

Archive Creation Project Description:

VHS Conversion and Preservation

Justin A. Barnett

USC Columbia, SLIS797

Profs. Edward Blessing & Travis Wagner

VHS Conversion and Preservation

If there was going to be a single takeaway from working on this project, it is that Hofstadter's law—everything takes longer than you expect—is putting in extra hours in the library and academic communities. Of course, some of that involves the fact that my initial plans were overoptimistic, but that only explains so much. Where I had hoped to have a significant portion of a 160-video-collection digitized by now, I have... four. And two will need to be redone, based on what we have learned of the process from doing those. Also, there are now 250 videos, and some unknown amount that have migrated their way to individual professors' offices over the last dozen years since the library stopped being the collection repository. Despite all that, however, I am actually pleased with the project, and confident that we are moving in the right direction. There is just so much I have had to learn about the intervening steps.

Our Goals & Background

Having approached this project as “what material do you have that you'd like to process, but haven't had the time to do,” it has grown into something substantially larger.

Theoretically, after recording the speakers and subjects from the video labels, we could have placed those into the catalog and called it a day. However, the recognition that their current format was both dangerous in regard to their lifespan as well as relatively unusable has resulted in needing to research, determine, and sometimes develop best practices for storing these VHS recordings in a sustainable fashion. As outlined in the

original proposal, the plan was to determine the best means of digitizing and preserving the VHS recordings, enact said means, and catalogue the videos so that they can be easily found and viewed.

Magnetic media—like VHS tapes—is a tricky area to work with. It belongs to the era of early electronic storage, and has been often neglected because, like print texts that were catalogued and shelved, there was something of an expectation that it would keep until needed. Fortunately, it is not yet *too* late for that misconception to be amended, but we are approaching that point. Magnetic media has a relatively short lifespan, as compared to print materials, decaying within 20-30 years even when kept in controlled conditions. Moreover, it requires specialized equipment to access—equipment that is rapidly disappearing, as manufacturers have stopped bothering with it and existing machines break down. While VHS is not a truly “digital” format, because of these traits, it shares many of the preservation concerns of digital material, and fully transitioning it to a digital format is the best way it can be saved, even as that brings along its own problems and concerns that must be kept in mind.

Things That Have Been Done

The original plan was to begin by reviewing past work and research on digitization, which went smoothly enough. My literature review includes much of the references for this, but was not enough, in and of itself—I needed to process that information and reconcile conflicts with best practices and newer alternatives, while juggling the resources the library and Cyber Forensics Unit (CFU) had available. The experience has resulted in a

number of lessons on the difference between ideal and practical. I have attached the initial draft of our processing guide, which focuses on background, expectations, and requirements (but not, as of yet, the step-by-step procedures) as a reference. While incomplete—because we are still refining the process—areas where determinations have been made are included.

With the CFU interested in using this project as an example of processing quality and potentially offering it as a provided service, my hope had been that we would be able to pull some funding from there. Based on the research, doing this well enough to showcase it requires VHS machines with s-video output and preferably a time-base-correction system built in, which seem to be going for anywhere from \$150 to \$500 on eBay; what we *will* be doing it with is a number of Panasonic AG-1350s that the library already has, output in composite video only, and result in some graphical artifacting. Given that, our capture device, a Diamond VC500, will be perfectly adequate. The software included with it, however, was entirely *not*.

EZGrabber, a proprietary application, was included with the device, which, based on several tests and conversions, uses a highly-compressed video codec for the captured data, wrapped in an AVI container. Unfortunately, the compression is too much to be acceptable for our archival purposes, and has necessitated testing out other options; currently, the favorite is VirtualDub, an open source piece of capturing software that can record losslessly. It has proven to require substantial tweaking, however, to keep it from skipping frames when recording on the rather weak PCs we have for this project (and thus why

some of the videos will need to be recaptured, as they occasionally freeze during playback because of this error.) Because I am continuing with this project, I will be testing out Avidemux—which I have had some better luck with but may be too resource-intensive for the machines—before making the final determination.

In the Planning Stages

One of the concerns with this project is the man hours of processing so much video. Because of the way VHS digitization works, it cannot be done any faster than the video itself plays, during which time there is little that I can do in regard to the project. Metadata notation can fill some of the time, but not a substantial portion: when each recording averages 90 minutes, however, there is a lot of time to fiddle one's thumbs. I have been working with Caroline, the librarian at GSU's Lane Library, to set up a very straightforward process (in the form of the processing guide attached) for undergrad student workers who are already in the library to follow when I am unable to be on campus, as the materials cannot be taken off-site and it is difficult to justify a day of sitting there for twenty minutes of actual work setting up five videos.

We have also been working with GSU's digital repository folks to figure out what they need for us to make the access copies of these recordings available for streaming. While the repository intended it to be used that way, in practice, they currently are unsure how we would go about setting it up; we will be, as it turns out, the first project trying to use it that way, and it comes at a time when they are already working to integrate two schools' library catalogues into a single system. The end goal is to have the video

descriptions and metadata searchable via the library website with links to the streaming format included. Several practicing professors have already expressed interest in using some of the video lecture series in their classes once we have made them readily available, which is buoying.

Finally, the primary collection we are working with is a series of faculty lectures sponsored by the university and open to the community that not only ran for the twenty-five years the library has recordings, but has continued to run to this day, and we would like to be able to integrate these archived materials with the current collection.

Unfortunately when the original head of the series retired in 2002, it seems that the position has been passed around on a yearly basis, with each faculty member handling recording however they saw fit and storing them in an equally haphazard fashion; for at least the last two years, the only recordings have been on the phones of some attendees. As such, the scope has grown somewhat—I have spent a bit of time coordinating with the current head, Dr. Lisa Dusenberry about setting up authorized recordings again going forward and attempting to figure out where past ones may be located. The administration at GSU is amenable, but slow-moving when it comes to passing along contact information or finding former faculty who handled it.

Current Status

Nowhere near complete. As noted, we have 4/250 (247, actually) encoded, and only two completely finalized, putting the completion percentage at a whopping 1.6%. What we do have, however, are three separate stations set up and (mostly) ready with both the

hardware and software for processing these materials (one in the CFU, two in the library) and have almost pinned down our exact process, enabling me to finish the processing guide and share that so that we can tackle the materials much faster. Getting the processing started (and then saved at the end) on each machine takes less than five minutes per video, which can be easily done by anyone on-site once the manual is completed. With three stations, that would mean we can get through the entire collection in roughly three months if each is only started once a day. Adding in days I will be on site and cataloguing the material, where I should be able to start 4-5 per machine, it should start going much faster, shifting the hold-up from the conversion process to the cataloguing one.

Once materials are being processed and converted, I will still have the somewhat daunting process of identifying the speaker and subject, then applying appropriate metadata. So far, all the initial tapes have had their author and subject, when identified on the label, noted, along with their date of presentation, and it is being stored in an Excel sheet (also attached,) but there are a number of inconsistencies, gaps, and missing bits of information that will hopefully be addressed within the recording itself. If not, it will potentially require checking yearbooks or speaking with faculty who may be able to identify the speakers. The additional VHS tapes we found while looking for additional VHS players have not all been catalogued, even by labels, yet, and are roughly doubling the size of the project.

Reflecting

It is somewhat heartening to have discovered that, much as I underestimated the time it would take to do this, so too have the people I am working with. The CFU is optimistic about selling this procedure to administration as a value-add, but unable to finesse the budget for purchases that are (I would assume) relatively minor, comparatively. In the library, we have found more and more material that nobody knew was there, and the scope has ballooned yet again... which I find oddly comforting. I am not the only one this happens to; hallelujah.

While there is still a significant amount of work to be done, it has now officially been signed off on as an internship, and I will spend a significant portion of the summer semester focusing on the doing, now that the majority of the preparation is out of the way. Hopefully, by the end of August, the collection will either be finished, or well on its way and able to continue without further assistance from me. I would very much like to see it through, especially as it pertains to the current faculty lecture series; while the current series has a page on the school site, it has no record of previous lectures available there, and both Lisa Dusenberry and I would be happy to have it both more accessible and, possibly, regarded by the administration. If getting through this project can not only salvage some interesting archival material and establish a procedure for handling other VHS collections, but also reinvigorate a program that has been neglected in the wake of institutional shake-up, it would be incredibly rewarding.

VHS PROCESSING GUIDE

VHS Digitization: Processing Guide (ROUGH DRAFT)

As part of the consolidation of 2018, the Armstrong campus of Georgia Southern has had to begin integrating its library systems with that of the Statesboro campus, and adapt to making use of the institution's digital repository. As a part of that, documentation of existing projects that require completion led to a look at the Lane Library's VHS collection, consisting not just of films that can be found in other mediums, but also hundreds of unique materials: recordings of various faculty lectures, research interviews, and historical events. While the Lane Library does, at the time of this writing, still have several VHS players available with which to view the materials, two of the machines have been retired in the last year, and they are currently considered a minor concern when it comes to replacement and maintenance. The material included on those VHS tapes, however, is of historical importance to the university and its legacy, and given the limited lifespan of the format and associated hardware, must be addressed promptly in order to ensure that it remains accessible in the future.

The following guide is currently a description of the best-practices we have been able to ascertain for accessing, preserving, and storing not only the VHS collection currently housed in the Lane building, but also any other material of the same type that the university has interest in keeping. It is built upon a mix of best practices followed by other institutions—New York University and Stanford, in particular—but also the National Archives' recommendations, technical specifications, the support and systems available at Georgia Southern, and our own work and experience.

Expectations

As a library and archive charged with housing irreplicable materials, it is important that the preservation work we do be of the highest quality reasonably possible; in addition, as we have partnered with the Georgia Southern Cyber Forensics Unit for this project, it is important to their department that we are able to provide documentation as to the state and unaltered nature of the materials for legal and criminal justice concerns. VHS materials preserved by this project should be:

1. processed and encoded to a format that loses as little information as possible, as per standard archiving expectations;
2. stored in a format that has minimal risk of further degradation;
3. easily accessible to library patrons and university personnel;
4. identifiable as primary transfers and tamper-resistant.

We have addressed each of these concerns as thoroughly as possible within our means and budget; the decisions by which we arrived at our process can be found in the appendix. This guide produces video encoded with as little information loss as possible, as well as access versions compressed for easy transfer and streaming; each video has a hash associated with the original conversion that can be checked against to determine if it has been subsequently edited, and is stored in long-lasting mechanical hard drives kept in archival cases, with the access version made available for streaming via the digital repository.

Equipment

We recommend the following hardware and software; variations may be usable (and are somewhat inevitable in the case of the hardware), but understand that the steps may differ somewhat based on differences in equipment. Any attempt at converting and preserving these materials requires the ability to play it back; notably, the last electronics manufacturer producing VHS players ceased in 2016, which means that hardware recommended and used via this guide may no longer be available in the future. While we have not suggested a large number of alternatives, we have attempted to explain why the devices were chosen such that subsequent preservation projects can best choose a suitable replacement among their options.

VHS Players

While you may have VHS players available—and in a pinch, anything will do—certain machines will result in a much cleaner final product, with less static, banding, and artifacting in the picture. Ideally, use a player with at least four heads, supports full-frame time base correction, and has s-video output. Combined, those result in a player that is less likely to damage tapes during the conversion process, corrects for bouncing and jittering, especially on non-commercial tapes (which applies to most of our collection), and outputs a signal that is significantly less degraded than a composite (yellow) video connection.

Recommendations: JVC7600/9600 models, Panasonic AG-1980 models.

Conversion Box

For standard-definition video, almost any video capturing device with composite and s-video input, and USB output, should work. A PCI capturing card would perform better, offloading much of the work from the computer processor, but is substantially more expensive and is only recommended if you have other, higher-definition media you wish to capture via the same general process.

Recommendations: Diamond VC500, AverMedia CE310B

Storage Media

Optical media—such as DVDs—lacks the capacity to store archival-quality conversions, meaning we require larger storage. While processing your collection, working with internal or networked hard drive space may be sufficient, but expect roughly 24 gb of space required per hour of footage. External devices are a short-term solution, but not an ideal long-term home for your material due to their susceptibility to damage. We recommend using a well-rated internal hard drive, attached to a machine via an enclosure, and then stored in static-free cases when not in use.

Recommendations: Western Digital Red (hard drive); Sabrent EC-HD2B or Inatech FD2005 (HD enclosure); ProStorage hard drive storage foam or Orico 3.5” hard disk protective box case (drive storage).

Encoding Software [UNFINISHED]

Currently, we are using VirtualDub [<http://www.virtualdub.org/>] to handle the video capturing and encoding from the usb input; it is an open source application capable of recording lossless video in an AVI container format. System requirements are unclear, and it may be nonviable based on requiring too many system resources for a sustained capture.

Integrity Checking

As part of the procedure, we recommend calculating a *hash*—a unique series of letters and numbers that describes a file and changes if the file is in any way materially altered—and storing it along with your metadata, so as to be able to verify it has not been changed in the future. Assuming a Windows machine, Microsoft has a *file checksum integrity verifier* tool (the FCIV) available for free download that will generate a SHA-1 hash value for each file as you complete it. Any tool, however, will work as long as you pick a specific type of hash and stick to it.

Recommendations: FCIV, any.

Material Processing

COLLECTION IDENTIFICATION AND METADATA