

UCLA Library Special Collections *Jump In* Survey Findings

Submitted by Gloria Gonzalez and Kelley Bachli, May 1, 2013

UCLA Library Special Collections is one of the largest research library repositories of archival holdings in the country, stewarding over 3,600 manuscript collections that are stored in a remote library facility. Like many of our peers, we had done little more than put disks in boxes to manage our born digital content, making the *Jump In* initiative the perfect opportunity for us to take our first steps. Without any additional funds or resources, surveying our materials initially felt like an insurmountable task. However, using the fairly thorough data that was available to us made this project manageable. While our survey is not comprehensive, it provides a good gauge of what collections have born digital content and gives us a great place to start more thoroughly surveying and caring for these materials.

To complete this project we used three different resources: an acquisition log, searchable finding aids on the Online Archive of California, and knowledgeable staff members. We started with the acquisition log, which documents the acquisitions from the former UCLA Department of Special Collections (Library Special Collections was established in 2010 and combined 5 different special collection units. Unfortunately our log does not include everything from all of these units, but does contain the bulk). We considered every collection acquired after 1980, which totaled 427 collections. We quickly ruled out collections that clearly would not contain born digital content because of the period or format of the materials, which further eliminated 32 collections. This narrowed the total to 395 collections that possibly contained born digital holdings.

Our next step was to search our finding aids on the Online Archive of California (OAC). 130 of the collections had no finding aid available online, and could not be searched. For the remainder, we used controlled search terms that accounted for various spellings of media types, and also expanded beyond only born digital media to encompass audio and video formats, or all “at-risk” media. Once we had established a list of collections with at-risk media, we solicited input from staff members who contributed a few additional entries. In total we discovered 42 collections that contained born digital content. We created a Microsoft Access database to record our findings and were able to generate a variety of reports that show the types of collections, media types, and location of materials. The data was also imported into Viewshare for further analysis.

The survey resulted in a variety of interesting observations. Early media were often seen accompanied by computer printouts of the content created by the donor or processing archivist. Many collections with born-digital material only included a few items, and only a few collections contained a large amount of digital media. Our findings indicate that the tipping point for hybrid acquisitions appears to be around 2001, with few collections with born-digital materials accessioned before and a consistent increase thereafter. Earliest dated born-digital materials located include IBM punched cards and magnetic tape outputs from the papers of Charles Nixon, a Professor Emeritus of Political Science at UCLA. The most recent born-digital media, compact disks with architectural designs in PDF format, are

located in the John Blanton papers currently being processed in the Center for Primary Research and Training. Other highlights from our survey include zip disks containing prisoner artwork in the Arts-in-Corrections records, 5.25" floppy disks of research files in the Southern California Women for Understanding collection, and an external hard drive in the UCLA Civil Rights Project administrative files. The examination relied heavily on OAC finding aids, so our next step will be a survey our backlog. Overall, the materials uncovered by the *Jump In* survey at UCLA Library Special Collections proved the survey to be a great addition to the foundation of our emerging program for born-digital archival management.

The timing of the Manuscript Section's *Jump In* Initiative greatly complimented the other recent strides taken in our efforts to manage born digital content in collections. These advancements include the construction of a born digital workstation fitted with forensic equipment and software for a variety of media, including floppies and hard drives. And in the coming months, we will hire a digital archivist to lead our efforts to develop standard methods for acquiring, processing, preserving, and accessing born digital material. We will also be partnering with UCLA Information Technology and Digital Library staff to migrate media into a preservation environment and to develop an access platform. To begin thinking collaboratively on how to manage born digital material, we are in the process of establishing a University of California wide "common knowledge group." In all, while we still have many goals to reach, we feel that "jumping in" has definitely aided in our tremendous progress.

Strategies used to locate and inventory at-risk media:

- Online finding aids were searched for the following terms: compact, computer, CD, digital, disk, disc, drive, dvd, electronic, flash, floppy, IBM, Mac, Macintosh, Microsoft, memory, pc, Sony, USB, zip, external, audio, video, cassette, tape, VHS, beta, reel;
- Use of Boolean search operators. Terms were enclosed in quotation marks to indicate exact phrasing. For example, this would prevent a search for "Mac" from returning all finding aids with the word machine.
- Microsoft Access categories used to track data: collection name, collection number, collection dates, accession date, born-digital media type, item locations (box numbers), maximum storage capacity estimate (collection-level), finding aid link, and notes. The notes field was used to include any explicit description or identifying information from the finding aid, and the number of each media type in the collection.

Summary of findings:

- 78 collections identified with at-risk material (including born digital, audio and video formats);
- 42 known collections with born-digital media with an overall maximum storage capacity measured at an estimated 3.9 terabytes;
- Storage media found include: punched cards, magnetic tape, 3½" floppy disks, 5¼" floppy disks, compact discs, DVDs, zip drives, and external hard drives;
- 133 additional collections without a finding aid on the OAC that possibly contain born digital content.