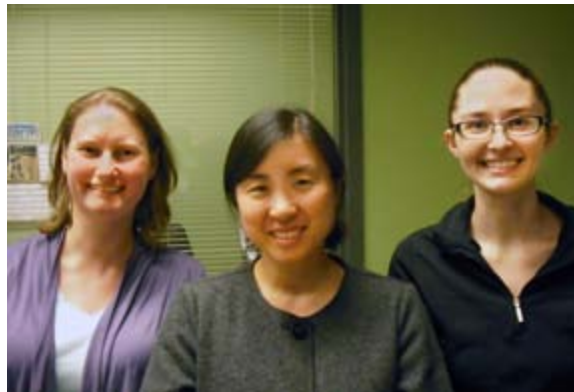


Jump-In Final Report for Emerson College

Introduction

In October 2013, the Emerson College Archives and Special Collections (ECASC) hired its first Digital Archivist (DA) to provide support in the areas of electronic records management, digital preservation, digitization, digital humanities and general systems administration. Prior to that date, the ECASC had been staffed by a “lone arranger” and the occasional temporary processing archivist or paraprofessional. Today the ECASC has three staff members: the Head of ECASC, the Archives and Records Management Associate (ARMA), and the Digital Archivist. As part of the assessment process for electronic records management and digital preservation, the Digital Archivist proposed surveying the electronic media currently held by the ECASC. This happened to coincide with the announcement for the 2014 Jump In Initiative. The Jump In Initiative appeared to be a good incentive to get the survey done in a timely manner and the recommended guidelines provided structure to the survey. The only catch: the DA is only funded for 24 hours a week to complete all the duties listed at the beginning of this report.



The ECASC staff

Survey Design and Process

The overall project planning, survey, and report writing process took almost four months (1/13 - 4/28). The project lead was the DA with support from the Head and ARMA as needed. As part of the planning process, the 2012 and 2013 Jump In Initiative reports were carefully read to identify any common challenges and strategies. The recommended guides were also used to inform the project design. Other resources included the Columbia and NYU survey tools

(http://library.columbia.edu/services/preservation/survey_tools.html and <http://library.nyu.edu/preservation/archivespreservation.html>) After completing the information gathering, the DA built a database in Google’s Fusion Tables because of the input form structure and the search/filter/editing functions. Field elements were selected based on a combination of the survey guides and input from the Head and ARMA. Since the Fusion Table will be used later for gathering statistics and locating materials, it was imperative that things like the source of the media and its location be noted.

After creating the survey template, the next step was deciding what types of media would be collected. Departments and donors tended to drop off electronic media in an ad hoc manner without any sort of metadata or note containing contextual information. Most electronic media arrived piecemeal since they were usually institutional records rather than part of a special collection. Although much of the media exists in multiple tape formats, it was decided that the survey would focus on media that could be easily read on a

desktop computer: 3 ¼" floppy, CD, DVD, USB, ZIP, and born-digital files were deposited into a networked folder set up by the ECASC. Anything that contained licensed commercial material or was not a primary source was not included in the survey. Items that had not been reviewed and accessioned were also not surveyed.

The first objects to be surveyed were the random piles of CDs and DVDs deposited on the DA's desk as well as the born-digital files in the networked folder. These first entries served as a trial-and-error process that led to revisions of the Fusion Table template. It also led to the discovery that many disks were missing file extensions and contextual information. The born-digital files were more straight forward but the sheer volume of files in a single folder made it impractical to record a precise count of objects. This later proved to be recurring theme throughout the entire survey. After the orphaned media were surveyed, the DA and the ARMA made a brief scan of the storage room for media. An intern from the Fall 2013 semester had created an inventory spreadsheet of all the multimedia objects in the storage room but the lack of detail made it difficult locate all the items he had noted. The DA decided to start by surveying the Master's degree theses, then go through the shelves of three of the four aisles (the last aisle did not contain relevant media). By this point, all the entries in the database were inconsistent in terms of formatting and detail. One reason for this was that the amount of information accompanying each disk varied greatly. Another reason was that time was simply running out. No photographs were taken along the way since the process would have taken much more time than was available.



A small sampling of our overall collection

Findings

The following table summarizes survey results:

Format	DVD	CD	Born-digital	3 ¼" floppy	ZIP	USB	TOTAL
# of items	385	302	15	12	5	1	720

Size	613.25GB	59.32GB	107.36GB	8.99MB	140.2MB	320MB	780.38GB
-------------	----------	---------	----------	--------	---------	-------	-----------------

One of the challenges of the survey was simply scheduling enough time to enter information. The DA's daily rate of entry ranged from as few as 10 items in a day to more than 50. On days where the processing rate was lower, the DA had to attend to other matters such as meetings and tasks for other projects. It was only days that were completely dedicated to entering information that a higher rate per hour (average 15 objects/hour) that more progress could be made. There was also the fact that the health and content of a media object affected how long it took to enter one record. The theses that were on DVD were sometimes very quick because they contained the same kind of files, similar number of files, and usually loaded very quickly. CDs often took much more time because they would contain hundreds of files with several levels of folders. Blank, unreadable, or problematic disks raised a red flag since the ECASC is not equipped for any kind of digital forensics work.

Next Steps

After reviewing the survey results, the following tasks were identified as necessary:

1. *De-duplicate and discard unnecessary media*
2. *Forensics*
 - a. *Is the media really unreadable or blank?*
 - b. *Is the media damaged?*
 - c. *Create summary profile of what is needed for media that is kept*
3. *Re-housing*
 - a. *Place media in binders with notes as to what collection it belongs to*
 - b. *Discard old cases*
4. *Preservation and conservation*
 - a. *Assess the content value and condition to determine whether migration for preservation is needed*
5. *Description*
 - a. *Summary of contents*
 - b. *Technical, preservation, and descriptive metadata*
 - c. *Where should these descriptions live?*
 - d. *Should there be a general finding aid for the loose content?*
6. *Access/Discovery*
 - a. *Assess the content value and condition to determine whether migration for access is needed*
 - b. *Determine what can and/or should be made accessible*
 - c. *Method of delivery to be determined later*
7. *Standardize entries and locations*
8. *Label the storage room shelves and binders*

While this work would normally be the domain of the DA and the ARMA, the progress would be very slow due to the College's request that the ECASC focus on developing its records management surveys and scheduling. The DA is also tasked with configuring and maintaining up to three new software platforms for archival description, digital preservation, and digital assets management. It is hoped that an intern can be brought on board to assist with the tasks listed above since they would be able to focus solely on that

project. At first glance, the tasks seem mundane, but the amount of critical thinking, research and writing that the description task requires will prove to be challenging as will any assessment-related activities.

Conclusion

In general, it is good to know the extent of our media collection, what needs de-duplication, what is damaged, how much space is needed for preservation storage - and to have it all recorded in a single location. It is quite likely that most of the content on the physical media will be migrated to a central storage location so that the media can be discarded to make space in the storage room. Most of the content will probably go onto an external hard drive while only a select number of items will go into the Preservica system for long-term digital preservation. Overall, the survey results will provide a foundation for the creation of other digital program services at Emerson College.