

ACRL-DIGITAL CURATION FOR BEGINNERS:
AN ARCHIVAL VIEW FOR LIBRARIANS

Edited on APRIL 28, 2014

* * * * *

Communication Access Realtime Translation (CART) is provided in order to facilitate communication accessibility and may not be a totally verbatim record of the proceedings.

* * * * *

>> Okay. So the recording has started. I want to welcome everyone to our Webinar, Digital Curation for Beginners, an Archival View for Librarians with speaker Lisa Snider. Thanks everyone for coming today. My name is Susanna Conrad, and I'm the recorder for the ACRL Digital Curation Interest Group, and I'm so glad so many of you could make it. If you experience any issues with audio, try troubleshooting with the audio wizard. Also, we'll have questions and answers at the end of the session, but Lisa's also said it's fine to post questions along the way.

So I'd like to thank Lisa Snider for presenting today. Lisa Snider is currently the electronic records archivist at the Harry Ransom Center at the University of Texas at Austin. She has an MAS and an MLIS from the University of British Columbia, and has worked as a digital archivist for the last 3.5 years. So, and now I'll turn it over to Lisa to get started.

>> LISA SNIDER: Great. And can everybody hear me? Let's start with that. If I can get some feedback. Yep. It looks like it's good. Okay. Awesome. Thanks so much.

Super. Well, thanks, everyone, and thanks for coming today. And let's get right into it. And actually first, I know some of you are working in groups and some are individuals, but if I can get an idea of where everybody's coming from, that would be great. If you can just let me know in that chat area on the

bottom right side, you can tell me if you're a librarian, an archivist, both, or if you're neither. Let me know what field you're coming from. It just gives me an idea of the fields. Oh, got lots of librarians and archivists. Excellent. Super. Private school. Excellent. Super.

So I got a wide range. Perfect. Excellent. Well, thank you very much. It just helps me know who's out there. And like I say, I know we have groups listening as well, and I know you guys are from all ilks, so wherever you're coming from, I think this talk will be useful for you.

So first of all, I just want to thank the Digital Curation Interest Group. They've been great to work with for this webinar, and especially Susanna, who's my handler today and hopefully I won't get out of line or do anything weird. So she's there to stop me if I do. So thank you, Susanna.

So, just a one-second primer about me. Before I came into archives in 2009, I was a Web developer for over 17 years, and I ran my own Web development business for 14 years. And I got my degrees last year, and from 2009 till today, I've worked in different places doing different things. I've been what I call an analog archivist, working with paper, photographs, that sort of thing. I've been a records manager. And I've done projects for public library systems. So I've done a lot of different things. But really, for the last three and a half years, I've

been focusing on digital archives and being a digital archivist, and here at the Harry Ransom Center, I provide long-term preservation storage, and, soon, permanent access to the born digital materials that we have here. And for the past three and a half years, going on almost four years, I've worked with all aspects of the digital archives sphere, as it were. So I've done a lot. I've seen a lot, but boy, every day I learn something new, because with digital, hmmm, every day you'll get something new.

And I've also worked with physical materials like floppy disks, that sort of thing, and with online materials. Things like Facebook, twitter. Those are what we call cloud materials. So that's just a one-second primer about me. But what I'd like to do first is start off with a definition. Now, I love -- definitions are so boring, and whatever, but let's try to make it interesting. So let me define for you what I mean by digital curation lifecycle. So I'm going to break it down. And that first word is digital. Well, there are two things that the digital encompasses. The first are digitized materials. Now, if you've never come across digitized materials, basically we scan objects or items, and they are made into a digital image file. And here's a great example for you. It's what I like to call the cat in the dress. This was an old black and white photograph, not really old, it's from 1980, and our friends at

Digital New Zealand greatly digitized this for us, so that everyone can see this lovely picture. Doesn't every cat need a tiara and a robe, and of course for librarians and archivists, I had to get a cat in there.

So this is an example of a digitized image file, okay, it was paper and now it's digital.

We also, digital also includes born digital materials. And in case you haven't run across those, those are a digital file that's created with a computer device. Now, I have to add the word "Device" in today because unlike, well, probably about eight, nine years ago, we just had computers, and now we have all sorts of devices that help us create born digital materials. Anything from your iPad, your iPhone, your Blackberry, if you've got Google glass, your desk top computer, that sort of thing. And as a digital archivist, I get these files either on legacy media, and that's like, think of old, floppy disks, okay, those are legacy media. I get them on new media. Think about an external hard drive that you use in your home or your office. And, more and more I'm getting materials from online or in the cloud, materials like a Facebook page, I've archived twitter pages, that sort of thing.

Now, to give you an example of born digital materials, well, you're looking at one because this is a PowerPoint presentation and that is a born digital material. Same as a word document

and same as any digital photograph that you take on your -- whether your iPhone or your digital camera. And thankfully, somebody took a picture of this lovely dog, who really stayed still and let its owner put it in this costume.

But this is a born digital image. And if you want to remember it, the way I remember it is Lady Gaga, you know, born digital materials, they are born this way, right? So this image that you see here, which is a piece of art, I believe, from Lady Gaga, one of her albums, old school, this was likely created on photo shop, so it is a born digital image. Okay? So it's born digital material. So now when you think of Lady Gaga, you'll think of born digital and the vice versa. I'm sorry I did that to you, but there you go.

So one question I get asked in my travels is whether content matters. And no, content does not matter. It's the digital that matters. You can have content, whatever you like. So as an archivist, I deal with manuscripts, I can deal with sound, video, E-mail, Web sites, blogs. As librarians, many of you, from what I see in your introductions, many of you will be dealing with maps, maybe photographs, and also researcher data. You know, researchers coming to you saying -- or you're going to them saying hey, we really want to, you know, archive your data, you know, give it to us, and also university records, eBooks, et cetera, et cetera. So remember, content doesn't matter.

So we've done the digital. That was fairly straightforward.

We've got digitized materials and born digital.

Now, what is curation? Well, as many of you know, and we laugh about this sometimes on twitter, today, curation is in the mainstream, right? I could give you 40 examples of this. One of my favorites is my fellow Canadian, Jeanne Beker, the fashion guru, who now curates a shoe collection, so curation has come into that mainstream use. And those of you who use pinterest will see curation come up again and again. But what is it for us, for librarians, archivists, you know, the people who work with digital material? This is one of my favorite definitions, and it's from our friends at the Digital Curation Center in the U.K.. I really like this definition, even though you see here, I've added things, I've taken things out, but it gets the point across. So I'll just read it quickly for you. Digital curation involves the selection/acquisition, maintenance, preservation, storage and access of digital research data and/or digital donor material through its lifecycle.

So basically, in plain terms, that definition is saying, digital curation is what we do with digital material from when we first receive it or make it to its preservation, storage and access.

So basically, digital curation is what we do with the digital material. That's all it is.

And a lot of people just say digital curation. Some will say the digital curation lifecycle. I've seen both used interchangeably, you know, but that's fine. For me, the lifecycle part represents all the parts of the digital curation process, or if you will, all the pieces of the digital curation pie.

Now, I've used a pecan pie there because I'm in Texas -- or I'm in Austin, and pecans are big here. I always get hungry looking at this slide, so now I'm going to be hungry and now everybody else is going to be hungry so I hope you've had lunch.

So we're going to go over those parts of the digital curation process in a second. So before I start on that, any questions about the definition that I just gave? It's pretty straightforward. We're basically -- it's what we do with the materials, the digital materials that we get, both born digital and digitized.

So any questions there? No. I don't see any. Awesome. So feel free to ask along the way. I'll try to catch them as I can.

So now that we have the definition, next let's look at all the major parts at that digital curation pie. And one of the most popular diagrams that's used, and I will warn you now, it is the scary ball. Okay. Cookie Monster, I know Cookie Monster is singing here, but I think Cookie Monster is also kind of

screaming a little bit saying, oh, it's a scary ball! It's a scary ball! But here it is.

This is from our friends at the DCC. I must mentioned their definition. They also created a digital curation lifecycle model. Now, if you haven't been exposed to digital curation or you've just done a little bit, this ball is kind of confusing, right? I don't know. When I first looked at it, I was like, okay, where does it start, where does it stop, what happens? And it can be kind of scary when you first look at it and of course I always think, hey, you know, it's kind of scary, let Cookie Monster eat the scary ball because it kind of looks like a cookie if you think about it. So let Cookie Monster eat that ball for now and what we'll do is I will take you through all the major parts of the Digital Curation lifecycle, and in doing so, I'm going to explain that scary ball, so it's not going to be as scary for anyone out there. Although I suspect many of you already know a lot about this, seeing your introductions. So let's get right into it. Okay?

So we've got that pie, and I'm taking out that first piece of the pie. And all of you know when you dig into a pie, you're the first one, you never want to be the first one, right? When you do it, that piece, boy, that can be difficult sometimes, depending on what the pie is made of. And that piece is crucial, right? Because as Martha Stewart will tell you, you

know, you have to cut that really well in order to get the next pieces to go smoothly. So with this, the first piece of our pie is the digital material. And yep, it's the same digital material that we just talked about. The born digital material. Remember Lady Gaga, you'll never forget, and also the digitized material. Come on, the cat in the dress, you'll really never forget that.

So my saying here is, without digital materials, we are nothing. Okay? Because it's true. Without them, I would not have a job, we would not have this webinar. We would be nothing. So the digital materials always come first, and they're the first piece of the pie. So we have digital materials. Great. That's all good and fine. So then we take out our second piece of the pie because boy, that pie is good, and we want more, and that second piece is selection and acquisition. Because we have to get those digital materials from somewhere. Right? So as an archivist, I acquire digital materials from donors. Okay. And as librarians -- now, librarians also acquire things, but a lot of you will select digital material. So we do the same thing, it's just in slightly different contexts. Plus, both librarians and archivists select material to be digitized. Right?

So, great. We've got our digital material. We've acquired it from somewhere, or selected it from somewhere, and now, the third piece of the pie, boy, are we hungry. That pecan pie was

good. We're now into appraisal. And yes, think of Antiques Road Show appraisals, okay, because -- I love this picture -- it's a similar concept. We're just doing it for very -- usually different reasons and for different purposes. So if you've got the concept of Antiques Road Show, where they appraise materials for people who come in, then you'll get the concept of appraisal. And with appraisal, my -- archivists do appraisal. I know librarians do it, too, but I think archivists do it more. I think we just have to do it more, whereas librarians maybe do it in a focused way, not as much. I mean, that can be debated back and forth. But no matter who's doing it, when we're appraising those digital materials, we pretty much ask the same questions. We first ask, what is this. Right? With that woman in our last slide. What is it? And is it important? And why is it important? The "Why" is actually the crucial part there, right?

So the why is, well, was her -- I'll go back to her because it's such a great picture -- was this lovely object worth a million dollars? Well, from her, you know, from the look on her face, it probably was worth a lot more than what she thought. Is it unique? Well, I would say that object very well could be unique. And is it neat. Like, is it really cool? Does it have real -- does it have no financial value? But is it just a

really neat object? In that case, I don't know that I would call that object neat, but it certainly is interesting, right? So we look at that, and then we ask, does this fit in to what we have already in our collections? And I'll give you an example. This is from Golda Meir, the prime minister of Israel, late '60s, early '70s. This is from the archives in Israel. This is Golda Meir's recipe for chicken soup. Okay? Awesome piece of history here that we're looking at.

So let's do a very quick appraisal to give you an idea of what archivists do. We first say, what is it? Well, hey, great, we've got a title. It tells us right there. Golda Meir's recipe for chicken soup. Awesome. That was easy. Now, this was a piece of paper that was typed up and somebody digitized it, right? They scanned it. So my feeling is okay, they scanned it. It must be important. And in this case, why was it important? Well, was it worth a million dollars? Probably not. Maybe a couple hundred bucks.

Is it unique? Well, it could be, or maybe there are just five in the world. But there wouldn't be a thousand of them.

And is it really neat? And that is, yes, I would say this is really cool because this shows Golda as a woman, and a mother, and not as a prime minister. And this is actually part of her being a prime minister as well, if you know her history.

So okay, it's important. Right? Now, the third major thing we ask is does it fit with our collection? So if here at Harry Ransom, let's say we're collecting recipes, yep, it fits.

Are we collecting famous women, like stuff from famous women? Yep. It fits.

Are we collecting, you know, the history of Israel materials? Definitely fits.

But, if I'm at an archives, and I'm collecting bottle caps, okay, this is not going to fit. And then we have to decide, hmmm, will it fit into our collection? Should we keep it?

Okay. So that's just one example of appraisal for you. And that's the third piece of the pie. And so we've got the digital materials. We acquire them. We appraise them, and now, boy, we're still hungry, we still want more of that pie, you know, it's looking good, so the fourth piece is that we make sense of our digital materials for ourselves and others.

And let me explain that for you. Because here's where archivists and librarians, we do slightly -- we just call things different names.

I'm an archivist, so I arrange and describe. So when I get a piece of digital material, the first thing I do is arrange it. Basically, I look at that material, and I organize it in accordance with archival principles. That can be provenance, original order, that sort of thing. And basically what I'm

doing is I'm organizing. I want to be able to find things in this collection and I want my users or researchers to find things as well. So I organize things in to collection. Then I organize it into many series, maybe some items, maybe some folders, and also with the collections, that can be a falls as well for the Canadians out there.

Now, just so you know, this is really difficult to do with digital material. Okay? This is not easy. Because I can have a terabyte external hard drive that you buy at Staples, Office Depot, and that can have a million documents on it. And think about that. Think about how to organize and even look at a million documents. It's a tough job. It really is. And it's not easy. So just so you know, it's not an easy part of what we do.

Now, for those of you out there who have never done -- who have never been in an archives -- you know, worked with archival material, and this is not a great analogy, I know that. The archivists are going to look at me and go -- but it's a weak analogy, I'll just warn you.

Think of a professional organizer, okay? What they do is they create order from chaos. And we're kind of doing the same thing, but for very different purposes. Okay? So here is order, or should I say chaos, and then order.

And then as an archivist, after I've organized the material, then I get out, oh, let's go with the soup analogy again. I get out my soup pot and I put my arrangement in it. I put in my materials. I put in provenance. That is, where has this material been? Who's owned it? Where does it come from? The original order, what order did I get it in? And all the research I've done, because archivists do a huge amount of research when they -- usually when they work on collections. Put it all in one soup pot, and what we get instead of soup is a finding aid. And if you've never come across a finding aid, basically, I call it a road map for researchers or users. It gives them an idea of what is in that collection or falls. And it's also used by the archivists themselves, and librarians. It's very useful. And again, my weak analogy of the professional organizer, if that professional organizer did what we just saw that they did, the description would be them writing a report saying, yep, I worked with this material, I put it here because of this. I got it from this person, and all those kind of details.

Okay? So that's what an archivist does.

Now, librarians, in my view, librarians also arrange and describe. I think we just don't call it the same things. With librarians, librarians arrange by indexing, creating facets, creating subject headings, that sort of thing. And they also

create catalog descriptions. So really, really, I think we're doing the same thing. We're just calling it, you know, slightly different terms, and, okay, it's slightly different when you look at it. But in the end, it really is arrangement and description.

So we just finished that piece of pie, and I kind of tricked you here, because that piece of pie that we just finished, this piece could have been added on to the one we just talked about. Because really, they go together. But it was such a big piece of the pie that I thought, oh, it's looking pretty big. I don't know. I don't know if I can eat that, so I'll split them up. And this fifth piece of the pie is metadata. Oh, yes, don't we all love metadata? Well, metadata, I think almost all of you here, from your introductions, I think all of you here know about metadata. Okay? I hope so, and I hope you've dealt with it in one form or another. And really, if you haven't, metadata is data about data. It gives us information about that digital file. And metadata, like what we just described with arrangement description, cataloguing and indexing, it's crucial, because without the fourth piece and this fifth piece of the pie, our materials are invisible to both us and our researchers and users. So as I hope we all know, metadata is really crucial, and it's really, really important.

So as an archivist, as a librarian, we're using standards. Okay? We've got metadata standards that we -- most of us use, not all, and we use different standards and sometimes differently, but we try to standardize this so that we can talk to each other, so that our systems can talk to each other. And as I say here, pick your acronym, and in fact, as most of you know, there could be a hundred more acronyms here, right? As a digital archivist, I deal mainly with METS and PREMIS for my metadata. I also deal with Dublin Core in some cases. And for the librarians out there, you'll be dealing with probably Dublin Core and MARC, and even the archivists, we, almost all of us deal with MARC in one form or another. And, you know, metadata really is crucial. I think we all know it. I'm probably, as they say, preaching to the converted, but here's an example. Okay. I had to put in another cat. Come on, we're librarians and archivists. So here's a cat. A very cute cat. So if you take away the information on the right side of your screen, that really only becomes a really pretty picture of a cat. Right? It has no value other than, boy, that's a cute tabby cat. Right? So as we all know, metadata gives our digital images, gives our digital files, gives even our analog materials value because then we can say well, this picture was actually taken by this gentleman here, and he probably titled it, "Standing Tabby Cat," and that's the title it's been given, and our wonderful

friends at Digital New Zealand have also given us subject headings that we can use, or subjects, where they've put all these lovely pictures. You, too, can go find them on Digital New Zealand and they're under the subject of cats. So that metadata gives us value. But remember, too, and this is one of my favorite phrases, metadata is a love note to the future. Right? It's crucial. It's crucial. It's really crucial for the past, the present and the future, but it really is us saying, this object, whatever it is, digital whatever, is important. Here's what you -- here are the basics you need to know about it.

So remember, when you're doing -- when you're creating your metadata and you're spending tons of time on it, remember you're writing a love note while you're doing it.

So now, before I get into the sixth piece of the lifecycle pie, does anybody have any questions? Any questions up until now? Or is it pretty much straightforward? You know, you've been working with analog material, many of you, and many of you with digital material, too. Does that all make sense? Any questions on that one?

Okay. I'll just give a second here, and I can just take a breath. All good. Okay. Great.

And like I say, keep writing questions in case you have them along the way.

Okay. So now the fun part. Oh, yes, and this is the part, this is the sixth piece of the pie, and this is the hard one to eat, I'll tell you right now, and it was the hardest one for me to do in this presentation, even though I do it every day. This is long-term preservation. Now, I'm going to say something up front because it drives me crazy, and I've seen it so many times. Long-term preservation is not storage. Okay? And I see this a lot with both archivists and librarians. Storage is just one part of the preservation process. It is not, in itself, preservation. I had to get that off my chest. If you remember nothing else from this presentation today, please remember that. Because storage, as we'll see, is just one part of this piece of the pie. This piece of the pie is the tough one.

So when I get in, I'm a digital archivist, and when I get in an item, such as this hard drive here, okay, the first thing I want to do -- well, maybe not the first thing, but in terms of preservation, what I want to do is clone it. I want to make an exact copy of it, okay, whenever I can. And sometimes I have to make an exact as possible copy because sometimes I can't clone this exactly like this, right? I can't clone everything that's on this to make a copy. But I try. I try where I can.

So think about this process, where you have the original, and then the copy, or the clone. Think of this if you need, you know, an example to remember, think of twins, okay? These are

my favorite twins in the world. One twin was the original, and the other twin was the copy. Which is which? I don't know. But you can decide. So think twins when you think of the first part of the preservation process.

So we have our original, and we make a clone of the original. Okay? Then from that clone, we want to make two copies. We want to make a preservation copy, and that's what I call my forever copy, that we try to never touch. Now, in practice, is that true? No. But we really, really try never to touch that copy.

And then we make an access copy, and that's the copy that our researchers and users are going to be given. If you have kids, think about this giving the access copy to your kids, because they can abuse it, use it as will, whatever. If they break it, it's fine, we can go back to that preservation copy and make a new access copy. Okay?

So we've got the cloned original and then we're going to make a preservation copy and access copy from it.

Well, that's easy, right? We just clone, we take the cloned original and we clone a preservation copy and we clone an access copy. Well, as in everything in digital archives, oh, nothing is that straightforward, and this is the reason why I have gray hair. I did not start out with gray hair but I have it now.

And this is the complex part, and this is -- I'm going to

explain it to you bit by bit, as it were. A little digital preservation humor. Okay. So I have that cloned original, right? And I want to make a preservation copy from it. So luckily I have software that looks at every file on the cloned original, and as it's copying it to the preservation copy it asks for every single file, it says, uh-huh, exactly, is this file in a stable and open file format that can still be used and read in five to ten years? Okay. So it asks that for every file. And if it looks at that file format, okay, the file format of that file, and if it says, no, that file format is really weird. It was used for one year by this person. No. It's not going to be stable and open and be able to be opened in ten years. There's no way. Then it says, okay, fair enough, can we convert it to a file format that is stable and open, and I'll answer your questions in two seconds, just so you know. So it asks those questions, which is a great piece of software, right?

Now, I'll tell you right now, that bug -- I won't even say buggy, this software -- I won't even go into it, but it's -- some file formats we can do that with, and some we can't, as in everything with digital archives. Things are complicated. So we try to do this for every file. The word emphasis there is on "Try." And I just want to mention that this process of conversion of the file formats is called normalization.

Now, I just wanted to mention that. I'm not going to really mention it again, just in case you ever see that term, that term means conversion of the file format.

So what is the software used to do this, Megan asks. Well, there are two pieces -- there are different softwares that you can use. Pronom is one of them and there's one from Britain and one from the states, and Fits is the other. They're great. They're wonderful. We're still working on them. Okay? It's a huge, huge field. And also, as I'll show you in a second, I'll show you other software that can help you out, because what I -- I call Fits and Pronom kind of manual ways to do things and I'll show you why in two seconds.

I want to give you an example of the preservation -- a preservation copy, what I would do because I think it will help you understand what I was just explaining.

So a donor gives me a floppy disk, and it's got one file on it, and boy, am I excited because it only has one file. And that file is a photo shop file. The ending, the file format, says PSD.

Now, that format is widely used by graphic designers, Web developers, but as an archivist, and the people who made up this software that I just mentioned, who are archivists, we go huh-uh, we don't think that's a stable and open format that's going to be usable in five to ten years. We really don't.

We're going to change it. We're going to change it. So in this case, we change it to what we think is going to be a stable and open format, and that's called a TIFF. Okay? It's when you see the ending of a file, it's dot-TIFF. So our preservation copy will be confronted into a TIFF. And I'm going to answer your question right after I talk about the access copy just to get this concept across.

So remember we dealt with that preservation copy, and now we have to make up an access copy. Well, guess what? We do the same thing as what I just described, but because we're digital archivists, yep, we're going to make it more complex because it has to be. The complexity with the access copy comes in from the fact that it's just for access, okay? This is going to be the copy that's used by researchers and users, hopefully five or even maybe ten years down the road.

So what we do here is, as very good people, we want people to be accessing this, we want to look at, well, what kind of software viewing programs are they going to use to use these access copies? So that comes into the access copy as well. Okay? So it's kind of complex, as you can see. There are a lot of steps that our software has to take.

So again, let me just give this example, and then I'll come to your questions. So again, we come back to the photo shop file. It was turned into a TIFF, okay, the file format was turned into

a TIFF for preservation, but as archivists, we go, hmmm, I think a JPEG is going to be the best thing for an access copy. And my reasoning why we do this is that JPEGs can be read by almost every piece of software out there. They're great. Because almost -- not almost -- almost every piece, but a lot of software can read JPEGs, whereas TIFFs, you might have the odd issue. So in this case, we turn that access copy into a TIFF -- or, sorry, into a JPEG file format. So our preservation copy is the TIFF and our JPEG copy is for the access copy. Okay? So we convert file formats, and there's a link here if you want more information. I'm coming to your questions in one second.

So you're looking at all of this and you're saying well, why change file formats? Why convert them? What's the point?

Well, okay, here's my great analogy. Think of dating. Okay? What you're dating, you want to date somebody, you want to look at them after the third date, and you want to say, hmmm, I want -- I think you're going to be stable and open with your feelings five to ten years into this relationship, what's hopefully going to turn into a relationship. Right? You don't want to be really -- maybe you do -- want to be with somebody who's closed, or, you know, leaves or does whatever. You want somebody who's going to be stable and open. Okay?

So if you think of that dating analogy, hopefully that makes some sense of why we want to convert file formats. And no, I'm

not even going to mention the word normal from normalization here. I'm just not going to do it because I could do all sorts of things to get me into trouble.

So before I touch migration, I'm just going to look at your questions. Actually, I'm just going to save those questions for right after this slide, if you don't mind.

With migration, I just want to mention it. I'm not going to go into it. You will see this term around. Okay? What happens is as a digital archivist, you've just seen a lot of work that I have to do. Okay? It's huge. Okay? Times that by a million files and you can see what fun we deal with.

I, on a regular basis, or regular intervals, usually five years, I should be going back to those preservation and access copies and saying, hmmm, are those still around? Are they still okay? Are we still using them as best practices? What's happening? Because think about it. I think it was only eight or nine years ago that the iPad came out. Okay? Technology is changing so quickly that file formats may change quickly.

Now, so we always want to look back at our work, and then do, if we have to, migrate the files to new file formats, or you can call this renormalization. Right? But you'll see this called migration. So I just wanted to mention that, because we're digital archivists and we like to have a lot of fun, we never want to sit around and drink coffee, here you go. We have to

migrate files as well. We create them, and then about three to five years later, we should be looking at them again and saying, hey, what about these file formats? Are they still around? Even the ones that we chose for the preservation and access copy, are they still there? And I can just give you one example for video. I know the video format, the preservation format that we've used, and I think the access copy as well, both of those have changed twice, I think, in the last two years. Like nobody really agrees on them. So just know this is not an easy process. We still have a tremendous amount of work to do on it. So I'm going to answer your questions now, and then I'll go to this last slide.

So, Megan asks, are you also keeping the originals? Yes. You are also storing the originals. But the originals are not touched. Like, the original files, you've got the original files, then you've got the cloned files, okay. Some archivists keep those original files, some don't. I keep them, as well as the cloned originals, okay, because I like to have backups upon backups, and I have the storage space to do it. So yes, I would keep all those for me together. And so Megan asks, are you keeping the originals so you have an original file, a preservation copy and an access copy? Yes. That's what I do. And, in fact, I also keep -- and I also keep the original -- sorry. I meant the original media. So let's say a floppy disk,

I also keep that as well. I keep that. I keep the cloned original. I keep the preservation copy and I keep the access copies. And yes, I would recommend that you always keep everything because as you say, as Megan says, things get lost in translation, and they get lost with server failures, and I've seen everything happen as a Web developer, so I know things happen big-time. So thanks for your questions.

Erin asked, exactly when a file is saved as a TIFF, you remove the layers in photo shop, yes. And yes, that's true. Erin points out that with that photo shop file that I talked about, those of you who have used photo shop know that it works with layers, and the TIFF -- now, I have seen things on list serves that says that the TIFFs keep the layers. I haven't seen that myself in the work that I've done, but I'm investigating that now. Because people tell me that you can keep the layers in the TIFFs, but I haven't myself seen that. So I'm still playing with that. So as you can see, it's not an easy process, right, because you're -- a photo shopped file is a 3D, in a way, image, right, because it's got bunches of layers that make it up, and you're trying to put it into a 2D format and that can be difficult.

Megan asks, do you use a file management system to help you keep track of when you need to go back to check files? So to migrate them. Do you use something to automate this process? Yes. And

this slide, I'm just going to do this slide right now because your question went right into it, and then I'll get to Amy's question. Yes. I highly, highly recommend using software for this. Some people do do it manually. Some of you may have heard the names DSpace and Fedora, those programs can be used but for me my preference as a digital archivist is programs that were made specifically for this task, preservation, and those are something like Archivematica, Preservica, I think there are some other ones, but those are the two main ones out there. So yes, those are the programs I use to keep track of things, I do not want to do it manually, but you can. You can. And people have the last few years, but it's just easier because it automates the process for you. So Megan, check out these programs if you haven't already done so.

And Amy asks, for those creating born digital objects, where does planning for curation and preservation fall? I would say it should be done what you're creating the objects because many people will digitize things, okay, or create born digital materials and then go, okay, now what. Right? Which is fine. You can do it that way. Nothing wrong with it but you want to try to plan as soon as possible and I would say if you can plan first, great. And if you can't, as soon as you're creating those objects, I would say get on it. Because the preservation planning will effect how you create those blind digital

materials and it will help you create ones that hopefully can be kept for a long time. So I would say do it as soon as humanly possible, but I know many cases, I've talked to many people, sometimes you can't, but just try to think about it as soon as you can, get those committees going, talk to people as early in the process as humanly possible. And in some ways, preservation could be the second piece of the pie and not the sixth. Right? So yeah, it's a great question. And by the way, if I haven't answered your questions fully, just type in there, and I'll get back to you.

Oh, and as Miles said, thank you, Miles, there are layered TIFFs that one can use. I myself haven't used layered TIFFs in the work that I've done, but I will be likely this year. So thank you. I've wondered about layered TIFFs versus TIFFs, and I know they've been around for a long, time, and yes, one can use them. So thanks, Miles, for jumping in there. Because I knew they existed. I just had never used them myself. Because we didn't need it in the circumstances that we were working.

And Amy also asked what conversations should we be having with digital material producers to better enable curation?

Well, just talking with people. Just communicating. I would say, and I know some of you are doing, like, records management or you're working with researchers, and just trying to start out with -- I don't mean this to sound bad, but simple

conversations. Just saying, hey, this is the process we go through, you know, it's quite complex, you know, and we want to make sure that your material is safe for the long-term. You know, can you help us out with that? And I know where I was in my last position, I -- we did a couple of presentations for people in open access week, and it was great, because we got across that information. It takes -- I'll tell you right now, it takes a lot of time and patience. You have to do it over and over and over again. They're conversations you have to have many, many times, which is good because you develop relationships with people by doing that as well, and they get to know your work better. You get to know theirs. That's what I recommend, is to talk about things as soon as you can with the producers of the material. And just start out slow. Start out slow, build up, and I think you'll eventually you'll definitely, you know, be able to make this an easier process. Eileen asks, would you distinguish digital stewardship from digital preservation from digital curation? Great question. I actually had slides for that that I had to take out. Here's what I recommend. Or what I think and what I recommend. For me, digital stewardship is the same as digital curation. Okay? I know other people would disagree, but for me, they're the same. Whereas digital preservation is only one part of the digital stewardship and curation pie. So that's how I see it, where

digital preservation is just one piece of that puzzle, of that stewardship and curation puzzle. So that's my view. And I know they can be used interchangeably and you'll see them everywhere, and it's very confusing, but that's my view of things.

And Martin asks, do we need to do anything to preserve the physical medium used to store a file? For example, do we need to refresh the magnetic charges that make up a file on a hard drive? Yes. And I've been looking into that. I haven't done that myself, because I haven't been in the field long enough to do any of those refreshings or migrations, but yes, you should be looking, if you can, if you have the time, not everybody does. If you can look at the physical medium, yeah, Martin, great idea, and I'd like to see more talk on that. You know, in articles online and that sort of thing. So please get that out there because I would love to see more talked about it. So great questions. Thank you.

So I'm going to go on to our second from last piece of the digital curation pie. Because that was the tough part, the preservation. And I know you might have other questions, and I'll go over them just in a couple of minutes. Won't take as long to finish.

So remember, as I said, storage is just one part of that digital preservation plan or program or process. But storage is very important. Right? Because as all of you can see, and those of

you who I see have worked with it, you know, it's a big process to do, whether you've got software to do it or not so we need to store all that work that we've just done. And best practice, just so you know, if you haven't come across this, is to have a copy of our digital files, all of them, okay, original, clone, access and preservation, on three different types of storage, with one of those being outside your work area, okay, because disasters happen. All you have to do is look at the news today, and you see it. Disasters can happen.

And some of you may have seen this term. It's LOCKSS, lots of copies keep stuff safe. Can't say it better than that. It's true.

And I just wanted to give you just an idea of what we're using today for storage of digital materials. I'm not going to go into it in any great detail. I'm just going to give you the quick, two-second version.

You can use a server to store your digital materials on. For those of you who haven't encountered servers before -- I know many of you have -- think desk top computers on steroids. Okay? That's the best way I can explain it in a two-second answer. It's basically just a trumped-up desk top computer and they work great. And there could be issues, but many, many people use them.

Also, you can use your external hard drive. Yep. That's the same one you probably have in your office or your home.

You can also use what we call a RAID. Now, I'm not going to go into the ins and outs of a RAID. Basically it's a redundant array of independent disks. Here's the simple explanation. You take five or six -- or you can have less -- external hard drives, okay, like the one I just showed you. You link them together, okay, like the kindergartners you see below who are holding on to that rope, you do the exact same thing with external hard drives, and by linking them together -- and yes, this is a simplification -- you make your external hard drives into super drives. They actually allow you to make redundant copies in one place. Okay? Which is awesome. It really is a great way to store things. And here's what it looks like, because you won't get kindergartners brought into your place. This is what it looks like. It's a box, and it has bays for five or six hard drives. You put them together, and it's a great way to store things.

Another really good way to store your digital materials is tape.

And yes, it's called LTO today. Yep. Think of your old-fashioned cassette from the '80s. Okay? Not the same.

We've improved. Huge amount. And it's different. But think of that and you get the concept. And, in fact, from what I've been reading, LTO, or tape storage is very stable. And I was

surprised because I thought, oh, tape, it's always been wonky, you know, but it really is very stable. So look at that option. And also, another option that you'll see a lot, especially on the lists, is the cloud. Okay. What is the cloud? And Martin, I'll answer your question in one second, right after I deal with the cloud. The cloud is basically you storing materials on someone else's server. That's all it is. You're taking your materials and putting them on somebody else's server, and you have no clue where those servers are in the world. Okay? You really don't. They could be anywhere. And I just want to mention quickly, because I'm not going to go into it any further, cloud storage can be very cheap, but there are major, major issues that you have to consider if you're looking at it. Okay? So just Google it. You'll -- or you can ask me, you can E-mail me. You've gotta look at some issues there.

And we're at our last piece of pie, but before we get into that I'm just going to hit the questions quickly here.

Martin asks, do we need to do anything to preserve the physical -- oh, we've already done that. Thanks, Martin. It's been a long day.

Miles, awesome, have you used Sony's optical disk archive? No, I haven't, and I was just reading about that, actually, last week. Sony has brought us out some new technology -- well, not really new technology, but how would I say it, probably a muscle

car of storage. That's how I'm going to say it. It's the muscle car. It's like an old Mustang or cold Camaro of storage. And people have been talking about it because it's another way to store things. But, as you'll notice, I didn't mention any DVD or CD storage. That is possible. And yes, you can do that. But for me, you can use those as one of your three storage options. I just -- I wanted to give you the other options because I think if you can use those -- and maybe have a -- you can use the Sony optical disk if you want, you know, once it comes out and we've got it down to a price point that you can afford, sure, you can use that as well. So just so you know, you can use DVDs, you can use CDs. I just didn't mention it there because those are the other options that for me, I like them a bit better. But you can really use anything. You really can. As long as you use three different things. That's all I'm fine with, you know, whatever you use, go for it.

Okay. We're on to the last piece of the pie. Boy were we hungry. And boy was that good pie. And luckily, our last piece is access. It's something we've all dealt with because as you all know as librarians, as archivists, as people who work in other fields, access is crucial, and we think it is, too.

But it can get tricky sometimes, okay, as an archivist, I deal with restricted materials, I deal with copyrights and rights issues, and librarians, you also deal with copyrights and rights

issues, too. So sometimes we want to provide access, but we have restrictions, and it sometimes does happen. But when we can give access, we're happy to do so. We can either do it manually, you know, putting files -- digital files on a memory stick, plugging it into a laptop, having a researcher come to a reading room. Or we can use some pretty nifty software programs that have been developed for digital -- for researchers to access digital material. I think the most common one, you'll see the names here, the most common one is content DM. We use it here at the Ransom Center and this is one of our digitized materials for Dr. Cavendish and his flying cage and canaries. So software is great. It's come out and it's made our lives easier.

So, guess what, that scary ball, that's pretty much it. It shouldn't be as scary to you now.

Now, did I go through the entire scary ball? No. As you read this note, you'll see, I've touched on the other things that were in that scary ball, but for this talk, I didn't focus on them specifically, but if you want that link, if you can't get it now, just E-mail me or tweet me, and I'll get it back to you. We've covered a lot of the other parts of the scary ball. So here you go. The scary ball. The digital material, that's right in the center, right, everything revolves around it. We

create that digital material, or we receive it. And then we -- or we can select it or acquire it. Okay? Receiving is either. We appraise it, right? Especially as archivists, we appraise that material. And also as librarians, you can select it. We describe it. Archivists arrange and describe, librarians catalog and index, for the most part.

And we preserve it. And as you saw, and as all the questions came up, you can see preservation as many aspects associated with it, as well as migration there on the corner. And, of course, there's storage, and storage is part of that preservation process. Remember, it's not preservation itself. I hope I got that across.

And finally, access. We know access is important. That's why we're all doing this. Just so you know, I didn't cover these two things because really, as a digital archivist, I will likely not deal with disposal or reappraisal for another five to seven years. Okay? So it really doesn't come in to most realities right now. Some of you may be doing it, but most of you won't. So, there you go. Congratulations! You now know the basics of the digital curation lifecycle! So you now know all the parts to -- or all the pieces of the pie, and now, any more questions? Because I know I went quickly through that. So we've got some time for questions here.

Great. Glad the archival connection is coming through, and I tried to make it as librarian focused as possible today as well. And again, thanks for everyone -- oh, yes. And a copy of the presentation. We're going to be providing a recording through ALA, and if you want a copy, I'll probably post the PowerPoint, too, if ALA is okay with that. I'll post it. And just, if you're on twitter, follow me on twitter, or just come to my twitter feed. I'll post it in about a week. And if not, just E-mail me if you want a copy of the presentation and I'll PDF it for you. No problem.

So I just want to thank everybody for coming today. I really appreciate your time, and keep asking questions because we're here for another few minutes, so if I can answer anything else for you, or comments. Anything like that. Great. Super. And I appreciate all the kind words. That's great.

>> This is Susanna. I just want to, while we're waiting to see if there are any additional questions, I'd like to thank you again, Lisa, for such a rich presentation. Also thanks to ALA and Allison for facilitating the webinar, and in about a week, we'll post a recording of this presentation to the ALA connect page and we can also post a copy of the presentation, the PowerPoint or PDF, whatever you prefer, Lisa. Also, look out for an announcement, soon, on our next webinar for April, which is an April the 30th at 2 p.m. central, and it's practical data

management presented by Kristin Briney from the University of Wisconsin Milwaukee.

>> LISA SNIDER: Super. That's great. So I'll post the presentation, then, with the recording. I'll just make it into a PDF, and, you know, let me know. E-mail me, let me know what you thought. How you felt things went. Or if you need any more information.

Super. Well, looks like no more questions. So thanks, everyone. Thanks for coming, and if you have questions after the talk, feel free to E-mail me or tweet me on twitter. I'm archives matter.

>> Thanks, Lisa.

>> LISA SNIDER: Thanks so much.