What the Heck Is This?!

Jumping into Managing Born-Digital Content at the University of Arkansas at Little Rock

By Chad Garrett and Shannon Lausch

The Jump-In team at the University of Arkansas at Little Rock Center for Arkansas History and Culture (CAHC) comprises an archivist and the director of technology. When we first received the call for participation via the SAA email list, we were struggling to find our way. We had come up with a workflow for digitizing photographs, audio, video, and so forth. We had each separately attended workshops and sessions on born-digital materials at conferences. We had even attempted to evaluate our operation using the AIMS (An Inter-institutional Model for Stewardship) Model of Born-digital Stewardship. Still, we were in need of direction because we had not focused our attention on what we actually have—beyond a few items that we knew had come in within the last two years, that is.

The day that Chris Burns sent out the call for participation, the two of us emailed each other with excitement. One of us wrote, "It's something we need to do anyway, so it's win-win." And indeed it was. CAHC is now taking its first steps in acquiring physical and intellectual control over its born-digital holdings. The Jump-In Initiative propelled our first step: conducting a survey of the born-digital materials in our collections.

Our Process

Before we began our survey, we first established what we meant by "born digital." We defined born digital as any digital record without a physical surrogate and any storage device containing records that were computer generated, whether that means CDs and flash drives or the computer itself. We did not include audiovisual materials unless they met the aforementioned conditions. We also excluded digitized materials with physical surrogates.

To locate relevant collections, we first tried to identify digital media by doing a search across all of our finding aids and inventory lists for terms such as "CD," "computer," "disc," and "disk." This method did not return as many results as we expected. We realized we could not necessarily predict how, or whether, digital items were described. We then developed another method. We immediately ruled out collections with date ranges ending before the 1970s, because we believed that digital storage containers would likely not have been donated to us prior to that decade as they were not yet widely used by the public. We then took a line-by-line look at the finding aids and inventory lists including items from the 1970s and onward. From there, we were able to identify the collections holding digital materials.

Our Inventory Spreadsheet

Once we determined which collections included digital containers, we then physically examined these materials and gathered information from them for our survey. We described materials at the item level unless they were part of a sequential set of data (for example "Interview with John Smith, 2 CDs").

Our survey was modeled from the example given on the Jump-In Initiative webpage. However, we did stray from that template in a few areas. We created this survey in Microsoft Excel instead of a Microsoft Word Document so that we could take advantage of Excel's features in calculating our numbers and sorting our data. We did not assign inventory numbers as our standard naming system is to assign a number based on an item's location. Many of the collections that contain born-digital materials are unprocessed, so we would not be able to create a permanent inventory number using our location-based naming system. We also added a field for date to ensure that it was recorded, if available.

In the course of surveying our materials, we also made the unfortunate discovery that four hard drives from computers we were storing in an off-site facility had disappeared. Within the past two years, the archives unit was reorganized under new management, and this inventory was moved before the current archivist and technology director were part of the team. In the confusion, a central IT unit was asked to look at the computers, and it failed to maintain physical control over the items that resulted in the hard drives being lost. This embarrassing situation reinforced why it is critical to have a publicized policy that ensures everyone on staff understands our mission and commitment to preserving digital materials.

Number of Media

We were surprised by the extent of born-digital media found even in older collections. Our surprise is likely due to incorrect assumptions that we made about the point at which donors began to realize the value in retaining their own digital *and* paper content. We were aware of disks, hard drives, and whole computers that we had received as part of recent donations, and our capacity planning took those into account; however, the known capacity only accounted for 90GB. The other 6,410GB were a total surprise. We counted over 180 born-digital items.

Media Labels

Recording the information from labels proved to be a worrisome activity. We are lucky that the majority of media we inventoried had some form of description written on the media itself. It is worth noting that the older the media we found, the better the description on the label. Once we moved into more recent collections that included masses of 3.5" floppy disks, we found far less attention to providing descriptive information on labels. For example, the few computer reels from 1970 include detailed notation of the tape contents. The 8-inch IBM floppy disks found in a collection from the early 1980s also include short names for the contents and close attention to dates. Once we move into the 1990s, though, we encounter an abundance of 3.5" floppy disks, over 14% of which are blank. We could not help but think back to all of the media we ourselves have used over our lifetimes and note with embarrassment the number of times we had failed to label the media.

A small percentage of the 3.5" floppy disks found included software manufacturer labels (e.g., "HP Laserjet 6MP") that indicate these were disks used to install software. While this is a definite sign that these are candidates for weeding, we realize the need to verify first that the contents match the label as it was a relatively common practice at the time to reuse software diskettes for storage (in fact, we were

surprised that we did not encounter a single AOL installation diskette since they were so ubiquitous at the time).

Moving Forward

Now that we have jumped in to knowing what we have, our next steps include two separate broad tasks that will likely occur in tandem. Obviously we need to select and prioritize which collections we will work with next. To prioritize, we will consider the value of the collection, the age and condition of the digital containers, and the uniqueness of the information digitally stored. Once we have a workflow for digital preservation, we can then begin work on preserving these materials.

Developing that workflow is another critical and involved tasks going forward. We need to determine the technical framework for acquiring, processing, and preserving the items in perpetuity. We are already in the process of reviewing several overarching models to find the best fit for our work. Through our reviews, we are seeking a framework that supports several concerns:

- Anticipating technical needs and workload by having a thorough inventory and preliminary assessment process.
- Identifying a system that supports an OAIS-based workflow.
- Providing the appropriate interface for user access.
- Acquiring adequate storage.
- Developing processes that support dark archive copies, long-term preservation through migration, and disaster planning.

To reinforce our commitment going forward, we are currently in the design phase of a Digital Services Lab that will support digitization of analog materials and the capture of born-digital materials. The institution has responded positively, and we are confident funding will be available when a current renovation is complete.

Once those two broad tasks are under control, we will then be able to focus on two more important aspects of our operation. First, we are in the earliest stages of staff training, currently focusing on building up specialized skills for the two staff members central to the initiative. As the work progresses and we are able to address the technical framework and open the Digital Services Lab, we will be able to develop a training program that supports employees within our institution and our partner institution, as well as the archives community in our region. Second, the survey has also given us a taste of the range of born-digital materials we may receive via donations in the future. This experience will help us to craft an acquisition policy regarding born digital materials and to advise archivists on adequately describing born-digital items in inventories and finding aids. By seeing what is often donated to us in a digital format, we can nudge donors to focus the materials they contribute on what truly fits our collection policy.

In the end, the Jump-In initiative was definitely a win-win for us. It helped give shape to the work we had been doing and will allow us to make decisions informed by real data rather than presumption. It

opened our eyes, made us a little scared of the enormity of the task ahead, and gave us a chuckle or two as we encountered strange new formats of media that made us ask, What the heck is this?!