

>> Jaime Mears: And I've seen a lot of memories. I've had the privilege of seeing a lot of memories walk through the lab in the past couple of months. We have photographs. We've had people bring in 16-millimeter film, zines, paper. Slides are really popular. We also have magnetic media for those 80's babies out there. So, you can see here this is my VHS-C collection, and the audio cassettes there are actually a part of our punk archive at D.C. Public Library that were donated. And then increasingly we have memories on things like this: So, cloud platforms on computers; Instagram; floppies, which we heard a lot about before. And another thing that I think we need to recognize is that it takes not only knowledge but money to be able to give people control of their memories and have them be able to steward them, which is something that I think Lauren Algee at DCPL Archives and Nick Kerelchuk of the labs – I think they had a lot of foresight in creating this project that I was lucky enough to be able to take part in this year. The DCPL has a history of labs and continuing education. We have a fabrication lab that has 3D printers in it, a laser cutter that you can use for free, a studio lab for musicians to record inside of; and I think that they saw this problem of people stewarding their archives as something that could be answered with a public-facing lab, giving people the tools to be able to do it, the knowledge, and also connection to obsolete hardware that's more and more expensive for people to purchase. So, the Memory Lab is three things. It's a place; it's a group of resources; and it's a series of programs. So, this is a picture of the Memory Lab if you want to see it in your local. It's at Martin Luther King, Jr., Memorial Library in Chinatown. Some facts about the place: You can go in for three hours at a time. It is do-it-yourself. You use a set of public-facing instructions, and we have some lab staff that are available to answer questions if you run into trouble; but for most of it you're just in there on your own. You can transfer eight formats. You have to provide your own storage. We offer twice weekly orientations, and so far the lab has been booked for over 650 hours since it opened at the end of February and it is the number one most used lab in our suite of labs. It's also a group of resources. So, this is the lib guide. This is kind of the heart of the project. This is where people go to look at those public-facing directions. It's had over 5,000 views. It's the most viewed lib guide that DCPL currently has but it also has other tabs that I thought were really important for the project. So, how do you take care of your physical stuff? How do you take care of your digital stuff? And, lastly, a tab on how to build your own memory lab because we've had other organizations interested in knowing how to do this. So, here are some of the resources that we made over the year. On the left you have this already. It's a zine that I made with a librarian who can draw because I can't. And this was a way to make personal archiving accessible to people and fun, which is a big challenge. My blog as well has been followed I think by other people that are thinking about perhaps having a transfer station. I've had 1,600 viewers in 65 different countries, and on the right is an example of a wiring diagram. I created some public-facing wiring diagrams that you can find on the site because I found being new to AV transfer that I had no idea how to hook up any of the machines and I couldn't find understandable documentation out there for a layman. So, I made some and I hope they're

useful. And then you'll also see a resource that was created by a librarian, which I'll talk more about in a second. We did 12 events teaching people about personal archiving at 7 different library branches. We had 6 partners and overall we had 411 attendees. So, one of the greatest parts of this project – and some of the staff are here today – was I was able to teach 50 librarians and library assistants a three-part series on digital preservation because even though the transfer lab is in the central branch, we really needed people with knowledge to be able to answer questions in the neighborhoods and teach classes. So, I developed three class packets with slides instruction that they can use and we have some of those programs being taught by librarians starting next month in several branches. These are some of the librarians here learning. One minute left. So, recommendations going forward for the project: We need to expand the lab. It's frustrating I think for some people because they try to book it and it's already fully booked a month in advance. Always increasing accessibility would be wonderful. A mobile archiving unit is something that I'd like to recommend. You'll hear Caroline Catchpole talk later about their mobile scanning project. That would be great. Maybe reading floppies like Jason was talking about. Having a van pull up to your neighborhood and grab your floppies. Continued regular programming on personal archiving is necessary. We discussed doing a meet-up group at the central branch monthly around personal archiving where we brought preservationists in the area together with archivists. I think this would be a great partnership with local universities. If anybody wants to learn about AV transfer or digital preservation, they have a hands-on lab they can try it out in. And then we really need to think seriously about measuring impact over the next several years. How is this lab and how are our programs really helping people who live in D.C.? So, that's it. If you have any questions, talk to me later, and next up is Valerie.

[Applause]

>> Valerie Collins: Alright. Hello. So, I am Valerie Collins and I am very excited to speak to you today about the work that I did at the American Institute of Architects. So, this is a photo of the Octagon, about 100 years ago the original headquarters of the AIA. This is taken from New York Avenue, Northwest and 18th Street, and I could make a comment about how quickly the landscape changes, but this is just to give you something pretty to look at while I talk for the next five minutes. So – Oops, sorry. So, I was tasked at the AIA, when I was working with the AIA archives, to help build a repository specifically for the born digital materials, and these were the materials that staff members were creating. And we were focusing only on the born digital records at this point because that was where the critical materials were being created by staff members and where there was no system in place to preserve them, and this was not for any lack of effort on the part of the AIA archives. That was what I was there to help them work on and to help fill this gap between the creation of these records and preserving them. We knew we would not only have to choose and implement a repository system, but we would have to work closely with staff members and that without the staff members of the AIA buying into the

system and making the necessary changes in their work habits, we would never be able to close the gap between the creation of a record and its deposit into a digital repository. And instead they would get lost in the middle piece on shared drives and hard drives. And so while the project involved ten months of like meeting with departments and as many people as we could get a hold of, pitching the repository to them, inventorying the state digital records at the AIA, and determining what the final organizational structure for the repository would be, I'm not going to talk a lot about that. That was an important piece of the project. It had its own challenges and its learning moments, but I want to talk a little bit more about the observations and experiences I had with choosing, testing, and implementing the actual repository system that we used. That was what dominated the final half of the project and it was also the place where I found I was most challenged with keeping an eye on our end goal, with all of the requirements that we had for what the repository should be and sort of all the shiny distractions that kind of happened along the way and how to work within the constraints of what we had. And so kind of the first lesson I had, I suppose, is that in this process of choosing a repository to work with visual materials is that you can never pull the people out of the system and that we had to think about how people are going to use this every step along the way. And so it still moved around again to how we were going to use it and how staff members were going to use this repository, which meant that my project really essentially rested on what do people do? How are they going to interact with the system? How should they interact with the digital objects? Or at least making my very best guesses about how we can anticipate people to use the system, particularly when I had no additional guidance, when our organization had no legacy system that I could look at to see where the frustrations and the pain points between users and the system was. I could run as many focus groups or user experience testing on the system we did choose if they want and I could get a lot of "great", "that looks good", "I think I'll like that"; but that doesn't always translate into how people will actually feel about the system when it's part of their daily work. So, we ultimately went with Preservica for a repository system, and the process of learning and testing the system was equal measures of a great challenge for me filled with unexpected pitfalls that I had to navigate. And so the AIA doesn't have an in-house IT team that could deal with some of the problems that occur when you're trying to deploy a repository, and so that meant that any issues that we had it went to the really wonderfully responsive Preservica support team or I solved them, and those were challenges that I was eager to have and eager to work with. For example, one of the things that started small, what should have been kind of easy to get around, at least I thought, was that how are we going to make the system display the metadata that we wanted. And so this was sort of an educative process that took a good chunk of my time, working with all the different interfaces and really trying to understand how the system worked. And perseverance paid off in the end on this particular instance. There was a lot of frustration and hair pulling along the way, but many institutions do not have the luxury of having staff members or residents like me who have the skills and the time to help with these bumps

in the road that go along with a system deployment. And so kind of looking back at what we've done over the past year – and one of the hardest things I came up against during this project was that, you know, sometimes despite our best intentions, the methods we have of providing access to digital material is governed solely by the systems that we have available to us. And so although by the end of my residency, we had created a repository organization that met the needs of the archives that was intuitive to staff members – This is sort of a quick screen shot of the staff interface, which is not the same as the archives interface, and this is not the home page but a different part of the rousing mechanism, and so this is sort of like a – I did something [laughter]. And so also one of the hardest lessons to learn was that sometimes extreme effort that goes into these building systems kind of translates into this apparent simplicity, which is great because you want it to be simple for your users. So, I was able to at the end build a box to put all the things into, and that whether they're born digital or made digital, these items now have a home at the AIA; and I guess that means now it's time to preserve them. So, thank you, and next will be John.

[Applause]

>> John Caldwell: Good morning. My name is John Caldwell and I am the NDSR resident in the U.S. Senate Historical Office and I'm going to try to explain the nuances of how the Senate is organized in 30 seconds [laughter]. So, Elizabeth, Karen, and Heather, hopefully I do this well. So, I work with two different communities within the Senate, the committees and the member offices. Committees have a statutory mandate to preserve their permanent records and at the end of every Congress, transfer them to the Center for Legislative Archives at NARA. Members own their own records and can determine the disposition of those records when they finish their service in the Senate. And so in these two very different worlds, I was tasked with studying the existing digital preservation practices and work flows in the Senate, test a variety of tools that could work in our environment and improve digital preservation across the institution, and then make recommendations for improving digital preservation and standardizing the process of transferring committee records from the Senate to the Center. And so that started with a lot of information gathering. I developed a survey to ask systems administrators, archivists, staff assistants, anyone that had a role in preserving electronic records. Questions about storage, about processing, email archiving, social media and web archiving, use of mobile devices to generate records material. And I met with 18 committees and 27 member offices or, as I just found out, 72 different people across the Senate to get their input on all of these issues; and then taking that information from the committee archivists, I mapped out everybody's work flows. And as you can see, there's a lot of variety here. We have a couple that are very complicated and some that are much more straightforward, and so this was very good information and really had a lot of bearing on the tools that I looked at, which took up the bulk of my time in this project. I tested 26 different utilities which focused on File Fixity; and, of course, DROID also does file format identification. Data visualization, email archiving, and conversion with the case of the CYK

parser and the Smithsonian Institution Archives' dark mail program, which is really cool if you guys get a chance to play with it. Other accession software, such as the Duke Data Accessioner and the Curator's Workbench. Copy utilities, directory lists, file rename utilities, and BitCurator – specifically, the bulk extractor feature for identifying PII and also potential uses for disk imaging. So, once I had all of these I had to make a new work flow and clean up all of the ten different ways that Senate archivists were doing things. And so this is what I came up with. It's a little cleaner and it really tried to leave a lot of flexibility so that archivists could within their own office environments do the description and the processing work that they need to do for their own committees while still implementing things like Fixity, identifying the file formats, and just getting everything a lot more streamlined, also encouraging some data visualization and email metadata capture. But work flows aren't enough. This is a very pretty picture that was a pain to make in Excel but it's a pretty picture and it needs a little bit of help to explain some of the more technical aspects. So, I created a series of guides and check lists to add additional information; and with the check list, it will allow the Historical Office moving forward as new tools come in or as policies change, either on the Center's end or internally, to tweak what needs to be tweaked while still leaving the process itself alone. And I actually learned this the hard way. The Center switched from MD5 to SHA-256, and I had MD5 all over all of my work flows and everything and having to go back in and change all of those – you only make that mistake one time [laughter]. And I'm going to show you just a few of the recommendations that I'm making for the Senate. One is to establish archival control much earlier in the life cycle. The Senate Historical Office and the archivists have always been very good about capturing descriptive information but not necessarily technical, metadata especially, and so trying to add things like Fixity and more accurately identifying file formats will really help get a better handle on the material as it moves through the transfer process. Stronger collaboration with technical staff. As we know, preserving digital records isn't just the bits and the bytes. It also has to do with understanding the storage implications and transfer protocols and hardware and software. And so my project really helped open the door to working more closely with systems administrators across the organization and, in some cases, actually led to office archivists and systems administrators collaborating during my interview with them on how to troubleshoot some of these issues. We've always had a wonderful partnership with the Center for Legislative Archives. Both our office and them are constantly working to improve guidance, improve their descriptive work; and we have been sharing with them all of the information so that everyone is on the same page. Periodically reassess tools, open source note feed tools change all the time. Fortunately, Congress ends every two years so there's a built-in reassessment period. And then finally, this idea of a Senate Information Manager. Digital archiving is much more than just trying to describe and preserve, but it is also managing the information from creation all the way through its active use; and the Senate Information Manager, as I hopefully envision this, would also capture the tasks and knowledge of the office and really act as a true information manager and

resource. So, that's my project in a nutshell, and on to Jessica.

[Applause]

>> Jessica Tieman: Hi. I'm Jessica Tieman. I'm the NDSR resident that's been at the Government Publishing Office. The project that I had for a year was preparing GPO's FDsys for an ISO 16363 audit. That being said, rather than talk about GPO's audit, I actually sort of want to share what my greatest takeaways are and lessons learned from this process, some of the things that I think the digital preservation community can sort of learn in general about repository certification and through other repositories doing this process. So, the first thing that I want to talk about is how we talk about standards. This was something that I learned really early on going into my residency program. I feel like when we're in graduate schools and graduate programs for library science we were introduced to like the OAIS framework and were constantly inundated with the idea that it's a framework. You know, your repository doesn't have to look just like the OAIS or so forth. But the ISO 16363 standard and TRAC and things like this, they're not talked about in the same way very frequently, at least for me it wasn't. It's very prescriptive. It's a list of things that you have to do, but even though standards are not theory, I do think there's a lot of flexibility in how they can be implemented and definitely the look and feel for all repositories is different in how they implement a standard like ISO 16363. One of the ways that I came to this conclusion was through interviewing other repositories that have done assessment. I interviewed six repositories that did the TRAC audits with the Center for Research Libraries and I also interviewed six repositories that are OAIS compliant but have not pursued TRAC or ISO 16363. I asked them these series of questions. There were a few other questions, but in general, "Why did you do assessment? What type of assessment did you choose? How did you prepare for it? What did you learn from that process? What were the biggest benefits?" And so forth. "Do you have any recommendations for me?" Across the board, all six repositories that had done TRAC – a lot of consistency. They all said the same things, gave me the same feedback. By and large, these were the three main things that I took from that experience. First of all, before you begin an assessment like TRAC or ISO 16363, anything this big and large scale and time and resource intensive, it's really important that from the very beginning you clearly define why it is that you're doing this. You have to anticipate that on a daily, on a weekly, for a whole year you're going to have to be explaining why you're doing this audit of a box basically, like Valerie said. So, it's really important that you can clearly define what there is to be gained from it, what it means for you specifically to be trustworthy; and some ways the only way to define that is to define it by saying, "What's at risk by not being trustworthy or claiming to be such?" So, part of this is also having a really clear definition of your designated community. All of the six repositories that I spoke with or who participated in my survey said that the designated community is much more important than I think people give it credit for, and within these standards often it's defined as like the users, but it's actually much more than that. In the ISO standard,

there are multiple criteria where the measure of how successful your repository services are – it's only measurable against how well it's meeting the expectations of your designated community. So, if you don't have those expectations clearly defined, it's hard to measure, you know, how adequate your repository is; and you can't say that your user community – or that your designated community – is everybody. Like, for example, for GPO – GPO serves the American public, but that's way too broad. So, we have stakeholders in Congress. We have academic users. We have the library community. It's very important to identify all of those sort of sub-groups and then also define what their technology skills are, what their resources are, what they like the interface of the system to look like. Do they understand what your services are? You know, are your services transparent? And then when you're sort of acting as auditor, like I did for a year, you have to have a consistent way of objectively evaluating your repository system. So, the most simplified way of looking at that and being consistently objective requires you to sort of think about these things. For all of the aspects of your repository, do you have documentation for it? If you have documentation, is there evidence that what you're doing actually matches what the documentation says you should be doing. Are both of those things – both the documentation and the evidence – are they transparent to your designated community? Do they actually understand what's going on or is it all just like a secret magical thing that they don't really understand? And even if you have the documentation and the evidence and you're transparent, is it actually good? Is it meeting their expectations? Is it adequate? And then what do you define as success in terms of adequacy? How are you measuring that essentially? So, it's really important that you can do this, and for the ISO 16363 standard you do that 109 times. And why is this important? So, I kind of go back to this experience was important for me. It's clearly important for GPO to sort of perform this audit, but at the end of the day, what about repositories that are not actively pursuing this audit process or can't have a resident to do this for a whole year for them? There are still very good reasons to do an assessment like this or similar to this. And these are the three main reasons I think. First and foremost, the ISO audit requires that you share out the audit findings. It also requires that you're transparent. So, it's an opportunity to put your policies and work flows out in the public, and for the visual preservation community this is really important. As I was doing this audit, I found that there were very few repositories to sort of go to for examples. It's becoming more and more common for sure, which just within the past year more repositories have definitely been doing this. But when I started, there weren't very many to go off of. Secondly, there's definitely going to be silent busting when you do this. You're going to be required to reach out to your IT department. You're going to be required to talk to upper management. You're going to be required to talk to groups that didn't even know you were doing digital preservation until you started talking about auditing it. So, that's always a positive thing. It's good to be sort of forced into that sort of networking activity. And, lastly, repositories are really dynamic. Your repository system is changing probably year after year. Your policies are updated; your systems change. The type of

content that you want to ingest probably changes a little bit. Your metadata changes. Staff change a lot. So, it's important to sort of do an assessment like this to identify, you know, maybe you forgot to update that one thing or how are you making sure that it's still adequate as it once was? And so secondly, even though those are all the reasons why it's important, you might think, "I can't do that. It's expensive; it's time consuming." There's a lot of other alternatives and a lot of things that I think are really worth looking at. First of all, I think a good alternative is a DRAMBORA audit, which is more of a risk assessment. It's not as inclusive as the ISO 16363 audit but it's still really satisfactory and it's not as time consuming for sure. Secondly, even if you don't want to do an audit, you can at least do somewhat of a knowledge sharing session. I did interviews with the staff at GPO, sort of like pretend audit interviews and it was an opportunity to get like systems people who work with our infrastructure environment sort of being interviewed about our policies and so forth. There was no – Like I wasn't keeping score on their answers, but it was an opportunity to realize, "Oh, I thought I knew that until you asked me exactly and then I had to like think about it again." Another really fun thing that I learned about is that there are repositories out there like the University of North Texas and the University of Florida libraries, they are actually doing their own internal TRAC audits and then they're sharing with each other and sort of doing like a peer review audit system. And I think that that is an excellent model. It's way more affordable obviously, way more doable for academic institutions, and it's another opportunity for universities and repositories to share out their policies so that we have more examples to go off of. So, please go do some audits or assessments and have fun, and next is Nicole.

[Applause]

>> Nicole Contaxis: Hi. So, I don't want to get in the way of your lunch because no one wants to do that. So, I'm going to try to work through this pretty quickly for you. First, I am Nicole Contaxis. I've been at NLM for the past year working on NLM developed software as cultural heritage. I just want to point out this piece of software that you see a screen shot of is called, "How to Grateful Med". My project is sort of a pilot program. So, this is the piece of software we ended up sort of experimenting with for the whole year. I don't know if you can tell. I love these graphics. But this is actually the building you're in right now. So, that's kind of exciting. Alright, I'm going to just quickly survey the goals of the project and then for this discussion I'm actually going to focus on why we picked the piece of software that we did. First, I surveyed other cultural heritage organizations and looked at how they were preserving software or, as in most cases, not preserving software. Then we looked at NLM's developed software through archival research and interviews. NLM is kind of a funny place. There are people here who have worked here for 30 years. They've developed and worked with the software I was studying for a very large amount of time. There is one staff member in particular whose office was a treasure trove for me, who had file cabinets in front of other file cabinets of stuff that was very helpful for me in my project. So, that was great. Then we had to identify software selection

criteria and pick a program. That's what I'll be talking about more in depth in one second. We also created a draft curation work flow using this piece of software and then I worked with the digital repository working group to prep these files for ingest. Let's move forward. So, selection. I basically understood this in two separate ways, and don't worry; I'll get to why Doogie Houser is up there. The first is historical. So, I understood this is basically one very simple comparison, which is there is some software that is historical in and of itself, and there are other pieces of software that sort of demonstrate historical things. So, Doogie Houser in one episode features Grateful Med. Grateful Med is the program that was sort of a crowning achievement here at NLM. It opened up NLM's bibliographic data to a much wider set of users. Any time I've gone to a conference and my mentor has gone to a conference and people hear what we're doing, they ask if we're preserving Grateful Med. It might seem weird that a search interface would be so popular among so many people, but it is something that people have used for years and had a big impact on how people did medical research. It also had a cultural impact, like Neil Patrick Harris is on camera talking about Grateful Med. That's amazing; it's really fun. So, I obviously wanted to preserve Grateful Med, which brings me to the second criteria, which is technical. The reason why we're doing How to Grateful Med and not Grateful Med itself is because the technical difficulties with Grateful Med far outweigh the importance of its historical – Well, I shouldn't say far outweigh. What I will say instead is that How to Grateful Med could demonstrate how Grateful Med worked. It would have been great to preserve Grateful Med itself, but for several technical reasons it was too difficult to do within this year. Now just to give a quick overview of Grateful Med. Grateful Med functioned as a search interface and it used pre-internet networks to access NLM's data. Now that data has been lovingly cared for by NLM since that point. The piece of software I'm working with is from 1988. Since that point, every year the medical vocabularies are updated; the format is updated. It is taken care of so that every doctor can access it and do their research. What that means for me is finding a copy of the data in the right format that a program from 1988 could read it was kind of out of the question. We had a bunch of different strategies to try to fix this but in the end, what we decided was with How to Grateful Med we had sort of example searches built into that tutorial. So, then at least historians could see how the program functioned and they could get a better understanding of it even if we couldn't make the program itself function. Oh, one last thing and some stuff you have little control over. I want to point out that through my inventory of NLM developed software, I located about 120 programs between approximately 1962 and 1999. Out of those approximately 120, we located about 9 and we had working copies of 9. So, that obviously really affects your selection criteria, and we didn't have control over that. It was just what we found and what we could work. I also want to take a quick second to talk about the process of selecting this piece of software. Doing this sort of software preservation and building up a new work flow in an organization like this requires so much group work, and so when we talked about this, we had a lot of meetings where I talked about the history of all the pieces of software that we had access to and sort of gave more

of a background on technology history so the people could understand how these different pieces of software fit into larger historical narratives. We got people in the room from all over NLM to give us their opinions and to give us their experiences of the software because, again, some people have worked here for a long time. And we were able to get a lot of input and it was great. But I will tell you there was one other really great thing about it, which is doing those meetings and having people give their input about which piece of software we should preserve was a great way for me to find really enthusiastic staff members. Everybody here has a lot of work to do, and I get to come in for a year and I know that the organization was really excited to have me but I'm like extra work. So, I was able to find the people who were sort of enthusiastic about it in their own ways and for their own reasons and they have been really instrumental to this program and to my success on this project. And I think including them in the selection process was a really great way to find who was enthusiastic and who would be willing to, you know, sacrifice their own time for me. And that is it. We are only running two minutes late. I think we will skip questions. At the panel discussion at the end of the day, if you have questions for the residents, we would be happy to answer them, although we won't be sitting on the panel. I'm going to move right into logistics announcements, which is lunch. There are two cafeterias close by. One is directly downstairs; the other is across the street. If you're worried about getting lost on NIH campus, we have little maps where you all signed in, and you can take it and it will show you which way to go. If you decide to go off campus for lunch, just note that you will have to go through security again to come back. So, I would suggest staying on campus and going to one of the cafeterias. We will be reconvening at 1:15; so we look forward to seeing you then. Thank you very much.

[Applause]