Ethan Gates, Genevieve Havemeyer,
Karl McCool, Blake McDowell, Ben Peeples
December 8, 2014
CINE-GT 3403
Assignment 2/Assignment 1 Revision

GROUP TWO
VAGABOND PUPPETEERS COLLECTION PRESERVATION PLAN

INTRODUCTION

The following document was constructed to guide the design and implementation of a digitization project, to be carried out by the authors in the fall of 2014 as part of their coursework for the Video Preservation I class in NYU's Moving Image Archiving and Preservation master's program. The course, and this project, were overseen by Mona Jimenez, Associate Director of MIAP, and Erik Piil, Digital Archivist at Anthology Film Archives and adjunct faculty member of NYU.

The materials chosen for this digitization project were taken from the personal collection of Mona Jimenez, and relate to an uncompleted documentary film started by Jimenez in the late 1980s. The film was inspired by the involvement of Jimenez's mother, Mary Walton Jimenez, with a puppeteering troupe, referred to as the Vagabond Puppeteers, who in 1939 toured upstate New York in support of a contemporaneous strike by the Dairy Farmers' Union of the State of New York. Mona Jimenez planned to create a film about the Vagabond Puppeteers through a series of interviews with members and supporters of the troupe (including Mary Jimenez, Jerome (Jerry) Oberwager, Harriet Holtzman Lansky, Pete Seeger and Al Kuchler) and an upstate road trip with her mother to sites where the Vagabond Puppeteers had performed. Though Jimenez's film was never finished, she retained possession of the production materials, which represent an invaluable account of the 1939 strike through the first-hand testimony of eyewitnesses.

For the purposes of this course, a selection was made from the total collection of production materials to target for digitization, owing to the equipment and time constraints of the classroom. The available materials were then further divided evenly between two groups for efficiency. This plan will thus cover the digitization and storage of the following video interviews, both conducted and filmed by Mona Jimenez: of Mary Jimenez, at the home of Jerry Oberwager in Great Neck, NY in August of 1988, and at her own home in Carrboro, NC in the spring of 1990; and of Harriet Lansky and Jerry Oberwager, again at Mr. Oberwager's home in Great Neck during August of 1988 and June of 1989.

Since their creation, these tapes (all 3/4" U-matic) have been in the possession of Mona Jimenez, and stored under office conditions. She possesses participation releases from her mother,

Oberwager and Lansky that assign Mona and the unfinished documentary's co-producer Gina Murtagh all rights to reproduce, copy, exhibit, publish or distribute the content of these tapes.

Following the execution of the project parameters outlined below, a final report will be drafted to summarize the process and results of the project's implementation. The ultimate goal will be not only to deliver preservation and access-level digital files of these Vagabond Puppeteers Collection tapes to Mona, but provide an outline that could be followed for the remainder of the collection not included in this sampling.

PROJECT SCOPE

Selection:

As discussed above, there is more to the Vagabond Puppeteers Collection than the 14 U-matic tapes that the two groups will be digitizing. Group 2 has a total of seven U-matic S tapes, which hold a maximum of 20 minutes per tape. Since at most there is 140 minutes of material, all of the tapes will be digitized for this project.

The relatively small amount of material provides our instructor, Mona Jimenez with a good sample. Digitization will allow her to more easily view and assess the content of the footage. Should the U-matic S material prove worthwhile, a larger digitization project can be done for the remaining material.

In choosing the order and preference of digitization, priority will be given to the three U-matic S tapes that contain footage of Mona's mother, Mary Jimenez. The four tapes including interviews with Jerry Oberwager and Harriet Lansky will be digitized once the Mary Jimenez tapes are successfully transferred.

Order of selection:

VPC_12 - Mom #1
VPC_13 - Mom #2
VPC_14 - Mom #3
VPC_08 - Harriet #1
VPC_09 - Harriet & Jerry #2
VPC_10 - Harriet & Jerry #3
VPC_11 - Harriet & Jerry #4

Metadata:

Preservation metadata (descriptive and technical) for source tapes has been collected and recorded into a spreadsheet titled "Video Preservation 1 FA2014 Vagabond Puppeteers

Collection", filename VideoPreservation_1_FA2014_Vagabond_Puppeteers_Collection. For each tape digitized the following metadata fields are being recorded:

- Client
- Project Name
- Item Number
- Item Title
- Label Annotations
- Container Annotations
- Media Type
- Source Format
- Stock Brand
- Manufacturer's Length of Tape
- Original Record Date (yyyy/mm/dd)
- Physical Assessment
- Digitization Station Operator
- Preservation Master Target Format
- Preservation video codec
- Preservation video wrapper
- Preservation audio codec
- Total Running Time
- MD5 Checksum
- Audio digitization adjustments
- TBC digitization settings
- Transfer Notes
- Source Deck (Name)
- Source Deck (Version)
- Source Deck (Type)
- Audio Transfer Original Channels
- Audio Transfer Preservation Channels
- Video Capture Card
- Video Capture Software
- QC Checked By [name /initial]

This .xlsx spreadsheet containing the technical and descriptive metadata for all media will be placed in the root project folder: Vagabond_Puppeteers_Collection.

Destination Formats:

Three files will be made for each digitized tape; preservation masters will be created during the digitization process using Blackmagic capture software, and derivatives - mezzanine and access files - will be created using FFmpeg. The preservation master will be a Quicktime-wrapped 10-

bit uncompressed file. Mezzanine file will be Apple ProRes HQ 422, which is an agreed upon high quality codec for video editing, and a lower-bitrate mp4, which is a good codec for use as a web distribution or access copy. The Smithsonian Institute recommends a bitrate of 8Mbps for a high quality online streaming file.

For preservation files:
container: QuickTime
video codec: uncompressed
chroma subsampling: 4:2:2
frame size: 720x486
video bit depth: 10-bit
audio codec: PCM uncompressed
storage size: 131.85 GB/ hour

For mezzanine files:
container: QuickTime
video codec: ProRes HQ 422
chroma subsampling: 4:2:2
frame size: 720x486
video bit depth: 10-bit
audio codec: PCM uncompressed
audio bit depth: 16-bit
storage size: 17.91 GB/hour

For access:
container: MP4
video codec: h264
chroma subsampling: 4:2:2
frame size: 640x480 (this aspect ration is for web access)
audio codec: aac
audio bit depth: 16-bit
audio channels: stereo (2)

Storage:
The created digital files will be stored on the network attached storage (NAS) in the MIAP Lab room at 665 Broadway. Since these files will be the personal property of Mona it is unlikely the NAS will be the long-term storage for these files. For the time being, they will be stored using a G-Drive 1TB external hard drive. We recommend creating a second, mirrored 1TB drive as backup storage. Given that there is a maximum of 280 minutes of material (140 minutes per each

group), the maximum amount of storage the 10-bit uncompressed preservation files will require is 615.30 GB and the amount the ProRes 422 mezzanine files will require is 83.58 GB, for a total of 700.88 GB. The required storage for the access files will be negligible in comparison.

Timeline:
The group will hold three two-hour sessions for the digitization process of the seven tapes. Each session will begin with calibration of the U-matic player and of the tapes to be digitized (covered in more detail below).

Although there is only approximately 140 minutes of footage to be digitized, extra time is being given for troubleshooting. As observed during testing, at least one of the tapes has audio problems. For unknown reasons, the timecode signal is on channel 1 of the audio, meaning there is no way to separate it. We know from class discussion that these tapes were hastily recorded. Given that, other unforeseen problems are likely and it seems prudent to allow additional time to deal with other audio and video issues.

WORKFLOW

Preparation:
Collection of descriptive source metadata was done during visual inspection of the tapes. We assigned titles to each tape based on label and container annotations, and recorded technical metadata such as media type, format, stock brand, length, capacity, record date, and notes on the physical condition of each tape in the spreadsheet referred to above.

We assigned unique identifiers that allowed for a range of up to 99 items using a naming convention based on the collection name, Vagabond Puppeteers Collection. Mona has confirmed that there are no more than 99 tapes in the collection. This digitization plan refers to items VPC_08 through VPC_14.

Unique IDs:
Mom #1 = VPC_08
Mom #2 = VPC_09
Mom #3 = VPC_10
H&J #1 = VPC_11
H&J #2 = VPC_12
H&J #3 = VPC_13
H&J #4 = VPC_14

Naming convention for digital instantiations of each tape will be as follows:
Preservation master: VPC_01_p

Mezzanine copy: VPC_01_m
Access copy: VPC_01_a

Source Characteristics:
As noted above, the source materials our group will digitize for preservation and access consist of seven 20-minute U-matic videotapes (five Sony brand XBR20 and two 3M brand MBR20 U-matic tapes). 20-minute tapes such as these were typically used for field recordings. These are first generation, original recordings. They contain talking head interviews intended for use in the never-completed documentary. As noted above, these recordings date from between 1988 and 1989, making the tapes at least 26 years old at this point. For this reason, signal loss is a risk: U-matic tapes from the 1980s have been reported to be at greater risk for signal loss. Also, simply due to sheer age, SSD (sticky shed syndrome) and other physical deterioration of the tape is a concern.

At the time of this writing, of these seven tapes, one tape has been thoroughly viewed and some distinct characteristics and potential technical problems have already been observed. The 20-minute tape viewed had been reused: its roughly 13-minute interview having been taped over a longer, previous recording, seen on the remainder of the tape. On the tape viewed, very high luminance levels were observed (peaking at 110 IRE, exceeding the legal limit of 100 IRE). The two audio tracks on the tape viewed consisted of one carrying the soundtrack and another carrying the timecode. One problem of particular concern is that timecode pulses were found to be bleeding through into the audio track and audible at certain points in the recording.

Technical Environment of Transfer: Hardware and Software Used:
The following is a list of the hardware and software present in the MIAP lab that will be used during this project. See diagrams below for more details on the video and audio signal paths for calibration and digitization.

U-matic Playback Deck, Sony BVU-950
Time Base Corrector/Synchronizer, Digital Processing Systems, Inc. DPS-290
Waveform/Vector Monitor, Tektronics 1740
Video Monitor, Sony HR Trinitron 1954Q
Video/Audio Generator (Test Signal), Kramer 811
Audio Delay Box, Datavideo AD100
MacPro 4.1, running OSX Mavericks 10.9.4
Capture Card, Blackmagic Decklink Studio 2 Version 7.9.3
Video Capture Software, Blackmagic Media Express Version 3.3.1

Setup and calibration:
Prior to digitization, all scopes and monitors will be calibrated with the use of a video reference signal, provided from a Kramer Video/Audio Generator (see above). The U-matic deck in use

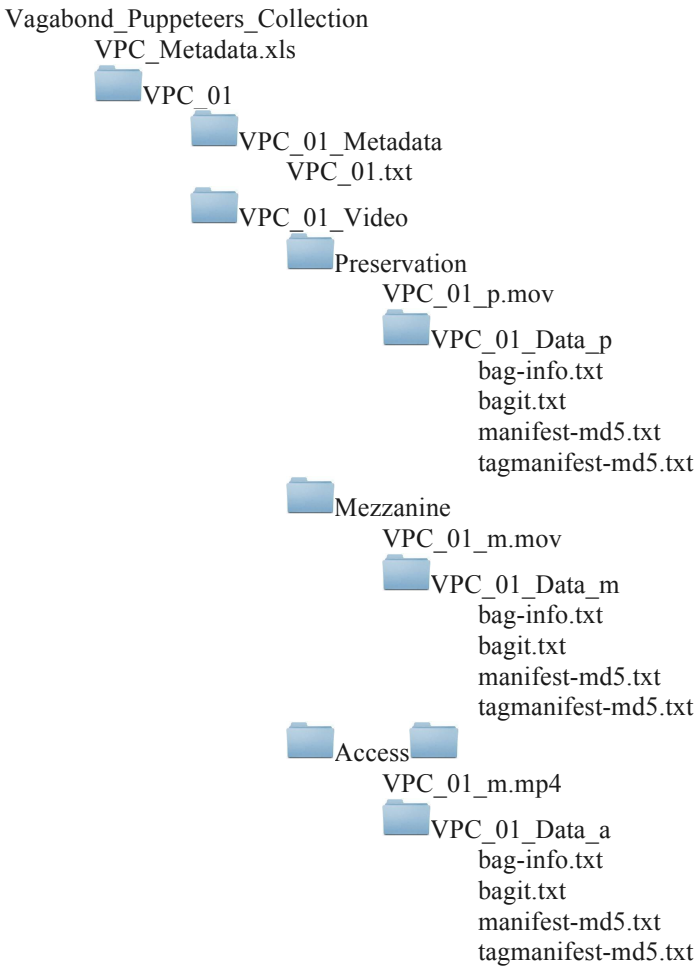
will be set up according to a calibration tape provided by DuArt and the Museum of Modern Art. Based on the signals observed on the calibrated monitors, adjustments may be made to keep the luminance and color levels of each tape broadcast safe, prior to digitization. Audio levels will also be checked using the built-in monitors on the U-matic decks and in the Media Express capture software as the tape is digitized, and any problems on the audio or video levels during digitization will be noted. In extreme cases, the tapes may be digitized into separate files to keep audio and video levels balanced.

Signal path:
For diagrams detailing signal path for equipment calibration and digitization, please see the accompanying PDF attached to this document:
14f_3403_Gates_Havemeyer_McCool_McDowell_Peeples_signal_path_diagram.pdf

Digital conversion parameters:
Digital conversion for these tapes will call for some very minimal signal adjustment. The general approach taken in digital conversion will be to avoid signal adjustment when possible and generally to capture the signal as it come off the tape. However, as noted above, luminance levels observed thus far are quite high and beyond the legal limit of 100 IRE. Furthermore, in adjusting the signal to bring luminance into legal limits, far more detail became visible in the image (such as the textures and patterns of clothing, for instance). Therefore, without adjusting the signal to bring luminance levels down to 100 IRE, information will be lost. The signal adjustment of the luminance levels was observed and approved by Mona Jimenez. In having the original creator of these recordings present in preparing for digitization, our approach benefits from opportunities to consult with her. Therefore, our approach will call for as little signal adjustment as possible, while adjusting to bring the signal into legal limits with the approval of the original artist.

QA and QC - technical content, fixity:
After digitization, a member of the group will view the digitized files to check their quality. If time does not permit viewing a file in its entirety, a minimum of two minutes must be checked at the beginning of the file, at the midway point of the file, and at the end of the file. Furthermore, fixity will be ensured on all files through use of checksums. BagIt will be used to create MD5 checksums, which will then be used to monitor fixity over the life of the files.

Transfer to storage:
Migration of digitized files to their storage destination on the NYU MIAP networked attached storage will be carried out using BagIt to ensure maintenance of file integrity during migration. Within the NAS and long-term and back-up external hard drives, a collection-level spreadsheet will be stored within the VPC folder. Deliverables will be grouped into folder packages named after their unique ID containing preservation files, mezzanine and access copies, checksum reports, and a source metadata text file. The file directory structure is as follows:



TAPE / TRANSFER CALIBRATION & CHANGES

*Temp. folder on NAS = Group_2_F2014_captures

VPC_08 (Mom #1):

TBC:

Black level: -019

Video level: -055

Chroma level: -042

Hue level: +003

VPC_09: (Mom #2)

TBC:

Black level: -019
Video level: -055
Chroma level: -042
Hue level: +003

VPC_10 (Mom #3):
TBC:
Black level: +012
Video level: -074
Chroma level: +009
Hue level: +008

VPC_11 (Harriet & Jerry #1):
TBC:
Black level: -002
Video level: +037
Chroma level: -005
Hue level: -006

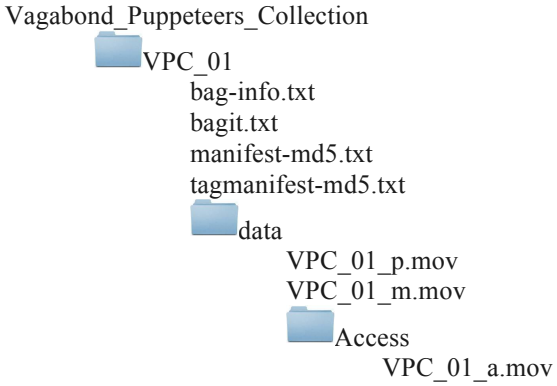
VPC_12 (Harriet & Jerry #2):
TBC:
Black level: -002
Video level: +037
Chroma level: -005
Hue level: -006

VPC_13 (Harriet & Jerry #3):
TBC:
Black level: -002
Video level: +037
Chroma level: -005
Hue level: -006

VPC_14 (Harriet & Jerry #4):
TBC:
Black level: -002
Video level: +037
Chroma level: -005
Hue level: -006

ADDENDUM TO ORIGINAL PLAN:

New file structure: We adjusted our file structure to align with the Baggit file structure, which creates a single data payload folder for the contents of the bag. This was done by keeping all three files in a single folder for each tape digitized. The checksums, therefor will be specific to each tape, and each bag will contain info for all three files, rather creating separate bags and checksums for each file, as previously planned. This was done as a practical measure, to save time during the transfer process. The new file structure is as follows:



Report on digitization process:

- There was a persistent and inexplicable minor (but visible) post-digitization increase in luma that we decided to accept because we were running out of time: While monitoring our very first transfer attempt (on VPC_08), we noticed a substantial drop in luminance on the waveform monitor when switching from the pre-digitization to the post-digitization signal (see signal path diagram). While the pre-digitization signal was adjusted through the TBC to peak at approximately 100 IRE, the post-digitization signal coming out of the Blackmagic capture card appeared to peak at 85 IRE, a difference large enough even to create a visible difference to the image on the monitor.

  After troubleshooting the signal flow we eliminated the tape itself, the TBC, DA, and monitors as the source of the observed problem, suggesting an issue with the capture card. Attempts to pre-emptively adjust the TBC and/or the capture settings in the Blackmagic software to account for an ensuing drop in luminance were ineffective: every time we started to capture in the Blackmagic software, the same low-luminance post-digitization signal was observed. On Erik's suggestion, the capture card firmware on the Mac station was updated, but again, the unstable luminance during transfer remained persistent. The only way to receive a post-digitization signal in line with our originally intended luminance level was to make a TBC or software setting adjustment on the fly once a transfer was already started, a less-than-desirable method considering the inconsistencies both within a captured signal and from one tape to the next. Ultimately (due in some part to time constraints), we decided to simply avoid signal adjustments if possible and aim for greater consistency between transfers despite a potentially lower luminance level in the digitized signal than ideal.

- We had to make several adjustments to our reference / calibrated TBC levels due to complications with waveform monitor described above.

- Head clog, leading to subsequent dropped frames, aborted transfer, and subsequent head cleaning: Several minutes into the final transfer (of VPC_14), the Black Magic capture card unexpectedly dropped frames (frames from the video were skipped during capture). Generally this problem results from a hard disk that cannot keep up with the set video sampling rate, and is usually resolved during hardware configuration; however our group theorized that in this instance the dropped frames may have resulted from a playback issue, and after aborting the capture tried cleaning the video heads on the U-matic deck (as they had not been cleaned after the successive capture of several tapes). This appeared to fix the problem and a second attempt at transferring VPC_14 was successful.

- Problems with waveform monitor/luminance & unstable (travelling) signal: Eventually, after a process of trial-and-error, including many of the changes outlined above, Erik disabled genlocking on the TBC, as the travelling signal seems to have stemmed from problems with the TBC synchronizing to an internal reference and causing feedback which was causing this instability on the waveform monitor. At the very beginning of our project, there was also an issue with our waveform monitor needing adjusting so that the signal generator color bars were hitting the appropriate levels for luminance and blacks. This adjustment was performed by Mona and Eric on the front of the waveform monitor using little screwdrivers to make the adjustments.

- Complications validating bags received on the NAS: Once files were successfully transcoded and spot checked, complete folder packages were created; files were arranged into the new file structure, nested inside their respective ID'd folder, inside the collection folder. Checksums were created for each complete folder package using Baggit and validated in their original location on the transfer station; each package passed validation (validations returned 'true'). The packages were then migrated into the collection's parent folder on the NAS. Validation of the checksums for each bagged package after migration, or upon receipt by the NAS (a "received" bag), would have required the installation of the associated software Bagger. Due to shortage of time, the group resigned to leave the folder packages for the time being unvalidated on the NAS. Copies of all files were also transferred to the external hard drive, as per our original plan.