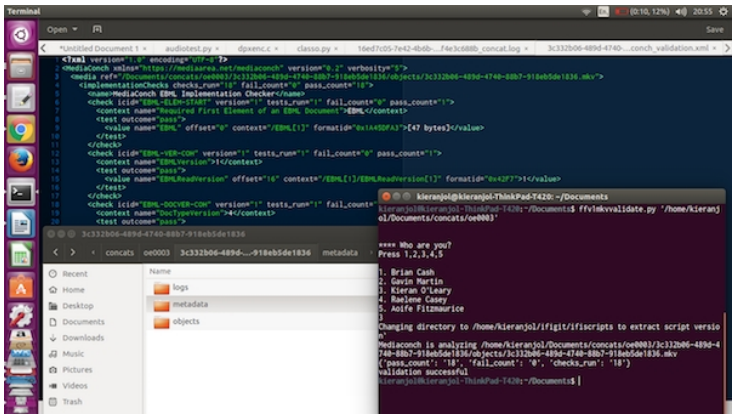
Interview with Kieran O?Leary



Kieran's workstation

Hey Kieran! Introduce yourself please.

Hi! I'm Kieran O'Leary, originally from a relatively rural part of County Cork in Ireland, now living in Dublin City, working in the Irish Film Archive within the Irish Film Institute. I've been fascinated by digital video since about 2002, when I got my first home PC (I was around 17 years old, so relatively late to the party). I started encoding my DVD collection to MPEG-4 ASP AVI files and became fascinated with the process. A few years later, I got into some terrible amateur filmmaking and photography, which ultimately landed me an internship with the Irish Film Institute and I'm still here, mostly working on code, workflows, metadata, digitisation/migration, facilitating access to our collections and other good stuff.

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?

My colleagues Eoin O'Donohoe and Anja Mahler handle deliveries of new, contemporary content, such as AS-11 broadcast files and DCPs/DCDMs that are delivered as part of overarching agreements with The Irish Film Board, the Broadcasting Authority of Ireland, and The Arts Council. They are considering implementing MediaConch to confirm that the deliveries are as agreed, but I mostly deal with reformatted tapes or scanned film. A lot of the tests that we do are based around figuring out if we did everything as best we could ? could we have migrated that tape better, were the correct settings definitely used, etc. As we transcode our uncompressed video to FFV1/MKV, the main checks that we do involve framemd5s to ensure that the FFV1/MKV file produces the exact same content when decoded. Occasionally, we are able to convince some vendors to provide us with checksums, so we perform fixity checks upon arrival. We recently had to use FCP7 for some tape capture, and the v210/MOV files in the Capture Scratch had no interlacement or aspect ratio recorded in the container. I wrote a simple mediainfo check to see if these conditions were true: Check if Pixel Aspect Ratio = Square (It should be 16/15) and check if no interlacement info is specified (it should be interlaced top field first).

I am currently working on a project where we received about 25 hard drives full of production material from a production company, Loopline Film. We will need a lot of checks here as the content is all quite heterogenous ? a large mix of P2 cards, various XDCAM flavours, DSLR, migrated tape, FCP projects, subtitle files and many more. There are a lot of duplicate files so we will have to perform various tests and checks to figure out which are the best candidates to move forward into our ingest workflows.

Generally, our ingest process involves registration, metadata extraction, packaging into a consistent folder structure, fixity, descriptive and technical cataloguing, and attempting to log every step in the process as best we can.

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?

MediaConch always makes me feel guilty because I know that I should be using it in more of our workflows. Currently, we mostly use the GUI when files are delivered from vendors. We are supposed to get a hard drive full of files with various attributes, and MediaConch automates a lot of this work via local policy creation, usually created from an ?ideal file'.

As we have a lot of FFV1/Matroska, it would be an important preservation event to perform an actual validation against the FFV1/Matroska standard via Mediaconch in order to ensure that we have valid, standard compliant files.

I recently started experimenting with using MediaConch's implementation checker for our FFV1/Matroska files. The pull request is here: https://github.com/kieranjol/IFIscripts/pull/201. It is designed around the package structure for our existing FFV1/MKV files. It finds an MKV file, launches MediaConch's command line interface, creates an implementation report in XML format and stores it in our metadata folder, then the script parses the XML to check if all tests were a success or not. These preservation events are logged in a growing log file in our logs folder, and then the checksum manifest for the package is updated to reflect the new/changed files. It really was lovely to be able to quickly integrate MediaConch into our workflows, as well as enrich existing packages. I think I'll probably get this process to run as part of our FFV1/Matroska normalisation script as well, so we have instant validation.

These events will ultimately be logged as ?eventType=validation' PREMIS events as well. Just in case you were curious, I ran the process on 290 files and all passed :) Each validation only took a few seconds, though one stubborn file took a lot longer. Dave Rice got to the bottom of the issue eventually though.

It has also been used intermittently for in house quality control. I recently had an epiphany where I realised that MediaConch HAD to be used for our in house quality control, because one of my scripts had a stupid bug that would have been caught much quicker with a MediaConch policy.

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

It's usually at an early stage. We use the Spectrum Collections Management standard here in the IFI, so the phase is ?Object Entry'. This is a pre-accession period where the files are still undergoing quality control measures, and they may be rejected.

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

It's actually been mostly used for vendor supplied files, which are usually v210/mov, sometimes prores/mov, but occasionally DPX or AS-11. As mentioned, it is in the process of integrating into our FFV1/Matroska workflow.

Why do you think file validation is important?

File validation is an important step in ensuring that we are putting the best possible files forward for ingest. In terms of the material that we receive from vendors, validation of the content ensures that we are getting what we pay for, and that there are no issues that could end up being a preservation risk. In terms of file format validation, it's really important that we can verify and document (as a validation Event in PREMIS!) that we have created or received files that comply with the file format specification. It's one of the reasons that I like FFV1/Matroska so much. I know a lot less about Matroska than FFV1, but it's good to know that we can figure out if we have valid files which ultimately should have the greatest level of interoperability going into the future.

Anything else you'd like to add?

Not much, just that I wish I'd engaged with the command line interface for MediaConch sooner. It is super flexible and I wish that there were more examples of use out there.

Thanks to Kieran! Check out all of Irish Film Archive's scripts here on GitHub.]