**Jump In, Too/Two Report**

**Adriane Hanson, Processing and Electronic Records Archivist**

**April 22, 2014**

**Institutional Context**

The Richard B. Russell Library for Political Research and Studies at the University of Georgia has been receiving a steady trickle of electronic records, mostly on floppy disks and CDs, for over two decades. We are expecting that volume of records to increase significantly in the near future as we begin accessioning papers from individuals who have been relying on computers for most of their tenure in office. The Jump In, Too/Two initiative provided me with the perfect impetus to get a handle on our current holdings of electronic records, most of which were received prior to my hiring in 2012, and to think about how to move forward from implementing electronic records processing projects to developing an electronic records program.

**Survey Method**

I surveyed all currently identified electronic media in our holdings, a total of 935 items that had already been removed from their collections. Most had been removed as part of a project undertaken by aprevious archivist, who also transferred some of those files to a server. No doubt there are other media remaining in unprocessed collections, but these represent the bulk of what we have received to date. I undertook the following steps, with the help of a student:

1. I compiled information from all available documentation into a single survey spreadsheet. This included accession records, a spreadsheet from the removal project, manifests of files copied to the server, and collection files. The fields in the survey included those suggested by the Jump In initiative, as well as the associated collection number and accession number.
2. A student labeled disks with a unique electronic media identifier.
3. We added information from the disk labels to the survey spreadsheet. A student transcribed label titles and IDs, while I recorded technical information.
4. We matched media to their accessions. A student identified all accession records which included a reference to electronic media and I reviewed them to match the accession description with the disks we have. In some cases, the accession records were not detailed enough to do so, and I had to label the disks as accession unknown.
5. I created a record in Archivists' Toolkit for each of the 72 accessions.

**Survey Results**

We have 935 items, received in 72 accessions, which are part of 52 collections. The combined storage needs for all of these is about 0.5 terabytes. The earliest accession was in 1994, and we have received 1-10 a year since then. Of the media, 135 (15%) are unlabeled. The remaining includes a wide range of materials, such as backups, data sets about voters, digital photographs, email saved in various ways, press files, files for websites, and word processing documents of various kinds. We currently have the hardware to read all of the media types. Most were 3.5-inch floppy disks, CDs, and DVDs. We also have 5.25-inch floppy disks, thumb drives, external hard drives, and zip disks. We do not, however, have the software programs needed to access all of the files. Among the files already transferred to the server, we have over 1,000 different file extensions.

**Lessons Learned**

**Resources required: staff and time**

I was pleasantly surprised with how little time it took to gain a basic understanding of all of the known removable media in our collections. We did have the benefit of these materials already being gathered in one place, which would add time to other projects. Altogether, I spent about two weeks on this project and my student spent 20 hours. Having student assistance made the process much more efficient. I was able to work more than twice as fast after having students join me on the project. To break down the speeds for anyone looking for accessioning metrics:

* Label media with ID and transcribe media label to spreadsheet: 60 disks/hour (student)
* Create accession records in Archivists' Toolkit: 15 minutes a record (archivist)
* Matching disks to accessions: 50 disks/hour (combined student and archivist)
* Project planning, including cleanup of legacy data, developing accession record template, and student work instructions: 10 hours (archivist)

**Maximum Size vs. Actual Size**

The current practice is to estimate your total storage needs based on the maximum storage capacity of the media you hold. Since we have already migrated the files off of many of our media to a server, I decided to test how accurate that estimate is likely to be. I compared the maximum size to the actual size for 25% of the 5.25-inch floppy disks, 3.5-inch floppy disks, and CDs that had been migrated, for a total of 127 items tested. Almost none of these were full, and on average they were less than half full. It would be helpful to have some other institutions run this test to see if my results are representative, but it does suggest we could estimate our storage needs at about half of the maximum capacity, making the estimate more accurate and less likely to overwhelm an IT administrator with limited server space to offer the archives.

**The Importance of Unique Identifiers**

Most of our disks had a unique identifier of some kind assigned to it, although in different formats depending on when it was assigned. If the documentation also referred to that identifier, the process of associating the files with their metadata was simple. When the documentation did not, however, it made for a lot of work to try to match them up based on media type counts or content, and I could not be 100% sure that they were matched up properly. There are cases where we can no longer determine what accession a disk came with or what disk some of the files on our server came from. This means a loss of contextual information and adds uncertainty to the chain of custody. So moving forward, we will write a unique identifier on each disk as it is accessioned and include that number in accessioning and processing documentation, as part of the folder title where the files are stored on the server, and in the description.

**Retroactively Accession Records: Are They Worth It?**

The short answer is yes, because having all the information about our electronic records in once place makes it easier to manage and analyze the data. We use Archivists' Toolkit for all of our accessioning. I made a separate accession record for electronic records received with paper so I can export the data about just electronic holdings into Excel for analysis. For instance, for any group of accessions I can add up the total file size, get the date range, see the main file formats, and more. How good my analysis can be is dependent on the quality and structure of the data, and to that end I added two fields to our records: associated papers and file formats. Having this information in its own field, rather than in a general text field, allows me to search by it. So by putting all our legacy media into Archivists' Toolkit along with accessions moving forward, I am able to analyze that data as a single data set without the inefficiency of maintaining a separate tool with the same information (such as keeping the survey spreadsheet up to date) for that purpose.

**Moving Forward**

My survey gives me an overview of the state of all our electronic records holdings, which suggests a few immediate actions to take. First and foremost, I identified 265 items that have not been copied onto our server. The server space I have is limited, so I had waited on transfers to leave room for future accessions. But now that I know the likely space needed for these materials (44 GB), I feel confident that I can copy them to the safer preservation environment while still having space for new accessions.

Through inventorying media and producing accessions records, I learned I needed to modify the accessioning workflow I developed last year. I have already added a few fields to the accession records to keep the data segmented and therefore easier to analyze. Next I will be testing different reports that Archivists' Toolkit can produce and adjusting the records to make the resulting data even more useful.

And finally, I can use the survey to prioritize future work. For instance, I can quickly see what collections are open, but still have unprocessed electronic records, to work on those first. I can also identify the most common file formats and research appropriate preservation and access formats for those first. I'll be doing a lot of work in the coming months to process, preserve, and provide access to these files, but having a general understanding about what we have from this survey gives me confidence that we'll be moving in the right direction.