Presented by

Chris Lacinak

4/30/2007
NARA 21st Annual Preservation Conference
Managing the Intangible:
Creating, Storing and Retrieving Digital Surrogates of Historical Materials

chris@avpreserve.com | www.avpreserve.com

This is a dynamic presentation that was modified to be represented in a static PDF document. For the narrated multi-media dynamic version of this presentation please visit www.avpreserve.com/resources.html.

Background

- President of AudioVisual Preservation Solutions
- Adjunct Professor at NYU MIAP
  - Video lab - Intro lab - Directed Internships
- Former CTO and VP of VidiPax
- Have worked with organizations of all types and collections of all shapes and sizes for many years on tackling preservation and access issues.
- Active Member of standards forming bodies including AES, SMPTE and ISO
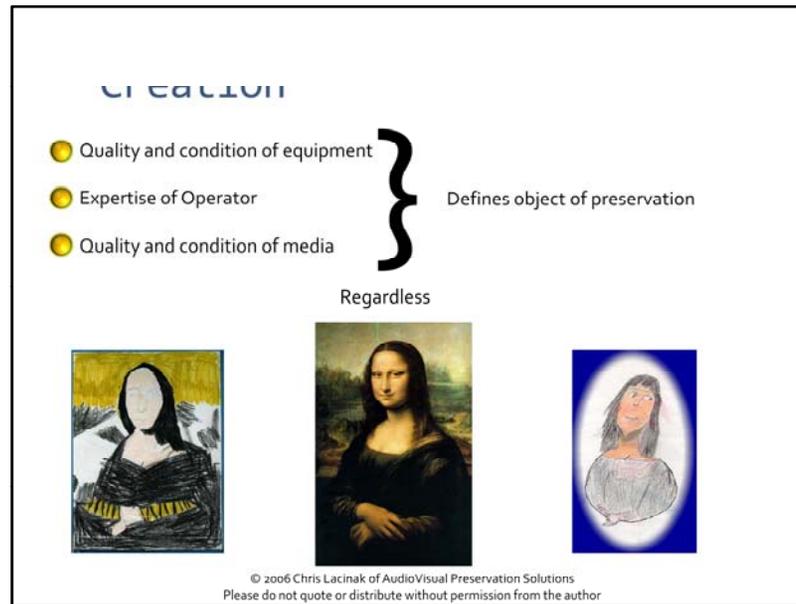- Chair of TC ARDL, and AMIA DAS

QL Pop Quiz

Due to rights associated with the content of the pop quiz it can not be included as part of this distributed presentation.

The purpose of the quiz is to demonstrate the principle that we can miss rather glaring things when we have our attention focused too narrowly or in the wrong place. My fear is that we as a field may currently be in danger of this in performing audiovisual preservation due to such a heavy focus on digital aspects. My intention today is to redirect our collective focus so that we're focused on the whole of the picture.

With audiovisual objects you can not talk about quality of the digital surrogate without talking about the system that goes into producing the object of preservation prior to digitization. Quality control in the preservation of audiovisual objects requires a holistic approach. To sum it up, we have little to no control over what we acquire with regard to quality and condition. A lack of preservation oriented standards results in even less control after we acquire the content. When you have fewer quality controls you must increase quality assurances to create an equilibrium of risk. Quality control and assurance is really about risk management. We assess risk and act accordingly. Low control equals high risk. To bring this into equilibrium we must implement increased assurances. Our reference level for equilibrium may vary, but should remain a reference nonetheless. This practice of obtaining equilibrium should be at the heart of every quality control and assurance design.

Static vs. Dynamic

vs.

=

© 2006 Chris Lacinak of AudioVisual Preservation Solutions
Please do not quote or distribute without permission from the author

The first thing we must understand is that audiovisual objects are dynamic, whereas objects such as still images or paintings are static. Here we see a painting as an object of preservation. All we need to assess the object of preservation is the light in the room and a set of decent eyes. The issues are readily accessible. In contrast, to assess an audiovisual object's condition we need a complex system consisting of, at very least a reproduction device, a piece of media, a skilled operator and a monitoring device. To further complicate the matter the relationship of the components that makeup this system and the precision with which they work in tandem directly influences the integrity of our object of preservation.

creation
- Quality and condition of equipment
- Expertise of Operator
- Quality and condition of media

Defines object of preservation

Regardless

© 2006 Chris Lacinak of AudioVisual Preservation Solutions
Please do not quote or distribute without permission from the author

To asses any object we first have to define a base reference for integrity. This reference comes from these factors. With a painting the definition of integrity is evident and easily accessible. With audiovisual objects the complex system required for reproduction of audiovisual signals makes this a much more complex task. Without adequate quality assurance and quality control measures such as those discussed in this presentation we have, at best a lack of certainty about the object we have reproduced and, at worst total failure and a detriment to the goal of access and preservation.
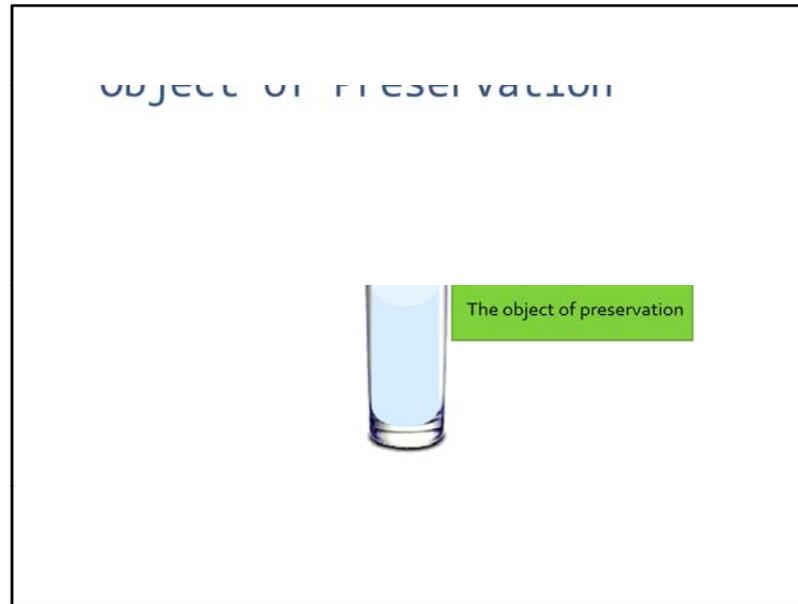
When I use the term "object of preservation" for audiovisual objects I am referring to the signal produced by the system I showed before. It is important to understand that the object of preservation is separate from the container. With this understanding the practice of audiovisual preservation shifts from a dependence on formats to a dependence on strategy.

To offer an analogy of this separation I like to use the example of a glass of water. The water is our object of preservation - an audio, video or metadata stream. The glass is our container - either a tape format or file format.

Scenario 1

We can transfer the contents of the source to a new destination.

Scenario 1

In both the physical and electronic/digital domain the container plays a primary role,

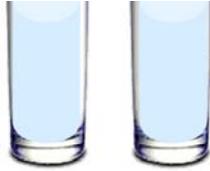but you can not judge the object of preservation solely by looking at the container,

just as you can not judge the quality of the water based on the outer appearance of the glass. Looking at a piece of media is only one indicator. It is not representative of the system required to produce our object of preservation.

Conversely, we can transfer the contents of the source to a new destination.

The object of preservation may be in good condition,

yet held in an unsuitable container. This exemplifies the shift from format to strategy in preservation.
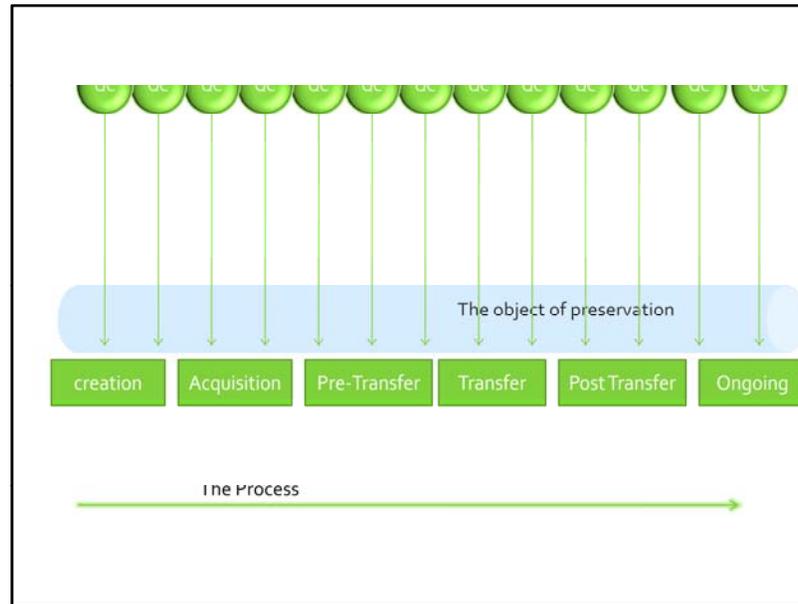
Scenario 2

The original becomes part of the preservation planning when our container fails to live up to the expectation we held of it. We can go back to the original to create a new surrogate if we have taken proper care of it.

We must think strategically to mitigate risk.
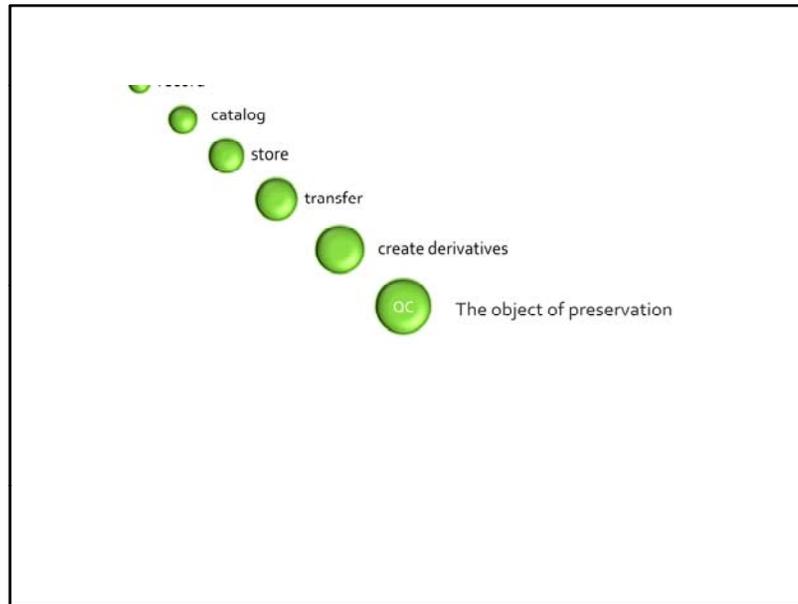
Here we see the all too often used end-of-life QC model…

turned into a more representative look at what QC should be. QC should follow the object of preservation through the processes that take place. This representation is helpful for exhibiting the "weakest link in the chain" aspect. There are hundreds, if not thousands of opportunities for failure in the reformatting process.
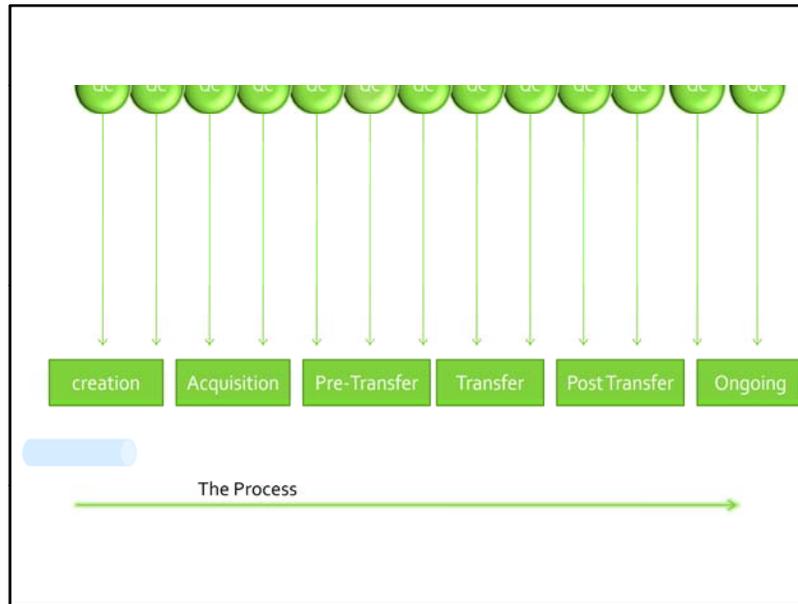
Here we see that a QC failure in the pre transfer process

The Process
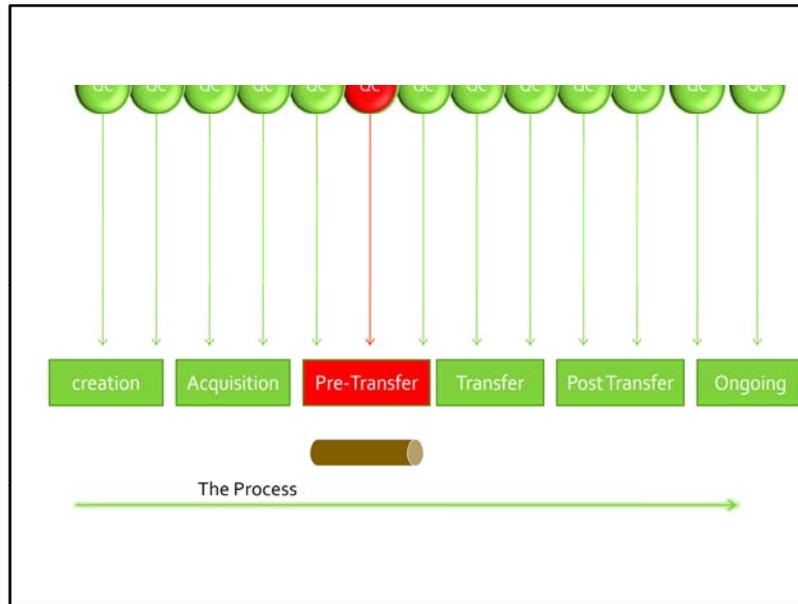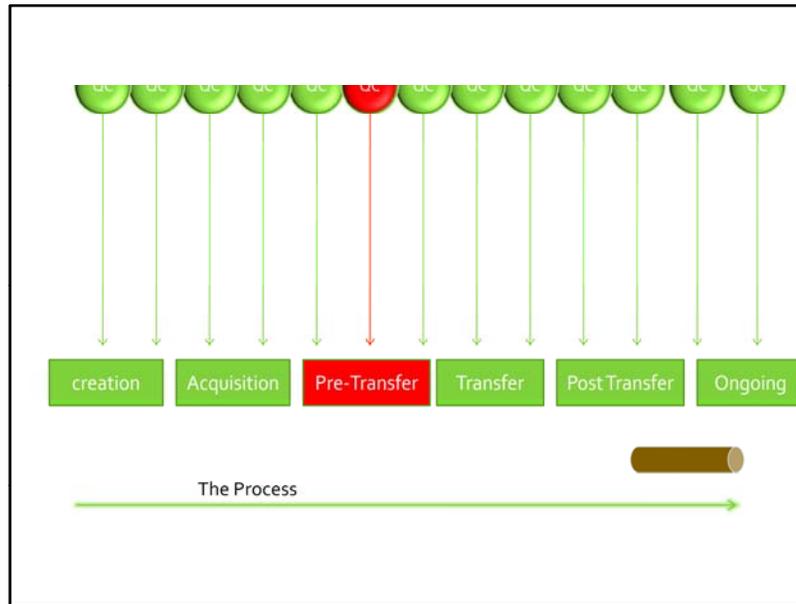
causes a detriment to the goal of preservation from that point on.

If this QC failure results in damage to the original it could be permanent. If it is an operational mistake that results in less than optimal reproduction and it is caught in the not too distant future it can be retransferred from the original.

Let's take a look at some examples, generously provided by these people, of quality failures that manifest as audiovisual errors in our object of preservation.

Here we see dropout. Depending on the media and the source of the issue the number of dropouts can be kept to a minimum and we may very well be able to attain an object of preservation with much greater integrity.
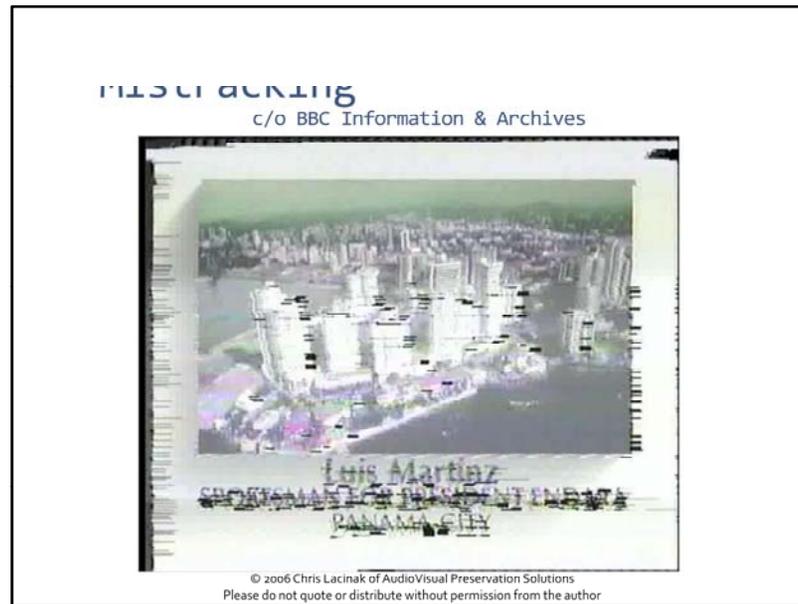
Here we see an example of mistracking of a videotape. This is most likely fixable through any number of methods ranging from use of different playback devices to alignment of the playback device, and depending on the problem may even be a media issue.

Tape Tension Error
c/o BBC Information & Archives

Tension Error Shows First at Top of Frame

© 2006 Chris Lacinak of AudioVisual Preservation Solutions
Please do not quote or distribute without permission from the author

Again, most likely resolvable through any number of remedies ranging from equipment based solutions to media based solutions.

Digital Tape Error
c/o Michael D

© 2006 Chris Lacinak of AudioVisual Preservation Solutions
Please do not quote or distribute without permission from the author

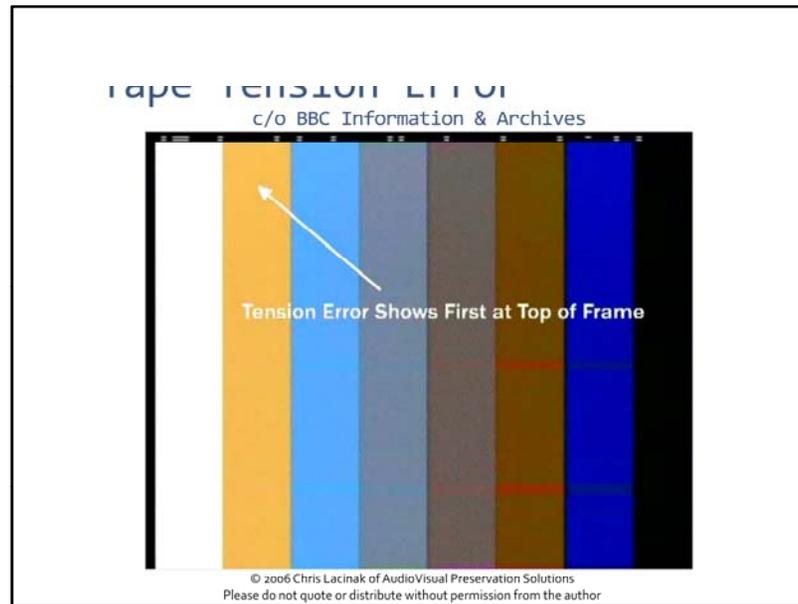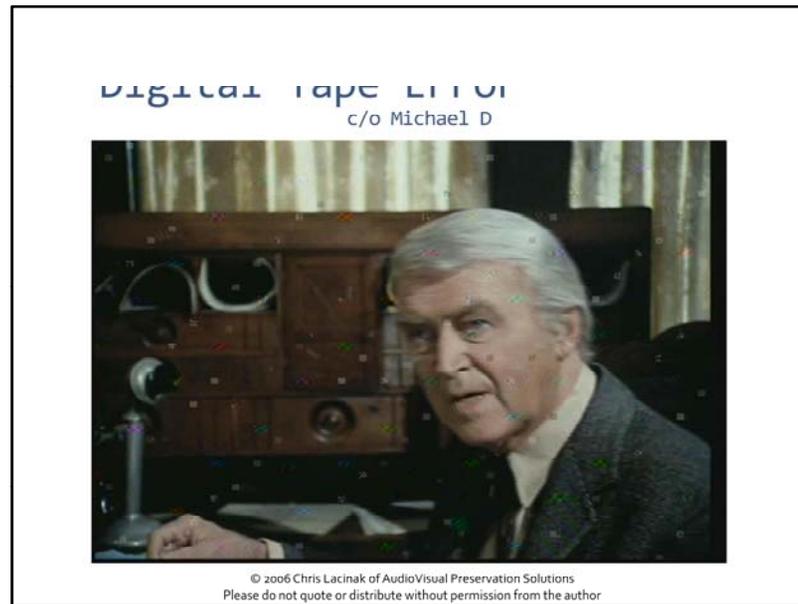Here we see a digital tape error caused by error rates exceeding the capability of the error correction mechanisms, potentially due to either alignment or media issues.

½ Open Reel Dropout
c/o BAVC

© 2006 Chris Lacinak of AudioVisual Preservation Solutions
Please do not quote or distribute without permission from the author
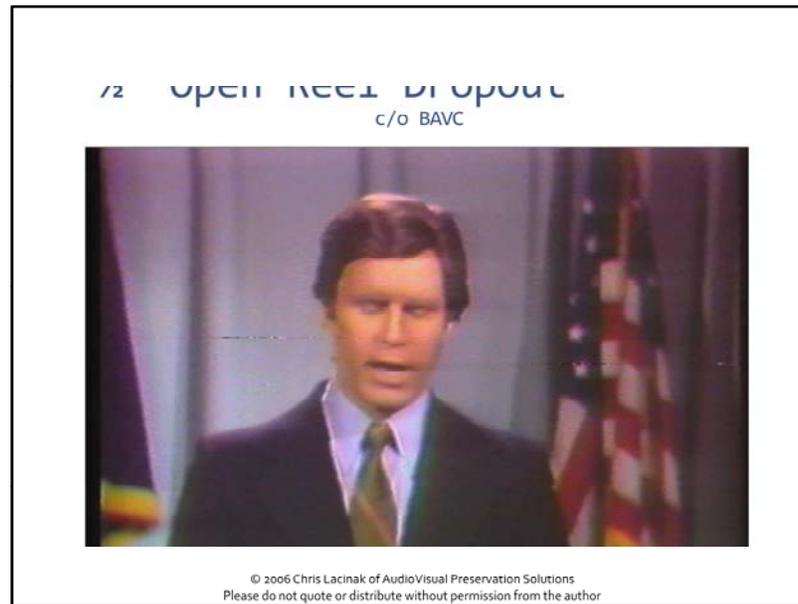
However this particular example brings up a very interesting caveat that leads us to a bit of a quandary. This is an image exhibiting ½" videotape dropout. The problem is that dropout existed even with brand new ½" videotape recordings. It was essentially part of the aesthetic. This brings up the need for an operator with a knowledge and understanding of aesthetic history that can accurately interpret visual cues based on various factors.

The quandary that this brings up is this. "How do we know that what we end up with doesn't have more artifacts than it should, if we know that artifacts are part of its aesthetic"? The truth of the matter is that there is no way for certain to say that you can. This further reinforces the need for strong QA and QC measures such as those I have listed in the handout for this presentation.
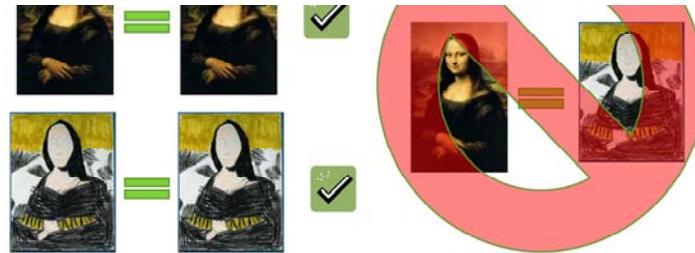
Even with this quandary what we can all agree on for certain is this.

Digital File Error
c/o Tektronix

© 2006 Chris Lacinak of AudioVisual Preservation Solutions
Please do not quote or distribute without permission from the author
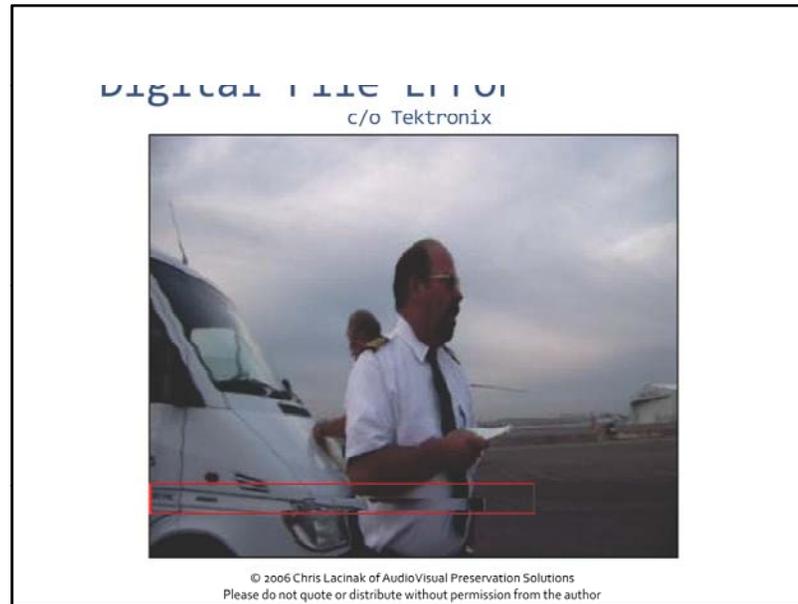
Here we see the result of a failure that took place post digitization, in the "ongoing" stage of the process. This image from Tektronix Cerify literature shows the visual result of a single bit error, which caused a number of subsequent errors in the stream of an encoded file. Some devices played this file back with this error. Others crashed when trying to play it back. This was just one incorrect bit in one video frame.
A checksum, such as an MD5 would catch this if the MD5 was created before the error occurred.

The Process

Analog Container/ Physical Format   Digital Container/ File Format

© 2006 Chris Lacinak of AudioVisual Preservation Solutions
Please do not quote or distribute without permission from the author

The ultimate risk is that lack of quality assurance and quality control leads to a quality failure prior to digitization...

creation | Acquisition | Pre-Transfer | Transfer | Post Transfer | Ongoing

The Process
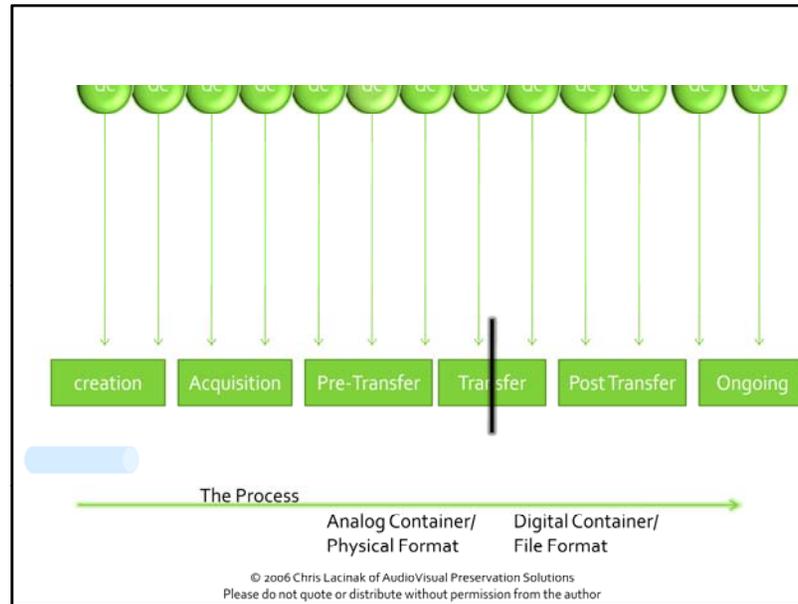
Analog Container/
Physical Format

Digital Container/
File Format

© 2006 Chris Lacinak of AudioVisual Preservation Solutions
Please do not quote or distribute without permission from the author

The Process

Analog Container/   Digital Container/
Physical Format   File Format
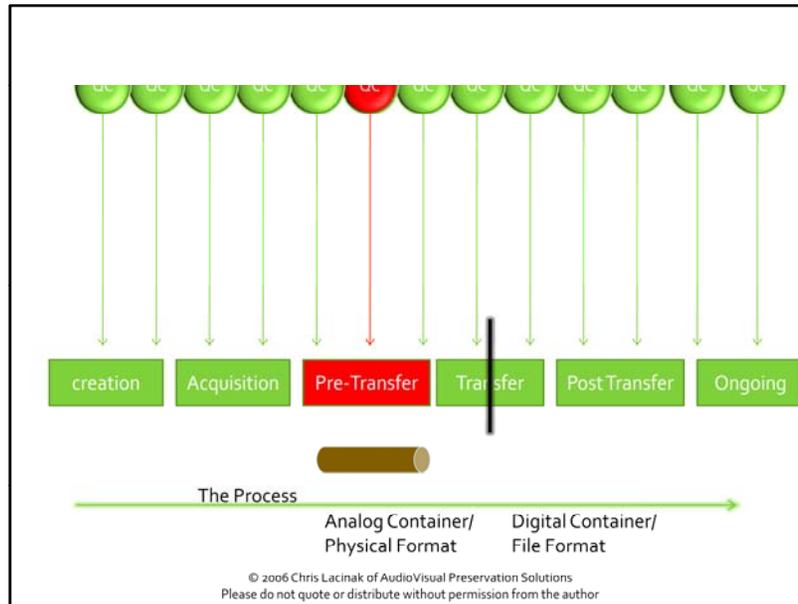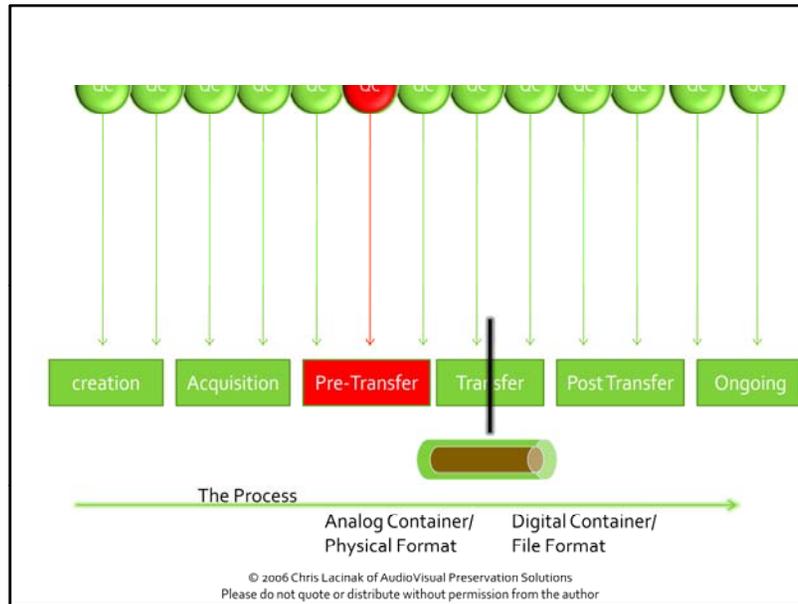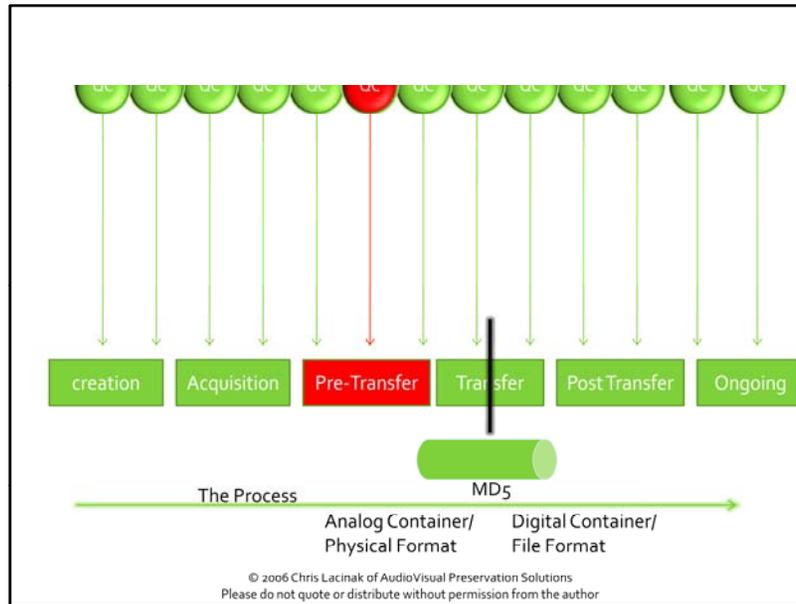
the mistake fails to be recognized and lives on through the digitization process being captured into a perfectly fine digital file format.

The Process

Analog Container/
Physical Format

Digital Container/
File Format

MD5 = MD5

© 2006 Chris Lacinak of AudioVisual Preservation Solutions
Please do not quote or distribute without permission from the author
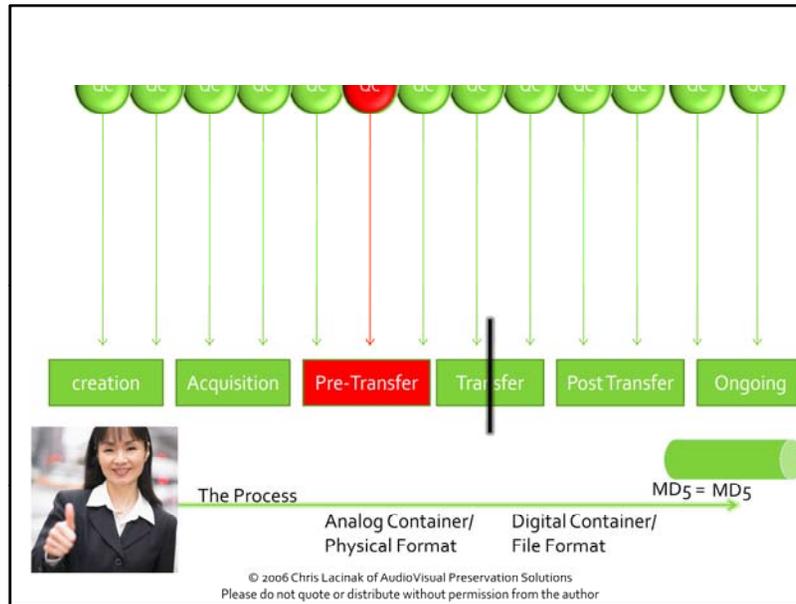
An MD5 check ensures us that nothing has changed, and the file is checked and verified as being valid.
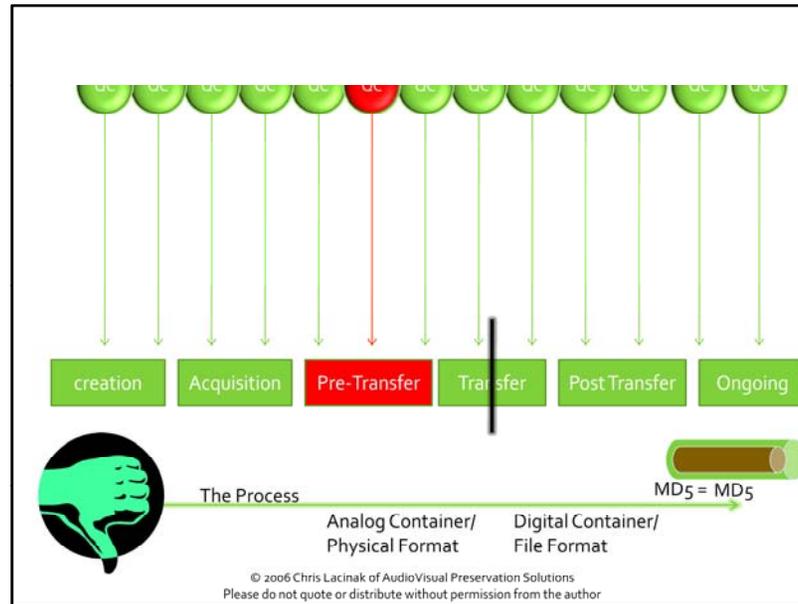
However ,when we go to access our essence "x" number of years later we find that the integrity of the object has been compromised.

Instead of looking at quality control and the object of preservation as trailing the processes as we are here

Equipment Media
Expertise
Care & Handling
Storage
Physical and intellectual control/
Collection management
All above
Diagnostic
Infrastructure Supplies
Align & Setup
QC
monitoring
Signal
Documentation tools
Transport
Analysis tools
Checksum
Validation
Accuracy
HDD monitoring
checksum
Metadata updating

© 2006 Chris Lacinak of AudioVisual Preservation Solutions
Please do not quote or distribute without permission from the author

let's take another look where QC and the object of preservation are considered central throughout the processes. This is more representative.

Post Acquisition
- Expertise
- Care & Handling
- Physical and intellectual control/ Collection management
- Storage

Pre-Transfer
- All above
- Diagnostic
- Infrastructure
- Supplies

Transfer
- Align & Setup
- monitoring
- Signal
- Documentation tools
- Transport
- Analysis tools

Post-Transfer
- Checksum
- Validation
- Accuracy

Ongoing
- HDD monitoring
- checksum
- Metadata updating

© 2006 Chris Lacinak of AudioVisual Preservation Solutions
Please do not quote or distribute without permission from the author

And gives us a more explicit view of the quality control and assurance considerations through this process.

Now we'll take a look at many of these considerations in practice throughout a sample workflow and process. It should be noted that this is a sample workflow and implementation of QA and QC practices for demonstration purposes only. This is not a specific recommendation. Every workflow is in need of its own QC/QA design and plan based on the risk equilibrium and factors.
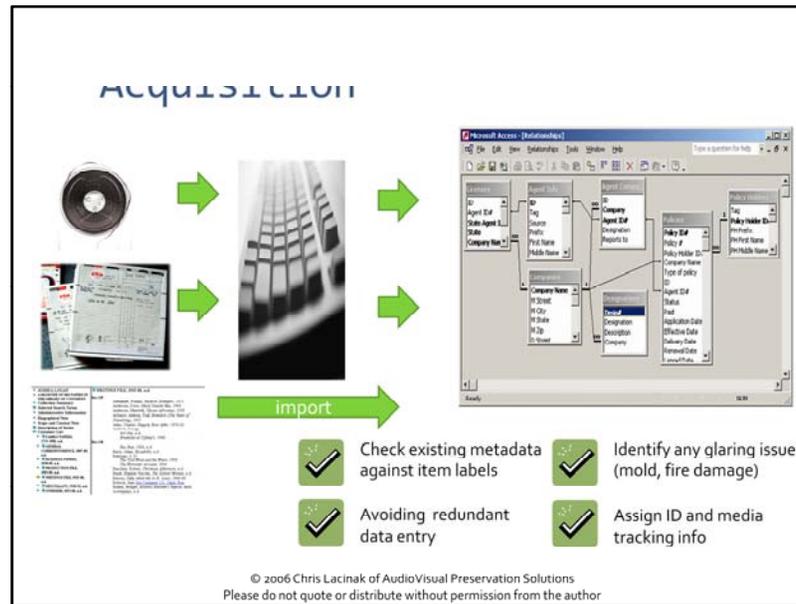
In the acquisition stage our primary concern is transcription, import and accuracy of metadata. This is significant since our aim will be to avoid redundant data entry and this metadata will be disseminated throughout the remainder of the process and the object's life. Therefore accuracy is key. Any big issues that are visually apparent will also be documented and flagged at this stage as needing expert evaluation.

This stage of the process looks at two primary aspects. One is to determine the toolset we need for proper reproduction. This is exhibited by the display of various ¼" audio open reel track layout configurations. We wouldn't want to, for instance, playback a ¼ track 2 channel tape with a ½ track 2 channel head. It gets much more complex with factors such noise reduction decoders, stylus size for a disk recording, presence of timecode, number of channels of audio on a videotape, etc…

The other goal of this stage is to identify any issues that may exist with the media that will interfere with stable transport and therefore proper signal reproduction. This would be through more in depth media evaluation and observation of physical transport.
And of course we're documenting the technical details of the object and our findings.

Treatment may be required as a result of findings from the inspection and diagnostics stage. Treatment is a fairly controversial topic which has to do with the fact that to a large extent it has been, and still is largely misunderstood by both practitioners and the general community. This is for a variety of reasons, but with projects emerging and underway over recent years we are getting closer to approaching it as a science and less as black magic. The main point is, just as with doctors, diagnosis is key. Provision of inappropriate treatment is at best a waste of time, and at worst fatal to the object of preservation.
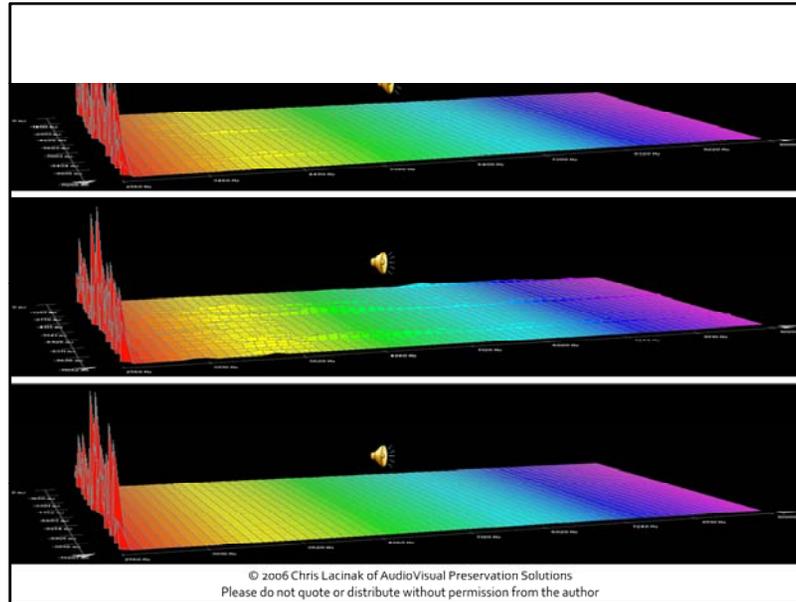
With both baking and cleaning of tapes we have seen serious misunderstanding result in serious misuse and, I fear a significant amount of loss and degradation of content. They are valid treatments for specific issues, but when used incorrectly or inappropriately they can do more harm than good. Unfortunately we have seen exemplified the words of a fellow Brooklyn, New Yorker and well know American Psychologist Abraham Maslow, who said "If the only tool you have is a hammer, you tend to see every problem as a nail."
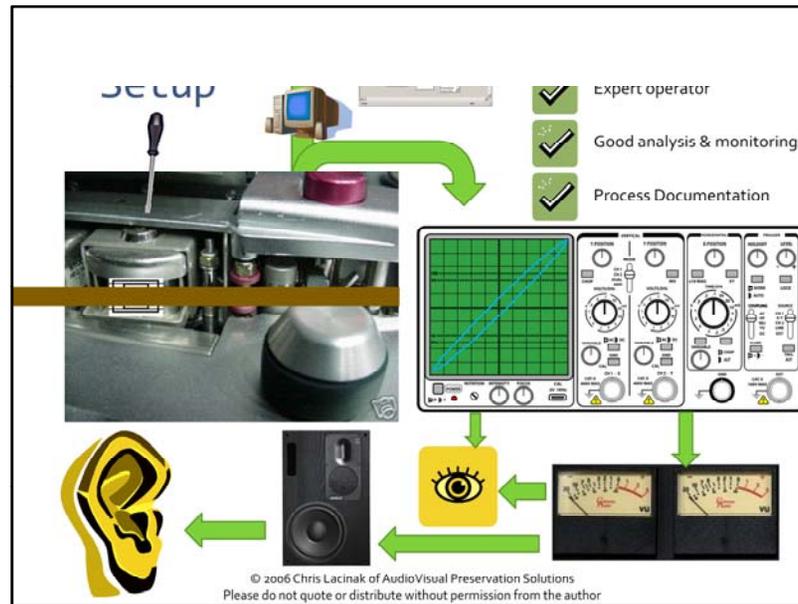
The practices involved with "setup" are to a very large degree representative of the activities I referenced earlier when I spoke of making sure that all the parts of the system are working together with a high degree of precision. The setup process is a make it-or-break it stage where we need all of the assurances that we see here. This is a also a very important step to fully document given the significance of our actions on the resulting signal.

Here we have a tape playing on a deck while we mechanically adjust the reproduction head to align with the tape so that we can achieve optimal signal output. We have automatic and manual generation of metadata being documented. We see an oscilloscope and meters being used to give us a mathematical and visual representation of the signal. We see high end equipment being used for reproduction and monitoring and anyone can tell that the golden ear and golden eye are experts.

What this slide and the transfer slide (coming up as the next process slide) leaves out, that also must be considered is the environment. This includes acoustics, lighting, electric, and more. All of which contribute to allowing the equipment and expertise to be realized. In other words if you're in a closet with all of this it's going to make it hard to do a good job still.

© 2006 Chris Lacinak of AudioVisual Preservation Solutions
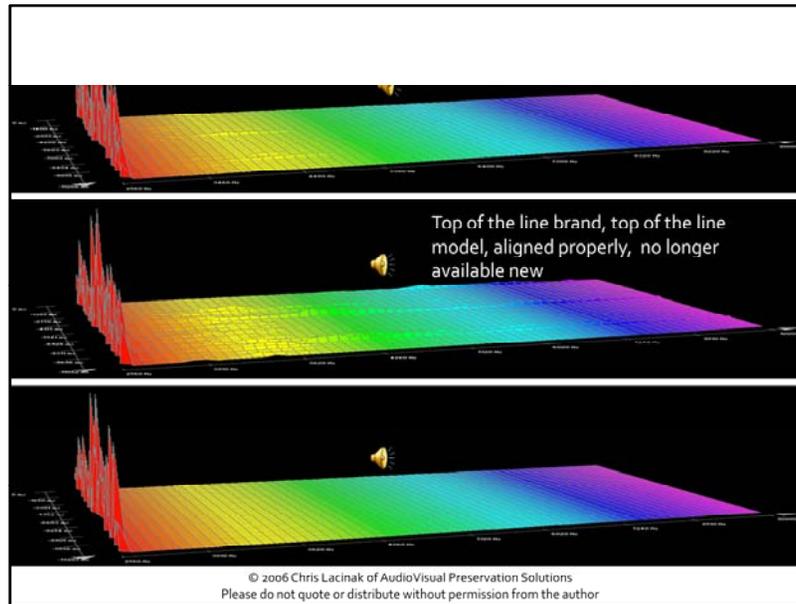Please do not quote or distribute without permission from the author

Due to rights associated with the content of this slide it can not be included as part of this distributed presentation. However the graphs are of the actual audio content and offer a visual demonstration of the point being made.

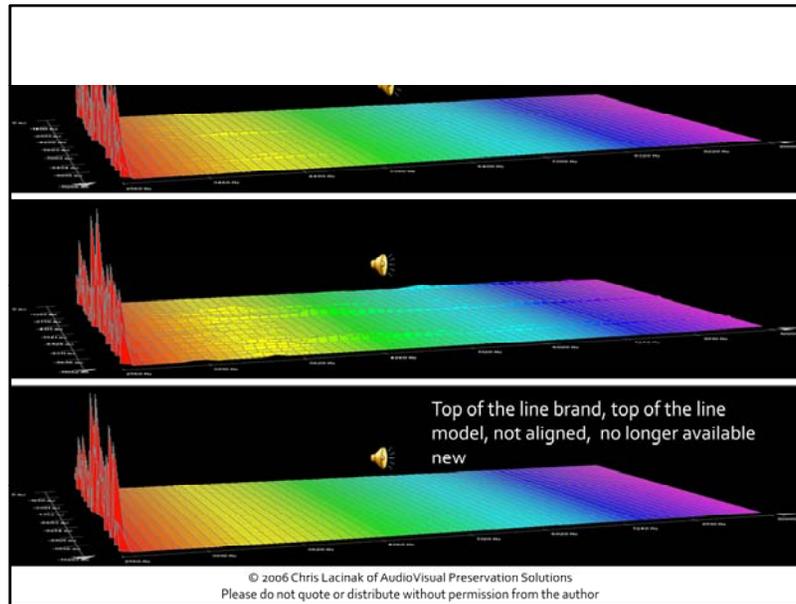This demonstration further supports a few points made so far.

1. Balancing expectations with QC/QA measures. This is a cassette. No one has very high expectations of a cassette. We all know the they're low quality and this might make us willing to accept lesser quality than we should. This points out that we should be aware of aesthetic history but not naively accepting.

2. The need for expert operators, and more than just good equipment. I know of many cases where an institution has bought what they know to be the best brand and model number of a piece of gear to start doing transfers internally without expertise or other QA measures to mitigate risk of loss, and the outcomes have been representative of the lesser quality side of what we hear here. It is not just about equipment, even if you do have great equipment.

3. The need for proper setup and following proper QA and QC measures.

Now let's take a listen. First we'll hear a brand new cassette deck that is the best current model from a company that once produced a top of the line model for preservation reformatting.

Top of the line brand, top of the line model, aligned properly, no longer available new

© 2006 Chris Lacinak of AudioVisual Preservation Solutions
Please do not quote or distribute without permission from the author

Due to rights associated with the content of this slide it can not be included as part of this distributed presentation. However the graphs are of the actual audio content and offer a visual demonstration of the point being made.

Now we'll hear the top of the line model that I referenced from that same brand. This model is no longer available as new. This deck was purchased, calibrated, aligned and brought up to spec. A huge difference is obvious

Top of the line brand, top of the line
model, not aligned, no longer available
new

© 2006 Chris Lacinak of AudioVisual Preservation Solutions
Please do not quote or distribute without permission from the author

Due to rights associated with the content of this slide it can not be included as part of this distributed presentation. However the graphs are of the actual audio content and offer a visual demonstration of the point being made.

And lastly we'll listen to the same top of the line brand and model that we just heard that is out of alignment.
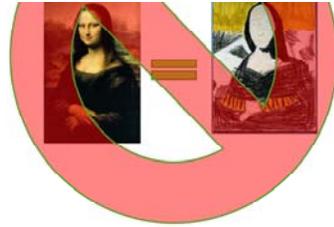
Either the first or the last of those recordings that we heard would be a clear demonstration of this scenario and a detriment to the goal of preservation and even access.
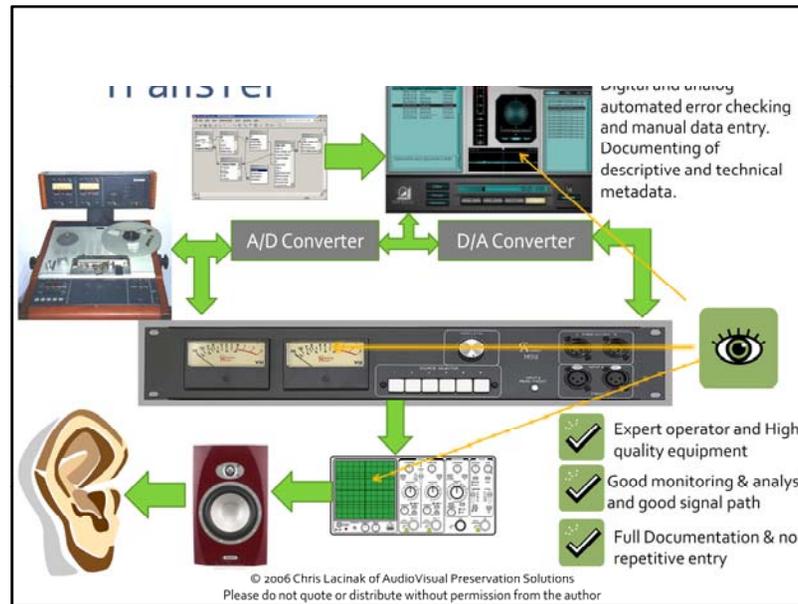
There are a few things that come up here that speak to maintaining a risk equilibrium. Once we know we have setup correct we have some latitude in the monitoring aspect during transfer. This can range anywhere from an expert only (i.e. audio or video engineer) to a less skilled operator supplemented with tools such as the automated documentation and analysis. We increase risk by lowering skill level and we decrease risk by increasing quality controls, thereby maintaining an equilibrium. Now our reference level for equilibrium can change based on factors such as the perceived value of the content, allowing us to operate with greater risk for less valuable materials and less risk for more valuable materials. Whatever the level we set should be maintained throughout by recognizing and addressing risk factors.

Notice also that we are importing information into our digitization station to embed into the file. Then we are generating all kinds of technical metadata.
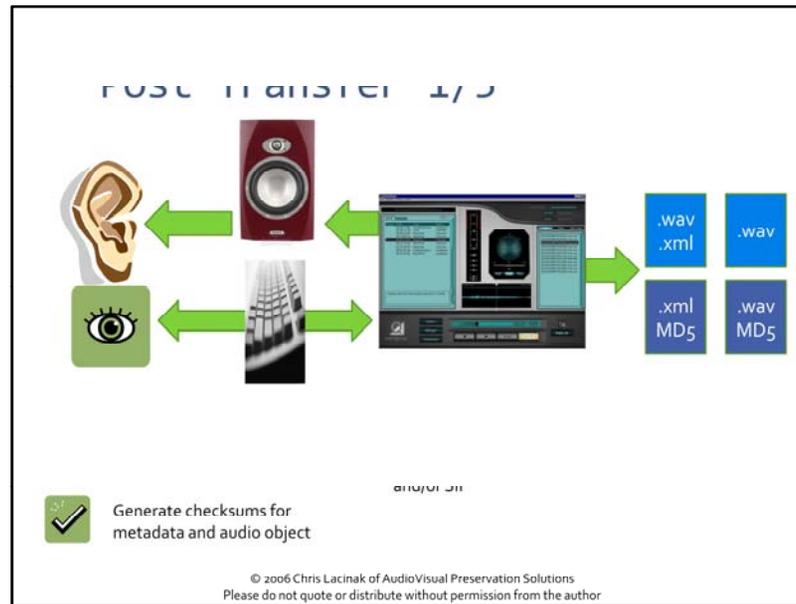
While we can lessen the skill level of the operator and offset that by adding automated analysis and reporting, it should be noted that automated metadata generation carries with it a time burden that most people fail to recognize. The reality is that automated analysis and metadata generation systems both under and over report errors. This effectively causes metadata noise and makes the resulting report less meaningful. We particularly see over and under reporting when we're dealing with content that is variable in type and in quality, typical of most archival collections. And while over reporting is a time burden under reporting is a risk. Depending on whether there is an operator present or not, and their skill level, under reporting may go unaddressed. So these lofty workflows where we see reported time savings of "x" hundred percent are generally along the lines of something more like a 30% time savings, and this time savings may very well come with increased risk. Not necessarily a bad thing if you know what you are dealing with up front, but a big surprise if you have expectations based on unrealistic workflows and time projections.

After validation and supplementation of the automatically generated metadata we will export the metadata that we have captured and generated as an XML file along with the preservation master file.

Then we generate checksums for both the digital preservation master and its associated XML document.

For those of you unfamiliar with checksums, a checksum is a computed value based on the specific blocks of data in a file. As long as the blocks of data in a file remain the same the computed value will rema in the same. If any of the data changes a new value will be generated. Therefore we can always run a check on the checksum to see if any data has changed. This should be generated as early on in the life of the file as possible and should be updated with each purposeful change of the data. The checksum should be validated on a periodic routine basis and at any point of transfer or engagement where the opportunity exists for corruption.
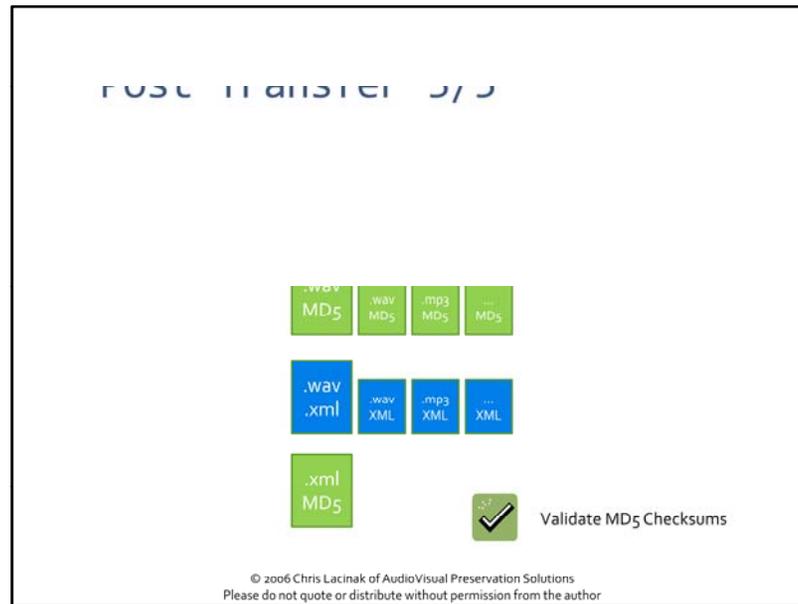
Generally checksums should be generated for any file that is considered significant and part of a preservation package.
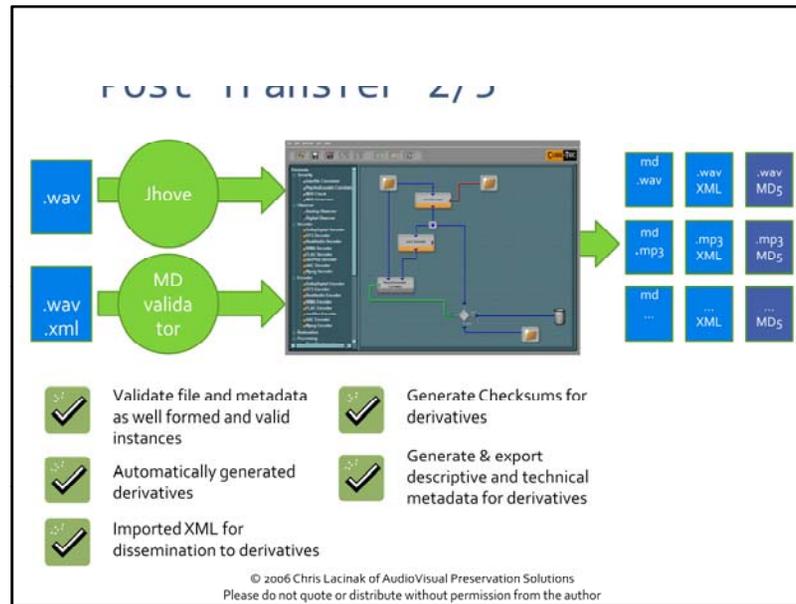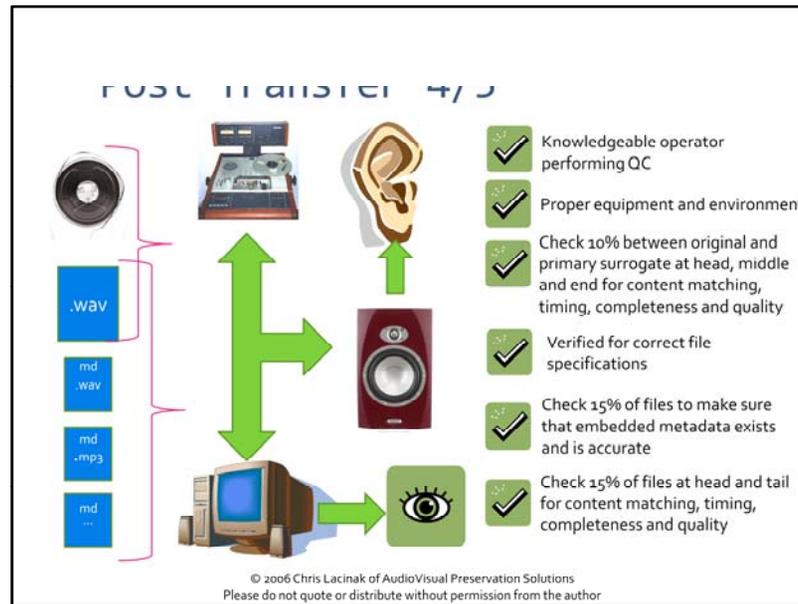
Next, before creating derivatives and relying on the preservation master and XML file we want to ensure that they are valid. This means verifying that the structure and data within the files conforms with the rules of their associated specifications. For the Broadcast Wave File we use Jhove. Jhove is an object/file format validation tool created at Harvard as an open source project available to the public. Jhove currently only supports a handful of file formats.

For metadata validation one would use or create a metadata validator specific to the schema and data dictionary used. Depending on the metadata standard or schema you use there may already be a validation tool in existence.

After the files are validated they are used to generate use and access copies via batch processing. Metadata is automatically ported from either the preservation master file or the XML file to populate the new derivatives. After the derivative files are produced they go through batch analysis to generate quality reports and checksums. The derivatives, their associated checksums and their associated XML documents are exported to their interim storage location.

Once all files are generated their checksums are validated using automated routines.

Post Transfer 4/5

- Knowledgeable operator performing QC
- Proper equipment and environment
- Check 10% between original and primary surrogate at head, middle and end for content matching, timing, completeness and quality
- Verified for correct file specifications
- Check 15% of files to make sure that embedded metadata exists and is accurate
- Check 15% of files at head and tail for content matching, timing, completeness and quality

© 2006 Chris Lacinak of AudioVisual Preservation Solutions
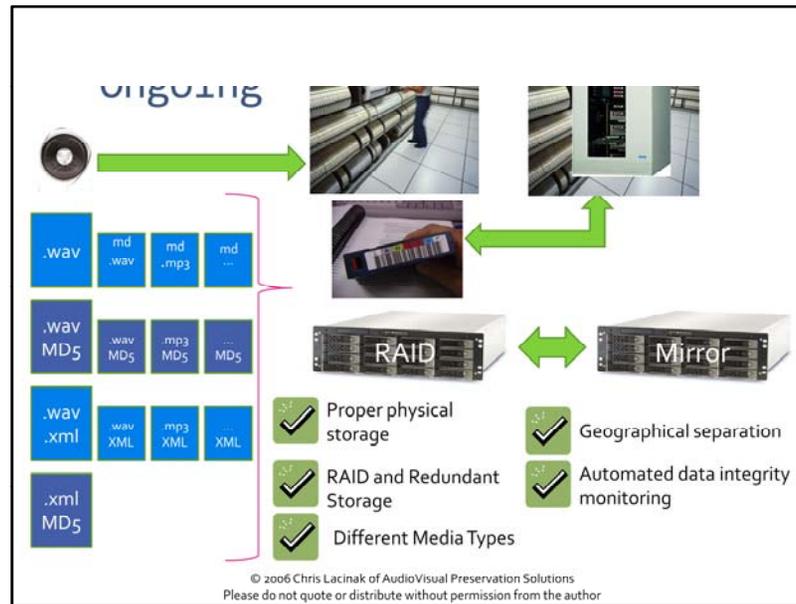Please do not quote or distribute without permission from the author

After ensuring that all files are valid there is a series of listening tests. At this point it is important to point out that the percentages listed here for quality checks are for example only. The percentages will vary with the other checks that you have in place. This sample workflow is a closely monitored process using both experts and automated quality control measures. We will also see in the next step that we run further quality checks that reduce any exposure to risk based on the lower percentages in this stage. Under different circumstances, depending on a number of factors these percentages could swing drastically up or down. I would also like to point out the danger in using percentages. Any specification should be clear in its use of percentages. If one states that 100% quality control is required it is important to state what this means exactly. This means different things to different people and leaves a lot open to interpretation. Does it mean full monitoring of the transfer with eyes on video and ears on sound throughout its entirety? Does it mean that all derivatives will be checked at the head, middle and tail or just the head? And so on. Any percentage should explicitly spell out the activities expected.

Again expertise is needed to accurately interpret what is being heard and in seen in order to make qualitative judgments. The quantitative aspects can be handled by a lesser skilled person or using automation as we see in the next slide. We still need quality equipment capable of accurately reproducing the resolution of the content and the room should not distort ones ability to make a good assessment. Too many times, QC rooms are built for quantitative analysis only and use equipment and environments that are entirely unsuitable for qualitative assessments.
We can either manually or automatically check to ensure that all files use the correct setting and client specifications.
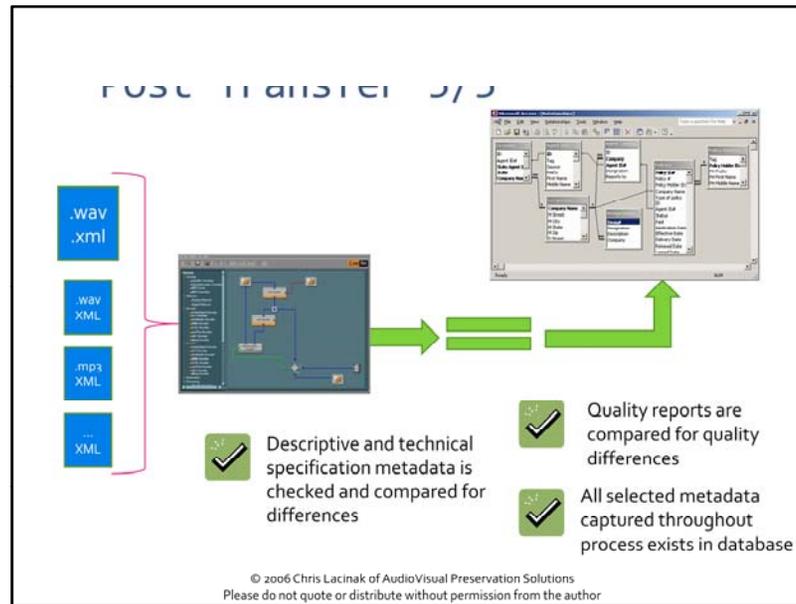
In this step we use the XML documents containing descriptive, technical and quality metadata from all of the audio files and run an automated comparative analysis. This will catch issues such as metadata mismatches within like fields of both descriptive and technical nature. It will also report quality differences that are beyond what is expected by setting thresholds for allowable variance. After the comparative analysis is performed we will import the metadata from the XML reports into the database where we now have metadata from the entire process up to this point.

© 2006 Chris Lacinak of AudioVisual Preservation Solutions
Please do not quote or distribute without permission from the author

The original should continue to be stored in a proper storage environment as part of the overarching preservation strategy. The digital files should be stored on multiple types of media. Here we are storing these files on a RAID NAS which has a mirrored RAID NAS in a geographically separate location. There is software available, or some NAS solutions come with the ability to enact a mirrored system. We also have our media backed up onto LTO tape and either stored in an LTO library or on a shelf in the proper environment depending on the quantity and resources of the organization. While it's not depicted here there should be system and data integrity routines and monitoring in place on an ongoing basis. What to do after this, Hierarchical Storage Management, Digital Repositories and Digital Asset Management systems are all topics for another presentation.

Without getting in depth into ingestion routines for digital repositories and asset management systems I will say that they generally employ validation and checksum tools upon ingest of content. The architecture of the repository will define what the submission information package (SIP) should consist of, and how the package will be managed and stored.
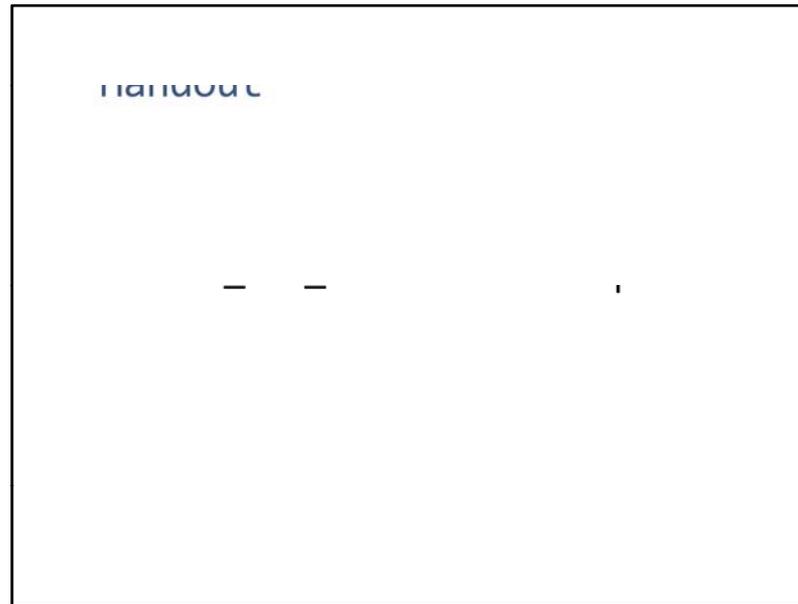
This handout embodies many aspects discussed throughout this presentation and may be found on the same page as the file that contains this presentation.

EVALUATING THE DIGITAL SURROGATE

Presented by

Chris Lacinak

chris@avpreserve.com | www.avpreserve.com

For more information regarding quality assurance, quality control and other matters pertaining to the preservation and access of audiovisual objects please visit http://www.avpreserve.com or contact us.