

Please stand by for realtime captions. [Captioner is on hold, waiting for event to begin.]Hello everyone this is Jamie at TPO, I'm doing audio check so that you can adjust your volume.

>> This is Jamie with TPO again were doing another audio check we will get started in about four minutes.

>> Good afternoon and welcome to the FDLP Academy webinar, my name is Jaime Hays, with me here today is Ashley Dahlen, handling any technical issues please feel free to message Ashley if you have anything, first Mary Moulton, join the NTL in 2011 following a career in the private sector, she leads on the enhancement of digital repository and discovery use of NCL resources, cochair of the [Indiscernible - low volume] alliance, for free public alliance, from 13 federal agencies, we also have Jessica team and, a preservationist library at GPO, and serves as the ISA 16363 -- Representing GPO, and data duration products, and a volunteer powered organization that serves the community. Jessica is a graduate of the national stewardship residency program, sponsored I MLS, holding [Indiscernible - low volume] in science from the University of Illinois, and Urbana-Champaign. I will walk you through a couple of reminders. Nothing out of the ordinary. If you have any questions you would like to ask the presenters or technical issues please feel free to use the chat box in the bottom right-hand corner of your screen. I will keep track of the questions at the end of the presentation Jessica and Mary will respond to the questions. We will email a link to the recording of the session for everyone who has registered to the webinar, and also -- Registered to today's webinar. And if anyone needs additional certificates please email FDLP outreach, and include the title along with the email addresses of those needing certificates. If you need to zoom in on one of the slides being shown by the presenter, click on the full screen button on the bottom left side of the screen come mouse over the blue bar so that it can expand and then click on the blue return button to get to defaults review view, -- We will let you know when the survey is available, and the URL will appear in the chat box, we appreciate your feedback, including your presentation style on the webinar, we will pass over the mic to Jessica who will take us from there.

>> This is Jessica, as Jamie said, I am a preservationist at GPO, I've been here for three years. A large part of my work here has been preparing GPO for external ISO 16363 repository audio audit, specific to the depository library,'s program community, if there is any questions on preparing for this process certification, because we already provided information about that it was brought to my attention that there are quite a bit of people out there who ask this question what is a digital repository in the first place, and versus any regular repository, what is the difference and what makes it a trustworthy digital repository, I will go through a few slides, before Mary's presentation my part of the presentation will focus on theory and background history of trustworthy repository, Mary's presentation will give more of a depth look, as self assessment for another repository. So to give practical use of it.

>> So the first slide. It is supposed to be an outline of those things I will talk about what it means to be trustworthy, what is the history of the repository certification? What I know in terms of why repository would pursue, and what makes it important, why is it worthwhile, why is that the cost and expense and time that goes into it. I will list off some of the digital repository efforts that are already out there that exist predominantly here in the United States. The first slide. It lists the three main things from my perspective which define a trustworthy repository. These are the requirements that the audience I hope takes away with them today, this first requirement digital repository, is going to be representing the actualization of long-term values and commitments to long-term preservation. I will go into more depth about why this is it, and what this means. The second part. The repository is going to be a realization of community transparency. Reliability and confidence. This obviously means your repository is considered to be trustworthy if your community understands what you are doing, are you transparent about your practices and your goals and commitments, are these reliable? Does the community have confidence in the repository itself and what you're doing?

>> Once it is met this first second requirement, it can be objectively verified against standards and best practices. To take these three requirements in depth. I will talk a little bit more about the history of that

first one. As I said the first one, this actualization of long-term strategic values to long-term preservation. This is really critical. One thing. It's an institution that if they have a repository or digital library, or does some kind of digital preservation. If the institution as a whole, all of the components of that organization are not really dedicated long-term to a digital repository. Technical infrastructure or ongoing management and policies. The hardware systems that are a part of it. You can't really call this a trustworthy digital repository. Purely from the fact. There is no guarantee that the repository itself will last very long, there is no guarantee that it will also be sustained in a part of the organized overall mission. I want people to take that into consideration perspective perhaps, you are an organization that is doing preservation of some kind. Or digital archiving of some kind. It helps to keep in perspective. Is this the type of preservation we are doing that is just a one off project, a short-term solution, or to me in the end, or a part of the organizations long-term strategic commitment? Is it part of the organizations mission? The history behind this. It goes back 20 years. Truthfully. 1998. There was a survey conducted on behalf of the research library group called digital preservation's needs requirement. It surveyed over 50 repositories. It found they actually did not have a defined statement about their commitment to digital preservation and they lacked long-term strategic policies. So it was one of those things. Over 50 repositories said we are doing preservation two thirds of them didn't really have an organizational commitment to it, didn't really have these bits and pieces that guarantee that this will be a long-term thing. After that survey. The digital library Federation did another survey along the same lines. 14 of the 21 libraries that responded to that library also didn't have digital preservation policies even though they were doing preservation activities. From these two surveys. Some of the groundwork to digital repository certification and best part is standards were put into place. By Ann Kenny, and Nancy McDonough -- Ann Kenny, and Nancy [Indiscernible - low volume], and to build out those policy frameworks. Those long-term commitments these things that are necessary for digital preservation to be successful. What they put out was a theoretical concept defined organizational stages of preservation, this diagram, that a lot of the librarians have digital library training, have seen iterations of, these leg it stools symbol or metaphor where digital preservation is the stool, then it cannot stand on its own if you don't have the technology or the resources. The organization support to prop up this statement, and this digital preservation, this is where we get the earliest language that says digital preservation isn't only how you manage a file, it isn't only that you are backing content up, or archiving it, describing it or talking about it as if it is archival thing. Digital preservation is this broad term applied when you have organizational peace as well as the technology, and the resources. It is one thing to say we will do this to preservation, but we will not have the budget that will allow us to do it 10 years down, if you do not have the budget, you are just parking it with the hope that you can continue parking it if that makes sense.

>> So the early paper that talks about three-legged stool, and the five organization digital preservation's, go into detail about how to create long-term requirements. How to develop a repository, according to standards, all of these steps. They built up this framework for early assessment. This is early 2000. Where we get this first language about this need for organizational pieces. The second thing. That I talk about. Digital repository. Necessarily is transparent with your community your community has to rely on you, have confidence in you. The first earliest places we see language like this, in the trusted repository audit and checklist document published by the research center for libraries. The track checklist came out in 2006. 2005 I actually have I apologize. This checklist that came out, this first real document that looks like a standard. It is in many ways a predecessor to the ISO 1363 that lives and looks like the national recognized auditing standard, this is the predecessor to that standard. Both of these standards talk about the significance of the community of users. The terminology specific to the standard the designated community, and an acknowledgment that the repository is only as effective as it meets the need of the community for example. In the ISO standard, if you are a digital repository preserving rock and mineral data. To some research. You have said your designated community is geologists. If you are

defining a designated community to be geologists, you have this commitment to this defined community that you have defined, you have a commitment to having the educational awareness of what those users need and expect from you. It's not enough to just part this mineral data in the repository, you have to provide a data dictionary and provide references to international standards for rock and mineral information, you have to make it usable in other words assessable, relevant to your user community. And you also have to to the best of your ability publish what you're preservation practices are in your commitments. You have to make that information forward spacing so that your user community can no and understand what you are doing with it, these are sort of for instance, if these researchers are not only accessing the repository to use the data, but they want to submit their own research to the repository then they have to know how long will this stuff last? What file format will it accept? Basically can I trust this and cannot guarantee the service will be around? Do I have confidence in this repository? It is that relationship. It's not a one-time thing. You have to continually communicate. You have to have information that is free and accessible, understandable to what you're doing. That is what it is all about. Oh if you do not have that piece it's hard to argue that you have a trust Worsley -- Trustworthy repository. It is dependent upon the community.

>> So the third thing that we talk about. Once you have that organizational infrastructure to support that repository, and that confidence and transparency. You are therefore ready to have a third party certification. I'm not personally going to advocate additional repository needs to be ISO certified to be trustors a Desmet -- to be trustworthy.

>> The community of users they define their needs for the level of trust. I think it is reasonable to assume an organization that serves international groups of high prominence, and lots of funding, they probably have to speak to a higher standard that serves a smaller group in a single town or a varied niche group, they can define what their level of need is to be considered certified or accredited, or to have confidence they really do find that. For those repositories that need to have the highest level of confidence and certification. Highest level of trust. There eventually going to have third-party certification. There is a variety of different ways to go about third-party certification. One of my later slides I will talk about those opportunities and what they could be.

>> This is a snapshot of the development of the ISO standards that exist, I will not get too much in the weeds, this is a recap for everyone and a visual representation of the development of these activities over the past decades. In 1985, this was actually one of the earliest things that came to be. A report from the committee, it's not necessarily a congressional report. But a research group established, somewhat providing recommendations to the national archives. The national archives in general. This earlier report in 85, this is where we had this first mention of switching to the digital age, there is going to be a lot more stuff coming, and we don't know how to respond to it. That was to set the president of more formalized research and standards, in 2002, you get the research library group developing processes for different repositories, and 2005 you get the checklist I talked about which becomes away to perform assessments, not necessarily providing certification but away to provide assessments. Between 2006 and 2012, the research libraries CRL, they perform these track assessments, and ultimately six repositories, four of these repositories, in the United States. Two of them are Canadian, assessed by using the track checklist. Really a lot of these assessments just started as a dry run CRL selected was one of the earliest repositories, it experimented in the way to show how you can view an assessment, and they publish that with the process and what that look like. What that first test that's when they started doing other repositories as well. As CRL was doing those assessments, a separate group from CRL, went forward and started working on taking all of the information that was created so far the checklist. The research library group and papers, all of these other groups that were contributing to this Montgomery, who moved to MIT. A lot of voices that were from the digital community where I play, this group went forward to create an international standard. This international standard become publish in 2012, once the standard was published, audits didn't start happening right away because in an

additional standard, this had to be created which created the process to accredit certification bodies to perform the audit. You have the standard for the audit first and then the standard to create auditors to do the auditing. It wasn't until 2014 that you could even make someone an auditor to do this. In 2015. This group they are called primary trustworthy the standard they actually tracked across internationally from different repositories hosting these high level training courses that were teaching repositories how to go to about certification but they were not performing audits at that time. It wasn't until 2016 that the accreditation board finally announces there is processes to create auditors. That was in 2016. Since that time. We have become an accredited certification body, this only group, that public announced it, not the only but the first to have accreditation. At this time PTAB has the knowledge of performing these public knowledge has been the group that GPO has contracted to become certified as well. So the final thing as an overview about why the repository would do this. Why is it so important? As I mentioned. My personal feeling. Your designated community defines what extent you should do this. That urge or call really comes from your user community as your primary stakeholder. So why if I'm going to say your community says how important it is, why else would you do it. Or what they consider this. These themselves actually summit up pretty well.

>> They have language that says the funders of the repositories in those who entrust those valuables encoded to them need to know whether their funds or faith is well-funded, the stakeholders needs to know that there repositories where the of trust, particularly for a government agency for example, it's very important that there money is well spent and put to use, they might not -- Another thing to keep in mind your community of users and stakeholders, your users and those who fund you, advocate for you, those who manage or administer the organization, they may know very little about digital preservation or may not understand what goes into it. They may have some sense of yes it is very important to digitally preserve stuff, but may not value the day in and day out activities that you do. They may know it's important at the very least they need to know everything they are signing off on, being carried out and efficient, they can trust you even though they don't not know much of what you're doing. Having a third party saying yes everything that you are putting money into, hiring people to do. This is successful and efficient. This third-party body comes in and says, all of these things that your user community, and the employer's say that they are doing yes they are doing them very well. I have the opportunity to interview the repository managers at the six repositories I mentioned earlier, who went through this process. I interviewed them about the process. The benefits that they got out of it and if it was worth it. And then the mostly everyone had a positive experience doing that practice. They all said number one things they got out of it was preparing for the third-party assessment allowed them to consolidate all of their documentations. It may have been non-homogenous or living on different systems, doing this system requirement organizing it and that was very efficient to them as an organization and allow them to objectively convey that they were doing best practices, to their stakeholders in public. It allowed them to convey credibility and visibility of the repository, I know especially for repositories that use this with the scholarly, a group of researchers, it could be hard to communicate what it is versus academic journal and for those repositories it's very important to have that extra to say this is the benefit of the digital repository that you can only get from us. This is very critical for them. In terms of what else is out there. A third-party assessment or in general. There are a lot of different flavors for certifications. How to go about being trustworthy or trusted. Some of these I guess most approachable ways to go about this is to do a self assessment. In general I recommend doing this before doing any third-party involvement, I recommend everybody looks at from my perspective this is a good example. The University of Texas did their own assessment. They have a public document that walks through how they meet each criteria and what is in the checklist, it is very thorough and repeatable method for other academic institutions that don't have the resources or the buy-in, for a formal certification but they want to do some sort of assessment to show what they are capable of, and that they are doing the right thing. The other thing to keep in mind. I've been mentioning this ISO 1363 standard, the recognized

standard, there are other certification processes, that include and make reference to this. There is the core trust seal certification it comes out through the alliance. The ICSU data system, where they can do this data approval and do I so certification, -- ISO certification.

>> It is not as comprehensive as ISO 1363, I know I have known of at least some institutions that have an approval, and those that are more archived that use it. In general, it includes 16 guidelines that are pretty specific to managing data research or datasets. I know the University Consortium and research IPSR -- And the Center for repository, those are all the ones I know of that have the data fill approval already. In terms of [Indiscernible - low volume], we anticipate being the first government repository, and anticipate being the first American to finish this certification process, the only repository that has been announced with this certification at this time. In January of this year, the national cultural archives NCAA, hosted for the Center for arts, this is the only repository I'm aware of that has it. In terms of what is specific to the federal organizations, as I mentioned, the USPS arrow repository, they have long-term interest in other certifications, and that's what they have now. This working group also shares information about these process and there is a lot of interest in self assessment and long-term certification, this is where I will transition to Mary to talk about the library and their experience with some of these systems.

>> Thank you Jessica.

>> I appreciate very much what you have just talked about. It is a great lead in to how we are trying to have some type of assessment in our situation which is very different from yours. Let's see. I'm mostly going to be talking today about core trust field requirements for self assessment. Right now we are involved in a self-assessment, recently migrating our repository, I will get into that in some detail as we move through the presentation. We are established in 1998 as a born digital repository. Where the newest of the federal libraries, and we are the only national library that is entirely a digital library. We have digital collections that include research report. Datasets. We are part of the Bureau of transportation statistics when we were founded in 1998, it was originally to archive and distribute the data publications and data products for the Bureau of transportation statistics. We provide data services. Provide reference and research services, and where the host for the national transportation knowledge network and also coordinate regional transportation knowledge networks. These are comprised mostly of libraries at state departments of transportation. University transportation centers. In some library transportation related of course that are in the private sector. We are an open access digital repository, that means all of our items are in the public domain, they are available for reuse without restriction. We operate according to three mandates. The first transportation equity act, that founded, we were founded in 1998. It established us as a national transportation library. As I mentioned mostly to collect archives and deliver statistical and other information. In 2012. There was a major transportation legislation called map 21. That really furthered our mission and our focus. It specifically mentions we are a clearinghouse for transportation data. That we are a central repository for the Department of transportation research results, that is internal research that we fund as well as external research lastly in 2013. The White House office of science and technology policy issued a memo commonly referred to as the Holdren memo. Required all executive departments and agencies that then spend more than \$100 million per year on research to ensure that the public would have access to peer-reviewed publication reports and digital datasets that were the results of that funding. We have been operating on these three mandates. It seems like 2013 was just a short time ago. We knew that we had some issues with our repository. I was hired at NTL in 2011. This was a story about why trustworthiness is important. I came on board in June 2011. Our repository had been down for about one month. It was not available to the public. We had a hardware failure. Our servers were so old that we had to go on eBay to get replacement parts to get the repository back online. I was hired right at the end of this episode. My first experience doing outreach and getting out meeting people. That we are stakeholders and researchers in the water transportation community. People would say to me you guys are so

unstable, why should we give you our research reports. You can't keep it together. How can you show us you are a reputable repository? They had every right to say that actually. So I talked to my supervisor. As brand-new with the national transportation library. She said you need to start learning about trustworthy repositories. That started me long journey of this road how can I communicate to our funders our researchers and the public that uses the content, how can I communicate to them we are a reputable organization providing a reputable service. Last year. This was actually several years in preparing for this. We migrated our repository, originally would have been on a homegrown Oracle database. It worked really well for us. It had a great search functionality. We had a great ingest system and cataloging. It was a very simple user interface however. It really did not support discovery. The other thing it done -- It didn't support LinkedIn communications we knew with our public access memo. We had to draft our own plan for public access, with the release of the Holdren memo in 2013. We knew that we would not be able to very well offer a platform for compliance with that memo to our internal researchers as well as the external researchers. We began looking around for a solution. We ended up partnering with the Center of disease control. On a repository re-platform called DACs, open source platform based on Fedora. It is a calm open source platform and the other partner in this endeavor is Noah -- NOAA -- And NOAA came afterwards, we went live and had a soft launch in June of last year our public lunch was around September. This was relevant what I will be talking about. I am very proud that we did this. We migrated. About 35,000 items. We had a tremendous amount of cleanup to do on our metadata every time you move data, you find out that you have to do something to it. This was no exception. We have a wonderful repository now. A great relationship with NOAA, the CDC, our users are very happy with this. It is a great way to showcase our collections and her wonderful work that the researchers are doing really good for discovery. We are super happy with this. So here is a little bit about the collection. Right now we have 35,000 digital objects. 10 collections. From all these various operating administrations and the Department of transportation and they are searchable. They are discoverable. We are very pleased with the outcome. I participated in a workshop last February 2017 that is. I work across the government with other science and technology agencies. In various groups. We all had been working toward our response to the public access memo from OSTI P. There was a lot of concern over how compliance would look. How really it's basically what we are talking about is digital content, and digital content is stored in the digital repository. How we will comply with this memo, and they have not had this exciting experience, where the repository was not accessible because of a hardware failure. There is a lot of catch-up going on right now with trying to communicate the quality of digital repositories to the public and to be accountable for what we are doing. The National Science Foundation I was involved in organizing the assessment in digital repositories one of my takeaways I will talk about in the next through the slides. One of my takeaways from this. Even though ISO certification was out of reach the best time to do this is when you are moving to a new repository, you can go through those requirements they basically layout everything you need to do. That is a pretty awesome thing. NTL we already have knowledge of what trustworthiness meant in the simplest terms, a digital repository a technical platform and the content and the people, the organization that supports this repository, and three things. The digital repository is not a website. Right off the bat. We know we need a different type of an assessment for then you would typically use for website. We need different analytics. We need that. And a method to measure and communicate the value of what you are doing, and the value of repository to your stakeholders. One of the takeaways from this workshop is that trustworthiness is a valid assessment. Digital repositories should be certified and in our case right now. We are not confident in having to do the resources for certification. We are trying to implement a repository standard. That can move towards that certification. That is what I will talk about now. Why would we do in this South -- Self-assessment -- During the South -- Soft lounge -- Lunch -- -- >> -- The soft launch -- We wanted to demonstrate following good product this is, and we would provide evidence that would mostly be documented, we would gain insight and further work that we need in the

repository and how we can further develop our trustworthiness to partners. Also following trustworthy guidelines create a benchmark that you can create for comparison in future years, you will be improving and knowing what your strengths and weaknesses are, and build user confidence into your repository and building a certification for a stronger foundation. What we started out with was looking at the data seal of approval, the world data system, it had identified 16 core require lens based on the ISO 16363. This management technology in some way to get guidance and gather feedback. Somewhere along the way last summer the world data system and the data seal of approval. RDA, the research data alliance, these folks got together and it was really nice when organizations talk to each other and decide, that they would endure the 16 requirements. They would call them the core trust certification, basically this is a self-assessment in providing public evidence. To peer reviewers and then periodic review every year. Jessica mentioned this is really a data repository based assessment, we are not a pure data repository although we do have data, and data products in our repository, I have no problem extending needs requirements to a repository that has more, the digital objects that you would define as a data set.

>> This is a slide I borrowed from Mustafa world, Desmet world crane -- World crane.

>> This is a condensed way to look at those guidelines. Jessica talked about this in depth. I am not going to cover these. She did a terrific job with them. Here's what we did last summer. Basically we went through three steps for preassessment. First we wanted to understand the core trust requirements. We talked to everybody we knew. We established a working group, had regular meetings with our group that uses our repository platform. We reviewed all of the core trust documentation. We updated and we have in our public access policy section on our website, we have guidelines for depositors, and we wanted them to understand what trustworthiness was. And also not to require that they use a trustworthy repository for their data, and use this as guidelines for the measurement of their repositories, at this point. We are okay with that. We are all kind of moving towards the same place. The second part of this. Collecting evidence. We did a preaudit. Involving and looking in places such as our website interviews with staff, and looking at technology platform, and the internal documents, measuring that against the federal mandate, lastly creating a scorecard, and tagging that with keywords and requirements identifying the gaps and requirements. This is what the scorecard looks like, this is actually bigger than this. What Erica did. She actually went through the 16 requirements made a spreadsheet of them. And put a documented evidence, whether it had to be updated, these are the outcomes as a result of the preassessment, we have updated our policies. We have a couple of them that are under review. I am pleased that Jessica mentioned the digital cure agency, we did not have one that was a definite outcome of preassessment, we had to update our metadata when we migrated. We also did a comprehensive review of our workflow. We have documented our standard operating procedures. Going through the preassessment helped us to communicate with CDC to help us communicate and what we needed from their technology. What we needed to be able to incorporate and identifiers in orchid which is a DOI for the author and to incorporate that into the records. We also needed the ability to repository, for the check a numerical string that indicates the integrity of the file. The file that is stored in the repository. Also what we need was documentation about the technical infrastructure. Those are things are now forthcoming. We are partnering with CDC on that documentation. We should be finished with this, in a few months. I put together a reference list for you. I would encourage you to go to our website, national transportation library where we have our documentation, it has to go through legal review, that is where we are right now stay tuned for that. Our repository, you can get in depth information about alternative methods of assessing trustworthiness, by going to their websites and their core trust seal websites, data seal of approval. The Big Data workshop I've been talking about, all of the presentations are publicly available. They are excellent. They give a lot of very good information about how, what, and why of assessing repository for trustworthiness. I will finish now. I think we have a few minutes for questions.

>> Thank you Marie. Thank you Jessica. -- Thank you Marie -- Mary -- And Jessica.

>> What is this repository and how to obtain the certification.

>> I will be honest I do not have the direct question -- The direct answer to that question. I believe it's important for them to do their market research and contact the groups that are involved with certification, see and find out if you can own your own estimate, because GPO is in the middle of contracted thing I can't speak to what we do or what we didn't spend on this audit. I think you will find perhaps, it is not as financially restricted as you might otherwise think. There is a lot of concern. We will never be able to afford an audit. I don't think the audit itself is the obstacle. That a lot of the repositories will have. I think it will become financially difficult perhaps in the sense if you do a self-assessment, if you find during that self-assessment, there is hardware or all these things that we need to replace are due, it will be very resource intensive. That will be the bulk of the resources used and spent not so much as a third-party audit.

>> We have a question from Robert, are stacks available as open source software or a community of developers?

>> Great question. Right now stacks have been made available to other government libraries. It is us and NOAA, and the CDC. The CDC has a small staff that does the development on the repository. You are probably all aware open source does not mean there is no cost. You don't pay for the software. You end up paying for the development. This they are using Fedora as a repository. It has a Drupal front end. I believe there is a potential for partnerships within the government, it has been developed for that purpose.

>> Okay. While we are waiting for anymore questions to Roland, I will quickly go over upcoming events in FDLP ACCME, we have a webinar 2 PM Eastern time, July 17 1 PM Eastern FDLP 101, avoiding the needle in the haystack, identify nonmetro communities, and then for natural hazards July 19 at 2 PM Eastern.

>> I can add a little more about the last question I had, in terms of self assessments, if they can seem resource intensive or perhaps difficult to try to get by in from administration, because of resources. I do think organizations and third-party will never be for us. Self-assessment at the very least provides opportunity to make the case for more staff. More resources in general. Still it is a worthwhile exercise. Especially if you can communicate. We are not meeting at standard because we need these things. Or a way to make those cases or communicate needs, if that is helpful to anybody.

>> Okay. This is Jamie again, I'm not seeing any more questions. Ashley is pushed out the link to the webinar satisfaction survey. We would appreciate if you would fill that out, that helps with programming at FDLP Academy, I would like to thank both of our presenters, Mary, and Jessica, thank you everyone for tuning in, we will see you at the very next webinar.

>> Thank you.

>> [Event Concluded]