University of California Curation Center (UC3)

UCLA Digital Library Program & UC3 Collaborate to Reaffirm a Preservation Solution for the Strachwitz Frontera Collection

Posted in [Digital Preservation](#), [Merritt](#), and [UC Libraries](#)

UCLA has long collaborated with the [Arhoolie Foundation](#) to provide digital access to and enable the preservation of the world's largest known musical collection of historical Mexican and Mexican American commercial recordings: [The Frontera Collection](#). Now, in a renewed effort and in collaboration with UC3, UCLA's Digital Library Program (DLP) has begun to enhance and expand the actual number of recordings that are being preserved in CDL's Merritt digital preservation repository.

Owned by the Arhoolie Foundation, the physical collection began in the early 1970s when [Chris Strachwitz](#), already a decade into his lifelong career as a record producer and collector of regional music, was introduced to [corridos](#), [Tejano](#), [Norteño](#) and other regional styles of Mexican and Mexican American music. This introduction, and Strachwitz's devotion to song catching, is the origin of the Frontera Collection. Over time, the collection has grown to nearly 170,000 recordings, has started to include Latin American music, and, in its entirety, spans the better part of the 20th century, from 1905 to the 1990s.

Over twenty years ago, the UCLA Library, the UCLA Chicano Studies Research Center, and the Arhoolie Foundation formed a partnership to digitize and make available online this unparalleled collection. Their collaboration has been funded through several grants provided by the National Endowment for the Humanities, as well as funding grants from the National Endowment for the Arts, the GRAMMY Foundation, the Fund for Folk Culture, Arhoolie Records, Mr. and Mrs. E. W. Littlefield Jr., the Edmund & Jeannik Littlefield Foundation, and others.
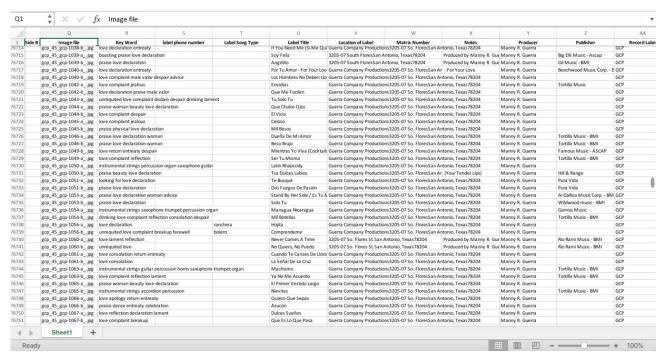
Recently, Elizabeth McAulay, head of the UCLA DLP, approached UC3 with a request to expand the existing collection in Merritt, which at the time housed more than twenty thousand Frontera digital

objects. However, since the initial deposit in 2006, the DLP has received thousands more song files from Arhoolie. Beyond providing access copies of these newer recordings on the DLP's Frontera [website](#) , DLP and Arhoolie are committed to ensuring their long term preservation.

Metadata & gap analysis

Working with Geno Sanchez, Digital Assets Coordinator at the DLP, a process was begun to obtain the latest metadata for all of the Frontera Collection's recordings, locate tens of thousands of master song files and their associated record album image files, conduct a gap analysis between the existing Merritt-based collection and metadata, and then finally assemble groups of files that would each comprise a digital object for ingest into the repository.

While the Arhoolie Foundation has delivered new song files and album artwork to the DLP over the years, it has also sent along spreadsheets of metadata to accompany every batch files. At the onset of our project, we reached out to the Foundation to request a spreadsheet complete with metadata from all of these batches, consisting of every master recording file name in Frontera. In late April, we received a spreadsheet with no fewer than 153,000 entries, each with fields that note information about the album artist, composer, catalog number, record label, producer, recording date, format, and recording file name among other bits. There are 42 fields for every entry, leaving out no detail of import. Even the equipment used to transfer each recording from vinyl or magnetic media to digital is listed.

Q1	fx	Image file									
	Side B	Image file	Key Word	label phone number	Label Song Type	Label Title	Location of Label	Matrix Number	Notes	Producer	Publisher	Record Label
76714		gcp_45_gcp-1038-b_.jpg	love declaration entreaty			If You Need Me (Si Me Qui	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra		GCP
76715		gcp_45_gcp-1039-a_.jpg	boasting praise love declaration			Soy Feliz		3205-07 South FloresSan Antonio, Texas78204	Produced by Manny R. Gue	Manny R. Guerra	Big Elk Music - Ascap	GCP
76716		gcp_45_gcp-1039-b_.jpg	praise love declaration			Angelito		3205-07 South FloresSan Antonio, Texas78204	Produced by Manny R. Gue	Manny R. Guerra	Gil Music - BMI	GCP
76717		gcp_45_gcp-1040-a_.jpg	love declaration entreaty			Por Tu Amor - For Your Lov	Guerra Company Productions	3205-07 So. FloresSan Ar	- For Your Love	Manny R. Guerra	Beechwood Music Corp. - E	GCP
76718		gcp_45_gcp-1040-b_.jpg	love complaint male valor despair advice			Los Hombres No Deben Llo	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra		GCP
76719		gcp_45_gcp-1042-a_.jpg	love complaint jealous			Envidias	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra	Tortilla Music	GCP
76720		gcp_45_gcp-1042-b_.jpg	love declaration praise male valor			Que Me Fucilen	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra		GCP
76721		gcp_45_gcp-1043-a_.jpg	unrequited love complaint disdain despair drinking lament			Tu Solo Tu	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra		GCP
76722		gcp_45_gcp-1044-a_.jpg	praise woman beauty love declaration			Que Chulos Ojos	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra		GCP
76723		gcp_45_gcp-1044-b_.jpg	love complaint despair			El Vicio	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra		GCP
76724		gcp_45_gcp-1045-a_.jpg	love complaint jealous			Celoso	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra		GCP
76725		gcp_45_gcp-1045-b_.jpg	praise physical love declaration			Mil Besos	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra		GCP
76726		gcp_45_gcp-1046-a_.jpg	praise love declaration woman			Dueña De Mi Amor	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra	Tortilla Music - BMI	GCP
76727		gcp_45_gcp-1046-b_.jpg	praise love declaration woman			Beso Brujo	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra	Tortilla Music - BMI	GCP
76728		gcp_45_gcp-1049-a_.jpg	love return entreaty despair			Mientras Yo Viva (Cocktail	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra	Famous Music - ASCAP	GCP
76729		gcp_45_gcp-1049-a_.jpg	love complaint reflection			Ser Tu Misma	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra	Tortilla Music - BMI	GCP
76730		gcp_45_gcp-1050-a_.jpg	instrumental strings percussion organ saxophone guitar			Latin Rhapsody	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra		GCP
76731		gcp_45_gcp-1050-b_.jpg	praise beauty love declaration			Tus Dulces Labios	Guerra Company Productions	3205-07 So. FloresSan Ar	(Your Tender Lips)	Manny R. Guerra	Hill & Range	GCP
76732		gcp_45_gcp-1051-a_.jpg	looking for love declaration			Te Busqué	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra	Pura Vida	GCP
76733		gcp_45_gcp-1051-b_.jpg	praise love declaration			Dos Fuegos De Pasión	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra	Pura Vida	GCP
76734		gcp_45_gcp-1053-a_.jpg	praise love declaration woman advice			Stand By Her Side / Es Tu N	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra	Al Gallico Music Corp. - BM	GCP
76735		gcp_45_gcp-1053-b_.jpg	praise love declaration			Solo Tu	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra	Wildwood music - BMI	GCP
76736		gcp_45_gcp-1054-a_.jpg	instrumental strings saxophone trumpet percussion organ			Managua Nicaragua	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra	Gamse Music	GCP
76737		gcp_45_gcp-1054-b_.jpg	drinking love complaint reflection consolation despair			Mil Botellas	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra	Tortilla Music - BMI	GCP
76738		gcp_45_gcp-1056-a_.jpg	love declaration		ranchera	Hojita	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra		GCP
76739		gcp_45_gcp-1056-b_.jpg	unrequited love complaint breakup farewell		bolero	Comprendeme	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra		GCP
76740		gcp_45_gcp-1060-a_.jpg	love lament reflection			Never Comes A Time		3205-07 So. Flores St.San Antonio, Texas78204	Produced by Manny R. Gue	Manny R. Guerra	Rio Rami Music - BMI	GCP
76741		gcp_45_gcp-1060-b_.jpg	unrequited love			No Quiero, No Puedo		3205-07 So. Flores St.San Antonio, Texas78204	Produced by Manny R. Gue	Manny R. Guerra	Rio Rami Music - BMI	GCP
76742		gcp_45_gcp-1061-a_.jpg	love consolation return entreaty			Cuando Te Canses De Llora	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra		GCP
76743		gcp_45_gcp-1061-b_.jpg	love consolation			La Señal De La Cruz	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra		GCP
76744		gcp_45_gcp-1063-a_.jpg	instrumental strings guitar percussion horns saxophone trumpet organ			Machismo	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra	Tortilla Music - BMI	GCP
76745		gcp_45_gcp-1063-b_.jpg	love complaint reflection lament			Ya No Me Acuerdo	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra	Tortilla Music - BMI	GCP
76746		gcp_45_gcp-1065-a_.jpg	praise woman beauty love declaration			El Primer Vestido Largo	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra		GCP
76747		gcp_45_gcp-1065-b_.jpg	instrumental strings accordion percussion			Nievitas	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra	Tortilla Music - BMI	GCP
76748		gcp_45_gcp-1066-a_.jpg	love apology return entreaty			Quiero Que Sepas	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra		GCP
76749		gcp_45_gcp-1066-b_.jpg	praise dance entreaty celebration			Azucón	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra		GCP
76750		gcp_45_gcp-1067-a_.jpg	love reflection declaration lament			Dulces Sueños	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra		GCP
76751		gcp_45_gcp-1067-b_.jpg	love complaint breakup			Que Es Lo Que Pasa	Guerra Company Productions	3205-07 So. FloresSan Antonio, Texas78204		Manny R. Guerra		GCP

Sheet1

Ready — 100%

Upon initial examination, it became clear that the recording file name would serve as the key field against which we would compare the content of digital objects in the existing Merritt collection. However, it also became evident there were many orphaned entries on both sides of the equation.

Through use of our [file analysis tool](#) , we were able to cross reference fields in the spreadsheet against those in a spreadsheet of data exported from Merritt's inventory database. Using the

recording file name as a unique way to associate each existing digital object in Merritt with an entry on the Arhoolie spreadsheet, we discovered that slightly less than a thousand objects (out of 27,000) were lacking corresponding entries. And of those, 578 did not include recording files – only album artwork. Attempts to manually locate recording files associated with the album artwork by searching on-premises storage at the DLP did not succeed, and it was ultimately determined that we would start with a fresh collection in Merritt to house all of the known song files and their art, rather than attempt to augment existing objects.

With a path forward based on the gap analysis, we dove into examining the metadata for each song in the Arhoolie spreadsheet and determined a mapping into a MODS metadata file to accompany each object. Since the metadata in the spreadsheet was so robust and meaningful, we decided to carry over detailed information about the album as well as include artist names, catalog number (i.e., matrix number), track duration, publisher, and of course donor.

Stepping back from the metadata, a complete digital object for a single song in the collection appears as follows in the repository. Note the front and back album artwork files, as well as a TIFF for the label placed on the vinyl record.

```
─
|  └── bv_33_318_a4 (digital object folder name and Local ID in Merritt)
   ├─ bv_33_318_a.tif
   ├── bv_33_318_a4.wav

   ├── bv_33_318_a4.xml
   ├── bv_33_318_back.tif
   └── bv_33_318_front.tif
```

Not all objects contain a superset of files like this one. At a minimum, only the .wav and .xml files are included.

Assembling what would become the submission information packages (SIPs) for each song took a tremendous effort to automate the location, movement and gathering of all required object components. At the same time, we decided to divide the full collection into batches to allow for easier transmission and management. The Arhoolie song list spreadsheet was divided into batches of circa 24,000 songs and we then transformed the spreadsheet for the initial batch into MODS XML files.

Submitting the first batch

With the first batch of objects assembled, we pivoted our focus toward enabling their submission to the repository. Submitting batches of content to Merritt is best done through use of manifest files. Manifests are csv-like files that record the web-accessible location of each SIP to be ingested, along with object-level metadata and an md5 checksum for verification purposes. Once each group of files was organized into a directory on local storage (one group per song), these were zipped up and an associated checksum generated. The URL at which the package could be retrieved was then recorded, along with the aforementioned metadata in a single line entry in the manifest.

With a very small subset of about a dozen or so SIPs, we tested submission into the staging environment for Merritt. This testing revealed we needed to make minor manifest encoding corrections to ensure support for diacritics, and that otherwise our process was successful. Shortly thereafter, a new collection in Merritt was established to house the initial batch of 24,000 song objects. We proceeded to split the primary manifest into four, six-thousand-object manifests for submission in turn – and within four days time, all objects were ingested into the repository.

Looking forward

As the entire Frontera collection contains close to 170,000 sound recordings, many more batches of content are yet to be assembled. At this time, the second batch is nearly complete and we hope to have it ingested before the end of 2021. Slowly but surely, Frontera will make its way to a safe new home, its master recordings ready to be retrieved and shared with generations to come.

As an extended team, UC3 and the Digital Library Program at UCLA have already begun collaborating on more projects like this one. We'll share our experiences as we move forward and look to hear from you with your thoughts and questions. But in the meantime, you may wish to do some song catching of your own while exploring Frontera – an extraordinarily rich collection of music that encompasses over 90 years of cultural heritage.