

**MODULE 15**  
**COLLECTING**  
**DIGITAL**  
**MANUSCRIPTS**  
**AND ARCHIVES**

**MEGAN BARNARD AND GABRIELA REDWINE**



**SOCIETY OF**  
**American**  
**Archivists**

---

## Appendix B: Case Studies

### Case Study 1: Working with a Creator to Select and Transfer Email

This case study focuses on the challenges that arose when working with a creator on the transfer of email files from his cloud-based account into the library's custody. For privacy reasons, both the identity of the donor and the receiving repository are not provided, but the case study relates actual events in an anonymous fashion.

The creator's collection contained both paper and born-digital materials, including a significant email archive. Starting in the 1990s, the creator saved much of his work and correspondence to floppy disks and other media, and, in the 2000s, he switched to using cloud-based email as his primary correspondence medium.

The creator engaged an agent to assist him in finding a repository for his archive. The agent was very experienced working with traditional, paper-based archives but was uncomfortable providing assistance with the born-digital materials. The creator and the library's staff were both committed to ensuring that the born-digital materials were included with the acquisition, and they had to work together, without the assistance of the agent, to facilitate the transfer. After the library and the creator reached an agreement about the acquisition of his collection, the materials were transferred in installments: first, the paper materials; second, an initial accession of cloud-based email, floppy disks, and CDs; and, finally, a second batch of email. The library will continue to receive future installments of materials for the archive.

Following the transfer of the paper-based portion of the collection, the creator needed some time to prepare for the email transfer. The creator wanted to separate his professional email from his personal email and requested the library's assistance in sorting and exporting messages from his cloud-based email client. The library's programmer wrote a batch script that identified emails based on the name of the sender or recipient, sorted messages into folders named for each sender or recipient, and created a Word derivative of each email message and saved it and any attachments alongside the original message. Each day for a month, the creator exported messages from his email client in eml format and transferred them via FTP (File Transfer Protocol) to a designated server at the library. The library's digital

archivist would receive the messages, run the batch script, clean up the Word derivatives by removing gibberish characters resulting from embedded formatting or attachments, consolidate folders when appropriate (e.g., merge “\_Doe, Jane” with “Doe, Jane”), and then transfer the migrated files and attachments back to the creator via FTP. The creator then used the named folders to separate his work-related email from his personal messages. It took a month to complete this process for approximately 14,000 email messages, which represented five years of his correspondence. Once the creator had decided which email messages and derivatives to transfer to the library, a staff member worked with him to encrypt a flash drive and send the messages to the library by postal mail.

The library and the creator encountered numerous problems throughout this project. First and foremost, the transfer and cleanup processes were much more labor intensive than either party had anticipated. The creator and the library were in contact throughout the process, often troubleshooting transfer problems, and corresponded by email daily for a month to complete the project. Second, the library’s server experienced performance and security issues. At the start of the project, the server kept timing out as files were being transferred, which resulted in discrepancies between the number of files sent and received. In addition, the library experienced a server crash during an overnight transfer, and, because the server had not been set up to automatically restart, library staff were not able to resolve the problem until the next day. Midway through the project, the library’s SFTP server experienced a denial of service (DoS) attack, which meant that for several days the creator was not able to establish a secure connection and had difficulty transferring files. The network speed also presented significant challenges and limited the number of files that could be transferred back and forth each day. Due to these ongoing challenges, the creator and the library mutually agreed that the final selection of emails would be copied onto an encrypted flash drive and transferred to the library by postal mail rather than over a network.

This selection and transfer process was so labor intensive and fraught with technical challenges that the creator and the library agreed to take a break and tackle the remaining emails at a later date. During this hiatus, the library had a personnel change, and a new digital archivist coordinated the second transfer of emails. When they were ready

to renew the email transfer, library staff reviewed the challenges that occurred in the first transfer, particularly those related to security, network infrastructure, and the labor-intensive nature of the work. The digital archivist and the creator discussed the challenges and agreed to take a different approach moving forward. The creator chose to sort his emails manually into business and personal correspondence rather than using the automated script employed for the first batch of emails. Once the files were sorted, the creator transferred them to the library through an FTP program. Although this method proved far less time intensive for everyone, the creator required substantial guidance and technical support from the library every step of the way.

In many ways, the experience with the acquisition and transfer of this creator's email files was an important learning opportunity for the library, and, throughout the process, library staff paid careful attention to the many challenges and difficulties that arose. The process revealed security and network infrastructure vulnerabilities that the library could then work to address, and staff were better prepared to anticipate and prevent such problems in the future.

One of the fundamental challenges was the amount of staff time devoted to a single acquisition before it even entered the building. In this case, the creator needed help to sort out technology issues and determine which emails he wanted to transfer to the library. If these had been paper materials, the library might have encouraged him to hire an assistant to help prepare his materials, and it is quite likely that the creator did receive such support from his agent in preparing his paper materials for acquisition.

Collecting and transferring an email archive was a novel endeavor for the library, and thus substantial staff time was allotted to this project because of the learning opportunity it presented. Yet it quickly became clear that this approach is not scalable, raising questions about how much staff time should be devoted to an incoming born-digital acquisition, as well as how much time is reasonable to spend interacting with a creator to facilitate a transfer. Manuscript repositories know how to quantify the supplies and staff resources needed to process a paper collection, but because born-digital collections often remain something of a mystery until the files are accessed, staff are less confident when measuring the value of born-digital materials in relation to the amount of staff time it takes to transfer and process them.

This case study also demonstrates the importance of having collection development policies and procedures in place to provide guidance for staff about the appropriate ways to approach born-digital acquisitions. In the absence of such policies and procedures, staff members must make difficult decisions on the fly, sometimes without the helpful vetting of colleagues, and the scope of tasks can quickly evolve in unexpected ways. The many valuable lessons the library staff members learned with this acquisition of email files are informing their ongoing development of stronger collection development and transfer policies and procedures.