MediaConch in action! Issue #1



Hey Eddy! Introduce yourself please.

Hey Ashley! I've recently become a ?Denverite? and have started a new job as an Assistant Conservator specializing in electronic media at the Denver Art Museum (DAM). Before that, I was down in Baton Rouge, Louisiana working as a National Digital Stewardship Resident with Louisiana Public Broadcasting (LPB). When I'm not working, I like to listen to podcasts, read comics, play bass, and stare into the endless void that is ?twitter? (@EddyColloton). I'm on a big H. P. Lovecraft kick right now so let me know if you have any recommendations (Dunwich Horror is my current fav, but At the Mountains of Madness is a close second).

What does your media ingest process look like? Does your media ingest process include any tests (manual or automated) on the incoming content? If so , what are the goals of those tests?

The ingest procedures I'm using for the Denver Art Museum are pretty different from the ones we worked out at Louisiana Public Broadcasting, for all kinds of reasons. The two institutions have very different types of collections, and they use their repositories very differently, too.

At the Denver Art Museum, ideally, material will enter the digital repository upon acquisition. Ingest procedures need to be able to be tailored to the eccentricities of a particular media artwork, and flexible enough to cover the wide array of media works that we acquire (websites, multi-channel video installations, software-based artworks, or just a collection of tiff files). With this in mind, we're using Archivematica for ingest of media into our digital repository. It allows us to automate the creation of METS wrapped PREMIS XML documentation, while manually customizing which microservices we choose to use (or not use) as we ingest new works. Some of the microservices I use on a regular basis are file format identification through Siegfried, metadata extraction with MediaInfo, ExifTool, and Droid, and normalization using tools like FFmpeg.

Things couldn't be more different at LPB. All completed locally produced programming automatically becomes part of the LPB archive. The LPB Archive is then responsible for preserving, describing and making that content accessible through the Louisiana Digital Media Archive (LDMA), located at http://www.ladigitalmedia.org/. LPB's ingest procedures need to allow for a lot more throughput, but there's much less variability in the type of files they collect compared to the DAM. With less of a need for manual assessment, LPB uses an automated process to create MediaInfo XML files, an MD5 checksum sidecar, and a MediaConch report through a custom watchfolder application that our IT engineer, Adam Richard, developed. That code isn't publically available unfortunately, but you can see the scripts that went into it on my GitHub.

Where do you use MediaConch? Do you use MediaConch primarily for file validation, for local policy checking, for in-house quality control, for quality testing for vendor files?

Primarily policy checking, as a form of quality assurance. At LPB, most of our files were being created through automated processes. Our legacy material was digitized using production workflows, to take advantage of existing institutional knowledge.

This was very helpful, because we could then repurpose equipment, signal paths, and software. But, using these well worn workflows also meant that occasionally files would be encoded incorrectly, aspect ratio being one of the most common errors. We would check files against a MediaConch policy as a way of quickly flagging such errors, without having to invest time watching and reviewing the file ourselves.

At the Denver Art Museum, we plan to use MediaConch in a similar way. The videotapes in the museum's collection will be digitized by a vendor. Pre-ingest, I plan to do tests on the files for quality assurance. After fixity checks, I will check to make sure our target encoding and file format was met by the vendor using MediaConch. I intend to use the Carnegie Archive's python script from their GitHub to automate this process. Once I know that the files are encoded to spec, I will be creating QCTools reports and playing back the files for visual QC. I've been following the American Archive of Public Broadcasting's QC procedures with interest to see if there's any tricks I can cop from their workflow.

At what point in the archival process do you use MediaConch?

Basically pre-ingest for both LPB and the DAM. When using MediaConch as a policy checker, my goal is to make sure we're not bothering to ingest a file that is not encoded to spec.

Do you use MediaConch for MKV/FFV1/LPCM video files, for other video files, for non-video files, or something else?

I use MediaConch for MKV/FFV1/LPCM video files and for other types of video files as well. At LPB we were using MediaConch as a policy checker with IMX50 encoded files in a Quicktime wrapper and H.264 encoded files in a .mp4 wrapper. You can find the policies I created for LPB here, and I talk through the rationale of creating those policies in the digital preservation plan that I created for LPB, available here (MediaConch stuff on page 24, and page 45). I'm happy to report that LPB is currently testing a new workflow that will transcode uncompressed .mov files into lossless MKV/FFV1/LPCM files (borrowing heavily from the Irish Film Archive's lossless transcoding procedures, as well as the CUNY TV team's ?make lossless? microservice).

At the DAM, we'll be using MediaConch as a policy checker with Quicktime/Uncompressed/LPCM files, and for validation of our MKV/FFV1/LPCM normalized preservation masters.

My understanding is that MediaConch is going to be integrated into Archivematica's next release. I'm really looking forward to that update, since at the DAM we have decided to create MKV/FFV1/LPCM files for any digital video in the collection that uses a proprietary codec, or an obsolete format. A lot of the electronic media in the museum's design collection comes from the AIGA Archives, which the DAM collects and preserves. A ton of the video files from the AIGA Archives were created in the aughts, and they use all kinds of whacky codecs - my favorite so far is one that MediaInfo identifies as ?RoadPizza? (apparently a QuickTime codec). Given that I don't want to rely on the long-term support of the RoadPizza codec, we're normalizing files like that to MKV/FFV1/LPCM through an automated Archivematica micro-service that uses the following FFmpeg script (which I cobbled together using ffmprovisr):

```
ffmpeg -i input_file -c:v ffv1 -level 3 -g 1 -slicecrc 1 -slices 16 -c:a pcm_s16le output_file.mkv
```

To be clear we are also keeping the original file, but just transcoding a second version of the file to be cautious.

Through that implementation we intend to use the Archivematica MediaConch microservice to validate the encoding of the video files that we have normalized for preservation.

Why do you think file validation is important the field?

I wish this was a joke but it honestly helps me sleep better at night. MediaConch and melatonin make for a well rested AV archivist/conservator, hah. I like knowing that the video files that we are creating through automated transcoding processes are up to spec, and comply with the standards being adopted by the IETF.

Also, using MediaConch as a policy checker saves me time, and prevents me from missing bonehead mistakes, of which their are loads, because I work with human beings (and possibly some aliens, you never know).

Anything else you'd like to add?

Just want to offer a big thanks to the MediaConch team for everything that they do! I know there's a pretty big overlap betwixt team Conch, CELLAR, and the QCTools peeps - you're all doing great work, and regularly making my job easier. So thanks for that.

To read more from Eddy, check out his Louisiana Public Broadcasting Digital Preservation Plan