



## Musings on 500 More Years of Film and Digits

By Ken Weissman

You will recall that in *AMIA Tech Review - Issue 2*, I provided details of the Library of Congress film preservation program. Building upon that article, you will also recall that one of the main missions of the Library is to preserve America's memory for future generations of Americans, with no defined end point. The goal is to have our moving images and the sounds associated with them available for our grandchildren's grandchildren and for as many generations beyond them as one can imagine – literally for hundreds and hundreds of years.

There has been a lot of research on the long-term storage of motion picture films, so the issues surrounding the deterioration of that media are pretty well defined and understood. While you'll find some people taking issue with that research, most people agree that under the proper storage conditions, a film will still be around and usable after 500 years or longer. The storage environment needed to achieve these life expectancies is relatively modest, somewhere in the neighborhood of 40 degrees Fahrenheit, 30% relative humidity. None of this is currently very controversial, so let's accept this as one of our starting points.

Further, one of the beauties of film is that the machine it takes to look at the images is amazingly simple. All you need is a light source, a transport mechanism, a lens, a screen and voilà, you can see moving pictures. Neither the technology nor the expertise to recreate it is ever likely to go away. Contrast this with hard drives or other types of digital storage. You can't look at the files in and of themselves, and you will need a lot of technology to bring the images back for viewing.

We are now at the dawn of a new era where the preservation of these images and sounds might be accomplished using digital methods that successfully capture all of the resolution inherent in the original film elements themselves. Theoretically these digital copies will then be managed by tried and true information management migration techniques that ensure that the contents will always be available even as the technology to access them continually improves and evolves.

So let's ask ourselves some questions. Perhaps the first should be: how long do we keep the film elements themselves after we have made digital masters from them? Well, it seems we first need to define the issues before we can even begin to answer that question. So what are they? Certainly the goal of keeping the images and their associated sounds safe and available for future generations is an issue. Some films are also valuable as artifacts while others aren't, so we may need to save some for an even longer period of time. For the sake of sanity in this discussion, let's put a 500 year limit on it. This is the equivalent of some 50 to 100 migrations of the digital data. We should know by then if we are doing digital right!

While maintaining our traditional photochemical workflows is important, we recognize that digital scanning for preservation purposes makes sense and so we need to plan for a future without film. As mentioned in the last article, our first foray into that world involved a specialized project that targeted the Paper Print collection. We learned a lot from that pilot project and now feel confident that digital scanning of films and other images has matured enough for it to be considered a preservation quality technology. But we remain cautious.

As our lab moves further into digital technologies, the plan (for now) is to scan the images from only a small subset of all the films that make their way into our film preservation laboratory, and then to preserve the scanned files or restore those using digital tools as needed. For an even smaller subset of those films, we will record them back to film, and put them away into cold storage at 25 degrees, 30% relative humidity where they should last practically forever (2000 years +). Our reasoning for this is pretty simple: we can make high quality film-to-film preservation masters along with access copies for viewing in our reading rooms or for use in our loan programs much faster and much cheaper than we can currently achieve using digital methods. This is likely to continue as long as motion picture laboratory raw stocks are available for purchase.

Our cautious nature aside, we recognize some of the inherent strengths of digital workflows, just as we recognize the weaknesses. In digital you don't have the degradation you get from even the most controlled processing when you go through *multiple* generations of analog film-to-film duplication. Actually you don't have to go through many film generations at all before the generation degradation becomes noticeable to even casual viewers. Clean up of dust, scratches, and even tears is pretty straightforward using digital techniques and these techniques are becoming more and more automated. With digital, stabilization of images is a snap, as is recreating tints and tones often found in silent era films.

But if we consider *staying* in the digital realm, we should ask ourselves, "What happens if we don't do film-to-film transfer as the main preservation workflow any longer? What's the impact going to be on the amount of data we create and our ability to handle it? What about our digital infrastructure?" This is when the numbers become really, really scary.

Speaking very broadly, with 4K scans of color films you wind up in the neighborhood of 128 MB per frame (including an IR channel “dust map”). Figure that a typical motion picture has about 160,000 frames, and you wind up around 24 TB per film. And that's just the raw data. Now you process it to do things like removing dust, tears, and other digital restoration work. Each of those develops additional data streams and data files. We've decided, based upon our previous experience, that it is best to save the initial scans as well as the final processed files for the long term. Now we are up to 48 TB per film. In our *nitrate* collection alone, we have well over 30,000 titles.  $48 \text{ TB} \times 30,000 = 1,440,000 \text{ TB}$  or 1.44 EB (exabytes) of data. Just in case you were wondering, this is another reason why we don't plan to scan everything!

One of our IT people was asked about the data impact if we were to digitally preserve a typical feature film. A typical scanner might have 5 to 10 terabytes of direct storage associated with it. Once you fill that, then you have to offload the media to a *Storage Area Network* or SAN for post-processing and then, once you're done with post-processing, you have to put the data into a deep archive. If you want to do anything with it later on, you have to pull it back out of that deep storage archive.

The process of pulling out a single terabyte of data from the deep archive obviously depends on the speeds of the digital infrastructure. But right now, I'm told that moving one terabyte from our robotic tape-based system to a SAN, where we could do some more processing on it, would take between 3 to 5 hours to complete -- and that's with 10 Gigabit interfaces. The numbers are staggering.

Everyone is pretty much agreed that you should migrate that data after five years to the next latest and greatest thing or you risk losing it. And of course, you want to have a backup copy. I've even been at several conferences and meetings in the last couple of years where people are saying, "No, no, no, no, you want to have at least TWO backup copies." On separate servers, separate geographic locations, the whole bit, because a single backup that you make might not be able to be restored. You want the second backup, just in case.

And you've committed to migrate it all, every five years. That's not going to be cheap.

There have been conversations recently among AMIA members weighing what it would actually cost to store a film for 500 years versus migrating data for 500 years. There are legitimate arguments on both sides, but suffice it to say that cold storage, 48 degrees and 30% humidity, costs money too. It's not necessarily simple to do that for 500 years.

There may be no way that you can actually calculate it, but I can't help but feel in my heart of hearts that the simple solution is usually the best. And film is a pretty simple solution.

We have had a very robust preservation program over the years, and we still have a long way to go. We have only been able to restore a small percentage of the films that we are preserving, and we're collecting new films all the time. We acquired a privately owned collection of films about three years ago that added almost 15 million feet of nitrate film to our collection, and we are currently negotiating for another of similar size.

In the meantime, we have our prototype project for digital archiving in place, and we are moving forward from there. But we also realize that we might need to be the last film lab standing, and we're fully prepared to be that as well!