Reto Kromer: Thank you very much and good morning, everybody. It's a great pleasure and an even greater honor to be here today opening the workshop. To come here English is not a language in which I am comfortable but I do my best to be clear.

I would like to present to you today a field report on some digitization issues and this report is intended to be a technical one but I'm trying hard to keep all the stuff as simple as possible. I hope this will not be boring for you. I have to add that this presentation will not be comprehensive. The field is too large for it. It's a kind of powdering with all the subjects. As Brian [Graney] said please ask questions if something is not clear.

I would like to start with some technical issues, issues that come from the conservation and restoration field general speaking, not specific about audiovisual. Then I use these words conservation and restoration in the same way Mike Mashon has done it yesterday. My company tried the principles that we try to follow every time, the first one is that the probability should increase something that something is held in the future. If the probability decrease you certainly are doing something wrong. If the property remains the same then simply you lose your money.

The second point is that all options that we have today should be available to restorers in the future. That's a very important issue that in the audio visual field is often not considered but in our opinion it is essential that the future generation can go back to the element as close as possible to an original.

That is important especially if you do something dangerous. You have sometimes to do a chemical treatment that can affect the longevity of the element. It can destroy the element but it's the only way to scan an element that is in very bad shape. In this case you have to be very careful and to keep, to try to keep open all these doors and windows for the future generation at least in the digital artifact you are making.

The last thing related to that is you have to carefully document everything and by everything I mean really everything. That's the state of the original element before taking action, all kinds of restoration you have done and how and why, the state of the original element after taking action and, again, which element has been produced and why and how.

So let's go into the major part of my presentation, the digitization. It's a kind of going from an analog to a digital world. You are going from something that is continuous to something that works by steps. You have something infinite and you go to a subset in a way. Often persons that are considering only the first part, how many pixels you use but really the question is about how far these pixels and we try to focus on the quality of the pixels more than of the quantity. But the big question mark is what digital means. Often a person thinks that digital, that's part of digital but that's not only digital. This one is also digital. That's, for instance, it could be the codification of life done by God with a basis of four you have c, d, t, a. You can code the land life on this earth, that's the DNA, but we are speaking today only about this. So digital is not necessarily binary but in the IT technology it's finite.

It's difficult to know what this 1's and 0's mean without the key that allows us to understand this. The easiest parameter you can use is the quantity of pixels. The image size sometimes called resolution but resolution means many things and I try to avoid it. And about pixel size we use them as an aspect ratio of four to three the academy format, the old silent film format and the old TV format its four to three. You have these numbers of pixels. The first one is 483 as format and here you have the HD format if it's properly done inside a pin up box and 2K and 4K you have a certain number of pixels. You see that again shortly.

The important issue which I wish to explain today is about the quality of pixels. One of the parameters is the pic bit depth, as much as pixel you have, that's better, you have an image. For instance, if you use only two colors, black and white, you have this first image. If you increase up to four type of colors, half color channel, then you have 64 colors in the second image and so on. We will work a little bit with this representation eight or two power eight 256 steps color channel that gives 16 million colors you have on the computer, for instance. That's the thing that's called LGB 24. That's a lot of numbers and these numbers represent a color. It's not a color but it represents a color. I should change this slide for it so you can see why but it's not in the style of the rest.

Here you have direct channels for 1's in green and blue for zero. So you have full green, full blue and you have the mixed colors. If I go back to my slide perhaps this is information about my red channel and my green channel, my blue channel. That's the first pixel coded this way. go to the second pixel, to the third pixel and this could be the information inside the file but one of the kind of information about the quality of a pixel.

Another issue is about the color space. I go back here, it's white, it's only 1's. All of the elements of 1 but only the element 1 doesn't mean very much. I will try to illustrate that with a program. That's the curve of full colors of 280 to 680. I don't need coordinates but I wish to see the primaries, that's the green, blue and red in my presentation. This presentation is PAL and I wish to know this white point, that's the 1,1,1,1,1 we have seen. If I increase the contrast I can cancel all the others outside the triangle. So we can represent these colors from the visual spectrum coming apart. If I switch from PAL to the American format then you see this. You can notice that the elements are switching a little bit. It's going down. It's going up. But it's the same 1,1,1,1,1,1,1,1 that represents this element. It's the same 1,1,1,1 and then 0,0,0,0 that represents this. So you need to know more not only the values of the information. HD television is this space and photo that's even bigger with the spectrum we can see here. I'm going to go back and then change everything. You can see you have the blue that's outside the green and outside the field that we view as humans. That was a reaction to the second element, the color space.

I will go to the third element, the compression. If I go back to my table with the pixels and I do some maps using this RGB24 schema that's only 8-bit per color channel then I have this size per frame. Every frame in 4K I need 38 megabyte of space, of storage. If I haven't filled up this one hour long I have these figures. So it's a big, big, big quantity of space that is needed. For this reason people are trying to find ways to reduce information. One method to compress is method that usually it's not called a compression but in my opinion is part of a compression is the subsampling. If you have

an image you scan the image with the information of the red, green, and blue channel and you cut every second line on the red and every second line on the blue. Then you have an image this is 4,2,2 that you have disposed in your trash. So for information that's much smaller is usually done in TV production. TV production of works with 4,2,2 but you have to be aware if you have a good scanner to produce this image and you store this image you are cutting away one third of the information you have.

Then you have two families of compression, big families. One kind of compression is compression lossless compression and one family is the lossy compression. If we encode this image with four elements, white, light green, dark green and black, and I can encode this element with these letters that's not compressed. A very, very simple compression is to say if they have many times the same letter and put the digit and the letter together and I have a slight smaller amount of vagueness. Often the compression is done by using some information mainly by lossy, a lossless algorithm that I code. In this case you have some elements that are reduced to make a standard with the four elements. Here I have 6 metrics and they reduce the 6 by 6 metrics to a 3 by 3 metrics. I have only a quarter of the information after that. But I use some information. I don't have exactly the same information as previously. If you look at the image you can see a little bit it's not a huge loss in this case but you can see some change.

You have lossless compression and you have a family compression that is very efficient. Today you have even better algorithm, that's JPEG 2000. You can compress to this case 41% of the volume of the data we now use. The problem with JPEG 2000 is that it needs a lot of computing power so it's not easy to implement for a small archive to use this format. Often we don't advise people to use it. In this image the surface for instance is all black, real black, 0,0,0,0 and for this reason can be compressed really easily.

JPEG is often used and JPEG 2000 is often used for lossy compression. Here you have some examples of the original image in TIFF and I have a JPEG, a JPEG 2000 compression to 5% of the volume and about the same ratio for JPEG 2000. You see the difference. That's the original image and here is JPEG 2000. If you look at this child you see the original image is better, but in other parts of the image you don't see very much of the detail. This one shows some detail.

I want to say something about the image. Image can be many things as we have seen yesterday. It can be the actual image you see on the screen or the TV screen or the computer screen and that's the image used, for instance, in TV production. But, we have seen that we wish to have the last one in the archival context that's the full image with the age code and so on as information, as important information for the future.

We have spoken yesterday about 16mm film and we have often in 16mm productions we have B-Rolls. Often at the edge of the filmstrip you have information telling you how they fade in, fade out, cross face and so on has to be applied to a film. I have often seen cases in which this information was moved and operated as the copies that are really made after. So you cannot take this information absolute information. It's one person has a certain type has the idea to do this kind of effects but it's not to say that's the image, the laboratory has printed the copies this way.

I have one example in which it was on the AB Roll it was every time a fade out quickly and a fade in slowly. So you had a fade out over eight images and a fade in over 32. In the lab they decided to do everything on 4 and 24. In this case every copy people have seen everywhere represent the grading that has been made in the laboratory and not the wish it has been expressed by the cutter or by the director. So it's important to have this information but you have, if possible, compare with the real copies that have been made.

In the middle I have the image recorded by the camera. I know many cases of films that have been shot by older cameras with the academy format but the film was meant to be explored in a panoramic format. So the top, on the bottom of the image you have image but it's not image that was supposed to be screened. You can keep this image but you need the information of the right format that has been used on the screening that people have seen.

This technique to do an over scan is often used in archival world has some plus and some negative points. I wish to discuss a little bit the last one interpolation. If you decide to do an over scan it's better to do a real over scan, a big over scan. If you have 2K scan and you have some information on the age you need the image 2K production and you can create again on this part of the image the 2K that you need. You do a small interpolation and if you do a small interpolation you include in the new file big artifacts. So it's a good idea to have an over scan, for instance, in 3K as you work in 2K and not to have 2K involved in HD.

Brian Graney: Reto [Kromer], I'm also wondering about going back to the depth in compression. If you could go back a couple of slides to the TIFF and JPEG 2000. I'm just thinking about the middle point you had on the subsequent slide about the image recorded by the camera that apart from image recorded in areas of the frame by the camera that's maybe not meant for presentation. There's also in presentation optimal contrast that presents just a certain subset of what the camera records. So for example, on the far left in the uncompressed image depending on the bit depth of the scan you may be able to extract a lot more shadow detail from the figure in the center than is visible here by manipulating the image. So also considering over scanning that's a factor because when you're scanning just inside the frame you maximize the contrast with the idea that the end product is going to be just a presentation image of optimal contrast. But if you expand the scanning window then you also might have to increase the bit depth to maximize the ability to manipulate what's in the edges to see as much as possible there as well.

Reto Kromer: Yes, that's a good point. I'm not saying you have to scan in 8 bit. I am only using it because it's simple to explain. But when I come to the format I will discuss some other formats that have much more information to do work on the contrast and the colors and so on. My ideal archive you should scan the elements with one light and with format that we will see and this way you have the original codes to the original representation of the state today and you have many possibilities to work with in the future.

Then I would say something about the terminology. Today everything is called restoration but restoration is only a little part of the work. We usually use the word enhancement. It's not so sexy but it's more correct. Restoration should be based on objects, models, and things that you can explain scientifically than to do something small like an image fade out light image. This has a lot to do with the perception, I think, in future generation of the images we are producing today. Everything is kind of plastic film because it's over restored. Everything is frozen and stabilized. Everyone is looking at colors and spaces are often not correct. I hope that we will find a way inside the archives to have more elements that are more close to original intent or severity of the different color spaces, for instance, that is part of the pleasure you have seeing the films.

I have been especially upset by many of the restorations that have been done in the last years about color which the three layers have been applied exactly in the same place. One of the characteristics of Technicolor was that it wasn't possible to do that. For this reason the film has a little bit unsharp. That is a characteristic of Technicolor and especially the red was the most unsharp color. You can read literature at that time and you can read about what is called the, I can't remember exactly. It's something with…

Jan-Christopher Horak: Reto [Kromer], I don't know if you want to get into this discussion. I hear what you're saying and I totally agree that a lot of the technical restorations done in the studios look like they were done with color film or formats from the 1990's or 2000 and not from 1941. However, if you talk to the archivists in the studio, as you know, what they will say to you is, well yes, that's true. But if in 1941 they had had the technology to do it the way we do it they would have done it that way. Historical accuracy is not as important as image quality and therefore let's forget about history. Let's put out something because otherwise we're not going to be able to sell our product anymore.

Mike Mashon: And it's the dumbest argument in the history of dumb arguments.

Shola Lynch: Thank you.

Mike Mashon: It's always like, well you know, if Orson Wells had had CGI…

Shola Lynch: So let's redo.

Jacqueline Stewart: That's what they say about the Wizard of Oz. You couldn't see the wires, right, on the monkeys flying. If they could have taken that out they would have done that. But it's the charm.

Jan-Christopher Horak: The thing is the wires you don't see on a Technicolor print because it was out of focus, it's only when they sharpened it up into the digital realm. Of course, now recently they did it in 3D so you get it in every D and super sharp. The premiere was a month ago at the Egyptian, excuse me, at the Chinese.

Mike Mashon: Did you go see it?

Jan-Christopher Horak: I did see it, yes.

Allyson Nadia Field: And how was it?

Jan-Christopher Horak: It was like you would expect.

Mike Mashon: Flying monkeys coming right at you?

Jan-Christopher Horak: 3D movie. But what can I say? The little kids liked it.

Reto Kromer: I had a long discussion with Sony Pictures on the restoration of Dr. Strangelove Stanley Kubrick because on the new version they have stereo sound. The sound was mono and his argument was if they don't put the stereo sound on the DVD they can't sell it. My answer was if you put the mono sound on the DVD and tell nothing on your DVD 99% of the people will never know. Maybe 1% will notice it is happy to have the mono sound.

Shola Lynch: Did it work?

Reto Kromer: The other issue is about I try to make some comparisons with other fields of conservation and restoration in the old paintings of the 14th century perspective was not managed very well. Sometime you have a line that is not going the right direction. No restorer has the idea to correct the line. But the painter if he knew, I'm sure he would put the right line. So it's very difficult.

To the workflow I will not say much because I want to leave time for other things and discuss if you wish. I have only put one slide. In the digital world you have many things that are melting together, so you can take a clear separation between different elements. The best workflow I think is the workflow that permits to have as least as possible steps to achieve all the goals we need in the archive to work with one scan to produce the master for the preservation for using and access formats that's the best way when you have to scan everything twice. It will reduce time and money too.

The last part is about formats. Here I have a very unorthodox principle. In the past I have been more orthodox. Today I am more secular. So I think the most important things that the archive most have the possibility to work as most as possible inside the archive with the formats that are produced. There are many possibilities. All of the possibilities have pros and cons. The format I was mentioning before is openEXR. It's a format that has been created for special effects but in my opinion could be a very interesting format for archives. It works this way. It's working on 16 bit and it's working with 10 bits long. The exponent of the number is used tell to the computer where is the first one in this row. So that gives you a very large latitude of possibilities. You can scan the image in one light and if it's underexposed its on the top, it's well exposed its in the center and it's underexposed you are in this region and you have all of the latitude to work in the region in which you are. This has the advantage to document how the film is now and to permit to have all the possibilities to work with to do a good grading. I like this format because it's open. The library school work with this program and it's a good

way to keep open the doors for the future generation. If I try to have a scan with 10 bits length and I change whatever the lights to have the best result it's very hard to commit that properly for the future generation. If the future generation has the impression that this grading has been made during this scan. It's not right and they cannot correct it.

So, therefore, I think that ought to convince the technical commission. For the moment they don't agree with me but I think sometimes change is slowly. For instance, the change from a rectangular to square pixels take about 20 years. Now we cannot imagine to have on our computers not square pixels but it was a long move. Perhaps something like that could be a solution.

Jan-Christopher Horak: Aren't you taking the, this is basically a principle of analog reproduction too with film. When you are making a new negative you're scanning at one light to get the greatest possible leeway and then when you end up making new prints then you go to and do your specific color timing from image to image. You have all of the information already there.

Reto Kromer: Here you can work about with 70 spots and usually you have about 10 or more.

Even the older solution if your archive is working with Macintosh perhaps MOV would be a good solution. I would say that if you have uncompressed data the container doesn't matter really because today you have many tools to convert from one container to another really quickly.

I don't know if you wish to discuss a little bit about that or for later. That's one of the possibilities that can be done via folder in TIFF that you can use almost on every computer. But as I said, technique is changing very fast and if you wish I can play from a hard disc my laptop the PX files, not in real time but on any computer so the technology is advancing and has much more possibilities. If you go from one format to another you have five steps to consider. You have now a tool that was not made for archive but that gives you many, many possibilities to do these changes. For instance, going from a MOV container with DV content to guarantee a certain format. You can change it, filter it to the PAL European format, 2 bit H.264 and encode in MPEG4 codec very easily. If you change the container you go to multiplexing direct to the new format so you can change an uncompressed AV to an uncompressed mode and then quickly you very quickly you have changed the header and the footer of the file and all the extraneous one to one features.

So that's some ideas we can discuss more during the day.

Jacqueline Stewart: … doing about how film stops represent color and they were especially thinking about skin tones. I wonder if that figures into any of the discussions about how to handle color in digital preservation and restoration.

Reto Kromer: I think if you are in an archival world that works with , you need to have with films, with old video, old television technique you need to have a wider color space

as RGP. One of these possibilities for photo and this is because the standard RGP I have presented is made for the TV production and is made for these kinds of devices to display. If you have a film that has being shown in Kodachrome it's very hard to have the right colors and the right contrasts in this color space. It was difficult even in the analog field and the best results could achieve, in my opinion, using Fuji color and later it became Kodachrome. So I think if you are going to the RGP classical model of colors you cannot represent properly everything.

Allyson Nadia Field: I have a question to you about the ethical principles that you put out at the beginning. The second one that all the options that were there before taking an action remain available afterwards. How, and this is just out of ignorance, but how then would you deal with something like the three quarter inch tape that we talked about yesterday and Louis Belanger's work that had to be, that had to bake. Presumably then that's a one shot.

Reto Kromer: Yes, that's tricky thing and if it's not really needed I suggest you not to bake the tapes but to store the tapes in a cold environment and you have something between the polymers and the monolayers that are called the oligomers and that is, you can create oligomers and if the storage condition is cool and dry you can recreate these oligomers. Often you have enough stability to do the transfer without baking. Baking, the main goal of baking is to put away humidity from the tape and if you store it in a cool and dry environment you lose the humidity too. But you need to have time, much, much time. You need to store at about freezing level, a little bit higher than freezing temperature about six months or so. For archive it's often not possible to do. But in my life I never have seen one tape that could not be transferred this way.

Allyson Nadia Field: So it's a time decision.

Reto Kromer: Yes. The goal is to take away the humidity and by baking you do acceleration of the composition.

Allyson Nadia Field: Thank you.

Jan-Christopher Horak: We found for the quarter inch tapes we could not get a signal off no matter how often you baked.

Mike Mashon: We have, that's our most populace format. We've got 200,000 of them and we have the robots that do all of the digitization on them. Our failure rate, complete failure rate is very, very low. We've been surprised by that. We haven't tried this storing for six months at cool and dry. I might try that with a tape that we're not, I might try that experiment. That's a long time to wait. Sometimes you don't have the luxury of time.

Jan-Christopher Horak: I need to understand, you put it in…

Mike Mashon: He was saying six months cool and dry.

Jan-Christopher Horak: Cool and dry first.

Mike Mashon: Without baking.

Reto Kromer: It's because you keep out of the tape the humidity this way. In the past it was a taboo because it was a taboo to say you can store a magnetic tape cool and you can freeze even a magnetic tape but now we have scientific evidence that it's not a problem. The isonome has no big change at least.

Mike Mashon: I'm not saying that every, I will say every tape we have, every film we have is precious but that said most of the tapes that we bake is like the 200th episode of the 700 Club that was registered for copyright. I'm not laying awake at night worrying about that.

Reto Kromer: I know that baking is much easier to recommend in the workflow instead of storing six months.

Mike Mashon: And it works quite well with two-inch quad.

Reto Kromer: Often you can re-bake.

Jan-Christopher Horak: I have to ask this question because when my digital people talk to me sometimes my head just starts to explode trying to understand what they're saying. So maybe you can explain to me. You had a slide up in which you differentiated between container, codec and content. Could you run through that again and explain what the differences between container, codec and content? It's all bits to me.

Shola Lynch: It's all bits.

Doug Reside: The OpenEXR format, is there a software that interprets it that's also open? That OpenEXR format, it sounds like it's not terribly well supported format yet as far as software that can play it back.

Reto Kromer: Not directly but you have all the tools you need to write so you can interface very well with, for instance…

Doug Reside: So it's mostly archival format then like you would store it in the OpenEXR.

Reto Kromer: It's not an archival format. It's a format that has been developed for special effects.

Doug Reside: Right, but I mean that it could be used as an archival even though it's not yet. It could be used as a format to store things but not necessarily as a playback format? Is that what you are saying? Okay. I see.

Reto Kromer: Yes.

Brian Graney: While you're looking that up I just wanted to mention also going back to Ally's [Allyson Nadia Field] question about the second ethical principle. At AMIA this year there was an interesting example of how that is sidestepped as a matter of course in preservation in what was an excellent presentation by staff at George Eastman House of the recent restoration of a lost Orson Welles film, *Too Much Johnson*. The element that was discovered was a work print which had all of the editorial notations in grease pencil right on the element. So throughout the presentation they were showing these markings but those grease pencil markings were erased in the course of using the work print as a source for preserving the film as a film. Just as a matter of course you have to do that if you're going to wetgate it or run it through an ultrasonic cleaner. So they did record them just with a digital camera. But essentially they had to take an irreversible action on the preservation of the work print in order to preserve the film as content.

Reto Kromer: Sometimes we have first to do things that are dangerous. We are not sure if the film will survive after that. If you do a redevelopment of a film it's a tricky thing. Sometimes you can do that and sometimes you destroy the film.

Jan-Christopher Horak: Or if there is severe shrinkage and you have to use one of these re-dimensioning chemicals. You know you're going to get one pass after that and then the film is going to completely self- destruct.

Reto Kromer: The container is the file you see on your computer that envelopes everything, sound and image and metadata and everything you wish. The codec is the path that says which the ones and the zeros are arranged for the sound, for the image, and how they are combined. If you have a streaming format for the internet you wish to have every time enough image information and enough sound information to play the element. If you have a studio formats you can read a part in the memory and play them. So you have different formats that make this possible. Then this kind of format you have to the bit stream, the zeros and the ones. You need usually one codec for the image and one codec for the sound and one container to hold this together. It's the simple explanation I hope.

Brian Graney: In the interest of time I think we should move on next to Andy's [Uhrich] presentation. Thanks Reto [Kromer].