

The Trailers of Spectacle Theater

I. Introduction

The production of digital moving images continues to proliferate in both content and format. Producers of such content range from high end, big-budget film and television studios to amateur YouTube contributors with nothing more than a cell phone camera. As a result the task of preserving digital video is by necessity a multi-faceted undertaking. The actual tasks involved must be informed by and respond to the material at hand. Each preservation undertaking is in this sense a novel process, a process that must cater to the unique demands of the specific object of archival consideration.

This paper recounts the experiences of one organization in its attempts to recover digital videos that were lost due to a hard drive failure. Hard drive failure (and its attendant information loss) is by now a common experience for many at this stage in the 'digital era.' However, despite the wide acceptance of this inevitability for personal computing, it is still in my estimation a common popular assumption that digital information is somehow more durable, more secure than previous methods of information storage. Many still operate under the idea that once something is on the Internet it will be there forever. Another is the way in which many speak of the effortlessness with which perfect copies of digital objects may be made, since it is just a matter of getting down the correct sequence of zeroes and ones. These and other similar assumptions belie the fact that in many ways digital storage of information is the most fragile and unstable method of transmitting information yet devised. An excellent introductory resource for the threats to the longevity of digital objects is an online tutorial entitled "Digital Preservation Management: Implementing Short-Term Strategies for Long-Term Problems."¹ I must clarify here that by transmitting I do not mean from place to place but from time to time, that is passing down cultural production to subsequent generations. Digital communication technologies were designed to be very successful at place-to-place transmission, *not* transmission from the present to a distant, or even not-so-distant, future. As a result they are inherently very good at the former, and not so much with the latter.

It is not this paper's object to debate the pros and cons of digital vs. analog preservation, or to suggest that long-term digital preservation is not viable. It is, though, important to foreground this writing with the notion that digital archiving and preservation requires more active and vigilant supervision than precedent forms of cultural preservation. Take the example of preserving movies on film. If an organization has a new film print and has the capability to store it within the appropriate parameters for temperature and humidity, that organization can reasonably expect that film to survive

¹ <http://www.dpworkshop.org/dpm-eng/timeline/index.html>

with no unsalvageable degradation for as much as one hundred years. The only maintenance required would be paying rent and utilities for the storage location. On the other hand, the idea of leaving a hard drive on a shelf for more than 5 years without intervention would make a ‘digital archivist’ fairly nervous. Preservation of digital content is not wholly different from previous types of preservation activities, but it does exist on a much smaller timetable. All this is to say that any organization producing digital content needs to consider how long in the future it will need or want to access that content. If the answer exceeds a few years, then steps need to be taken. These steps need not be excessively expensive; again this will depend on the nature of the specific organization and the specific content.

The case at hand is Spectacle Theater and their collection of movie trailers. Spectacle Theater is a micro cinema located in the Williamsburg neighborhood of Brooklyn, NY. It seats 25 – 30 people. In operation since 2010 out of a converted bodega, their “programming encompasses overlooked works, offbeat gems, contemporary art, radical polemics, live performance, and more.”² From its beginning the people behind Spectacle have edited their own trailers for the films they screen. These are shown at screenings as well as online. Online streaming was at first available through YouTube at first, and is now handled via Vimeo.³ They switched to using the Vimeo service in the middle of 2011.⁴ In 2012 the Spectacle computer suffered a hard drive failure. For many of the trailers this computer was the sole site of storage for the original files and these files were lost. Their Vimeo Pro account allows them to store the original files and download later, along with the lower quality versions that get streamed through the Vimeo website. So trailers that were made after the middle of 2011 were downloaded from Vimeo.

The origins of this paper lie in a previous project I was involved with working on archiving a New York City area film screening aggregator website operated by Mr. Dieringer called Screen Slate.⁵ During this project Jon began discussing the problems he and another Spectacle contributor encountered in trying to recover trailers that were lost via the Spectacle YouTube site. The issues they encountered in this endeavor and the circumstances that led to the situation seemed to exemplify many of the potential problems individuals and organizations may encounter when they are unprepared for the threats to the survival of digital content.

Earlier this year the human rights organization Witness, also based in Brooklyn, released an extremely helpful resource entitled “Activists’ Guide to Archiving Video.”⁶ This work compiles the most recent sets of best practices for preserving digital video materials. The suggestions to Spectacle moving forward and many of the observations about their evolving workflows are based on this guide and the ideas contained within.

² Spectacle Theater on Facebook, <https://www.facebook.com/SpectacleTheater/info>.

³ Spectacle Theater on Vimeo, <http://vimeo.com/spectacletheater>, and Spectacle Theater on YouTube, <http://www.youtube.com/user/SpectacleTheater>.

⁴ This and other information specific to the Spectacle trailers come from a series of conversations and e-mails between Spectacle co-operator and contributor Jon Dieringer and myself between October and December 2013.

⁵ <http://www.screenslate.com/>.

⁶ <http://archiveguide.witness.org/>

The guide points to many efforts that are beyond what is necessary for the case of Spectacle, since as an activist organization one of the ongoing concerns is the guide is ensuring the authenticity of the digital video to such an extent that it can be used as evidence in courts of law. Therefore I will refer to it frequently but this paper does not seek to summarize all of its contents.

The mention of scope underlines one of the functions of archiving and preservation that can often be overlooked, i.e. the curatorial function. While it is lovely to imagine being able to save all of the productions of culture or a set of cultures for posterity, the reality is that this is impossible. All cultural heritage organizations are forced to make decisions about what specifically they will take under their guard to bring forward through time. This fact is not inimical to the archival function but instead integral. Although the means by which current generations can decide what subsequent generations will find most interesting or informative about the present are imperfect to say the least, decisions must still be made. Therefore I will state why I believe Spectacle's collection of homemade trailers made over the four years deserve preservation.

First, the collection provides an interesting historical account of the activity of one of New York's most unique and intimate filmic venues. In addition, the trailers themselves serve as exemplars of the types of creative works that have proliferated in Internet culture. They serve as remixes of the films themselves, and at times as satirical looks at mainstream film advertisements. By repurposing the content of the films and the songs used in the trailers' soundtracks the works are emblematic of digital creativity, which is constantly and vibrantly recycling older works to create new ones. In this way they also locate themselves and the theater itself in the reshaping of how copyright works in the digital age. Rights issues will be an important topic later on as one of the sets of problems encountered in trying to reclaim video material in YouTube. So the trailers also serve to signify the cultural changes surrounding how copyright is viewed and how old works are used. In their book on how fair use of copyrighted material should be expanded, Patricia Aufderheide and Peter Jaszi note that "remixers demonstrate everyday the notion that creativity is a social phenomenon more than the individual expression of any individual or individuals."⁷ All of these together serve to make Spectacle's trailers interesting pieces in their own right.

II. Issues Facing the Collection

To discuss the issues in the collection it will be helpful to separate them into four categories; creation, storage, institutional memory, and rights management. Creation refers to the earliest workflows for editing and uploading the early trailers, and will also describe how Spectacle has improved this situation. Storage deals with how and where the files have been stored. Institutional memory describes particular information problems encountered when Spectacle had to download their videos from YouTube. Rights management in this case deals will refer to particular grey areas involved in Spectacle's practices and in general the practice of harvesting video from streaming sites.

⁷ *Reclaiming Fair Use* (Chicago: The University of Chicago Press, 2011), 21.

III. Creation

At this point approximately 30 editors have been responsible for the over 700 trailers Spectacle has in its collection.⁸ Of these there are 5 or 6 very prolific trailer editors, and this includes Jon Dieringer. In its first six months two editors were responsible for all of the trailers that were produced. With so many producers over the past 4 years, there has been a lot of diversity in the ways the files are created, which can lead to certain problems.

Dieringer points out that earlier on “a lot of people didn’t... understand video compression or transcoding for editing... they learned how to use the editing functions in Final Cut but not what to put in and spit out to get the best quality.”⁹ This led to a significant number of files with varying original quality levels. The trailers that had to be reclaimed via YouTube are from the first year and so Spectacle was put in a situation where they were saving compressed versions of files that did not start out as optimally as possible.

However, more recently some of the Spectacle contributors include professional video editors, and many of the others have learned more professional workflows for creating video for streaming services. The important difference here is that they have targeted the use of the video, streaming on Vimeo, and catered the workflow to this use. The more professional and stable version of the workflow now underway involves Final Cut Pro 7 or Adobe Premier as the editing software. For Final Cut source footage is transcoded into a ProRes file so that Final Cut can edit natively. Premier users edit natively in H.264 .mp4 files. In both cases the outputted files use the video codec H.264 with AAC audio codec for the audio information, wrapped in either .mov or .mp4. The bit rates of the outputted files range from 2500 – 5000 kbps in SD and 5000 – 10000 kbps for HD. These settings adhere to Vimeo suggestions in particular, but also would work well for YouTube.¹⁰

Now Spectacle can better predict and control the quality levels of their streaming videos. The codecs are also good choices from a preservation perspective as they are highly used and well documented, resulting in a better chance of these files being renderable in the future.

IV. Storage

Spectacle’s storage for the files can be broken into two methods; that before the hard drive crash and that after. Before the initial hard drive failure on their main computer the storage was on the computer itself, a Mac Mini, on editors’ personal computers, and on

⁸ This and all other workflow information come from an e-mail from Jon Dieringer, December 5, 2013.

⁹ Ibid.

¹⁰ Vimeo, Video Compression Guidelines, <http://vimeo.com/help/compression>, and YouTube, Advanced Encoding Settings, <https://support.google.com/youtube/answer/1722171?hl=en>.

YouTube. This is an extremely problematic setup evidenced by the fact that when the Mac Mini hard disk drive (HDD) crashed original files for trailers were lost and could not be recovered. YouTube cannot be considered a storage location because there is no way to recover uncompressed video from the service. Storage on individual editors' PCs is also a problem. Individual editors for the trailers that were lost became less involved and deleted files from their computers to save space for their own uses.

All this was realized once the information loss occurred. Now the storage method involves a Mac Mini, a separate RAID (Redundant Array of Independent Disks) enclosure of hard disks formatted as JBOD (or just a bunch of disks), and Vimeo. Vimeo can in this context be considered a backup storage option because Spectacle uses a Vimeo Pro account and the Pro account allows cloud storage of original files in addition to the compressed versions that are streamed through the site. Therefore in the present iteration Spectacle uses its central Mac Mini HDD as a site for temporary and immediate access and screening, uses the JBOD for nearline primary storage, and the Vimeo account as backup. This is only applicable to the collection of trailers. Other information located on the Mac Mini and not backed up on the JBOD, or information on the JBOD that is not on the Mac Mini, is still in a situation where hard drive failure will result in unrecoverable loss.

Formatting the RAID enclosure as JBOD as opposed to a standard RAID level means that Spectacle is currently forgoing the ability of RAID to institute levels of redundancy in its storage. JBOD means that the disks are used for the total space capacity, in this instance 6 TB. If one of the drives fails that will lead to file corruption and loss, as opposed to some RAID levels, which would allow for the possibility of recovering data in the event of one disk's failure.

V. Institutional Memory and Rights

These two issues are very different and are only grouped together here because of the specific context. For the videos that were lost and only available via YouTube, recovering the lesser quality files from the Spectacle YouTube page was only one issue. The other steps involved being able to contextualize the files once they were harvested. Not only was content lost when the hard drive failed and files were lost. Important metadata describing the files was lost as well.

As hinted to in the introduction, Spectacle because of its small size is occasionally able to operate in legal grey areas in terms of screening certain versions of films found online or making its own remixes of films, and also in creating its trailers using the films and other appropriated material such as music. Due to this, when early contributors were posting trailers online, they were reticent to include much specific information regarding the files that might lead to copyright issues. Therefore videos were posted online with lacking metadata even as simple as the title of the film to which the trailer pertained, and in some instances purposefully cryptic and incorrect titling. This was done to avoid take down of the videos by YouTube and to avoid undesirable attention to the theater and its screenings. While this set of practices I think helps my earlier argument about the location of these trailers within a negotiation of copyright recently, it became a real problem when Spectacle had no recourse but to take these versions of the files to replace what had been lost.

Stripped of their original metadata or containing purposefully misleading metadata, Jon Dieringer and another Spectacle worker were forced to do detective work on their own in order to organize the files they had downloaded. This involved seeking out the first two Spectacle editors responsible for so much of the early content that had been lost. This process is ongoing and as Jon put it in an e-mail to me, “At the moment I’m essentially the most senior person who’s most actively involved so a lot of this is in my own head. The dudes who started the place aren’t totally accessible and have really bad memories.” Due to the problems with metadata resulting from rights concern and a lack of institutional memory, Spectacle is in a situation where even though they have reclaimed most of the content that was lost, albeit in lesser quality, there is still a lot of work in trying to identify just what they have recovered. Luckily, in the interim between when videos were posted to YouTube and when Spectacle began harvesting to recover files, only a few videos had been deleted from the site. The reason these videos were deleted in every instance was due to copyright issues.

VI. YouTube Harvesting

Jon and Spencer Yeh, the other Spectacle member most involved in recovering the lost files, used Google Takeout. Google Takeout is a service developed through the Data Liberation Front wing of Google and allows for export of information for 13 Google products. YouTube became available as one of those products in September 2012. In its documentation page Takeout states that the YouTube files that are exported are the original files that were uploaded.¹¹ However Jon was skeptical about this given what they received, so it may or may not be applicable to videos downloaded after a certain point. I haven’t been able to locate definitive explanation of this history. For Takeout to be effective, the uploaded videos must be linked to the Google account using Takeout. Files are exported in an archived file.

There are other options for harvesting YouTube video along with metadata from the original files. The *Activists’ Guide to Archiving Video* recommends using the YouTube API, youtube-dl, TubeKit, and WAIL. Using the YouTube API requires programming knowledge. This API is naturally the basis for Google Takeout since the Google now owns YouTube. Unless the user is familiar with programming using APIs, this option may take some time to master. youtube-dl is a command line utility that downloads videos and has options for getting metadata. This allows for more control but also requires user familiarity with command line interfaces. TubeKit is a set of tools for building one’s own YouTube crawler based on specific needs. It requires the creation of a MySQL database to store metadata, basic overall collection metadata such as a project title and directory for storage of videos and the database, scheduling crawls, and the selection of up to 16 attributes to crawl along with the videos themselves. Attributes include title, description, duration, published date, URLs for the video and thumbnail, as well as counts for views, favorites, and comments. WAIL is a GUI application and the simplest to use. It allows for one-click harvesting of specific web pages. It compiles a number of web crawling programs, Heritrix, Wayback Machine, and warc-proxy. Prior

¹¹ Data Liberation Front, Google Takeout, <http://www.dataliberation.org/takeout-products/youtube-1>.

experience has shown that Heritrix can have problems with downloading dynamic web content such as streaming videos, so this may not be the best alternative.¹²

VII. Recommendations

Spectacle has already done a lot to improve the workflow for the creation of their trailers from a preservation perspective. However a number of things should be considered moving forward.

While the storage situation using the Mac Mini, JBOD, and Vimeo storage is adequate from the perspective of the trailers themselves, Spectacle may want to consider adding another backup source for information that is on the JBOD or Mac Mini but not backed up elsewhere. Since the JBOD formatting forgoes the redundancy offered by RAID, information on it and the Mac Mini are still susceptible to irreversible loss. Since the overall space needed to potentially store is 6 TB, the limit of the JBOD, the best option is likely the purchase of an external HDD. High quality drives offering 3 – 4 TB of storage are available between \$200 and \$300, and in this situation this would be a good place to start since at the moment the JBOD is not close to being at capacity.

Minimum sets of metadata necessary for each trailer should be developed, and there should be instructions for all the editors on what metadata to include with the files when uploading to Vimeo. This will help in the event the JBOD corrupts so there will not be another loss of significant metadata, even if the essence of the files themselves is safe. If rights issues are still of concern moving forward, resulting in hesitance about uploading certain types of metadata to Vimeo, then a separate database or document should be created to track file names and locations for trailers and the films to which they pertain. If this is necessary then the storage recommendation above is very important, so that this document is backed up and not at risk of total loss itself.

A more general inventory or catalog may be necessary, especially if Spectacle thinks the collection of trailers are items with value beyond pointing towards specific screenings. The catalog will aid in storing important metadata in a central location such as recommended above, and also allow collection monitoring and management. For instance if H.264 and AAC codecs at some point become unfavorable, Spectacle will be able to identify and prioritize what trailers need transcoding. The *Activists' Guide to Archiving Video* has an excellent section describing good cataloging practices for digital video.¹³

VIII. Conclusion

Given the current uses for the trailers Spectacle has, its current setup is fairly successful. Original files are stored in multiple locations that are separated geographically. Videos can be streamed online and are accessible at the theater itself so they can be screened in-house. When Spectacle creates other works from the trailers, for example annual best of

¹² All programs suggested in *Activists' Guide to Archiving Video*. Links to documentation pages for all can be found in webography.

¹³ *Activists' Guide to Archiving Video*, Catalog, <http://archiveguide.witness.org/catalog/introduction>

compilations, the files are accessible. As the collection continues to grow some of the recommendations above will become more important. At over 700 trailers it is a fairly large collection of content for such a small organization. If preservation of all this content is something Spectacle believes in doing, planning for what needs to be done to effectively manage the collection should begin sooner rather than later.