-to-label conversion problem: the ProQuest folder tree is too deep, producing labels that are too long. The Bowker subfolder is copied over separately to shorten the label length.

label, defeating the purpose of the original folder arrangement.

The solution to this problem was to use the Gmail interface directly to create new labels with meaningful names. These labels would appear as folders in the Outlook interface, and e-mails could then be copied into them.

Performance also presented problems. Transferring a batch of e-mails could be

very slow. At times, some e-mail messages were dropped during the transfer, but the user was presented with a warning message when that happened. Finding the dropped e-mail could be very time-consuming, since most e-mails usually were transferred successfully and the dropped e-mail could occur anywhere in the batch. Finding the dropped e-mail involved carefully comparing the original folder to the folder in Gmail; this also demonstrated the importance of copying the e-mails rather than moving them. Once the dropped e-mail was found, it could be dragged to the Gmail folder, and it would normally copy without further problems.

Another intermittent performance problem occurred as several batches of e-mail were copied. Batches were copied one at a time, and a new batch would not be copied until the previous batch finished.

Nevertheless, after several batches were copied, failures would start to increase; either single e-mails would be dropped more frequently, or an entire batch would fail to copy with an error message saying the connection was terminated before the operation could be

completed. The only solution to this problem was time; allowing the interface to “rest” for a while seemed to give it time to catch up, and the transfer could be resumed without problems.

In the end, 5,314 e-mails were transferred from Outlook; slightly fewer, 5,015, were stored in Gmail. The difference occurred because Gmail automatically removed duplicates from the set

of e-mails; since all the messages were stored in one location (the Gmail archive), duplicate e-mails stored in different folders in Outlook were simply given more than one label in Gmail, and the duplicate was not transferred. This was another unforeseen advantage of using Gmail to store the archive.

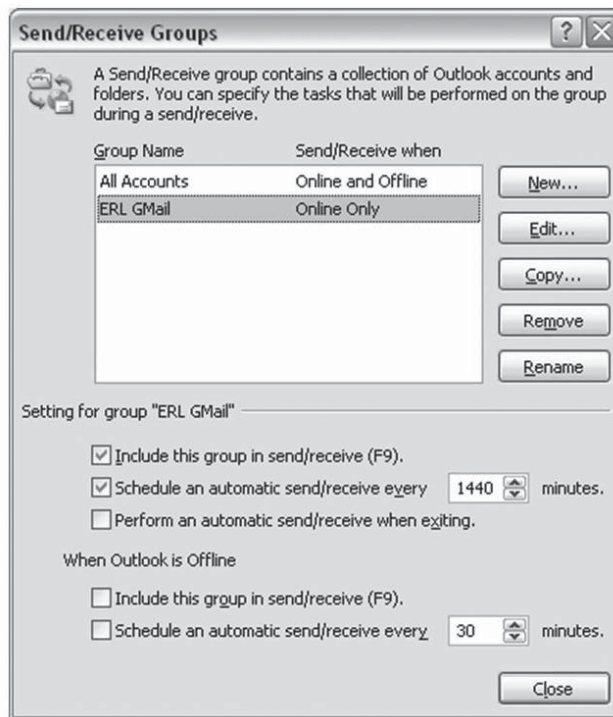
Experience so far

Some issues became apparent as soon as the transfer process began. The

most problematic is a significant slowdown in performance of the Outlook client. The cause appears to be in Outlook’s automated send/receive process; the slowdown occurs while this process is running. The solution involves creating a “Send/Receive Group” for the Google Mail account only. The automatic send/receive can then be set to run only once a day for this group, which is acceptable for an archive where little or no e-mail traffic is expected.

Another minor annoyance is the behavior of flags over the IMAP interface. E-mails that were flagged in Outlook when they were transferred to Gmail might appear multiple

(continues on page 645)



Outlook Send/Receive Groups, found under Tools > Send/Receive. Notice the long period of time in the schedule: 1,440 minutes is once a day.

lections from NSVRC and the Pennsylvania Coalition Against Rape. The site also provides information on activities of the center, including Projects, Publications (types and topics are numerous), Organizations, News, and Opportunities. The Publications section alone makes this a valuable site. *Access:* <http://www.nsvrc.org/>.



• **NCVC: Violence Against Women.** The National Center for Victims of Crime (NCVC) also provides information on violence against women. Domestic violence, rape-related posttraumatic stress disorder, and sexual assault information is available. *Access:* http://www.ncvc.org/ncvc/main.aspx?dbID=DB_ViolenceAgainstWomen155.

• **Rape, Abuse and Incest National Network (RAINN).** RAINN is the nation's largest organization dealing with sexual assault. This

easy-to-use Web site provides ways to Get Help, Get Info, and Get Involved. The News Room provides resources including multimedia, press releases, statistics, and speakers. *Access:* <http://www.rainn.org/>.

• **Toolkit to End Violence Against Women.** This site provides 16 chapters on combating violence against women. Each chapter discusses a specific issue from Strengthening Community-Based Services and Advocacy for Victims to The United States Within the International Community-Responding to Trafficking in Persons. Additional chapters highlight violence against women in the military, Native American women, and violence in sports. *Access:* http://toolkit.ncjrs.org/vawo_15a.html.

Note

1. United Nations Development Fund for Women: Violence Against Women. Retrieved from www.unifem.org/gender-issues/violence_against_women/. (Accessed October 25, 2009). ¶

("Gmail" continues from page 641)

times in the Outlook task list (the flags became stars in Gmail, but still appeared as flags in Outlook.) The ever-helpful How-to Geek also has a solution to this.⁶

At the time of this writing, three people are sharing access to the CDO e-mail archive in Gmail through the IMAP interface to Outlook. The main issues we were hoping to overcome have been resolved: space is no longer an issue, and we all have simultaneous access to the archive. The performance issue noticed during the transfer process has not become worse as more users access the archive.

Even so, the use of Gmail as a repository for institutional memory will need to be re-evaluated over time. There are risks involved in placing our valuable data in the hands of a third party. An in-house management system may prove to be a better long-term solution to the problem of preserving our electronic records and institutional memory, if time and funding allow. For now, though, Gmail is proving to be a satisfactory solution.

Notes

1. Google, 2008, "Getting started with IMAP for Gmail," mail.google.com/support/bin/answer.py?hl=en&answer=75725.

2. Mark R. Crispin, 2003, Internet Message Access Protocol-version 4rev1, [ftp://ftp.rfc-editor.org/in-notes/rfc3501.txt](http://ftp.rfc-editor.org/in-notes/rfc3501.txt).

3. Electronic Privacy Information Center, 2004, Gmail Privacy Page, epic.org/privacy/gmail/faq.html.

4. Tim O'Reilly, 2004, "The fuss about Gmail and privacy: nine reasons why it's bogus," www.oreillynet.com/pub/wlg/4707.

5. How-to Geek, 2007, "Use Gmail IMAP in Microsoft Outlook 2007," www.howto-geek.com/howto/microsoft-office/use-gmail-imap-in-microsoft-outlook-2007.

6. How-to-Geek, 2008, "Prevent Outlook with Gmail IMAP from showing duplicate tasks in the To-do Bar," www.howto-geek.com/howto/microsoft-office/prevent-outlook-with-gmail-imap-from-showing-duplicate-tasks-in-the-to-do-bar/. ¶