Event ID: 1895043
Event Started: 1/10/2012 Noon PST
Please stand by for realtime captions.

I am now happy to introduce Jacob Vidal.

Good afternoon, everyone. Thank you for joining me today. I am excited to see so many people here and hope you find this webinar useful. In the first installment that we did, we did a general survey of preservation. And talked about how things are preserved, how to think about preservation and that sort of process minded way that doesn't -- that lets you get on with the work of building digital libraries while also being safe and sound and doing good risk management.

Today, we will drill deeper into text and image file formats, which for many of us is not the entirety of what we have in our digital libraries, or certainly one of the most important formats. We will talk about the specific technical issues that exist with text image file formats and talk about things we need to be aware of in preserving them for the long term.

I will remind you that the first webinar in this series was archived on the Info People website if you wish to receive that one, and this will be as well. So, let's talk about text. Digital text encodings have their roots way back in telegraph coats. So, the ASCII format that you have heard of, the American Standard code for information interchange dates from 1968. So, these are not new on the scene. Many of these things have 40 or 50 years of history behind them. A good indication that they are preserved oval.

ASCII is a seven bit code, so that means that there are seven magnetic bits assigned to each character in an ASCII text string on your computer. Seven yes or no is. The ASCII format itself has 32 control characters, things that give instruction to the computer, and 94 printable characters.

There they are. That is the US ASCII code chart. So, this would tell a computer as it read the mechanics break or -- if it newly was reading ASCII text, this would tell the computer how to code it. If you wanted to know how to write a capital letter J and ASCII, you will program your computer to write 1010100, and that gives you a letter J, on and on and on.

This covers the basics of what you would need to communicate, as well as the only language that you wanted to use was English, and you do not want to get to elaborate with any characterization.

That falls short of what most libraries and archives need to do, obviously. So, the common format for libraries and archives are doing digitalization project is Unicode. And that is stored in a format called UTF8. UTF8 and ASCII is equivalent in the way that coats UTF-8. It could stand for Unicode transmission format, eight bit, just as ASCII was a seven bit format, UTF-8 is an 8-bit format. It has an additional bit. And in fact, it can have a number of octets. So, it can have up to four of those 8-bit bytes that make up a single character in Unicode.

The first 128 of those are the ASCII chart that we just saw. So, ASCII is tremendously forward compatible. So, ASCII is a very safe and simple format, but doesn't have a lot of range. UTF-8 expands on that by adding characters for a righty of different scripts, mathematical notation, music notation, there is a vast range of material that can be represented in Unicode and UTF-8.

It also has the advantage in digital preservation and being easy to identify. If you took part in the last webinar, we talked about decipherment and decryption, and we use an example of linear B and ancient language. One of the things that helped that was being able to identify the counting systems. So people know what numbers look like. Likewise, UTF-8 has the virtue of being easy to identify. So, there are simple search patterns that one can do that will better than 99% of the time identify UTF-8 text strings.

So, they are very easily characterized, which means if there is every catastrophe and you need to recover data from a corrupted database or a set of text files, UTF-8 if you -- give you a leg up on doing the recovery. One of the reasons it is valued as a preservation format. Also, the second bullet point is that it is the native encoding for XML. Increasing the XML is the way that we store, share and transmit data to one another. HTML language that is used on websites, all across the web, is a flavor of XML, if you will. And, XML is by and large we, one database export so that they can be moved into a new database.

When we think about migrating digital data, very often, XML is the intermediary format that we go from here to go from one digital library system, to a future version of it, for instance, you are deciding to move from your current digital library infrastructure to a new infrastructure, very often you'll pass through XML away. This is very important for preservation purposes. Using a format that works best with XML.

Finally, UTF-8 provides support for multiple languages. For almost all of the materials that we have in our collection, they have different languages and scripts and there is simply no way to effectively and accurately represent those in the ASCII text format.

Just to give you a little more dramatic, some snapshots of different parts of the UTA -- UTF-8 character specification. You can see the code here for Armenian capital letters, there are dozens and dozens of languages that now have UTF-8 representation, as well as the UTF-8 encoding for technical [Indiscernible] and mathematical operators are a good example of that.

UTF-8 is a character that provides a lot of room, and unless you're doing some very exotic dishes taste in -- digitization, powerful UTF-8 have all of the characters that you'll need. This is where the digital library world is that in terms of representing textual data.

Those Unicode character sets that you just saw were of course an image. And this is an important piece to remember, and an important point in our transition here. Computers don't read, they encode and decode. So, if you ask your computer to find the word preservation in a PDF file, it is really just looking for matching strings. So, a text string. In that text string is resolved to a series of ones and zeros. So, when we talk about text in the digital library world, we are really talking about the encoded text.

For we think about something like a digitized book, we are talking about a page image. Something that captures the visual presentation of that text, as well as the text transcription, and the metadata that holds all that together. So, this is an important distinction to make. As many of you have been involved in scanning projects, you know that step one is the easiest step on which is capturing an image of the page. The more complex work and things that create a robust digital archive is transcribing that image so that there is life, searchable text behind it. That is done through optical character technician programs, simply transcribing and typing text again.

The marriage of those two is really what creates a robust jewel object. -- Digital object. I want to pause for now because we're going to spend the next part of this and really the majority talking about how to create digital images. And I'm hoping people are basically familiar with these text concepts. I want to take a minute or two to see if there are questions that you want to address before we go on. We will also of course have time at the end of the webinar to take questions about the entire presentation.

Okay, good. Let's talk about what makes a good digital image. This is the format you have probably heard about most and will continue to hear about digital library circles. The TIFF format stands for tagged image file format, by a company called Albis. So even though this is relatively new, the TIFF image has been around better than 25 years now. It is in version 6, that version was published in 1992. And it has no IT restrictions on it. There is a legalistic point that we make here, that Adobe Inc. owns the file TIFF format, that they don't charge any licensing fees or restrict you should -- usage of it. It is an open, published -- I want to note that although we shorthand in library circles as TIFF a power uncompressed, open and viable archival format, TIFF specification is actually very broad.

It includes lots of allowances for compressed image data, it is important to be diligent, especially for outsourcing or contracting imaging work, to be diligent about using uncompressed TIFF. There's also a TIFF standard that allows for LVW, the guys who developed it, it is a lossless compression. That is for digital preservation. Of course, for preservation, we don't like to have any compression file formats, because there is a concern as we move forward in time, if we lose track of the information about how that compression with Donovan one point in time, being able to uncompressed that data later may be really problematic.

Whenever there is compression, there is some sort of data alteration, if not data loss. And recovering from a in the future at complexity to the digital preservation challenge. So, uncompressed TIFF is really the kind of plain vanilla standard archival format for imaging.

The other format you may be hearing things about, and maybe thinking about using in projects is JPEG 2000, which was developed by the joint photographers expert group, released in IFO standards, and again, with a no-cost license for its core component, similar to the way that Adobe owns the TIFF specification but has released it for use, JPEG is owned and maintained by JPEG the released by a no-cost file format for broad use. JPEG 2000 is what is called a wavelet-based format so it can hold several levels of compression within one format.

You can think of this as a pure metadata. So, there is an uncompressed player that tells you all of the informationin the JPEG, and then on top of that are different levels of compression and different levels of zoning.

The reason that this matters in his interesting to people is that essentially from one file format, you can serve the content in various ways. So, if you're asking for a thumbnail, the Web server can deliver just that layer of the JPEG file to a user. If the user then ask for a high-resolution web version, a larger set of data can be delivered.

And if somebody with appropriate depression after for the -- appropriate permission asks for the archival of the full-fledged version, that data can be delivered all from the same file. And so, potentially it lets you build a very elegant delivery and stored into the file. The problem is there is a sort of offering tools for JPEG 2000, it is a file format that looks excellent for purposes on paper but in practice can be difficult to work with. Image editing programs don't necessarily support the whole file format.

There aren't necessarily be told that you would want to deliver on all of the potential features that JPEG 2000 has. And also, because it is new, it is relatively common from one JPEG 2000 editing tool to damage the file created by another one. If they don't all of them in the file format inconsistent, reliable ways. There is great potential for JPEG 2000, and over the next couple of years, hopefully we will have a more robust support for it.

Hopefully, if you tune into this webinar again in five years, I will play the JPEG 2000 is a great way to go and it can provide so much to the work that you want to do in your archives. We are not there quite yet and at the present time, the recommendation still is to use TIFF images. They are the old standbys.

I also want to mention digital negative. There has been a lot of discussion about this, as digital cameras have become more robust in common. Digital negative is a format that was developed by Adobe to provide a nonproprietary way to manage what is called raw camera data. And raw data is essentially a readout of everything that the image sensor in a camera saw.

Photographers love this because it lets them take that batch of data and essentially virtually reshoot on their computer. So, if they decide that they had a few settings they wished to change, they can change those and modify the output of the digital negative. This is also, in some ways, looking good for preservation purposes. Because a little bit more closer to what the camera saw in the first place.

The problem exists is that each camera manufacturer creates and manages raw data in slightly different ways. So, the way that can in the output raw data from its sensors to the computer and the way that Nikon might do it are different. Some of you may have heard actually recently that [Indiscernible name] has gotten into some financial and legal straits, so it exit question of if but when this goes -- if Olympus goes under, will the tools that they make to support that raw data be available to us? Spivak digital --

---

Digital negative creates some problems because there are simply so many manufacturers,fingers in the pie. It is good to get one step away from that and store data in something like that TIFF format, that is not tied to any certain manufacturer. That said, for those of you who are thinking about digital photography, a digital negative may really be the thing to collect. Adobe has kept the format proprietary and it is a little closer to raw camera datathan you would get any TIFF image or in the JPEG image.

So, it may have a place, but I would argue against it foremost scanning and imaging projects for your converting from analog object to an image file. I think that TIFF remains the most viable file format.

Some other formats. The other JPEG format, not JPEG 2000, raw, which we talked about, portable network graphics. There is a file which is used to deliver file data on the web, and Photoshop document, STP files are something that we talk about in our library system. They all have some sort of problem. In the case of JPEG and PNG, portable network graphic, depression or size -- compression or size limits become problematic.

They cannot hold the full range of data that we would like an archival master. In other cases, the Photoshop document format and raw, intellectual property descriptions in manufacturer proprietary standards can become a problem. So that we may not be able to successfully decode the data in the future.

The other format, where you probably encounter images a lot as we are talking about this, is PDF. So, there are lots of PDF types, and these each have sort of varying levels of preserve ability to them. PDF is currently at version 1.7, and simplistically, a simple way to think about PDF is as a metadata wrapper for text and graphics content. So, a PDF file can contain almost any media.

You can have audio, video, text, forms that you input data into, and it can store both raster and vector graphics. Raster graphics are the type of image files that we are used to thinking about in our line of work, a grid of colored dots, a grid of pixels that present an image. Vector graphics are enough medical description of curves and hills. They are used to essentially paint electronically. The virtue of those is that they can be resized in these killed without losing fidelity. And if you work with architectural materials or digital illustrations, vector graphics are common in those fields.

So, for people who are doing illustration or animation digitally work and vector graphics. PDF version 1.4 has an offshoot called PDF/ A, forarchival, that is used for archiving. This is a subset of the PDF specification. It is more restrictive about the types of content that can be included and the types of data structures that can be in the PDF.

PDF /A is in fact very preserve oval, and it has been -- preserve oval and has been nice to see it adopted, current versions of office, for instance, allow you to directly go to a PDF/A file. Microsoft likes to make a PDF file as complex as they possibly could have. But, PDF does have a rolling preservation, and they certainly don't discourage its use.

By, be aware that -- but, be aware that PDF versions are not all the same, and the way that one tool creates a PDF in another tool creates a PDF can lead to substantially different output. But, there is a space in there. What you're looking for in terms of a very preserve oval PDF is PDF version 1.4 PDF/A format.

Having talked about the basic text and format used for images, I think it is good to talk about what actually goes into that image itself. We talk about images and future webinars we will talk about audio data and video data as well.

In terms of resolution and bitmap. Resolution is essentially a measure of how frequently this scanning board digitization occurred, and assembles the source material. In case of imaging, that means how many samples per physical image in the case of a sound recording of samples of second -- a second of time elapses.

For digital images, 300 dpi is our bare minimum records -- recommendation. 600 dpi has become very standard, and get some special services, people are doing imaging at a higher level of 1200 pixels per inch or even more. These numbers emerge from some studies about the way that the human eye resolved images.

---

300 pixels per inch is about the level at which the human eye can't distinguish individual pixels. Your mileage may vary. I probably can't distinguish more than 200 dpi, myself. You might do better than I will. Also, the distance at which you are reviewing something makes it different.

---

600 dpi is a standard that is emerging out of best practice that allows essentially some overhead to ensure that there is plenty of data there to work with before legibility becomes a problem.

The other aspect of this is colored. If you have 300 dots in each inch, you have to color each one of those dots to make an image. This is done in computers by including red, green, and blue values. This is the RGB color that you hear about. And in a bit, right, which if you go back to what we were talking about with text files, this is just 80 scan the ones which allows you to represent 56 shades.

So, in a 24-bit image, you have three 8-bit channels, 8-bit for red, 8-bit four green, and 8-bit for red -- blue. It provides options for each color, and three times options for each color, and 3x256 gives you about 15 million color combinations. It's which is a lot. It is enough to cover the majority of imaging we need to do.

Getting this data is limited in scanners by the number of sensors and the scanners array. So, it has a parkway that moves across the scan that, with scanners top to bottom and a motor that moves left to right. The finance of the array and if -- the motor control gives you the resolution. When looking at scanner specs, it is important to look for that hardware resolution, rather than pixels that can be added or interpolated in software.

In camera's resolution, it is limited by the physical prime, the height and width of the center -- a sensor in the camera, and the density of sensors on an imaging chip. So, when you look at the cameras, you will see a megapixel resolution which is really just telling you how many millions of pixels are available on a sensor chip.

All that said, there is still more to be known about digital color. And so, although our kind of common recommendation has been 600 dpi,24-bit colors, stored in a TIFF file, for that color to be really useful, it needs to becalibrated.. The eye, the image sensor, and the image rendering device, the monitor, projector or what have you, all have different color sensitivities.

None of these are going to be a perfect match for the source. For the actual object that you're looking at. And those resources will actually vary in their color output, depending on the kind of like that you shine on them. So, the best practice that has emerged is to calibrate every device in the imaging chain, and then not edit color on the initial capture. Something that I really discourage is, if you are in your imaging shop, imaging something and then looking on your monitor and making color corrections. Every good study that has been done shows that you only add noise to the channel, that color does not actually become any more faithful to the original capture.

In fact, the less intervention you do, the more systematic operation you perform, the better your color output will be. Having created a calibrated master, then one could create derivatives for each use case. So, when you create a Web server that is, you may want to see how it renders any browser and on very small matters. If you're going to send something to the printer, you will be able to print it on paper and see if the colors come out the way you need.

That source file should again be left alone. To give you a sense of this, I want to take a few minutes and talk about manic adaptation. This is being, you don't trust her eyes lesson. I will show you a picture in a moment of a fruit basket. You will see pretty obviously that part of it has been tended. And then I will show you a simple color image. We will look at that for 20 or 30 seconds. And without looking away. Then I will show you the original image again. I think you'll get a sense of why the imaging community is so sensitive about this question of not trusting your eyes.

So, there is an image. Just look in the center of this image, there is a little black dot there. And then I will put up a new color, if you look at that for 20 or 30 seconds. And just keep your eyes fixed in the center of this image on the black dot and in the moment, we are going to go back to that first image that we saw. Of course, what is happening right now, as you look at this is that your eyes are beginning to adjust their calibration to see the world as sort of the blue on one side and yellow on the other. And in just a few moments, when I show you that first image again, hopefully, and again, we are translating over the web, hopefully you will see a profound difference.

So hopefully, as that happens, you see the colors shift on that image. You can imagine that he spent a lot of time in a room with lots of different color and light sources. Okay, good. I got a wow in the chat. Your eyes begin to adjust. Right? Your eyes on India of what -- have an idea of what the world should look like and when your eyes begin to agree with that, they begin to crack. This leads to lots of problems in the digital imaging.

DI process is like -- but I process is done by two ways. One is hue saturation, the code in her eyes, versus the depth. So, like you to very dark blue, for instance. The other ways that your eyes process color is with the Roth, -- rods, which is luminous, how bright or dim it is. Computers and digital imaging devices don't see light that way. They break light into three color channels, red, green, and blue. It's been a fixed amount of data to each one of those colors. And then store that at digital data, ones and zeros for

---

It colors the end and it returns it in RGB digital devices, or in BM YK, blue, magenta, yellow and -- already, the way that you are I understand color in your computer does is slightly different. There are ways to overcome this.

So, our visible spectrum ranges from 390 to 715 nm. Below that is infrared or heat and above that is over Violet. Under different types of radiation, media reflects florescent ways. So, if we shine x-rays, we get about to find parts of the spectra. And those techniques are used, for instance, to be seen in these studies where we will look at say the underpainting of a growing, or the undercurrent behind a painting. By shining different types of late and different spectra at it. We can bring forward different image elements. That technology is by and large not built into her scanners digital cameras. Mostly, we just use the physical spectrum. But be aware that any object they were looking at is showing you more spectral data the more I can capture, and in some ways, then your senses can capture.

Just to give you a sense of what this all means, here is a Costco physical spectrum from right on one and two white on the other. Depending on your color sensitivity, he will perceive the edges of more less -- to give you another sense of how color can be captured, this is an outperform -- output from a video capture.

Digital images are stored as essentially black or white or grayscale images that are very high-resolution, and then to color and luminance channels that are much lower resolution, layered on top of it producing a final image. Producing again another way of sampling color data.

This is what is called the [Indiscernible] labyrinths model. That vertical column is the luminance channel that we talked about, and then sort of the color arrows are the color bands. So, this is a weighted imaging scientists will talk to each other about what an actual color is. So, a [Indiscernible] love value has a luminance from bright -- top to bottom, and that has a place on the APs kill for the color value. And then a [Indiscernible] love value can be mapped onto a particular RGB value.

Your for instance are the 16.7 RGB colors. To give you a sense of what that shows you, the large, sort of half teardrop shape are the colors that can be described in [Indiscernible]. The yellow quadrangle is what RGB computer devices display. And the blue depending goal is the print damage, the colors that can be printed using a CMYK color space. As you can see, there is a limit for all three of those things overlap. What happens in celebration and what happens in a good imaging workflow is that you maximize the overlap of those three spaces. You can get the most accurate possible color calibration.

Oh, I'm going to send a note. CIB lab Is the name of the imaging model. For color. So, what happens when you calibrate devices is that essentially, each device looks at gnome color values and reports -- looks at known color values and reports back what it sees. That is used to create a crosswalk or a map of how each device proceeds.

That lets you get accurate color. That data is stored in what is called an ICC profile. And that should be included in imaging workflows, so that when you move an image from one system to another, the color values can go with it.

One final little thing that I want to do is show you this. In the last several slides, I hope no one has noticed the background color changing too much. It actually started white, just after our little common adaptation, and ended in this sort of pale and she'll color. Pale eggshell color. This is to show you how your eye can adjust quickly. To give you a sense, those are all of this life that we have looked at, from the starting color to the ending color.

So, color is actually in some ways the most finicky part of image workflows. And, dissertation can and have been written about how to get the best digital color. What I would say to you is, be aware as you get into imaging products, but you need to have a plan for color capture. And the safest plan is, fortunately, the simplest. And that is to have a calibration system and users consistently.

If you're doing this in house, there are simple color calibration kits that you can purchase. And, you could spend probably as much money as you want to come a spending anything gets you a lot further ahead than trying to do this by eye.

If you're outsourcing projects, and outsourcing didn't -- digital image capture, this is something you want to talk to your vendors about an talk to them about how they plan to calibrate. And be aware that there is still a little bit of mythology and sink oil about color capture. It is a good area to pursue offenders to understand how their imaging workflows work.

Finally, what is important, digital collections. One thing you want to think about when collections are coming in is how color is represented. Right?So, if you're getting image files that were created on a computer for and usually Commander. in RGB color space, that is great. They're going to live in their native environment. If you're skimming from printed files, CMYK files, there may be some edge cases where you can't accurately represent color.

And that should be part of your thinking and planning for digital projects. With that, my slides are open. I see I have a couple of questions. I want to touch on those. Feel free to type in questions in the question-and-answer area, and it looks that we have 15 or 20 minutes here that we can talk about image and text format.

I have a note here. I probably was unclear in speaking about the DNG format. DNG is a way of kind of harmonizing trough camera data. So, every camera manufacturer has a proprietary way of managing their sensor data, in what happens with a DNG file is at that raw sensor data gets mapped onto a standard way of storing the camera data. So, the only risk there is really an that transition from -- in that transition from raw data to the digital negative, and that is fairly nondestructive.

But, a cautionary note, when you are using DNG for photographic data, it is an excellent file format. There is not a lot of reason that I am aware of that you would want to use DNG for capturing say case images, or other kind of reformatted materials.

So, I think it is a really robust format for porn digital photography. I don't know that it provides you a lot in instances -- for born digital photography. I don't know that it provides you a lot of instances for -- thank you, Mark has the right analogy, a PDF for raw data.

A question here, are all pixels created equal, and are all censors the same size? So, here is a conceptual versus technological issue. So, a pixel is just a picture element, an abstract of the unit of imaging. A sensor is an actual device that is going to measure electrical current. So, a sensor can be large or small, sensitive or insensitive.

And the number of sensors -- the number of essentially pixels that a sensor which can capture has to do with the actual electronic. And the number of pixels in an image is an abstract measurement. So, different manufacturers of sensors of varying quality and density.

And it is worth saying, too, that the way that signals are captured electronically on a photosensor habits -- have as much to do with the land and the lighting and the structure of the camera body as they do with any sort of abstract model of imaging.

So in general, a pixel is a pixel. But be aware that when you get down to the details of each individual camera, the actual physical science of it -- sensor array and sensors on it would have some impact on the quality of imaging that it can do.

Let's see, let me pick up on types of scanners while we are talking about this. And the question about images as a certain dpi. So, the long end of an image, when you hear this sort of reference, some scanners, especially slide scanners have essentially a fixed sensor array that takes a photograph of the slide. So if you think about a 35mm slide, it is already greatly reduced from the object that a photo is taken off.

So, the scanning of those life is really just -- of those slides is to take a sample of that frame which may be 6000 pixels on the long end. So, the actual resolution of the images 6000 pixels per 35 mm. And from then, you create an image.

When you are seeing that, oftentimes, you'll see that in slide scanning. You will see it in the same imaging where you're going to take a picture of a painting, 600 dpi would be an enormous file size, since the painting is 4 feet on its alongside, your imaging standard might be -- do something like 6000 pixels on the long edge, which would have to do on the size of your camera's sensor array in the pixels and resize the image that you're working with.

So, there is certainly a point at which we were talking about photographing, architecture, sculpture, painting, or your reformatting slides, for 600 dpi false -- falls away. Scanning, text and photographic print, 600 dpi is sort of an archival and library standard. The museum community works a little bit differently. In those senses.

In terms of purchasing a type of scanner, honestly, the quality of scanners has moved ahead quite a ways. Since the last decade. The question you really want to ask is, what is the hardware resolution of the scanner, and what type of image output? So, if your scanner can output an uncompressed TIFF image and it has good hardware resolution, that is the imaging sensor is actually typical of doing 600 dpi, you probably have a decent scanner.

What you will often see advertised is something like 4800 dpi, and the way that that is done is by using a smaller sensor, taking samples of the image and then using software to a pixels in between them to interpolate. If you look at the specifications or manuals that are available for most scanners, at some point, you will be able to find in the actual hardware resolution.

Yes, question about -- let's see, I want to see if there are any other equipment questions while we are talking about that. Oh, the calibration kits. There are a variety of manufacturers for these. Some things to look at, I will put them into the chat here, I1, great segment of -- Greytag MacBeth or, manufacturers, and they have a target that is imaged by each camera comes standard --, Skinner, what have you. And something that could be used to read back output. This is usually a suction cup gizmo that goes on the monitor or projector and what you would do is scan this target and then the target is displayed on a second device and the output is red -- read and is determined how much growth has occurred, also, for instance, you will profile a printer and scanner target, print the target, and use it to read the printed output. That is all set into the calibration software and create a profile for each device.

Prices range from $100 up to several thousands of dollars. And again, this is sort of these things were simply playing -- where simply playing into the system gives you the benefits, and it the point that you're thinking about spending more and more, you also ought to have a point where you have a professional photographing stuff making those decisions for you.

The use cases, for the more expensive and more detailed equipment, are more and more -- require more and more expertise. So, if you're trying to get good reliable color, getting calibration systems gives you the majority of those benefits. If you're trying to do color accurate enough for you can distinguish the French pigment types, hopefully you also have -- different pigment tends, hopefully you have a conservator that can make those decisions. I would encourage you not to get lost in the woods there.

Let's see, our people hearing me? I have a few notes that audio is dropping out. Great. Let me touch on a couple of questions here about image derivatives and selection of color, 24 versus 32 bit, for instance.

So, for derivative use cases, very often, when we think about an archival imaging flow, we create this sort of high-resolution master file, with well calibrated color. And from that we create derivative files. So, a common derivative is for the web command for a web derivative, you may have a very small sort of thumbnail sized file for browsing and indexing. And then you may have a high-resolution version that is still not the full uncompressed master file, but has as much or more data than anyone can comfortable see -- comfortably see on a standardized model.

So, say a 1600 or 2000 pixels along the long end. So that when someone looks at that full-screen on their computer, they see a really rich image. So, those would be common Web derivatives that might be compressed. He might use JPEG or ping as a compression. But, in some cases, they might be color calibrated. Probably not a concern for your thumbnail, but for the high-resolution version, you might use that color calibration file to create the derivatives that data gets sent to the computer, if somebody has a computer that is set up to do color calibration, they will get a better image experience rather than someone who doesn't.

You'll still have no control over their setup, but you at least allow that possibility. Also, at some point, you might wish to print image. So, you might take your master image file, it might appear as a printed version in your annual report, especially if you're in a museum, you might publish an exhibition catalogue. And so, from there, from the master file, use a CMYK master derivatives that would be used to print. And again, because CMYK and RGB don't quite agree, you might make some decisions. He might work with your printer to recalibrate your color slightly to get something that looks good to the ice. Between the original -- good to the eye, between the original and printed version. Those with the most common versions.

In every case, having the master file to go back to enables you to create derivatives. The mistake that was made early in digital library project was essentially to try to create and I -- a version that looks good to the eye, which often meant a version that looks good in print. And when you try to go back to do alternate digital versions, you have lost or altered color data. A story that I could tell on myself was from preparing a Christmas card, that a library sent out.

In doing the transition to the CMYK, there was a blue and a purple that became almost indistinguishable. This Christmas card from now, everyone who would use this special collection, and we got a response back from [Indiscernible] delighted and thrilled to see this image that he had never seen and amazed that we have this shade of blue and an object from this date and time because it suggested that all of these things were beginning to write his next book about trade routes. What really happened is that something it should have been purple came out blue, and to him, that's a pigment that should not have been available in one part of the world was. To us, it meant that we messed something up in imaging workflow.

Fleming also touch on this question of 32-bit versus 24 -- let me also touch on the question of 32-bit versus 24-bit color. 32-bit gives greater accuracy, for sure. And the only reason there is any skepticism in the digital archive community has to do with the software that that color is going to pass through. So, if, at a certain point, that 32-bit image has to pass through a 24-bit piece of software or piece of software that only understands 24-bit, we don't know how that image is going to be altered or concave it. -- Altered or truncated. One thing that will be seen throughout every encounter that you have in the world of digital preservation is a digital preservation recommendations always like behind the state of the art. Because we are trying to control for the unknown.

So, a classic example of this is in the wave file format that is used for audio data. Some programs are better than others at respecting the length of the wave file, so there was a problem early on in audio digitization where if you put any metadata into the header of the file, so if you add the title of the track with the audio file, it was then trimmed off that any bits from the tail end of the file.

The pieces of software that were not smart enough to know the file had gotten bigger with the addition of metadata. So, you would end up clicking the back end of the file. The same thing happened with a 32-bit image gets a 24-bit type in the workflow somewhere. So, if you can insure and make sure that you are -- your entire imaging chain is 32-bit all the way through, you in fact can get a much richer image. 24-bit is just a much safer recommendation.

Let's see, we talk about [Pause] there is enough here on supplementary materials. Sort of four -- sort of for e-books and published versions. So, I would encourage you to be aware of kind of mixing and matching the problems.

One is this question of kind of what is observable, right? So, all of those recommendations for supplementary materials, intellectual controls, really has to do with essentially metadata structures. Right? How to make sure that part a relates to part B., and that parts a and B. are seen as part of a larger whole.

Each one of those parts will still have to be represented, as text and graphic data. And so, there are sort of two levels here. On the one hand, you want to make sure that all of the information you're collecting is in good text format, UTF-8 is where we'd like to see everyone right now. And also, in the image data that is coming in is in viable formats.

There is a second question of how to arrange all of that material. Into a coherent digital object. This has become a problem in, for instance, electronic journal publication, where a journal article may have text from an associated chart, it may have a data set that goes along with it, and keeping all of those things together is intellectual and seen as a complex description or arrangement problem. The individual chart should be in reliable format, the individual images should be in the individual text as well.

Those that speak to the issue -- does that speak to the issue? Am I interpreting the question correct? If not, post something else in the chat and we can come back to it.

Oh, so there is a question here about JPEG 2000 readers. So this is one of the problems and JPEG 2000. Not only do you have to be good reader, but you have to have a good server. So, there are plenty of software that knows how to decode a JPEG 2000. But, for those features to be used, JPEG 2000 also has to be delivered. So, unlike the sort of standard model where you request a file, it is delivered to you in transaction. In the JPEG 2000 world, there is ideally a file on -- dialogue between the server and client.So, my client says, showing a server of this resolution in the server delivers that Commander. McLane says that, actually, give me a little bit more data and the server -- and my client says that actually give me a little bit more data and the server gives it back.

We don't have every piece in the chain linked together, yet. So, if you have the right software, you still might not have any content to look at. And people aren't creating content for people to look at because there aren't great readers out there to do it and there is sort of a chicken and egg problem that we are overcoming and JPEG 2000.

But many of the QuickTime pollutants -- plug-ins, there are a number of greatly available -- you still need to write server infrastructure to connect.

Well, we are at the top of the hour, Jacob. It looks like we are out of questions. At this point. So, just as a reminder to everybody, we will be archiving this, so anybody who had issues, you can go back and listen to the archive if you have any issues. And this is, as a reminder, the second in a series of four, and the next one is happening, when, Jacob?

Good question. February.

February. It is on the website. So, check the website. And you can find out when the next in the series will be. And, all of the websites -- all of the webpages will link to the previous and next ones. So, you'll always be able to follow the series along. There you go, number four is March 20, in February 7 is number three. So, hopefully we will see you all on February 7, thank you, Jacob, and we will see next time.

Thanks, everyone.

[Event Concluded]