Mediainfo, MediaConch and FFmpeg in the preservation of digital video

Source: PERICLES Blog

During late July 2016, Dave Rice and Ashley Blewer delivered two workshops at Tate Britain in London on the use of Mediainfo, MediaConch and FFmpeg in the preservation of digital video. These two workshops are a good example of the type of impact projects such as PERICLES can have and the resulting motivation they can engender, in this case leading on from the collaboration between Dave Rice and the PERICLES Project on consistent video playback. Both Dave and Ashley are moving image archivists and developers working collaboratively on MediaConch, an initiative within the PREFORMA project which would provide a sound input into the PERICLES approach of managing evolving ecosystems. Change management will grow over time and adapt to the tools available to provide the input needed to manage change. The discussions from these workshops has provided further opportunity to address the challenges faced by practitioners and developers and the knowledge garnered has resulted in valuable input back into the final research actions of the PERICLES project and is a direct extension of the Communities of Practice activity for digital video and software-based artworks.



Image caption: Tate's time-based media team and the specialists discussing a corrupted file, copyright Tate 2016.

The context for Tate's interest in the tools presented in the workshop, is in their application as workflow aids for the preservation of digital material within Tate's collection and Archive. Tate's time-based media conservation team is responsible for the care and management of a diverse range of digital audio-visual materials. These materials and objects are generated predominantly as a result of the migration of a file from tape as well as being produced directly by artists and accessioned by Tate as born digital artworks. Our general aim as the custodian of a work is to preserve the artwork in line with the artist's intent and also document the technical history of works were possible. With that in mind, one of our policies is to maintain all master materials in their original format, and ensure that any migration to a different format happens without changes to those files' significant properties. This approach aligns with the aim of the PERICLES project, to ensure that digital content remains accessible in a constantly evolving environment and where identification of significant properties is a main issue. The results of these workshops will inform the PERICLES dependency models, in particular the domain ontology.

The first workshop consisted of an internal session with Tate's Time-based Media Conservation team, focussing on specific scenarios and challenges faced by the gallery. The second workshop was open to professionals from the digital community with a

focus on introducing the tools as well as discussing broader concepts and applications. Among the participants of the open workshop were representatives of public institutions facing similar challenges in the preservation of their digital assets, such as the BBC, BFI, the British Library, the British Museum, Irish Film Institute, and the National Archive. There were also representatives of Artefactual, the Canadian-based developers of the Archivematica software, an integrated suite of open-source software tools that allows users to process digital objects from ingest to access.

Throughout the two workshops participants were able not only to see the demonstration of a group of tools but to also run some scenario-based tests, namely using MediaInfo, an open source software that displays technical metadata, as well as the related MediaConch and MediaTrace tools. Additionally, FFmpeg, Dumpster and Hex Fiend were also explored when manipulating video files.

MediaConch is an open source tool comprised of an implementation checker, policy checker, reporter and fixer developed for the preservation of audio-visual files, primarily Matroska and FFV1. However, given that MediaConch uses MediaInfo it can identify and do a basic analysis of all the media formats supported by MediaInfo. What it can't do yet is report and fix other formats. MediaConch expresses requirements as policies that can be created manually or based on files that are known to have the correct and wanted requirements.

One of the main questions addressed during the workshops was how MediaConch and the policy checker could improve our workflows. It was generally agreed that the policy checker will be helpful when migrating content from tape to file-based formats. After defining the specifications of the resulting files, a policy can be created describing them, this could then be sent to the vendor. The new files can also be checked against the policy by Tate staff to ensure consistency in the results.

Another example on how this tool can be of benefit to the team, is to ensure that relevant technical specifications are not changed in the process of creating new file formats for exhibition purposes for instance. A policy can be created to check the exhibition format against the original file, and easily compare whether for instance a colour space or the chroma subsampling is inadvertently changed.

We were also introduced to ffmprovisr and its community, which is a great resource for FFmpeg, a crucial tool for handling audio-visual files. The ffprovisr blog consists of a repository containing useful FFmpeg sample command lines and their descriptions including how they actually work. This is especially useful for new users who are less confident with the command line interface and scripting. The workshop participants were encouraged to contribute to pages like the ffmprovisr page on Github and to use its friendly user forum. This is a space where we can share questions about FFmpeg functionalities and also useful scripts to manipulate video files. These contributions help develop the open source tools by making clearer to developers the needs of the user community and, consequently making the tools more relevant for the users.

Specific video files, that had proved challenging to the team were used as case studies on the first day and we had the opportunity to test different tools and solutions, with different levels of success. Dave also demonstrated how to analyse those files, looking at their technical properties in MediaInfo, and finding what could be causing the problem.

This workshop was highly relevant for the Time-based Media Conservation team at Tate, allowing us time to rethink our workflows and to consider how the tools presented can be implemented to make that workflow simpler, faster, more accurate and less vulnerable to human error. Discussing these practices in an open forum and listening to the challenges faced by our peers in other institutes was rewarding and very valuable.