**Jump In, Too at the University of Pittsburgh’s Archives Service Center**

Ashley Taylor, Archivist

Motivations

In 2013, the Archives Service Center (ASC) at the University of Pittsburgh received the papers of Jason Altmire, who served Pennsylvania’s 4th Congressional District in the United States House of Representatives from 2007 to 2013. It came to us entirely on an external hard drive delivered by Altmire’s Chief of Staff, and included both born-digital and digitized material. This collection was a first for us at the ASC ; while we had received collections with electronic media before (the odd floppy disk here, a few CDs with PowerPoint presentations there), we had never received a collection that had no physical component at all. Being aware of literature on electronic records and professional best practices, we were able to establish a workflow for how to arrange and describe the Altmire “papers.” As a result of acquiring this collection, though, we realized that now was a good time to cement official policies about how we processed, preserved, and provided access to our electronic media on a larger scale. As a project archivist who would soon begin work on a much larger collection that included a significant e-record component- the papers of U.S. Senator Arlen Specter- I decided to dip my toe into the water by taking my first DAS courses in late 2013. After talks with my supervisor and colleagues in the University Library System’s Digital Research Laboratory (DRL), we decided that we should begin taking steps to establish an official, overarching policy for working with and preserving electronic records. With the announcement of the Jump In, Too initiative, I decided that this would be the best avenue for helping us get a better sense of the scope of the e-records currently in our collections.

Tackling the Survey

With the exception of the Altmire papers, practically every collection in the ASC’s holdings is based on analog materials: manuscripts, papers, audio-visual material, maps, artifacts, and so on. Any electronic media that comes in is usually only a disk or two that is processed and put in with the paper records physically where the record would have fallen intellectually. In collections where we have more extensive non-paper media, it’s often separated out into a unique series; for example, we might have a “multimedia” series with different subseries for “Cassette Tapes,” “CD-ROMS,” “VHS Tapes,” etc. This method allows us to store the multimedia items in better locations, such as more appropriate containers and in a smaller, more strictly environmentally-controlled area of the stacks. Sometimes, the presence of electronic records would be noted in the finding aids; sometimes, it happened to be a matter of knowledge on the part of the person who processed the collection.

In order to determine which collections contained electronic records, I began by searching through our finding aids using keywords. These keywords included “disc,” “disk,” “CD-ROM,” “CD,” “CDs,” “floppy,” “audio,” “visual,” “electronic,” “media,” and “multimedia.” As expected, some of these searches turned up material beyond the scope of my project, such as VHS and cassette tapes. However, I decided to err on the side of being thorough, rather than risking missing one lone CD tucked away in a folder somewhere.

Though this search got me a baseline of collections to start with, I knew from experience that one of the best ways to determine what was in the collections would be to ask my colleagues. I work in an office with seven other archivists, and we each have subject specialties and can be intimately familiar with some collections that the rest of us have never worked with. At a staff meeting in mid-February, I passed out a list of my preliminary list of collections to survey, and solicited ideas about which other collections to look in. This yielded more results, including in collections that did not have finding aids, and in some that were not finished being processed. To limit the scope of the survey, my supervisor and I decided to limit my survey to only finished, processed collections.

One of the more surprising benefits of asking my colleagues for their input was that it prompted them to think of e-records that they might have otherwise forgotten about. One of the biggest surprises was the discovery of two hard drives that resulted from staff and physical space changes over the course of a few years. Discussing e-records at a staff meeting prompted colleagues to remember the forgotten drives, and allowed us to re-discover that which may have otherwise been lost. It also got some colleagues to ask if we were going about processing these items the right way; should we be creating digital copies as soon as they come in? What would we do about items we couldn’t provide access to in our current reading room configuration (floppy and zip disks being the main focus here)? This discussion strengthened our decision to delve a little deeper into the world of electronic records.

Results

Using the list of collections I came up with using my keyword search of the finding aids, as well as the collections suggested to me by my colleagues, I was able to identify 17 collections (out of our total of 815) which contained electronic media (excluding, for survey purposes, the Altmire papers, which have already been processed and preserved). For the sake of organization, I divided my Excel sheet into two tabs, each focusing on the two main groupings of records within our holdings: UA (University Archives), and AIS (Archives of Industrial Society). I found six different types of electronic media, the breakdown of which can be seen in the following chart:

CD-ROMS led the way with 59% of the tally; Zip disks clocked in 2nd at 23%, due to one early-2000s collection in particular, which contained 68 of the disks alone. We had smaller quantities of 3.5” floppy disks and DVDs, along with two hard drives (Altmire, plus the two rediscovered drives) and two flash drives. The total maximum *potential* storage space tallied up to 240.3 GB for known material. However, having looked at the contents of many of the CD-Rs, many disks only contained a few smaller documents or images, using nowhere near their max storage capacity of 700 MB. Not included in these totals are the two currently inaccessible hard drives, which we are unable to determine the size of. If going by guidelines suggested on the Sample Inventory Template (<http://www2.archivists.org/groups/manuscript-repositories-section/jump-in-tootwo>) provided for the purposes of the Jump In project, we would estimate each at 1 TB; however, given their age (mid-to-late 1990s, we estimate that the actual size would be considerably less, even if we are able to recover any data that is on them.

After consulting with my colleagues and my supervisor over which direction to take and what collections to survey, I did the actual survey of the items on my own. This took me about 3 hours a week over the course of a month to pull each collection, identify the media, and enter the metadata into my inventory. Looking at the inventory, one sees that when the number totals of a particular media were very high (112 CD-Rs in the Western Pennsylvania Conservancy Records, for example), I did not enter title metadata for each disk. In the interest of time, this work can be done at a later date, especially once we decide on a course of action for these records. All of the material dated from 1991 to 2013, with one exception: analog photos that had been digitized and included on a flash drive and a CD in the Dr. Thomas E. Starzl Papers. These photos went back as far as the 1930s. Most of the media was in good condition; there were a couple CD-Rs with slight scratches, and a few floppy disks that were a bit dirty and scuffed. The major exceptions were the two aforementioned hard drives; they were taken to the University Library System’s IT department, where they are currently working to see if they can be salvaged or read.

Looking Ahead

This survey was at least mostly reassuring: we have a manageable number of electronic media objects, they’re generally in very good condition, and we have the means to access the data on all of them. Though we do not have a floppy drive or Zip drive in our office, we can get access to them through our IT Department. With the arrival of the Altmire papers, and the looming challenge of the Specter papers, we have previously worked with IT to secure server space for storing e-records, including a preservation, “deep storage” copy. This is similar to the way we work with the photos and manuscripts that we have digitized in the past. In conjunction with the DRL, we have an extensive record of metadata and workflow information that is kept for each object that has been converted to digital media and mounted on our various websites.

After the survey was completed, my supervisor and I met with colleagues in the DRL to discuss our plans for laying out a roadmap for the future regarding digital records. The DRL is working on implementing PREMIS and FITS records for any digital material going forward, partly in anticipation of moving all of our digital content into a Fedora back-end repository system. In addition, as the ASC preps to move our collection data into ArchivesSpace (from Archivists’ Toolkit), we plan to investigate the ways in which metadata can be gathered and shared amongst the different tools. Following guidelines such as Julianna Barrera-Gomez and Ricky Erway’s “Walk This Way: Detailed Steps for Transferring Born-Digital Content from Media You Can Read In-House,” Ben Goldman’s “Bridging the Gap: Taking Practical Steps Toward Managing Born-Digital Collections in Manuscript Repositories,” and the AIMS Born-Digital Collections 2012 report, we intend to work together to begin assembling policy and workflow documents to help us deal with the material we have, and plan for how to ingest any future electronic records. The Jump In project has helped give us concrete information about the importance of acting soon to ensure that we are good stewards of our electronic records, now and into the future.