by
laura
capell

The Digital Archivist

Building the Foundation:

Creating an Electronic-Records Program at the University of Miami

AS WE CONTINUE TO MOVE THROUGH THE PILOT PROJECT AND DEVELOP

OUR POLICIES AND PROCEDURES FOR BORN-DIGITAL CONTENT,

WE'RE ALREADY LOOKING AHEAD AT THE NEXT STEPS.

Developing and implementing effective strategies to manage electronic records (e-records) is one of the biggest challenges facing the archives field today, as we acquire growing quantities of contemporary records generated by an increasingly digital society. However, jumping into e-records archiving can be a daunting task. Think about the diversity of born-digital content that makes its way into heritage collections, such as photos, documents, videos, websites, databases, Twitter feeds, and emails. Then, consider the range of hardware and storage media produced over the years and the vast universe of software used to render all of this digital content and make it usable.

Remember, too, that we don't live in a static world, and digital content, hardware, and software continue to evolve. Given all of that, it's easy to get so overwhelmed by the challenge of archiving e-records that it's tempting to just do nothing and stick with the comfortable familiarity of the analog world. But putting those hard drives or dreaded 5.25" floppy disks in a box in the back corner of the shelf and ignoring them is not a viable long-term solution.

Similar to many institutions, the University of Miami Libraries (UM Libraries) recognized that we needed to get started with e-records sooner rather than later. So we decided to jump right in—or, at least, get our feet wet. We started with the understanding that our solutions may not be perfect, but they are the best we can do given our current constraints—and they're better than doing nothing. Processes will continue to change over time, so we want to make sure that what we do today will be adaptable in the future. With that in mind, we're following the principles outlined by Ricky Erway in OCLC Research's report, "You've Got to Walk Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media"[1]:

- Do no harm
  (to the physical media or the content).
- Don't do anything that unnecessarily precludes future action and use.
- Don't let the first two principles be obstacles to action.
- Document what you do.

Getting Started

In 2011, UM Libraries formed its first working group to research best practices and technology for managing born-digital content and recommend initial steps for getting started. Today, the libraries' E-Records & Web Archiving Committee guides the development of policies and procedures to facilitate the acquisition, preservation, and access of a growing variety of born-digital content found in our archival collections. The committee offers a range of perspectives from across the library, with representatives from the special collections departments, as well as digital production, cataloging and metadata services, and learning and research services.

A key first step in managing born-digital content is to determine what you have. So we decided to join the Jump In, Too/Two e-records survey initiative sponsored by the Society of American Archivists' Manuscript Repository Section in 2014.[2] Our Cuban heritage collection and special collections departments participated in the survey, and we trained student workers to examine finding aids and dig through boxes to note the location and format of digital media found in the collections as well as any descriptive information about it. We now have a better understanding of the types of media formats we have and how many of each, which is helpful as we establish priorities, develop workflows, calculate storage needs, and determine how to best integrate the work into our existing activities.

▶

A Pilot Project

We selected a pilot project to test out workflows and to get a better sense of the tasks and time involved in processing e-records. We focused on the born-digital content in Sen. Mel Martinez's papers, which includes 300 DVDs and two external hard drives from his time in office from 2005 to 2009. The DVDs feature Sen. Martinez giving speeches or appearing in news clips. The external hard drives were more of a mixed bag, though. One contains more than 80GB of content with 140 different file extensions, while the other contains roughly 4 million text files of correspondence. Due to the quantity and variety of the content, we decided to break the project into the following phases to keep it manageable:

1. Create a disk image of each piece of media for preservation.

2. Process and provide access to the contents of the video DVDs.

3. Determine the best approach to process and provide access to the hard drives.

To create the disk images of the DVDs and external hard drives, we used FTK Imager, a free disk-imaging

Sen. Mel Martinez (right) and an example of a DVD from the Senator Mel Martinez Collection (below)

tool that's easy to use. We're storing the disk images for long-term preservation, and we also use them to export working copies of files for arrangement, description, and access, without risk of modifying the originals. The disk image format that we decided to use for the external hard drives is E01, which is a compressed sector-by-sector bitstream copy that captures everything on the original hard drive, including the file system structure, unallocated space, and hidden and deleted files. The disk images for the DVDs are in ISO format, which is commonly used for optical media. The imaging process took longer than expected—especially for the hard drives—and we also discovered that some DVDs can be quite finicky, requiring a fair amount of troubleshooting to get them to image properly.

Next, we processed the videos from the DVDs. We knew that we wanted to provide access to the videos using the university's Kaltura video management platform, so our processing workflow involved generating MP4 derivatives, creating an item-level inventory with a description of each video, and then ingesting the MP4s and descriptions into Kaltura. The videos are now available online, but due to copyright issues for much of the content, access is limited to computers within UM Libraries.

Screen shot of FTK Imager showing the format options available for creating forensic disk images
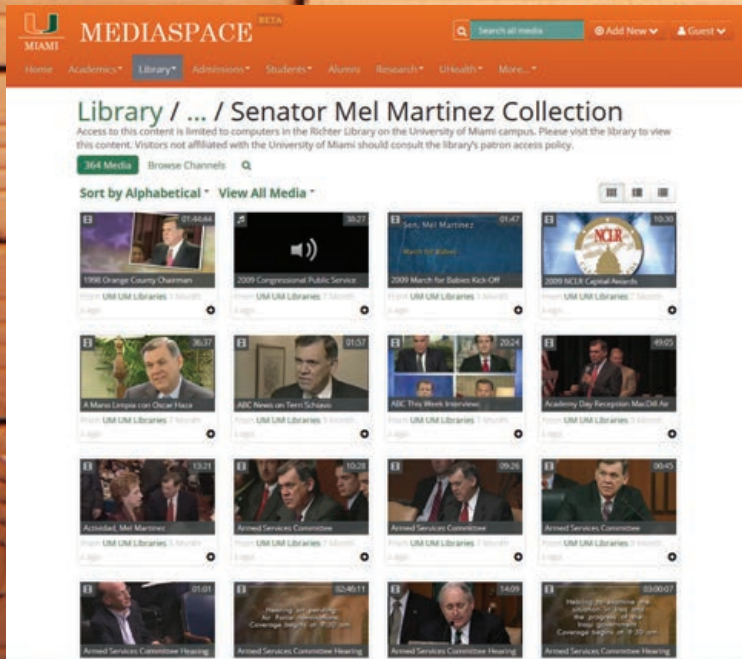
Senator Mel Martinez Collection in Kaltura, the university's video access platform

Creating the disk images and then going back to generate the access derivatives turned out to be time-intensive. Plus, we realized that we were primarily interested in preserving the videos themselves, not the structure of the DVDs that we captured in the disk images. For future projects involving videos on DVD, we plan to skip the disk imaging and just generate MP4 files directly from the DVDs. We'll then use the MP4s as the master files and as the source of any access derivatives. This will save time and still allow us to preserve and provide access to the video content.

The final step in the project is to process and provide access to the contents of the two external hard drives. We recently set up an e-records workstation that has FTK digital forensics software, and this will be our first project using it. We plan to analyze the files to get an overview of their contents and organization, as well as to look for personally identifiable information, such as Social Security numbers. We've yet to decide how to best provide access, but as a first step, we're considering making the con-

tents available via an offline use copy on a computer in the special collections reading room and making the photos available in our digital collections in CONTENTdm.

Working With Donors

Our E-Records & Web Archiving Committee is also developing policies and procedures for working with donors about the acquisition of electronic content. As more people want to donate a growing variety of digital content, we realized that we need to adapt our traditional donation procedures to clarify expectations, highlight potential privacy and rights issues, and learn about the nature of the digital content so that we can make plans to successfully transfer, preserve, process, and provide access to it. In order to improve communication with donors and ensure that we ask them the right questions regarding their e-records, we created a checklist so that we can get a better understanding of the creation, lifecycle, and nature of their digital files before we begin the transfer process. We also

drafted an e-records addendum for our deed of gift to clarify and document any rights, permissions, and privacy issues. Working closely with donors to better understand the lifecycle and context of their e-records will help us to be better stewards of them into the future.

Moving Forward and Looking Back

As we continue to move through the pilot project and develop our policies and procedures for born-digital content, we're looking ahead at the next steps. First of all, we want to build more robust digital forensics workflows, including exploring methods for more extensive analysis of our digital content and developing workflows to handle a wider range of media and formats. We've started on this by processing several new e-records acquisitions that have varied greatly in nature—such as a 1TB hard drive of videos, zipped folders of documents received via email, and even the Twitter feed about the U.S./Cuba policy change in December 2014. Second, we want to use the results of our survey to start processing legacy media in our collections. Finally, we want to explore more options for providing access so that we can effectively make a wide range of born-digital content available for research.

A KEY FIRST STEP IN MANAGING BORN-DIGITAL CONTENT IS TO DETERMINE WHAT YOU HAVE. ...

Our initial foray into e-records archiving has been largely successful, and we will continue to build on what we've learned as our program grows. However, the path has not always been smooth. We've encountered a variety of challenges along the way, and I am sure that new challenges will appear because the born-digital landscape is continually evolving.

▶

These are some of the broader challenges that we've encountered:

- Technology is a multifaceted challenge that includes dealing with various forms of obsolete technology that will continue to make its way into collections as well as staying up-to-date with new advances.

- E-records come in all shapes and sizes with varying rights issues, and there is no one-size-fits-all solution. So we are wrangling with the best ways to provide access to a wide variety of formats and content types.

- Working with e-records can be time-intensive, and it is one more element to balance in an active library. It's challenging to determine where it fits in the priorities, to decide who is responsible for which parts of the workflow, and to provide adequate training.

Despite the challenges, we've learned one valuable lesson: It's better to do something rather than nothing, even if that something isn't perfect. We're modeling our activities on the work done at other institutions, plus what we've learned from the professional literature and training opportunities. We're also documenting our activities and decisions, so that looking back, we'll know what actions we've taken. Archiving born-digital content isn't a static process, and as new tools and procedures are developed, we

DESPITE THE CHALLENGES, WE'VE LEARNED ONE VALUABLE LESSON: IT'S BETTER TO DO SOMETHING RATHER THAN NOTHING. ...

will adapt our practices accordingly. But for now, we're doing what we can, working step-by-step toward our long-term vision of a sustainable e-records archiving program. 💻

Laura Capell—The Digital Archivist's first invited guest author—is electronic records archivist and head of the digital production department at the University of Miami Libraries. She is responsible for managing digital projects for special collections materials and for coordinating the libraries' electronic records archiving initiatives. Contact her at l.capell@miami.edu.

Endnotes

1. Erway, Ricky. 2012. "You've Got to Walk Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media." Dublin, Ohio: OCLC Research. oclc.org/research/publications/library/2012/2012-06.pdf.

2. Information about the Jump In, Too/Two initiative sponsored by the Manuscript Repositories Section of the Society of American Archivists is available online at archivists.org/groups/manuscript-repositories-section/jump-in-tootwo.

Additional information about the e-records surveys is available in the May 2014 *Computers in Libraries*' Digital Archivist column, "Taking the Long View: Surveying Collections for Preservation and Digitization Priorities."