Mollie Echeverría

Prof. Monica Maceli

LIS-697– Database Design and Development

16 October 2016

Final Design: Canal St. Studios Audio Archives Database

Proposal:

For my LIS-697 final project, I would like to design a database related to my internship in the archives of Canal St. Communications Studios, the studio practice of artist and musician Laurie Anderson. Canal St. Studios is currently in the midst of a number of archival preservation projects, one of which is the stabilization and digitization of hundreds of audiotapes of Anderson's music. This work is currently being done off-site by a professional audio engineering studio, and has been ongoing since the summer. My project will be focused on creating a database that Canal St. can use to store data about their audio materials and about the progress of the digitization process.

At this point, Canal St.'s technical director has been keeping track of the audio digitization project using an Excel spreadsheet. Canal St. has been sending tapes to the engineering studio in a few large batches, after which they are sent back to Canal St. along with digital audio files. There are at least 760 tapes that will ultimately be digitized; around 550 have been digitized so far. The audio digitization spreadsheet includes a unique catalog number for each tape, along which metadata about the tape, like song and album titles, date of recording and format, and data about the digitization process, such as digitization dates and status, and free text fields like notes. Overall, the spreadsheet has 27 different fields, each of which are still being updated. Canal St. uses the spreadsheet to keep track of the physical locations of the tapes, the

digitization status of each tape, and which hard drives the digitized audio files will be kept on. A large volume of Anderson's tapes may be donated to a museum in the coming years, so the spreadsheet data will be what her studio will use to document the tapes and their contents when they are donated.

Having data about this audio digitization project stored in a database versus in a spreadsheet could help Canal St. Studios maintain more data integrity during the course of this project. Due to how many fields are currently stored in this Excel spreadsheet, maintaining consistency across all the different columns in the sheet may be difficult. The spreadsheet also stores datasets about a number of different entities, including the physical tapes themselves, the musical projects stored on the tapes, the musicians that played on them, concert venues where they were recorded, and the engineers now digitizing them. A relational format would make a lot more sense for storing the kind of many-to-many relationships that exist in the project data. A database would also make it much easier to query across these varied datasets.

In addition to allowing different datasets to be brought together in a more accessible format, a MySQL database would be easier to share with collaborators, and would allow for more secure storage of the audio metadata once the project is over. Excel files can easily get changed or corrupted if shared, and it may be hard to keep track of a single master version. Excel spreadsheets are also not suited for large volumes of data, so the spreadsheet format may become increasingly unwieldy as the project progresses. MySQL databases are designed to be accessed by multiple users, and they allow for much greater processing power when dealing with a large volume of data.

Finally, not only would a MySQL database be better equipped to able handle the studio's current data, a database would also allow for easier expansion in the future. Canal St. is also

digitizing materials like photographs and video, and is also taking stock of physical artworks currently kept in storage. Given the wide variety of mediums in which Anderson has worked, including music, video, performance art, and digital art, her studio may wish to one day keep metadata about all these areas of her practice in a single source. To expand the audio database to include information about digitized images or video, for instance, it would be easy to update the database with new tables, allowing all of the studio's archival data to be stored in a central database rather then spread across different spreadsheets. An underlying MySQL structure could also be used in a database management system like FileMaker Pro should the studio decide to invest in this kind of software in the future.

Business Rules:

Canal St. Communications Studios currently houses hundreds of archival audiotape recordings. In preparation for potentially donating these tapes to a museum in the future, all of the studio's audiotapes have been sent to an engineering studio for stabilization and digitization, after which they will be sent back to Canal St. along with digital audio files for each tape. Canal St. currently uses an Excel spread to keep track of the progress of this process, as well as to record metadata about each tape.

Once tapes are sent to the engineers, they are "baked" to prevent sticky-shed syndrome before being digitized. Canal St. keeps track of the processing status of each tape. This status is either left blank if a tape has not been processed yet, or marked 'Baking' (for tapes that are currently being baked), 'Baked' (for tapes that have been baked but not yet digitally transferred), or 'Transferred' (for tapes that have been baked and transferred to a digital format). Once a tape has been completely transferred, Canal St. notes the final date of transfer, including the day, month, and year, in the spreadsheet.

The engineering studio employs multiple transfer engineers. Each transfer engineer processes one to many different tapes. Each tape is processed by one to many transfer engineers. Canal St. records which audio engineers worked on each tape. This field is left blank if the tape has not yet been processed.

Canal St.'s tape spreadsheet contains extensive metadata about all of the tapes in the studio's archive. There are currently 766 different tapes listed in the spreadsheet. Each tape has a unique catalog number, as well as a QR code corresponding to the catalog number that is also affixed to the physical tape. Each tape contains audio from a different musical or artistic project of Anderson's. This may be an album, a live concert recording, audio from a performance art piece, recordings of a rehearsal or audition, or sound effects or samples used as part of a larger project. Any of these kinds of audio may have been recorded as a standalone audio track or as an audio track for a video. If the project contained on the tape is unknown or unclear, the tape's project is simply noted as its date of recording, or it is marked 'TBD', 'Unknown', or the project field is left blank. Each project is recorded on one to many different tapes.

Each tape also has a 'Song Titles' field, which records the individual tracks on the tape. These may be different songs, or, in the case of audio not from an album or musical performance, descriptions of the sounds or spoken word content on the track. Each tape has one to many different tracks. If the tracks are unknown, the field is marked 'None Listed' or left blank.

The recording date of each tape may be noted, including the day, month, and year. If the day or month is unknown, '1' may be used to represent the day and month. If the day, month, and year of recording are all unknown, the recording date is either noted as '1/1/1901' or left blank. There is only one date noted per tape.

The media type and tape speed of each tape is recorded. Tapes may be one of a number of different formats, including ¼ inch 2 track, ½ 2 track, and DAT. A tape may have one or more different speeds; it is marked 'multiple speeds' if the later. If the media type and tape speed is unknown, the field is left blank. The total running time of each tape is recorded when it is digitized, including hours, minutes, and seconds. If the total running time is unknown, this field is left blank. When a tape is digitized, its sample rate and bit depth are recorded. If a tape has not been digitized yet, the 'Sample Rate' field is left blank.

A tape may be either a master recording or a safety copy. A tape may or may not have been recorded in a recording studio. For tapes recorded in a recording studio, the name of the recording studio is noted in the spreadsheet. If a tape was not recorded in a recording studio, 'No' may be entered in the 'Recording Studio' field. If the place where the tape was recorded is unknown, the field will be marked 'Unknown' or left blank. Similarly, if a tape was recorded at a concert venue, the name of the venue will be noted, or 'No' or 'Unknown' will be entered in the 'Concert Venue' field.

A field called 'raison d'etre – Audio for Video' is used to record the reason why a tape was recorded. Each tape was recorded for one specific purpose. If the tape is the audio track for a video recording, 'Audio for Video' is noted in this field. If a tape is not connected to a video, then its purpose ('studio work tape', 'rehearsal tape', 'music for film', etc.) is noted in this field. If the purpose a tape was created for is unknown, the field is left blank or marked 'TBD'.

If the names of people besides Laurie Anderson who contributed to an audio recording are known, their names are recorded in the spreadsheet. If known, the sheet records the producer(s) and engineer(s) of each tape, as well as other musicians who played on the tape, and any non-musician collaborators, such as interviewers and orchestral conductors. Zero to many

people may be noted as have served any of these roles on a given tape. If the specific instrument or collaborative role of the person is known, this may or may not be recorded as well. These fields may include a person's first and last name, a group or company, or just a person's first name or initials. If the contributor(s) in a given role are unknown, the field is marked 'N/A' or left blank.

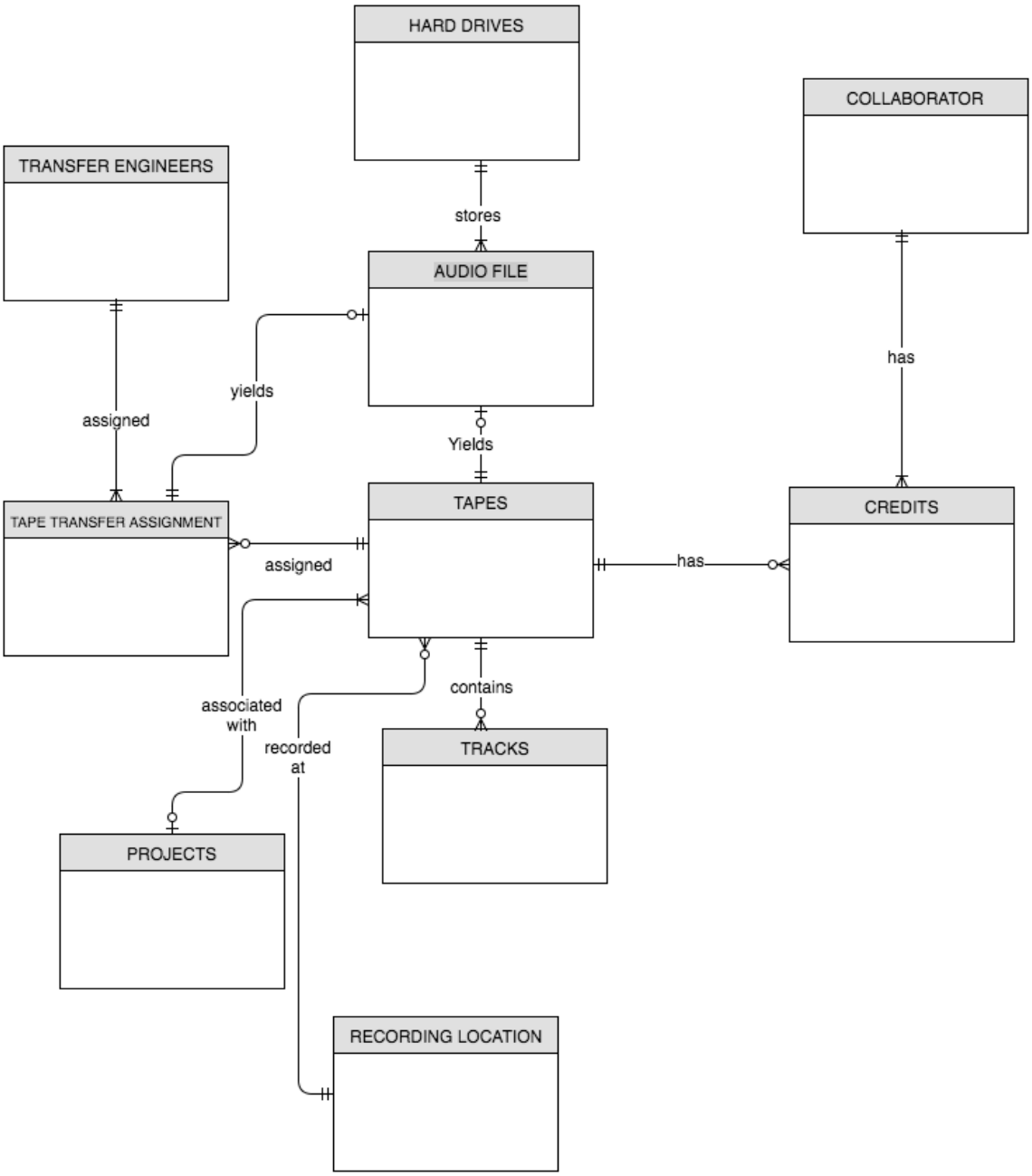
Each tape may or may not be assigned a quality rating between one and ten based on its audio quality. This field may be left blank if a tape has not yet been processed. Each tape may or may not also have a free-text quality note and/or a free-text archivist's note. The quality note describes the sound quality of the tape in more detail. The archivist's note may include commentary about physical aspects of the tape and/or about the audio it contains.

Once a tape has been digitally transferred and returned from the engineering studio, Canal St. scans an image of the container of each tape. These scanned images are stored in the same digital folder as the audio files for each tape. If a tape's box has been scanned, it is marked as 'In folder' in a field called 'TapeBox Scans'. If a tape has been transferred but its box has not been scanned, it is marked 'Pending'. If a tape has not yet been transferred, this field is left blank. If it still in the tape container, the original track sheet for the tape is also scanned. The tape's track sheet may also be noted as 'In folder' if scanned or 'Pending' if awaiting scanning. If the cassette box of a tape does not contain a track sheet, this field is marked 'N/A'. Either field may be left blank if the tape has not yet been transferred.

After a tape is returned to Canal St., the spreadsheet records that it has been successfully returned. The tapes were sent to the engineers in a number of batches, so they have not yet all been processed. If a tape is not yet back at Canal St., this field is left blank.

Once Canal St receives the transferred tapes back from the engineers, the tape's audio files, along with scans of the physical tape and its track sheet (if available) are stored on one of a number of hard drives. Each tape is only stored on one hard drive. Each drive contains many tape audio files. The hard drive where the digital audio of each tape is stored is noted in the spreadsheet.

Conceptual Data Model:



Entities and Attributes:

Entity	Attributes
Tapes	Catalog # QR Code Recording Date Media Type Tape Speed Master or Safety Tape Location Archivist Notes Raison D'être Transfer Status Quality Rating Quality Notes Total Runtime
Audio File	Date of Transfer Audio File Sample Rate Audio File Bit Depth Tape Box Scan Status Track Sheet Scan Status
Transfer Engineer	Name Transfer Engineer ID
Tape Processing Assignment	Assignment ID
Hard Drive	Drive ID
Project	Project ID Project Name
Track	Track ID Track Name
Recording Location	Recording Location ID Recording Location Name Recording Location Type
Credit	Credit ID Credit Type Credit Notes
Collaborator	Collaborator ID Collaborator Display Name Collaborator First Name Collaborator Last Name

Relationships:

- Each tape has been converted to zero to one audio files
- Each audio file is generated from one tape

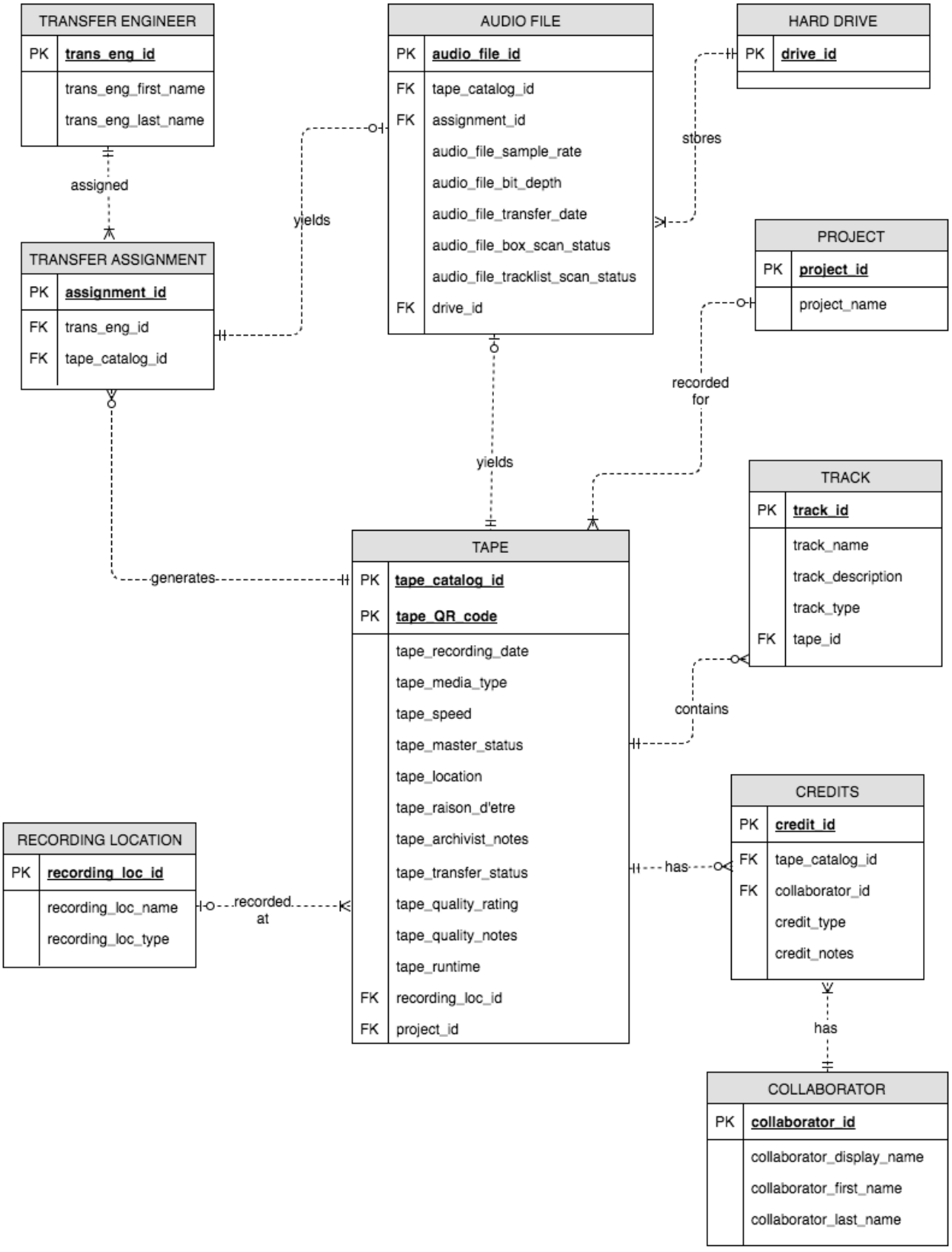
- Each tape may have generated zero to many transfer assignments
- Each transfer assignment is generated by one tape
- Each transfer assignment has produced zero to one audio file
- Each audio file is to result of one transfer assignment
- Each transfer engineer works on one to many transfer assignments
- Each transfer assignment is designated to one engineer
- Each audio file is stored on one hard drive
- Each hard drive contains one to many audio files
- Each tape contains audio from one project
- Each project is recorded on one to many tapes
- Each tape contains one to many tracks
- Each track is recorded on one tape
- Each tape has zero to one recording locations listed
- Each recording location has had one to many tapes recorded there
- Each tape has zero to many credits
- Each credit is associated with one tape
- Each collaborator has one to many credits
- Each credit is associated with one collaborator

Constraints:

- Each tape's catalog number consists of the letters LAA followed by a four-digit number (LAA0001, LAA0002, etc.)
- Tapes are given an audio quality rating of one to ten
- Tape speed is recorded either in IPS or noted as 'multiple'

- Sample rate is recorded in kHZ
- Bit rate is recorded in bits
- A tape can be either a Master or a Safety
- A tape's transfer status can be [NULL], Baking, Baked, or Transferred
- Each hard drive is named with the letters LAA and a number (LAA #1, LAA #2, etc.)

ER Diagram:


TRANSFER ENGINEER
PK trans_eng_id
trans_eng_first_name
trans_eng_last_name
assigned
TRANSFER ASSIGNMENT
PK assignment_id
FK trans_eng_id
FK tape_catalog_id
yields
AUDIO FILE
PK audio_file_id
FK tape_catalog_id
FK assignment_id
audio_file_sample_rate
audio_file_bit_depth
audio_file_transfer_date
audio_file_box_scan_status
audio_file_tracklist_scan_status
FK drive_id
stores
HARD DRIVE
PK drive_id
PROJECT
PK project_id
project_name
recorded for
yields
generates
TAPE
PK tape_catalog_id
PK tape_QR_code
tape_recording_date
tape_media_type
tape_speed
tape_master_status
tape_location
tape_raison_d'etre
tape_archivist_notes
tape_transfer_status
tape_quality_rating
tape_quality_notes
tape_runtime
FK recording_loc_id
FK project_id
TRACK
PK track_id
track_name
track_description
track_type
FK tape_id
contains
RECORDING LOCATION
PK recording_loc_id
recording_loc_name
recording_loc_type
recorded at
has
CREDITS
PK credit_id
FK tape_catalog_id
FK collaborator_id
credit_type
credit_notes
has
COLLABORATOR
PK collaborator_id
collaborator_display_name
collaborator_first_name
collaborator_last_name

Tasks Supported By The Database:

- The database can assist its users in keeping track of the process of tape transferring. Users can query the database to see whether a tape has been processed or not and who it was processed by. The database can also be used to determine the physical location of each tape.
- The database can be used to record technical and production details for each tape. Users can query for both the physical qualities of each tape and for information about the audio recorded on it.
- The database can be used to store information about digital files generated during the transfer process. This information can include technical specs about the files, information about accompanying image files, and the location of these files.
- The database can store information about constituents related to tapes, including the transfer engineers who processed them, as well as producers, musicians, and other personal who were involved in their creation.
- Since the digitization process is still ongoing, the database can be expanded as needed. If Canal Street eventually starts storing files across multiple hard drives, for instance, a new table for RAID storage information could be added to connect the audio file and hard drive tables.

Changes Between Initial Proposal and Final Design:

Since my initial proposal to create a database for storing art objects, I decided to shift the project focus to the storage of audiotape metadata. I first considered expanding an existing database used for linked data about objects at the Whitney Museum, but came to realize that the relational database model was not well suited for RDF triples. After asking Canal St. Studios

about their possible database needs, I was told a database to keep track of data related to their ongoing tape digitization project would probably be of most use to them.

I mostly maintained the same fields used in Canal St's Excel spreadsheet, but decided to merge or split some of them to avoid many-to-many relationships. The transfer engineer field, for example, originally had one to many names associated with each tape. I ended up creating a new Tape Transfer Assignment table to account for the fact that multiple transfer engineers might be assigned to the same tape. I also combined fields listing the tapes' producers, engineers, collaborators, and musicians into a single Collaborators table, connected to each tape by a Credits table. This was due to overlap in the roles; Peter Gabriel, for instance, is listed as a musician for one tape and a collaborator for another, and Brian Eno has occupied all of the four roles.