FELLOWSHIP HANDBOOK

Tanya Yule
San José State University
August 2018
SCOPE ¹:

- inventory and assess the material identified by the Host Station for preservation
- digitize the materials at the digitization station
- create detailed catalog records for the digitized material
- collaborate with AAPB archivists to implement the AAPB’s workflow for submission of the materials into the AAPB for digital preservation, including creating proxy files, generating preservation and technical metadata, and ingesting metadata into the AAPB’s Archival Management System
- research the significance of the collection and create a special collection to highlight the materials within the AAPB
- write a blog post about the collection for the station and the AAPB
- collaborate with the Faculty Advisor at the Graduate Program to document their audiovisual preservation work with the creation of a 3-5 page handbook and a video tutorial on use of the equipment for the benefit of future students

Public Broadcasting Preservation Fellow: Tanya Yule, MLIS 2018
Graduate Program: San Jose State University School of Information
Host Organizations: The Center for Asian American Media and the Bay Area Video Coalition
Faculty Advisor: Alyce Scott, Lecturer, School of Information
Local Mentor: Jackie Jay, Preservation Technician, Bay Area Video Coalition

INVENTORY
At CAAM
At BAVC

CLEANING
Tapes
Machines

CAPTURING
vRecord Configuration for digital Capture
SETTING LEVELS
DECK SET-UP
  U-Matic Deck
  Betamax Deck
  DigiBeta Deck
Notes

SCRIPTS
  Transcode Script for CAAM
  AAPB Proxy Script

AMS INGEST
Drive Load Workflow
  Transfer Logs
  MD5 Manifests

DELIVERABLES
INVENTORY

At CAAM

The CAAM .xls spreadsheet contains basic information about the items that are to be digitized.

In the Google Drive:
   Inventory Folder > CAAM Preservation Project Inventory.xlsx

- Assets with an **Instantiation ID** have been digitized.
- Assets highlighted in **yellow**, were unable to be located
- Assets highlighted in **pink**, were unable to be captured due to physical issues with the tape, the information about those items lives in the BAVC SalesForce record.
- Items that have instantiation IDs should be removed from the physical boxes and reshelved at CAAM (update spreadsheet to their locations).
- Update spreadsheet with items to be captured, and which box they are located in, make note of how many items each title has (Master version is best).

At BAVC

Bring all boxes, and CAAM hard drive to BAVC.

- Barcode and enter all tapes into the BAVC Salesforce system, in the PBPF_CAAM_1 record.
  - Pop or push down red tab enabling the ability to accidentally record over the item.
- For items that have yet to be digitized but have a BAVC Barcode, update them from “uncaptured” to “active”
CLEANING

Tapes
- Clean all U-Matic tapes, enter notes in Salesforce. For tapes that make noises, or show an error message
- Bake tapes that show an error or make noises

Machines
- Clean all parts of the decks that touch the tape with alcohol and a wipe or q-tip for hard to reach areas.
  - Clean before and after each pass through the machine, for every tape.

CAPTURING

System Equipment
Capture Hardware: Blackmagic Decklink 4K Extreme
Capture Software: VRecord 0.7.12
Computer: Mac Pro 2.66 GHz Quad-Core Intel Xeon
Operating System: 10.11.5 (El Capitan)
PreSonus HP2 Headphone Amplifier
Beyerdynamic DT770M 80 OHM Headphones

vRecord Configuration for digital Capture
This step will have to be performed before capturing or setting levels.
1. Open new terminal window
2. Type: vrecord -e
3. Hit enter

Configure capture settings:

Select file path for master files:
SimplyUltra → TransferProjects → PBPF_CAAM_1 → Ingest

Select file path for Aux files:
SimplyUltra → TransferProjects → PBPF_CAAM_1 → Ingest → vRecordAuxFiles

- Video Input: SDI
- File format: Quicktime
- Codec for video: Uncompressed
- Video Bit Depth: 10-bit
- Television Standard: NTSC
- QC Tools: No (these will be created in the scripts)
- MD5 Checksums: No (these will be created in the scripts)
- Audio Input: SDI Embedded
- Channel Mapping:
  - For U-Matic: 1 Stereo (1 & 2 Channel)
  - For BetacamSP or Digital Betacam: 2 Stereo (4 channel)
- Select View: Broadcast range

Hit Enter when done (This will allow you to do the following commands)

To set levels (entering passthrough mode):
  - Type: p
(See notes for setting levels)

To capture:
  - Hit enter
  - Type file name (use this convention):
    - BAVC[use BAVC barcode #]_FullNameOfFile_Try#
    - Example: BAVC1006412_YellowLotusDub_Try1
  - Type you name
  - Hit Enter to capture, Hit play on deck
  - Hit Esc. to end capture

DECK SET-UP
Specific to Station 1 rack and decks, this set-up is likely to change, use this as a framework for setting up.

U-Matic Deck
Deck:
  Sony BVU-800 U-Matic Deck

Cables:
  Mon A In → Deck 2 Mon Out
  TBC2 NTSC Out → TCB1 Comp In
  575 SDI Out → BM SDI In
  Deck 2 NTSC → TBC2 NTSC In

Digital TBC “1”: Leitch DPS-575
Levels set to default/unity
Use to adjust H-Position

Settings:
  Genlock (green)
  Autotrack (green)
  Video In: Comp
  Audio In: SDI

Analog TBC “2”: Digital Processing Systems Inc., DPS-295
Use to adjust levels
Settings:
  Function: TBC
  Auto Doc: Off
  Control: Local
  Freeze: Frame
  Input: NTSC 1
  Pro Amp Effects: Proc Amp

Betamax Deck
Deck:
  Sony DVW-A500 Digital/Betacam
  Adjust audio levels from deck

Cables:
  TBC 1 Comp In → TBC 2 NTSC Out

Digital TBC “1”: Leitch DPS-575
Levels set to default/unity
Use to adjust H-Position
Settings:
  Genlock (green)
  Autotrack (green)
  Video In: Comp
  Audio In: Analog

Analog TBC “2”: Digital Processing Systems Inc., DPS-295
Use to adjust levels
Settings:
  Function: TBC
  Auto Doc: Off
  Control: Local
  Freeze: Frame
Input: NTSC 1
Pro Amp Effects: Proc Amp

DigiBeta Deck

Deck:
Sony DVW-A500 Digital/Betacam
Adjust audio levels from deck

Cables:
No cables

Digital TBC “1”: Leitch DPS-575
Use to set levels
Use to adjust H-Position
Settings:
Genlock (green)
Autotrack (green)
Video In: SDI
Audio In: SDI

Analog TBC “2”: Digital Processing Systems Inc., DPS-295
Settings:
Function: TBC
Auto Doc: Off
Control: Local
Freeze: Frame
Input: NTSC 1
Pro Amp Effects: Proc Amp
Open BAVC Ingest Notes

- Take timecode notes for any information that should be reviewed or can cause issues during the QC process, including: bright flashes, film artifacts, dropouts, image errors etc.
- If there is an abundance of issues note that in the summary
- Review the timecodes after the video has been trimmed, and not the exact time they occur, finish each timecode with a semicolon to mark which have been reviewed
- Add the summary and timecode to the Salesforce record
- Content description should be brief to later add to the record metadata in AMS.

**Summary:**
Abupt start Bars and Tones. Bright flashes and abrupt transitions throughout video. Many native film artifacts in black and white footage. Multiple image errors on right side of clips. Four clips post credits with no audio.

**Timecode:**
00:01:25 - Film artifact;
00:29:13 - Native film artifact;
00:32:35 - Native dropout;
00:33:47 - Chroma error, resulting possibly from a crash record;

**Content description (no original description in DB):**
A documentary about the Cu Chi tunnels in Vietnam during the Vietnam war. The film contains interviews with survivors of the war who inhabited the tunnels, including archival footage, and photographs. The filmmakers visit the tunnels in current day to explore the way so many Viet Cong lived during the war. The documentary features an interview with Robert Bas (American P.O.W captured in 1968, who was at the “Hanoi Hilton”), and the Viet Cong member who captured him, explaining the events and emotions surrounding the event.
SCRIPTS

Transcode Script for CAAM

*Run scripts at end of day*

1. Open Terminal window
2. Drag and drop `transcodeEngine.py {space} -i`
3. Drag and drop folder from SimplyUltra → TransferProjects → PBPF_CAAM_1 → RunScripts
4. Hit Enter

*Run through series of prompts for creating derivatives:*

01. How many derivatives?: **2**
   a. Which codec for derivative 1?: [2] *prores/mov*
   b. Derivative 1?: [2] **Leave interlaced**
   c. Map audio for Derivative 1?: [1] **Keep original**
02. Which codec for derivative 2?: [1] *H.264/mp4*
   a. Derivative 2?: [1] *De-Interlaced*
   b. Map audio for Derivative 1?: [1] **Keep original**
   c. Frame size: [1] **640 x 480**
03. Move files to presraid: **yes**
   a. Name of file/folder: InProgress/PBPF_CAAM_2
04. Create QC Tools: [1] **yes**
05. Create checksums: [1] **yes**
06. Create MD5: [1] **yes**

- After script has finished running:
  - Open: SimplyUltra → TransferProjects → PBPF_CAAM_1 → RunScripts → mediainfo.csv
  - Enter information into SalesForce
  - Check box: added to presraid
  - Dropdown: pass
- Move all the files to: SimplyUltra → TransferProjects → PBPF_CAAM_1 → QCReady

*Note: The AAPB Proxy files will be created after the AMS Ingest.*

In Salesforce add the:
- MedialInfo
AAPB Proxy Script

*Run this after AMS Ingest, load AAPB drive with master files, once drive is loaded change file names with GUIDS (see AMS Ingest), create proxy files for AAPB.*

AAPB PROXY SPECIFICATIONS

Create video access proxies for all digitized files, to the following specs:
- h.264/AVC (Codec ID avcl), 711 kbps, 480x360, 4:3 aspect ratio, YUV 4:2:0, 8 bits, audio AAC 48.0 KHz/128 Kbps in an .mp4 wrapper

The ffmpeg commands used to create these files are:

1. **Individual files:**
   a. Open Terminal:
   b. Copy and Paste:
1. At the beginning of the fellowship enter the basic required metadata into the **AAPB AMS .xlsx** file, keep the record in Google Drive.
   a. Do not include the **Instantiation ID** at this point, this will be created after you have digitized the item and barcoded using CAAM’s barcode ID
   b. Hold off on uploading until you are nearing the end of the project.

2. After capturing records:

   **To transcode an entire folder using Bash script:**
   a. `create_video_proxy_AAPB.sh` – to transcode a batch of audio files or video files at once.
   b. To run one of these bash scripts:
      i. Make sure the script is installed and executable on your computer by typing the command:
         - `chmod +x [drag and drop the script: Create_video_proxy_AAPB.sh]`
      ii. In command line:
         - Drag and drop the script into the command line
         - Drag and drop the folder containing the files that you would like to transcode
         - Drag and drop the destination folder.
            a. Before you hit ‘ENTER’, the command should look as below:
            - Hit enter

---

```
ffmpeg -i [drag and drop the master file here] -vcodec libx264 -pix_fmt yuv420p -b:v 711k -s 480:360 -acodec libfdk_aac -b:a 128k [name of the output file].mp4
```

```
```

---

**AMS INGEST**

---

AMS INGEST
a. Barcode the actual captured item with the CAAM Barcode, this is the instantiation ID.
b. Enter this information into the AAPB AMS .xlsx
c. Remove any records that had been created initially, that did not get captured.

3. Export the record from Google Drive
a. Rename the file: caam_pbpf_inventory_3.xlsx
b. Open the file using Apple Numbers
   i. Export → CSV → Unicode (UTF-8)
   ii. Retain the same naming convention

For the AMS upload you must use Firefox as your browser, Follow the directions below:

AAPB Inventory Ingest Documentation

1. Sign into the AMS at ams.americanarchive.org.
2. Go to the Records tab.
3. Under the Operations button, click Import Collection.
1. In the dialog window that appears next, select the organization you are importing records for, and click Import Collection again. This will bring you to the MINT interface.

   ![Image of Import Collection window]

1. Click Choose File, and select the .csv of the records you want to ingest. Click the checkbox next to “This is a CSV upload.”
2. Leave “Contains header” checked.
3. Leave “Define the field separator” as comma.
4. Change “Define the escape character” to “\”. By the time you’re done, the MINT upload form should look like this:

   ![Image of MINT upload form]

By the time you're done, the MINT upload form should look like this:
1. Click Submit.
2. A list of recently imported datasets will appear. Select the dataset that you just uploaded from the list. (click the .xml icon.)
3. When you click the database, several icons will appear underneath it. Click the mapping icon, in the middle of the three.

4. To Create new mapping:
   - In the dialog box that appears, click the “Add New Mapping” icon on the far right, which looks like a document with a green + over it. Name your new mapping something distinctive. A good format might be the name of the organization or collection followed by the date.

5. When your mapping opens, you will see the source schema down the left hand side, the target schema down the right hand side, and the mappings in the middle. The page will look empty at first. To see all of the spreadsheet column headings that you will be mapping, click the plus signs on the left, under ‘Source Schema,’ to expand the view. To see the fields that you can map to, click the pbcoreCollection button on the right hand side, under ‘Target Schema’.
6. In the mapping section, drag the column headers from the left hand side to where it says “unmapped” in the mapping section next to the corresponding element or attribute. Since PBCore is a hierarchical standard, some elements are nested under their parent elements, and will not appear in the mapping area until you click the “+” next to the parent element. This screenshot shows the pbcoreRelation element, expanded out to show its child elements; the pbcoreCoverage element, which is still nested and therefore does not show its child elements; and the pbcoreAudienceLevel element, which has no child elements.

As a note, all instantiation elements exist within the ‘pbcoreInstantiation’ container, and will not appear in the mapping section unless that container element is expanded.
Any attributes that can be applied to an element can be seen by clicking the small @ symbol next the element. Here is a screenshot of the same three elements shown above, now with the attribute fields expanded out.

7. For repeating elements, click the green plus sign on the right of the element you’d like to repeat. It will appear at the bottom of the mapping section. The screenshot below shows the pbcoreCreator element, the elements that come after, and the duplicated pbcoreCreator element at the bottom.
● It’s best practice to map going in descending order from the top of either the source schema section or the mapping section. The first time you map a source field to a mapping field, the source field will highlight blue. But if you map another source field to that same mapping field repeated, it will not turn blue. So the source fields turning blue is not a reliable indication that the field has been mapped, which is why going down in order is better to keep track of what’s been mapped already.

● When all of your source fields have been mapped, click the Preview button at the top of the screen. You will be able to preview the data here, but the most important part is to check that the validation tab says that the xml is valid. If it’s not valid, this tab will list your errors (which could be in the mapping or in the data). For example, this validation shows a failure to provide a ‘source’ attribute for ‘pbcoreIdentifier,’ which is required by the schema.

● Once the validation tab tells you the xml is valid, click “finished” along the top of the window to save your work. This will take you back to the screen showing recent datasets.

● Select the dataset you’ve been working on again, and click the transform button, on the right of the three icons that appear.

8. A new dialogue screen will appear which offers a drop-down menu to select a mapping. Select the mapping that you just created, and click submit.
9. When the transformation has completed, click the approve button that appears next to the dataset.

10. The data has been ingested into the AMS and should begin appearing if you search for it in the records tab, as it is processed.

Generating spreadsheet with original data and new GUIDS

1. Log into the AMS at ams.americanarchive.org
2. Click the Records tab
3. Filter to the records from the organization you just added records for by identifying the correct organization in the search options on the left-hand side of the screen.
1. Once you have the set of records that you’re looking for, click the Instantiations tab on the top of the screen.

1. Click the Operation button, right under the ‘Instantiation’ tab, and select Limited CSV. A dialog box will appear asking if you’re sure you want to export records; click ‘yes.’ Either download the csv when it presents the option, or wait for the AMS to email you a link for the download (which will only happen if it’s a really big file).
1. When you have the csv, open it.
2. Make another copy of the original spreadsheet, and add “_withGUIDs” to the end of the filename of the copy. Work in the copy.
3. Open the copy in Excel. Create a second sheet in the workbook, and copy the GUID and the UniqueID columns from the csv that you just downloaded into the second sheet of the workbook (preferably the UniqueIDs in the first column, and the GUIDs in the second column).
4. Select the first sheet of the workbook.
5. Create a new column to the right of the column with the pbcoreIdentifiers.
6. In that column, run a VLOOKUP, for the GUID, against the 2 columns on the next sheet, printing the 2nd column, specifying FALSE for possible match. Make sure you mark all of the column and row numbers in the reference table with $.
   a. Example: =VLOOKUP(A1, Sheet2!$A$1:$B$300, 2, FALSE)
2. Copy and paste the VLOOKUP formula down the entire column. It should print all of the GUIDS in the column.
3. Create a new column as the first column of the spreadsheet. Copy the GUIDS from the VLOOKUP column, and use a paste special (values only) to paste the GUIDS into the first column.
4. Delete the VLOOKUP column and the second sheet.
5. You now have a copy of the original spreadsheet with the new GUIDS added.

Drive Load Workflow

Drive loads should generally be started at the end of the day. Too many drive loads on a single station can slow the station’s SAN connection, and QCing during a drive load can be very slow.

- Drive Formatting
  - Attach labels to all important parts of the drive
    - Box
    - Drive
    - Power Brick
    - Computer end of USB Cable
  - Plug Drive into computer
  - Format the Drive

*ONLY PERFORM THE FOLLOWING STEPS IF THE CLIENT DOES NOT HAVE MATERIALS ON THE DRIVE ALREADY*

- Open Disk Utility
- Click on the corresponding drive under External on the left-hand side of the window
Click the Erase button at the top of the window

- **WARNING: THIS WILL *DEFINITELY* ERASE THE DRIVE :-)**
- Enter the name of the Drive. You can only use 10 Characters, so make them count!
- Set the Format to ExFAT
- Click Erase. This process should run rather quickly.
  - Drag the BAVC_Preservation folder from the desktop of your computer to the root level of the hard drive
  - Once finished, you can either start loading to the drive, or place it on the PresSuite shelf if it will not be used immediately.

- **Drive Loading**
  - Plug the drive into the load station
  - Open a Terminal Window
  - Type in the following command
    - `rsync -rlgtvD --progress`
    - Press the space bar
    - Drag in the contents on 05_Loading from the opportunity folder on the SAN
    - Drag in the BAVC_Preservation folder on the Drive
    - The resulting string should look something like this:

    ```
    rsync -avv --progress
    /Volumes/SymplyUltra/TransferProjects/OpportunityName/05_Loading/Files_Access
    /Volumes/SymplyUltra/TransferProjects/OpportunityName/05_Loading/test/Files_Preservation
    /Volumes/ClientDriveName/BAVC_Preservation
    ```
  - Press Return and watch the files start to load!
  - When complete, move files into appropriate (Pres or Access) folders.
  - Click the "Loaded to Drive" box on each record in Salesforce.

---

**Transfer Logs**

Transfer logs are included on the client's drive along with their preservation master files. For more information about client deliverables, go here:

- In Salesforce, go to **Reports**.
- Navigate to the report called under **Preservation_2018 → ClientLog_Audio** or **ClientLog_Video** depending on the format (they have different schemas).
- Run the report and click the **Customize** button.
- Click the **Edit** button next to the Opportunity Name filter
● Type in the first few letters of the opportunity name of the project you are completing
● Click OK
● Run the report
● Check that all the fields have consistent and correct information
● Click Export Details
● Export as an .xls file
● Open the .xls file in Excel and remove the text at the bottom about time/user that ran the report.
● Highlight the sheet and reformat row height to 0.19
● Scan through all content checking for inconsistencies or missing information

● Save on the client’s drive as [OpportunityName_TransferLog] for example "PAP7_OregonHistoricalSociety_TransferLog."

**MD5 Manifests**

Once the files have been updated with the GUIDS (after the CAAM drive has been loaded), create the MD5 manifests:

Look for hidden files:

- Open Terminal
- Type: `find [drag and drop the directory containing the master files] -type f` 
  - Hit Enter

*Hidden files are okay, but it is preferable to run without them.

  a. Master file Manifest:
    - i. Open Terminal:
      - Type: find [drag and drop the directory containing the master files] -type f -exec md5 {} > [name of md5 manifest] \

**Example:**
`find /Volumes/PBPF_CAAM_1/BAVC_Preservation/Files_Preservation -type f -exec md5 {} > /Volumes/PBPF_CAAM_1/BAVC_Preservation/Manifest_MasterFile.csv \
`;

  b. Proxy files:
    - i. Open Terminal:
      - Type: Find [drag and drop the directory containing the master files] -type f -exec md5 {} > [name of md5 manifest .csv] \

**Example:**
`find /Volumes/PBPF_CAAM_1/BAVC_Preservation/Files_Access -type f -exec md5 {} > /Volumes/PBPF_CAAM_1/BAVC_Preservation/Manifest_ProxyFile.csv \
`;
**The directory is the external HD that has been loaded for the AAPB containing the master and proxy files.

DELIVERABLES

CAAM (keep naming convention that BAVC uses):
- Master files
- Mezzanine files (ProRes)
- Access files (640 x 480)
- SalesForce metadata report (BAVC)

AAPB: See file name requirement
- Master files
- Access/Proxy files (360 x 480)
- MD5 Manifests:
  - Master files
  - Access/Proxy File

Confirm that the files copied exactly onto the external hard drive
1. Create a second MD5 manifest for the files on the drive, using the same command
2. Compare the MD5s to make sure that they are identical

1. Email the MD5 manifests to Rebecca and Miranda at rebecca_fraimow@wgbh.org and miranda_villesvik@wgbh.org

2. Mail the drive containing the files to:
   Rebecca Fraimow
   WGBH Media Library and Archives
   1 Guest Street
   Brighton, MA 02135