11-1-19 ACM-NE Archiving/Metadata Presentation Notes

Slide 3 - Amherst Media

- Began archiving w/archive.org in 2010
- Vast archive of VHS, ¾ & beta tapes going back 30+ years
- Determined that we needed to preserve Amherst Media's programming, as it was a part of the town's history
- Began the process of obtaining a Community Preservation Act grant in 2012 to fund equipment and staffing to digitize and archive our content - the town agreed that this was important and town meeting voted to approve our grant request
- Equipment
    - Two iMac's - each one connected to a VHS deck via a Blackmagic thunderbolt hub
    - Allowed us to capture two tapes at a time
- As of 2015, we had digitized over 1500 VHS tapes with the help of work-study students

Slide 5 - Master metadata spreadsheet

- As we go through my process, you'll see that I learned a lot about what to do differently after starting the Amherst Media archive project. My spreadsheets are set up quite differently now.
- No solid plan or understanding of how the archive.org batch upload tool or metadata fields worked
- We had already captured 8 drives worth of content by the time we started our first upload to archive.org, which meant we had to use this spreadsheet plus create separate spreadsheets for each batch upload that aligned to the necessary fields for archive.org
- *Quick overview of each field in the master spreadsheet*

Slide 6 - Archive.org batch upload spreadsheet

- Two collections on archive.org - one for current programming and another called Amherst Media Archives, which was for the content that we were digitizing.
- Same basic metadata as the master spreadsheet, but this one is cleaned up and better organized for archive.org uploading
- The master spreadsheet contained a number of fields that were mostly metadata for us to use in-house, such as tape number, format (VHS, ¾, beta) and who captured the tape.
- Speaking of who captured the tapes...Why log who captured the tape? When you're working with multiple volunteers, it makes it a lot easier to narrow down where the issue is if you have someone who is repeatedly logging the metadata incorrectly and, believe me, it happens.

Slide 7 - Archiving at BATV

- BATV has been providing the resources for the community to create content since 1987, but we haven't been holding on to all of that content.

- Programs produced prior to 2014/2015 were not digitized and were being “destroyed after five years in-house if the producer no longer wanted them. The only tapes/DVD’s we keep indefinitely are town meetings. Our town meeting tapes go back to the 1990’s.
- We have a lot of DVD’s instead, some of which are already digitized and stored on our Thecus NAS server for archiving. We’ve been housing files on the Thecus since about 2014.
- Since we already had several hundred digital files on the Thecus, I decided to start there with our archiving.

Slide 9 - Current archive project

- Having completed the process of uploading all of the digital files from the Thecus, we’ve now moved on to bring in some of those VHS tapes that were “destroyed” over the years.
- Tapes were never actually destroyed as we had a resident/producer who was not too keen on seeing them get thrown out and has offered to take them over the years. We’re convinced he has a VHS tape fort in his basement, but no one has been able to prove it.
- Bob brings one box of tapes at a time and I digitize them, gather the metadata, and return them back to him in exchange for another box of tapes.
- *Brief overview of our archive station setup*

Slide 10 - BATV Batch Upload Spreadsheet

- Went into this process with more knowledge and an opportunity for a clean start with my spreadsheets based on what I learned from my AM experience. Having gained some knowledge, I also added some columns to my spreadsheet over the years. This will allow me to add all of my content to the same archive.org collection instead of having to have two in order to keep new programming separate from the older stuff that’s being archived. I’ve done this by adding a “date” column to the spreadsheet. The date that I use is the date that the program was produced. If I at least have a year, I usually pick January 1 of that year. This means that when you look at our archive.org collection, you’ll see the most recent programs first, since I always include a date with my uploads, including those that get uploaded through the web uploader.
- Why didn’t I add a “date” column to the Amherst Media spreadsheets? There were a couple of reasons. We didn’t always know when a program was produced. The nice thing about having a longtime resident/producer involved with the project is that Bob was often involved in the production in some way, or he knew who was, so he can help me with the big question marks when it comes to metadata. Amherst Media’s metadata wasn’t always complete and we didn’t want to take a complete and total guess as to when a program was produced. If I at least have a year, which I typically do, I’m happy with that. The other reason was that I didn’t realize that we could add a date column at the time. Even if I had known, without always having a date to enter into that field, I didn’t want the old files to get mixed up with the newer programming, so at the time, it was easier to just create a second collection
- *Brief overview of the spreadsheet - talk about commonly used keywords*

Slide 11 - The creator field

- For individual uploads via the web uploader, we use the same metadata we use in the spreadsheets for the batch uploads
- I always put the producer's name in the creator field. In this example, it's Sandra Giroux.
- Again, if I don't know the exact production date, which is rare with the current programs, I typically use either the first of the month or the day that I received the program as the production date. In this case, the program was filmed on October 2, 2019.
- Topics or subject - If you're using the spreadsheet, you can see that I used my same commonly used keywords: billerica, batv, billerica access television, and massachusetts (a recommendation made by John during last year's workshop)
- For this particular program, I also like to use the name of the program, our town live/otl, and include the names of any guests/organizations that were featured on the show.

Slide 12 - The Creator Field Continued

- Why is it important to use the creator field? It helps distinguish producers within a collection. All of the programs that we include in our collection are produced for BATV, however, they are produced by a variety of people and BATV does not take ownership of programming produced by our community members.
- We assign a producer to programs entered into our cablecast system, so we should be doing the same for our archive.org collection as well.
- It makes it easier to find programming by a particular producer. All I have to do is click on Cosmo Cavicchio or George Fennell to find all of the episodes of Cosmo's Journal, the Yankee Doodle Parade coverage, or Gabfest within our collection.
- Also makes it easier for producers to share all of their programs

Slide 13 - Metadata Bookmarks

- Have a series that you regularly upload to archive.org? Did you know that you can enter in all of the repetitive metadata, such as keywords and basic show descriptions, and bookmark them for future use so you don't have to re-enter that info every time?

Slide 14 - How to create a metadata bookmark

- Enter all of the repetitive metadata
- Click on "Save this metadata" on the bottom right of the page
- A window will open with a URL that contains all of the metadata you entered. Copy and paste that link into a new browser window and load the page.
- Create a bookmark. I like to create a folder of archive.org upload presets and save all of these metadata bookmarks in that folder. If you use Chrome, you can also export that folder of bookmarks to share with co-workers or load onto other computers in the facility.
- Never have to enter repetitive metadata for series programs again!

Slide 15 - Questions/Contact info