

# Jumping into Born Digital:

The Weinberg Library Tackles the Connors Collection CDs

---

*A **Jump In Two/Too** essay for the  
Society of American Archivists Manuscript Repositories Section*

by

**Kristen Yarmey, M.L.S., M.A.**

Associate Professor and Digital Services Librarian  
University of Scranton  
Weinberg Memorial Library  
Scranton, Pennsylvania

May 1, 2014

This essay describes the [University of Scranton Weinberg Memorial Library's](#) Terry and Paula Connors Collection and the actions taken so far on its CDs by staff and students from the Library's Digital Services Department: Preparation, Accession, and Inventory. Also outlined are our next steps (Imaging, Selection, and Description) and our plans for the future (Access).

### The Connors Collection CDs

The **Terry and Paula Connors Collection** consists of photographs taken on assignment by local professional photographer Terry Connors (working in partnership with his wife Paula) from the mid-1960s to the present. With clients like the Greater Scranton Chamber of Commerce, the Lackawanna Bar Association, and the Diocese of Scranton, Terry's photographs document the public face of Lackawanna County for more than four decades, serving as a nearly comprehensive photographic survey of significant people and events in our community.

The University of Scranton itself was one of Terry's major clients, making this collection highly relevant not only from a Special Collections local history perspective but also valuable for our University Archives. Terry regularly took public relations photographs for us, documenting the University's major activities and accomplishments but also capturing the "best foot" that we aimed to put forward - the story that we wanted to tell about ourselves to our prospective students, donors, and community members.



**A few sample photographs from Terry's 2005 University of Scranton assignments**

The Connors Collection is a hybrid - it's made up of both physical and digital materials. As a photographer, Terry switched from film to digital around September 2005, shooting in digital ever since. In 2008, the Connors donated negatives and slides from Terry's assignments dating 1966-2005 to our [McHugh Special Collections Library](#), and they expressed interest in donating digital photographs (dated 2005-2013, all stored on CDs) to us as well.

So we had an active, interested donor, a clear chain of custody/provenance, and the chance to complete a comprehensive and unique collection of high relevance to our University and local community. Appraisal-wise, seems like a no-brainer, right?

I hesitated for a while, though, in order to think through the challenges involved in effectively and sustainably accessioning and preserving this part of the collection. My Digital Services department only dates back to 2008, and up to this point we've primarily focused on digitized materials. We'd

dabbled in born digital, with documents like [theses](#) and [University publications](#) and a [fledgling web archiving program](#), but never at this scale, and never with digital photographs. We were also rapidly running out of space in our jerry-built digital storage setup and needed a little more time to transform it into the lovely, full-grown preservation repository that it is today (thanks, [DuraCloud!](#)).

I've known from the beginning, though, that I wouldn't be able to keep my head forever buried in the born digital sand. I know there are a bunch of other legacy CDs and DVDs slowly aging away in our Archives and Special Collections, and I know that we'll likely acquire more digital media (for example, CDs of images from the University's Public Relations office) in the not-so-distant future. And no matter where I hide, I haven't been able to escape the ticking clock and the growing urgency of transferring important digital files off of CDs that are now pushing nine years old. It was time to buckle down and do it – and thus the Terry and Paula Connors Collection CDs became my first real born digital project and a test bed for the future of born digital collections at the Weinberg Memorial Library.

### **Preparation and Planning**

One of our first steps was to meet with Terry for a preliminary assessment of the CD component of the Connors Collection. We're lucky to have a very interested, supportive, and accessible donor! Terry talked us through his work processes, brought us sample CDs, showed us his appointment books and worksheets, and explained some of the shorthand abbreviations he used to distinguish certain types of shoots or images. He even invited us to call him with any future questions we had. (He also told us some pretty fantastic stories – but that's an essay for another day.)

We also needed to update our donor agreement for the Connors Collection, which had originally been written only with negatives and photographic prints in mind. After conferring with Special Collections and the University's General Counsel, we worked out a revised version of the document for Terry and Paula to sign. The updates were mostly about setting shared expectations – explaining that we would transfer the digital images from the CDs into our repository and take reasonable and informed steps to preserve the images for the enduring future. Terry and Paula gave us permission to discard the CDs themselves in the future. The Connors also generously transferred any rights they held for the materials to us, which will make it a lot easier for us to preserve and provide access to the collection over time.

Next, we started getting ready. I did a lot of background research, referring to sample inventories from previous *Jump In* participants and reading articles about accessioning born digital items (especially, of course, the OCLC reports *You've Got to Walk Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media* and *Walk This Way: Detailed Steps for Transferring Born-Digital Content from Media You Can Read In-House*). I also found some helpful case studies involving CDs, like Stanford's [STOP AIDS project](#) reports.

Based on these recommendations, I started to prepare for accessioning. I set up a project documentation folder on the University's shared file server and drafted out a rough processing plan. I also created an [accession log](#) and [inventory spreadsheet](#) and worked out a naming schema consistent with our existing digital collections. Along the way, I consulted with my colleagues in Special Collections, since they're working on inventorying Terry's negatives – both to benefit from

their experience and to ensure that our work would be consistent across the physical and digital components of the collection.

### Accession and Inventory

Early in April, Terry brought us our first three boxes of CDs (dated 2005-2007) as well as several appointment books and worksheets documenting his assignments. The CDs were stored in plastic containers, roughly in chronological order.



**A peek inside the first three boxes of CDs**

Almost all of the CDs were stored in some kind of protective housing (mostly envelope, sleeve, or jewel case), and most had descriptive information recorded on the CD housing and/or the CD label. Many of the envelopes and sleeves held photo index cards in addition to (or, in a few cases, instead of) a CD. I came to think of these items as CD “packages” – that is, a CD, its housing, and any included photo index cards.



**The whole package – CD, envelope, and photo index cards**

While I had drafted out an inventory process before anything came in the door, we ended up making several tweaks along the way. There was a little bit of trial and error as we came across slightly different types of materials. (Even in the digital world, it turns out, sometimes you just have to get your hands dirty.) Here’s an overview of our process so far.

We started out at the container level:

1. We photographed and labeled each item we received. Each item got a unique identifier, following the collection naming schema:

**tpc-b###** = Box of CDs

**tpc-ws##** = Assignment worksheet (yellow legal pad)

**tpc-ab##** = Appointment book (red bound volume)

**tpc-cd####** = CD or DVD disc

2. Each item was then documented in our [accession log](#). We noted the date of delivery and added brief descriptive information (e.g., estimated number of CDs, approximate date range) for quick reference. We also logged the status and location of each item to help us keep track of things.

Then we did a more comprehensive inventory of the contents of each container:

3. Each CD package received a unique identifier and a label, following the naming schema (tpc-cd0001, tpc-cd0002...). Preprinting these labels saved a lot of time!
4. We took photographs of each CD and its housing (whether envelope, jewel case, or sleeve) using a digital camera and tripod. Once we had a workstation set up, this task went pretty quickly (about 30-60 seconds per CD).

It was unexpectedly difficult to get good shots — CDs are shiny! — and it took a while to figure out appropriate camera settings for capturing detailed labels. Our overhead lighting was also problematic for glare, so we'll try a new location for photographing the next boxes of CDs that come in.



Fabulous student worker April Francia '15 at our CD-photographing workstation (left)  
Digital Services Librarian Kristen Yarmey (yours truly) inventorying CDs (right)

5. Then I logged each CD package in a detailed [inventory spreadsheet](#).

There was an abundance of descriptive information on each CD package (usually dates, clients, and sometimes subject of the assignments), and sometimes there were minor

differences between information on the CD label versus the CD housing. For the time being, I decided to briefly summarize this information rather than trying to transcribe it with 100% accuracy. Since we captured photographs of each CD and its housing, we can add more detailed descriptive information in the future.

Terry had given us a heads up that there might be personal photos mixed in with professional photos on the CDs, and we'd agreed that personal family photos would be weeded out of the collection prior to ingest into our digital repository. Terry often marked the presence of family photos (with the abbreviation "Pers"), so as I inventoried the CDs, I recorded whether any (or, in some cases, all) of the images on a CD might be personal.

I've been debating quite a bit about how to handle the large number of photo index cards that accompanied the CDs. They provide a really useful quick glance at what might be on each CD, but they're voluminous, and their thumbnail images are small and difficult to interpret. I'm still thinking through the best way for us to take advantage of these cards, but for the moment, I recorded a count of photo cards, their date, and any other label information (like an order number) in the inventory for quick reference.

## **Results**

One of the unexpected (but happy) surprises for this collection was the amount of physical material that came with it. This ends up working well for us, since it's a hybrid collection – these items fit nicely with the negatives and other physical materials our colleagues in Special Collections have previously accessioned. So far we've received:

<b>Material</b>	<b>Count</b>	<b>Contents/Notes</b>
Plastic containers	3	Held a total of 364 CDs and 1 DVD, mostly stored in paper envelopes, sleeves, and a few jewel cases. Many CD envelopes also included photo index cards (a total of 1,195).
Appointment books (bound)	4	Continue a series of books previously donated to Special Collections. Each volume covers a calendar year.
Assignment worksheets	4	

Turning our attention to the digital media:

Media Type	Manufacturer/Label	Estimated Storage Capacity per disc	Count	Maximum Storage Requirement (GB)
CD	Kodak Picture CD	700 MB	289	197.56
CD	Maxell CD-R	700 MB	2	1.37
CD	Memorex CD-R	700 MB	15	10.25
CD	Walgreens Digital Photo Album	700 MB	3	2.05
CD	WalMart FujiFilm Digital Album	700 MB	54	36.91
CD	Unknown	700 MB	1	0.68
DVD	Maxell DVD+R	4.7 GB	1	4.70
		<b>Total</b>	<b>365</b>	<b>253.53</b>

One of the pleasant surprises in this collection is that the CDs received so far seem to be in pretty good physical condition, despite their age (mostly 2005-2007). Most of the discs had a few light scratches, but in general they had been well protected. I think this bodes well for our success rate during imaging.



The more the merrier? A selection of CDs, including several flavors of Kodak Picture CD

The presence of commercial CDs (Kodak Picture CD, Walgreens, and WalMart) was also a little unexpected. Among the Kodak Picture CDs especially, there seemed to be several variations. This is the first time I'll be working with these proprietary formats, so I'm doing some research to try to understand precisely what information each CD stores and how that information is structured.

## Next Steps: Imaging

Since this is a pilot project, I wanted to take the opportunity to set up a quarantined workstation where we can accession and process born digital files safely and effectively. Justin Goreschak '15, the second of my two extraordinary student workers, has been taking the lead on this task. Justin wiped one of our retired desktop computers (known affectionately as the "Old Beast"), set it up to run Ubuntu, and installed [BitCurator](#). Right now, Justin and I are experimenting with BitCurator's forensics tools and hashing out a step-by-step procedure for imaging the Connors CDs.



**Awesome student worker Justin Goreschak '15 setting up our BitCurator workstation**

At this point, our plan is to fully image all of the CDs as soon as possible, prioritizing the oldest first, to make sure that we have the best chance possible at successfully transferring the bitstreams. (I'm considering temporarily skipping over CDs that appear to be entirely personal photos and returning to those as a lower priority.) As we go, we'll use the [CD inventory spreadsheet](#) to document imaging status, calculate rate of failure, and estimate the time needed to image the full CD collection.

I'm hoping to use BitCurator's forensic tools to better understand how each CD variety is formatted and what information it contains. I especially want to do some additional research and testing on the proprietary Kodak, Walgreens, and WalMart CDs before processing the CD images. Thanks to good advice from the [Digital Curation Google Group](#), I plan to use [IsoBuster](#) to review the files present and will try to run the software to understand if there's metadata or other information available.

## Next Steps: Selection and Ingest

Once each CD has been imaged, we'll move on to selection and ingesting the selected materials into our digital repository for long-term preservation. Already, a few challenges are on the horizon, especially for those pesky proprietary CDs.

Ideally, I'd like to preserve the original image files from Terry's camera, any embedded metadata, and any edits or additional metadata that he might have created when the CDs were generated. If possible, I'd like to weed out any ancillary software or automatically generated thumbnail derivatives that may also be on each CD. More on this after imaging and testing!

Personal family photos also present a selection challenge. We do want to weed these out, both to protect the privacy of the Connors family and to focus our efforts on preserving materials of archival value to the community. Many of the CDs and their housing note the presence of personal photos, but we may have to work CD by CD to review the contents.

I also expect that we'll come across many instances of duplication or near duplication. Some variants may be valuable - for example, Terry said he sometimes did light touchups for color correction, lighting, and cropping on selected images, so the collection includes edited as well as unedited versions of some images. These we're likely to keep.

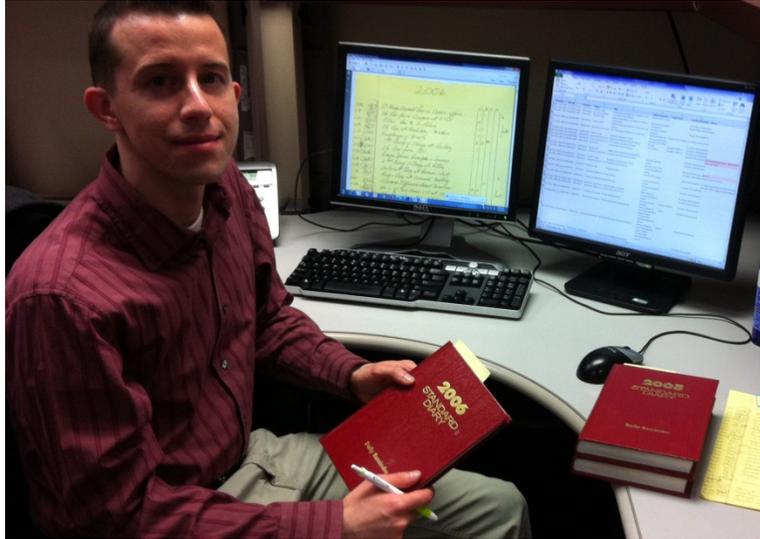
Terry also said that on many of his assignments (especially group photos), he'd take several very similar shots, sometimes adjusting settings on his digital camera, to ensure a quality result for the client. We'll likely keep these as well, though they will add to the bulk of the collection. I'm considering weeding out blurred or overexposed images from a shoot, however, especially if there are several good quality images present from the same assignment.

Once we've selected material for long term preservation, we'll transfer the files into our onsite digital repository. Thanks to our shiny new [DuraCloud instance](#), Terry's photographs will also be preserved offsite, in Amazon's S3 and Glacier storage. DuraCloud will also help us monitor the fixity of the image files over time.

### **Next Steps: Description**

Due to the significance of this comprehensive collection, it's important to me that users (especially University and local community members) be able to discover, browse, search, and engage with Terry's photographs – so we'll need really good metadata. Luckily, we have heaps of descriptive information to work with, much of it thanks to Terry and Paula's diligent record-keeping over the years.

Terry's appointment books and assignment worksheets generally document the time, date, and client for each assignment, as well as the subject or event being covered. Digital Services Assistant David Hunisch has gotten us started on description by tackling this material. Using digitized copies of the worksheets for quick reference, David's transforming the sheets and books into a searchable spreadsheet, attaching controlled vocabulary terms to establish relationships with materials in our other digital collections. We'll then be able to use Terry's notes on the CD labels and housing, as well as the 1000+ photo index cards, to help us match up CDs with specific assignments.



**Digital Services Assistant David Hunisch, working with appointment books and assignment worksheets**

I also anticipate finding a wealth of information stored as embedded metadata in the image files themselves – not only the date the photographs were taken, but also information about the camera and settings Terry used throughout the shoot. As we image and then process CDs of each format, I'll be testing several methods of extracting [Exif](#) data and preserving its relationship to the other content on the disc.

The big challenge for description will be recording and reconciling all of this information efficiently and effectively, and doing so in a way that's consistent with how Special Collections is describing the physical portion of the collection. Currently, Special Collections staff and students are inventorying the Connors negatives, arranging them into series based on client, and they've also started a database to record information about the assignments represented. As we move deeper into descriptive work, I'll be verifying that our fields and content standards match up. Ideally, we'll be able to merge our streams of information together into a single finding aid, describing negatives, prints, digitized images, and born digital images in a unified, user-friendly way.

### **Future Plans: Access**

Further down the road is the challenge of providing user access to the images. Again, having fantastic donors is a real plus for this project. Terry and Paula were willing to transfer their share of copyright in any images to us, allowing us to take any actions needed to preserve the images and hopefully publish them. That said, we believe that many of the images may fall under "work for hire" copyright restrictions, so I'll be thinking through the implications of this.

One of our priorities for access, of course, will be University photographs, and luckily copyright is in our corner for those assignments. At this point, I envision publishing selected images from University assignments via our [CONTENTdm instance](#), with references to additional images from that assignment that will be available to users upon request.

How to select which images go live? We might just have a secret weapon for this task – a select file! After a shoot, Terry would save all of the images from his camera onto his computer, backing them upon to his CDs. Normally, he then sent only the best images (a “select file” of the finished product) to his clients. In this case, the client is our own University Public Relations Office, and conveniently, they’ve got several binders and CDs of their own with “Terry Connors” labels on them. So I think (and hope) that we’ll have our selected images pre-selected via PR, saving us time and also helping us avoid publishing sensitive or private photographs of University people and events.

### **Reflection: On Jumping In**

If there’s anything that I’ve learned in six years of working with digital collections, it’s to expect the unexpected. No matter how much you prepare, every project brings unanticipated challenges – but also new discoveries. When my Digital Services department decided to tackle the Connors Collection CDs as our first major born digital accession, I knew I’d have a lot to learn, but I had no idea how exciting and enjoyable that learning process would be. Who knew born digital media could be so much fun? In the words of the Dorothy Sayers heroine Harriet Vane, “I’ve broken the ice... and the water wasn’t so cold after all.”

Of course, if there’s anything *else* I’ve learned in the past six years, it’s that when estimating the time needed for a project, one should carefully calculate an end date based upon all the relevant factors – and then double or (more likely) triple the result. As we started work on the first three boxes of Connors CDs, it became clear that we have a lot of important decisions to make in the near future, and we’ll need to do more research and experimentation to inform those decisions. I’m grateful, though, to the Manuscript Repositories Section and the Jump In project for pushing me out of my cozy nest and encouraging me to take a stab at flying.

Like most of our digital collections projects, this is a work in progress. We’d welcome any questions, comments, or suggestions sent via email to [digitalcollections@scranton.edu](mailto:digitalcollections@scranton.edu).