

Event ID: 3053176

Event Started: 10/18/2016 8:28:37 AM ET

Please stand by for realtime captions.

Good morning everyone it is a little after 8:30. I have one or two quick announcements. And you can pass these along to other folks that are in our other great sessions this morning. For the online folks I want to remind everyone if you are watching the screen, if you are here that is great but there is a second that covers the events of the Washington ballroom. Those will also be recorded if you are touring between participating here and there. I wanted to let you know that there are options if you are a virtual attendee to see some of the other sessions as well. The quick announcement for those in the room, please pass those around, this afternoon, one of the local restaurant that we go to for lunch, California pizza kitchen is doing a fundraiser event. That is great. There are details on the board. It means there is a very limited menu from 11 AM until 12:30. Either plan to go later in the break time for lunch of that is your destination or, read the details of the fundraiser which I have put on the board which is a great idea as well. Those are my announcements. Will turn over to our folks who will talk about digital stewardship. Thank you.

Good morning. I am Beth Williams for the Stanford Law Library. I'm very pleased to be talking to this morning about digital stewardship. Anyone who has undergone a digitization project, large or small is quickly confronted with technical considerations that in a world of multiple standards seems more than a little daunting. I'm extremely pleased to introduce such a large group of experts who speak to us about their progress in creating a preservation plan for GPO along with practical considerations about implementation of the program which will hear about later today. This morning's program we will explore an important piece in the puzzle of the GPO future. Developing partnerships with us digitizing government information in a way that provides, not just ready access to government documents, but long-term preservation of this could ago information. Another way for depository libraries like us to 60s and reduce the footprint of our physical collections, while making and sharing larger digital collections, we need practical guidance. Thankfully our panel members are here to help. Please join me in welcoming Cindy Etkin Senior program specialist. Fang Gao to the RCM library technical services. David Walls preservation library. Jessica debt -- Jessica demon strategy and technology expert consultant. Lisa LaPlant FDsys program manager. Thank you.

Thank you to counsel for having a session on digital stewardship. We are pleased to be here to share a whole bunch of information with you, and thank you to the audience for coming to the session when we know that there are other sessions going on at the same time. We are going to go through the different panelists and then we will hold questions until the end. For those of you who are watching online, this is me Cindy and I will go first. I was on vacation in September and I got an evil confirming that I would speak on this panel. And I thought, I am not on this panel, but okay. I will talk I'm here to talk about how digital stewardship and the digital preservation fits in with our national plan, and how we are moving towards our vision of making government

information available when and where it is needed. We released the national plan earlier in the year. On the side are the outcomes that relate to digitization. We want to more online official government content successful for depository libraries to use and for the public at large. We want federal agencies to have more awareness, and could you be content to the Federal depository Library program and to FIPNet which David will talk about in just a bit. We want to the catalog of government publications to be the most comprehensive index of federal government information. And we want our users to be able to be confident that the content they are using is the physical -- official, digital and secure. That it is complete. And Jessica will be talking about the trustworthy digital repository audit. And how we are going to be achieving that in the end, the national collection of US government is just information will be accessible to the future generation. Some of you may recall that we had three strategies for success. Three strategic priorities. Two of them relate to what we are talking about this morning. The first one is lifecycle management of government information. Working to build a whole lifecycle management process library services and content management. The third strategic priority is services and it was mentioned a little yesterday that we need to have more of a service environment. This is one of our strategic priorities it to become more service oriented. User oriented and to provide the users with what they need.

While people have been out talking about FIPNet at conferences and in different kinds of settings and we have been encouraging people to consider the Taliban partners riveting things internally that aren't necessarily visible to you all. That is that we are starting to build that administrative structure so that we can administer the FIPNet program. What we have in the way of new documentation that supports fast, have public policy statement. These are available. Tooth -- 2016 -- to divide the content of scope for GPO of online access. And have a policy 2016 -- or that defines permanent public access through preservation. So we now have documented policy that the content that will be available in our system and that preservation is indeed a way that we can provide permanent public access to future generations. Just this last week we made available a collection development plan for GPO system of online access. We have not done an announcement yet or a news alert yet. We will do that. For those of you who are here and listening through our lifestream, it is available on FDLP.gov and you can find under the trustworthy digital repository page. This is supporting documentation for our efforts in the TDR certification. These are some of the actions that are in the national plan that relate. This is probably pretty small for you to read, but it is developing a preservation program within the LS DM business units. It is digitization of information international collection. It is about continuing to but is patently normative standards and guidelines for David will share a little bit of that with a. It is about the authentic and digital content and FDsys.gov/info and Lisa will share a bit of that with you, as well as visit the. And of course it is very important to us and you all have told us it is important to you, that we do have the TDR certification. We're going to be working with agencies to identify their content, digital as well as tangible. On part of this whole lifecycle management we are transforming Ellis DM from a print centric to a content centric organization. Ultimate goal is achieving the vision and to provide government Ajit -- education where it is needed which of we take a look at the strategies and actions, we are moving on all of those that we will achieve our goal for those of you who know me, I really like hockey. For those of you from St. Louis I'm really sorry. But here is a really nice picture that I captured of a winning goal against St. Louis. But we too can have a winning goal. Our ultimate goal is achieving the vision. Only have laid forded the plan and the actions that we are working on, everybody in the folks in

programs technology system we are going to do well for you, I think. That is my pulling it all together and how it relates to what we are doing and where we are going. And I will turn over to David.

Good morning. Thank you for joining us so earlier -- so early in a Tuesday morning. As city whither the national plan, what it wants to do is call out some very specific things in the national plan that are within the national plan for access, are important outcomes to support digital stewardship. Access to more complete authentic content. That is something that we all want any community has been asking for for some time. Access to content shared by federal agencies as well as content divided by FIPNet partners. Their collaborative partnerships made by the FIPNet initiative. And content that is accessible for future generations because we know that there is no such thing as long-term access without digital stewardship.

Within the national plan for access there are specific obligations that we recognize through the federal government that we have an obligation to preserve our information and specifically that means the FIPNet program and salvaging the partnerships. Developing a program within library services and content management. Racy preservation of government information to a very high level and make it a national priority. To digitize is much as possible through partnerships the tangible collection of publications. And of course maintain the inventory of the copies of record that we do have, and continue to participate in guidelines and standards. What actually is digital lifecycle management. It is present technology and workflow. Workflow processes. To ensure long-term sustainability of and access to digital objects and metadata. Is not the active stewardship of digital assets over time their policies, staffing, resources, and technology. Preservation is a strategic initiatives and programs like FIPNet and processes designed to maintain useful access to information assets serving the information needs of both the present and future generations. I like to talk about publications and digital content as Condit -- as a because that is a word in our culture that immediately triggers a sense of value. Revalue assets. We manage them carefully. Sometimes working in a large library the publications just become so much stuff. But if we did of them as assets that is something that will develop policies and processes things to preserve and guard against. In the preservation federal network, to remind people of what it is, it is a collaborative strategy of information professionals working in various partner roles. We do not have the content, the libraries to. We want to work with other stakeholders to ensure access to the national information. The various partner roles that we can seems of that people have responded to and said they want to participate in are on the right. With them until -- digital entity content hosting cataloguing of metadata, Web harvesting, tangible preservation stewardship, the preservation steward partners. Clutching care and conservation. Within GPO and the FIPNet community in the preservation program where developing but have caught in the slide is the things on the left are things that we are doing to build our own in-house infrastructure within library services and content management. The things underwriter things that we are specifically going to also support the FIPNet program and partners, and the outside committee of stakeholders. And that is to continue to support the partner roles for the community colleges of the fight the documentation, develop guidance for partner roles. We will talk about that this morning. And to provide opportunities for partners to discussed and provide the future webinars and forums, and to engage in the Trinity better in stakeholder participation and problem holding. We're also hiring new position within LST M2 sport preservation and finishing plans for not only collection development but also for lifecycle management. You are improving processes

for metadata extraction and requirements for developing digital quality assessment someone content comes in the we can appropriately assess it. We are looking for improved stewardship of web content and are large task is and is could talk about the trust of digital positive or assessment is to be able to respond to those things were some improvements are needed and of course to maintain the digital repository system of record for digital stewardship of our digital assets. As far as guidance goes, there is a link on the conference publication page for the new guidance document for digitization prop partners. That document basically recognizes these three principles that digital imaging partners provide to us and that the increase public access to legacy and historic federal government information in digital formats. They are providing authentic digital services federal publications that preserves the life of tangible copies for use and their preserving authentic digital copies within the trusted digital repository framework of future generations within a guidance document, we will take questions about that if people have looked at it. The guidance document itself specifically discusses the principles of digital imaging, what you are trying to achieve and eight digital imaging process, what are preferred and acceptable master imaging formats, we decided to give them options. A process for content already digitized or resulting from a current digitization efforts. Part of the basis is to recognize that libraries are recognizes content not necessarily for us but for their own collection development requirements for preservation needs and things like that. Where providing an opportunity for the libraries to give that content to us and store within FDsys so that it can be part of the national collection of digital government information and receive the security and access that it needs. We want to talk to people about conventions and things like this. There is minimal technical metadata specifications in the document. And we can talk about content packages of the digitization imaging partner can provide them. I will turn it over to Jessica.

Hello. My name is Jessica. Part of my participation on this panels to provide some context as to how internal units in GPL work together to achieve digital stewardship best practices and prescribing to certain standards. And also I'm going to bridge together the bits and pieces of this a little bit. I think that by this point in time a lot of people have heard about the TDR standard have heard that we are striving for digital certification. There is the new preservation program. There is FIPNet. There is document public policy statement. There is a lot of new stuff going on in it may not appear to be completely related. But all of the bits and pieces are definitely related and a lot of them are sort of running and Max a lot of the recommendations in a come out of the digital repository assessment I performed. In November 2014 is when DPL first announced that they really wanted to certify the repository FDsys against an international standard. And the international standard is ISO one 636. I will refer to it throughout the rest of the day I have a session at 245 today that I will really go very deeply into my actual internal assessment and what the next ups for GPO will be to achieve external certification within the audit process. November 2014 is one GPO announced that they wanted to go through this direction and I came to DPL in June 2015 as part of the national digital stewardship presidency program. This is a program that is administrative through the library of Congress and funded through in that. Essentially I want to GPL and was there to dedicate my full time and attention to an internal assessment against the standard. Just as an overview of what I did is as resident performing this assessment, I worked very closely with the ISO Standard which is a standard and list of 109 criteria and with 109 criteria that are grouped across the three in areas. There is an area of criteria that are organizational -- organizationally for structures digital object management and infrastructure and security management. These 109 criteria are not proportionally equitable thrall to the main areas.

For example digital object management is the bulk of the ISO Standard and was about 60 or so within the hundred nine that are solely about digital object management. That is really a large core that assessment process. With an organizational infrastructure and security risk management have about 25 or 30 criteria each. Part of my role on this panel is to make sure that everyone understands that GPO as an institution will be could undergoing this assessment process because there are many groups and parties involved in the success of the digital repository. So while the library of services and content investment are playing a very predominant role in preservation policies and preservation guidelines, and establishing some workflows, particularly for reformatted content for digitization of content and from FIPNet partners in some -- so forth because it was what is available is the diminishing of party over the FDLP even though they play a prominent role there are many other groups in GPL that really contribute to the repository. When it comes to who I am evaluating come I realize the text is small, but when it comes to organizational infrastructure, the parties that are really involved include LSD M but also the business unit programs which is actually worth under large majority of my time and that is also where I've situated Daewoo situated during residency program. It also includes plant operations. Includes acquisitions. Our IT and security departments also provide a lot of services and support for our repository systems. We also have a lot of vendors and contractor development staff that really support all areas of our repository. When it comes to digital object management the large majority of those activities and responsibilities fall upon program strategy and technology. PSDý sort of means the daily operations in the development of the repository itself. All of the way that we handle our digital objects, a lot of our infrastructure systems that support that and a lot of our automated processing and parsing of the content, all of the responsibilities fall on PST and Lisa LaPlant will go into more depth into how that content is adjusted unpreserved right technology point of view. Specifically, PST includes staff that are dedicated specifically to our search technologies, our repository, our content management system and even the government website is managed right PST. When it comes to security and risk, PST also plays a predominant role. PSDý governs a lots of the risk register for our repository. For example if our file formats are at risk of not being [Indiscernible] five or 10 years from now they then have the conversations about Kama Sutra be transforming our what would we do, so they manage things like that. They also sort of do the planning for the financial planning for our storage cost and things like this that support our infrastructure systems and always the because PSDý serves a predominant role in infrastructure for the repository preservation services come avails of interface a lot with IT and security.

What that looks like in terms of very specific responsibilities, and terms of evaluating organizational infrastructure, the types of things that are in ISO 6136 are things like policy documents. Operating procedures, collection development , it includes analysis of our statutory authority. Our design documentation. Its newest landing our financial security. Our configuration management documentation. And also documentation of things like service-level agreements. This is not an expensive but -- extensive list but the example of what types of things that I look at when I'm assessing organizational infrastructure to support preservation. When it comes to digital object management the types of things that ISS are metadata policies. High authentic or how we are measuring the integrity of our digital objects. Are file leaving an object dedication. Re: checking for file type. Discoverable is our content. Is there digital repository conforming to standards such as the open information system. And when it comes to infrastructure and security risk management the types of things that ISS is how well do we monitor our systems. To have

redundant backup double parity storage and so forth. Are our service level agreements content of. We properly modeling our storage costs, are environmental monitoring, resource processes, disaster planning and mitigation, things like this are what ISS for infrastructure and security risk management. And in terms of determining the processes that GPO does are successful or not, and in compliance with the ISO 6136 standard the way that ISS that something is good essentially is how well it means the needs of the community. The designated community is a concept that is defined in the OAS standard and also in the preceding documentation that chartered or led the way for the OAS standard. It is also spoken about in depth throughout the ISO 6136 standard and states that a designated community is an identified group of potential consumers who should be able to understand a particular set of information. These consumers might consist of multiple communities that are designated by the archive in might change over time. Earlier I know it is conference it was brought up the basically the success of preservation is not just that you can access the content but that it is actually understandable and reusable, and relevant. So everything that ISS is within this context of, doesn't meet the needs of our designated committee. And for GPO the designated community includes the American public which is very broad. But more specifically includes the FDLP federal partners and stakeholders and includes our content originators . Constantly making sure that are metadata or collections and these are things are actually run rubble and usable. And that our processes are understandable as well. There has been an increased effort to provide more documentation of the system on the government website have more "rotation about what our services are and forms like this. And very briefly, I have been asked to reiterate why it is ISO 6136 is the best or why GPO is choosing this sort of assessment process. Suggest a brief history, and the spring I gave a similar conversation and went more in-depth to the history. You can go look at that on the FDLP website recording. But just a history, the repository is modeled off of the OAS framework and Lisa will go into more detail about that but this is the standards of the repository that is built off of. And that was as recent as 2003. About the same time that OAS was written a document was being released this is still a very key document despite it being over 10 years old. Come almost 15 years old. Ultimately this document defines all of the things that digital needs to be doing and claimed that there needs to be a way to certify Digital's a positive -- apothecaries to know that when they are claimed urging preservation, are they really and how you confirm that? Ultimately, that grew into the track standard and they outperform several repositories under track and track with a checklist of responsibilities and things that repositories need to be doing. But it was not necessarily an internationally recognized standard. It was more -- it was a very strong and useful document. And CRL performs the daughter's and those repositories really paved the way. The six repositories under the certification in the process of CRL really paved the way for what repository audit could likely, or should look like. And ultimately track evolved into the ISO standard that is now a standardized process. And if you listen to my previous discussion in the spring you will know that no repository has been audited under ISO 6136 because the standards to certify auditing bodies to perform these audits wasn't released until 2014 so it is not at this time it is assumed that they are probably auditing bodies that were certified to perform the audit, but none of them have publicly stated -- stated that. This is an opportunity for GPO to potentially be the first repository to undergo the certification process. I know that this slide says to 30 but I think it is that 2:45. So if you come to that presentation I will be going more in-depth about my internal depth specifically. And I will turn over to Lisa.

Good morning. My name is Lisa LaPlant and I'm the FDsys program manager. I've been involved with the FDsys program since 2004 and it's a pleasure to speak to you again. I know it's only been a couple of years since I've been able to speak with you all. I'm going to give you a brief technical overview of the FDsys system in relation to digital object management under the OASIS model. For some of you this is a refresher. We will talk about packages and models. For some of you it will be a new exposure to some of this terminology.

First, background. The digital system was launched in 2009. It was a replacement for the old GPO access system. Some of you probably remember the old GPO access system. It provides access to over 100,000 titles from the site of, executive, and judicial organizations. And we get around 40 million retrievals or document downloads per month. Another big piece of information, news related to our site was the launch of govinfo we launched as a public beta on fiber in 2016 and it eventually will replace FDsys. I encourage you to come to the session this afternoon. Think it is out for Cochran after dusk is session. We will give an update on the content and some of the updates related to govinfo.

I know you have probably all seen the site before but we will get into it. The system can be thought of as three distinct components. It is a content management system that securely controls digital content throughout the lifecycle to ensure authenticity and integrity. It is a preservation repository that follows archival system of standards to ensure long-term preservation of and access to digital content. And it is a public access website that combines extensive metadata with modern search technology.

How about a deep dive into the OAIS model. FDsys is what's built on the open archival information system model. It is an ISO standard. And the reference model addresses functional activities for and just, archival storage, data management, administration, preservation planning, and access to digital content throughout its lifecycle. In addition the reference model provides a set of the best functional requirements for preservation systems. Want to go through each one of these things doesn't think it will help frame our discussion, especially in terms of digital stewardship and lifecycle management. The first requirement is to negotiate for and accept appropriate information from information producers. Second, to obtain sufficient control of the information provided to the level needed to ensure long-term preservation. Third, determined by either itself, or in connection with other parties the community should become the designated community. And therefore should be able to understand that the information provided. Force, ensure that the information to be preserved is independently understandable to the designated community. In other words, the community should be able to understand the information without needing the assistance of the experts who produced it. This, it should follow document of policies and procedures which ensure that the information is preserved against all reasonable contingencies and which enable the information to be disseminated as authenticated copies of the original or traceable back to the original. Six, make the preserved information available to the designated community. It really captures the entire lifecycle from production to management, to be made available out of the designated communities. The reference model also provides the unique vocabulary for how information is packaged, described, and presented. The model uses the concept of an information package. The package includes digital objects, metadata required to describe the objects, and the packaging information that associates the content and metadata to gather. In terms of content lifecycle the preservation process within

FDsys begins at in just one in archival information package is created from a submission information package. Now I will go into more detail about packages in FDsys/govinfo. I first package is a submission information package. It contains files to be similar to the system and it may optionally contain metadata. It can be collection specific or it can meet a minimal set of submission guidelines. Here is an example of a submission information package in our content management system. When I log into the internal interface of FDsys/govinfo and I go to the Federal Register collection, and I go to the FIP area, this is what I would see. You will see it starts out a little bit small. It is a very different file folder. There is a graphic cemented. There is a PDF cemented. There is as -- Avenue TML summative and text some of the. This is what I would the ready for estimate the Federal Register package to the system. It includes all required file formats, or renditions and it also includes optional file formats for a valid FIP in the Federal Register collection.

Our next step in the process is the creation of an archival information package. And archival information package contains all the mitted formats or renditions plus a mods XML file which contains constrictive metadata, a premise eczema file which contains our authenticity and preservation metadata, and an AIT.eczema file which is are met file and put the entire package get together. This tells us everything, all of the parts and pieces it makes the package self describing. If I log into the preservation repository, what will I see? I will see all of our submitted renditions. I see locator PDF post script graphics, text as TML. I fear multifile premise file, and our AIP Mets file.

And access content package is specific to our system. It is not defined in the OAIS model but it is something that we use in order to be able to manage the content and metadata internally within the system. It contains both a copy of the Smither preservation files, and any access derivatives. For example, a digitally signed PDF is in and access content package. The PDF that is unsigned is in our AIP in the preservation area. Another example for the Federal Register collection is the XML that we create from the mitted SGML files in the ACP, but not in the presentation area in the AIP. We do this in order to be able to continually update and create different types of access derivatives. We can run and AIP through our processing steps in order to create new access derivatives as technologies change and that the needs of our designated community change. You also notice that there is a file new the bottom called FDsys file. We managed that chooses to manage as we move it from component to component the system. This is what our ACP looks like in our content management system. You will see that we have all of the folders and renditions with the -- cemented. Those are the files in the renditions that came into submission. And then we also have the access derivative. Have the HTML that is created from the text. We have a PDF, digitally signed PDF from the submit a PDF. Have an XML rendition that is created from the SGML. Finally, our dissemination information package. This is what is available on the public website. It has all public access renditions plus mods, premise, and months. While you can get all of these pieces individually on the public site, you could download a PDF. You could download a mods file or premise file. We also make it available as a zip file so that you can download the entire content package if you would like. If I click on the zip file on the public site I will see the zip file has HTML, PDF, XML and also help our Mets file, mods and, premise.

Slightly shifting gears, as many of you know, there are two separate instances of our system. One is our production instance, and the other is our continuity of operations instance. When I say

instance it means to suffer copies of the entire system. We will get into that a little bit more. At each instance, have a set of servers and filers. We have storage filers which I will come back to in a second. We have database servers. We have content management system servers. Processor service. Search servers and admin service. Application service, Web servers and virtualization service. Just to give you an idea of the scope for search servers, we have 14 of them at each instance. We are not talking just a single box or server. This is a very large system. And that is duplicated at each instance. If I come back to the storage filers you notice the purple thing next to it. That is a brain. Filers can be thought of as a storage brain that manages storage, performance, synchronization, failures, without impacting the individual availability of storage volumes, files, or bits. When you hear filers, think about it as the big brain that manages your storage. We have two of them, encase one of the brains feels a little flaky one day. Just to slightly touched on storage, we do have quite a bit of storage that manages that is the base layer for where all of the content resides. If you have specific questions about storage I would be happy to address those, or come see me after the session and we can do a really deep dive on storage. My original slides had the super deep dive that we will hold off on that for now.

Also, at each instance, we have a multi-local area network. We have a web tier and application tier. A database tier. At the gavage that as layers of separation. Items that are in the web tier can really talk to items in the web tier and items in the application talk to application. Items in the database tier talk to database bequerels of a set of routers and switches. That is really the snapshots of what we have at each instance. Everything that I just went through, we have a that production, and we also have it as our continuity of operations instance. This is how we keep it in sync. I will not going to every single box and dotted line, but we have a set of software and hardware solutions where we are able to keep those instances in sync edit supports both our disaster recovery and are processes. -- Failover processes.

I have gotten this question recently about the number of copies of files in a system and I wanted to kind of run-through how many. We talked about the instances and the different packages. Copies of files in a package. We are both our production and to instances. We are production, AIP in the preservation repository. Production ACP on the content management system. Production AE CPLD production site. That is one instance. We have the same thing at Coop and our weekly backup. At any point in time you can see that within the whole FDsys/govinfo ecosystem there are seven copies of a file.

And to wrap up on authenticity and integrity, we record the 256 hash value for all content files metadata and encourage you to go to the public site whether FDsys/govinfo and take a look at what is in the premise file. You able to see the hash value for all content files that are in the system, and there are a number of open source tools, even desktop tools that you can check that hash value against a value that is in our premise. We also record the significant events in the life cycle content in our MS metadata files. Checkout what is in the premise. And finally put digital signatures in her PDF document. We talked a little bit about authenticity and integrity of our content. I will introduce Fang Gao and she will talk about making it happen with metadata.

Thank you Lisa. So far we have heard about national plans your Greenford about strategies for success. And also FDsys being a system of record. This is our content management system and preservation repository, and also has a user-friendly interface. All of this we have to make it

work with metadata. How is this going to effect our work? My unit, library of technical services is a units within LCM. Library services and content management system. We are involved in the whole lifecycle management and functions for government information we acquire and identify government information. Reclassify ad catalog preserving government information. We also make them available to our public and FDLP especially. Lot of our work you already know my staff create records in CDP, the were -- catalog. But you probably haven't heard these things like creating metadata for FDsys. Some of my catalogers have been created records for all congressional publications which live in FDsys. Apart from those we are also creating metadata for historic content in FDsys. For example, with the Treasury Department count publication, one of the titles is official register of the United States by Civil Service Commission. And also annual reports from the GPL publication. We're not only creating records, we also sometimes create collection of eight record. Example for at is [Indiscernible] collection. We have a zoning information Center collection. Is ingested and it is available from 1951 through 1999. There is a projects we have learned a lot. Let me just give you an example from the most recent from the Panama Canal commission publication. The process is not an easy one. We have staff, catalogers as is the librarian project and program staff involved in the whole process. Will have a title, we make sure the records in CDP. If it is not that we will create a preliminary record. That our system librarian will crosswalk it into FML and the we will have the project and program staff getting the process started with the record. So there are quite a few back-and-forth until he get the process completed. Of the we go back and complete updating the metadata record. Having the Provo directly to the publication. That is basically the whole process. Let me give you an example. For which URL we will use in our record. If you look at this search, which is a little bit small, if you are creating a record for -- my staff does a search for the title and want to find the URL from the Pro to direct it to, after the search for per luminary record we have dozens of hits. I was the that would not be a good URL to use. However, the next screen that you can see when we do a search now nine records come up and it is all this collection. So it has a precise result for this search. How did that happen? It looks like this would leave useful and perfect URL for our Perl to direct to. How did that happen? The trick is to have the system number that it in the month record. To this project we have learned quite a bit. Moving Forward, what we will need to look into is to designate a single and authoritative source of metadata for which metadata were performed for mutt multiple purposes. And also we need to explore automated ways to assess metadata between different systems, CDP and FDsys info. If you search in CDP when we have the records and referrals, it goes directly to FDsys. If you search budget for this example the Panama Canal, the record comes up and on the left-hand side you will see a CDP record. If you click on it will take you to Eckley to the record -- it directly to the record in CGP. Sometimes you click on the record and click on the view record and it takes you to the FDsys homepage. It doesn't get you to be specific record. We need to do more about that and having the two systems talking more. Another way is to develop metadata exchange requirements and procedures one at a new digitized content and FDsys info. With the data we need to equip our staff and decrease our efficiency and XML coding with a different metadata standards and crosswalks. Have heard about improving input structure to better meet the needs of our processing US government information in a digital environment. Things that are in the works we are working on collecting the development library and position and library positions. And we would definitely want to collaborate more with other business units within GPO but also custom service. This is some of the things that we will be working on. That is all for my session and we are ready for questions.

I have a simple question. If I understand the process correctly, if an organization follows the guidelines that GPO has provided, and GPO decides to it and just the information, you are going to do an internal audit to make sure that the data cohorts with standards. If my scan is in-depth somehow and -- would that be distant visible from your own digitized product?

The guidelines at that GPO will be providing will be informed about some ISO standards and non-ISO standards. Best practices. The content that comes from FIPNet partners, content itself isn't necessarily compliant with an ISO standard or it could be dividing a what standard you are looking out for the repository then it is more about our systems and organizational infrastructure but we will be able to differentiate from where content comes from because we do preserve information about the content in our premise records.

Melissa Bronstein from the University of Utah and actually my comments more than a question. I just want to say that this presentation was actually helpful especially -- I'm not the most technical person in the world and I think this is really useful and you guys later everything out very clearly and I just wanted to say thank you.

Daniel Cornwall Alaska state library have a comment and a question. First I really want to commend GPO for putting out specific digitization guidance of libraries for getting this is a really important step in moving forward. And one question aside from -- I'm not sure if this is on the institution Sydor GPO side, but will there be at some point they check for completeness and human readability.

Yes. This is David Walls GPO. There will. We're building a quality assessment routine and from the content that we get with partners that we would have to go both through of the graphic completeness as well as technical specs to be sure that it is what it needs to be long-term access and digital preservation.

Cast harlot University of Washington. Along the same lines, has content flows in, as we hope it will, from partners how does GPO anticipate staffing or other capacity built into it to meet that?

Right now we have a digital curious and librarian position that we are hoping to have approved and go out for someone to apply for. We are also trying to look and see what capacity we do need. It is hard to judge that at this point. We're trying to streamline a process to bring the content in in a more automated way so that with automated processes, there is fewer chance for error and more efficiency. If your hands have to touch it. And as we analyze those processes and as the partnerships develop, we will then be able to see what other staffing positions that we need.

I just want to add something to what David said. One of the priorities that was given to LSDM this year was the field positions and you heard reported how many were down from our ceiling. When people retire or leave, or move on to other positions, we're looking at positions, we are looking at the skills that we need, the knowledge that and skill set that we need for other kinds of positions as we move forward in the digital age and becoming more content centric. We are maybe filling a position with something else than what it previously was. That is part of what David mentioned with the preservation or the digital curious and library and the collection

development library and we'll be looking at other needs as well as we start into this process and get into our regular work streams and work processes. We will be able to figure out more. But we are not just automatically filling positions as they become vacant. We are looking holistically at what we need as an organization to support the digital age and the processes for it.

Rich University of Florida. For partner digitized material that meets the digitization specifications is the feasibility of feeding that in the bibliographic metadata be added to it not necessarily by the person, the partner digitized does not have all of that detail the process could flow in and the additional work could be done by either another partner or by GPO.

This is David Walls. Yes because there are a number of issues within what you said that I kind of here which is that libraries may not have complete set of things so there may be content thing out of a serial that may require talking to other libraries about filling in the gaps. We try to set the bar precise, but somewhat low for the amount of metadata that a library could provide. Would like to have a marked XML companies the content some libraries would be able to provide easily and some cannot. The purpose of all of the guidance is to provide enough of a firmer to show specifically only want but also begin a conversation with a partner to see what we can negotiate. With the capability is and where there might be gaps that need to be filled in.

Just a reminder, if you have a question that you would like to ask through the webinar, please move your mouse up to the center top and click on chat. That will open up the chat window and just type in your question or comment and send it to all participants. Thank you.

Give me the definition and I can hand you a bunch of packages. If I want to go all the way the other direction.

We had to kind of write them to recognize that there is a variety of people out there with incrementing capability, and understanding. If you regularly create content submission packages for your own turn internal digital suppository and can do that with all of the metadata we can give you the content submission package guidelines.

There is a wide variety of opportunities for people.

We were trying to create the table where everybody would feel like they had a stake in a place at the table.

Cindy. I just want to add to that. You will not see in the guidelines that are draft all of the information about how to create the content packages. What we have all there is all of the digitization part of that. If you then want to talk to us about in dusty your content, and we encourage you to. Into FDsys. Then we can offer the conversations about, Kenny package or not package. And work that out. And as David said, the beginning of the conversation.

Cast harlot University of Washington. I may be confusing CGP, iOS, of info, that could you not have a little indicator in there as well as to whether this was an item that was Arty for preservation partners for the tangible? Or does that information not go here?

Cindy at DPO are you talking about in FDsys/govinfo ?

That is in the CGP we are investigating put of all of the information into the catalog of government populations and CGP so that when somebody searches they will know where they preservation copies are located. We had not thought about where that might go in FDsys.

In FDsys we can create field to put the partner of who cemented the content into the metadata. That is one option. And then there is also internal discussions about what might be the path forward for that as well. You can put -- that was something that came up earlier. About who is the partner, who provided the information's. We do that now as an indicator and what is contained in the FDsys/govinfo That is part of what disco mentioned as tracking the Promenade. But that is a little bit different from identifying the preservation stored. One of the things that was pointed out in one of the things that we are looking at Moving Forward is doing more automated extraction of metadata and because that information will be in a field in a marked record in the CGP, that could possibly then be one of those fields that is extracted, and then would become part of the metadata record that is in FDsys/govinfo But not yet.

It would be nice if that record is able to be analyzed. Of one preservation partner. You could actually assess which ones don't have preservation partners and which ones still need additional preservation partners with a counts that you could do on the way it is entered.

Any questions from the audience?

Diane from Northwestern University. One underlying repository software are you using? We are using Fodor. I'm just curious.

We saw from GPO. Documentum is our content management system we have eight set of college some Java that we used to process and parses the content. We have a series of custom Java Web applications that is on the front end. Does that answer your question?

Noss computer the. In the back and in terms of managing your objects, and content types, what are you using?

In terms of file format for our custom reporting for integrity. When I talked about the 256#because maturity to report that we run against all of the hash values to assist whether there has been any changes in the values and assess the objects.

[Indiscernible]. I got a little bit lost in the middle with the alphabet soup. And I may have missed a basic premise of this. This is sort of a premise question. But a prefatory question. It sound like we're talking about taking a partner digitize material and authenticating it. Is that what we are talking about? That is a huge change from what we talked about in the past. Just wanted to make sure that I was understanding what we are talking about.

David Walls, GPO. That is what we are talking about. Why not? You have been managing the tangible collections in partnership to the FDLP with GPO for use and the prominence on the content bibliographic it says that the government publishing office created it and printed it. You

have been launching over it in the library for years. If you want to digitize that publication and do it according to current preservation best practices, we can authenticate that digital content bibliographic we account for in the catalog of government publications, and make it available to the American public, just as you did for years in tangible form it on yourselves.

I would like to record -- the record to reflect puppy heart glitter and jazz hands. Happy dance.

Thank you for saying that.

This is Lori Hall. I know the two are excited but I just want to put a little reality check on things because that is what I usually do. For every collection that comes in it is different. There are different things that have to be potentially done. If it comes in as wonderful as Scott says that his is in we don't have to do a whole lot. If it doesn't, it requires an additional work on a lot of our parts which may mean our staff doing it. It may be that we need to contract staff to do some of it. Master, you have the digitization guidelines to drive. Is not gone up to 10 nine yet?

They are on FDLP any could've for the documentation with the conference for the handouts.

Everybody take a look at that. My reality check is, the collection plans have a list of priorities that we have listed different levels of priorities with collection type. We get random offers of collections from agencies that we have to assess and prioritize Lisa's staff and contractors are not thousands of people. There are tweaks to the system they need to be done for different collections, different assessment of metadata. I don't want to put a damper on it but we are getting all of this stuff prepared to be able to do this and we have a lot of processes in place now, guidance, staff working on this. Hopefully that didn't burn anyone's bubble.

I will give one jazz hands. James Jacobs Stanford University I'm wondering if you could talk about fugitives and how they fit into the process other than digitized content. I'm talking born digital collections or individual documents that members in the FDLP a collecting and sending to you.

From GPO actually yesterday we talked about it in our updates and we also have a poster, we have done our internal review of the document discovery process. We have several channels adding the submission and we have -- right now we are working on two processes. We have some backward as thousands of titles. While my staff are working of the dialogue, we are also having another process set up working on the new submissions coming in to the inbox. So far, for the backlog we have already made the determination and cataloging of more than 700 titles. For the new things coming in a, we have catalogued probably more than 400+ titles. While we are working on this temporary process we are also have come up with a plan looking at a long-term solution, may be having, working on maybe a software technology solution and to also make some of the processes because we need to have better systems to track when we get the submissions, who is working on them, where they are in the whole process. So definitely we are aware of the issues and we have been working on it. And if you would like to have more information you can talk with me and also we have a poster on the other side of the hall, there is more information there also. Thank you.

Cindy from GPO. I just want to add something to it she is talk about all the work that they are done on fugitive. I want to let you know something that we will be entering into with the Federal research division -- with elaborate Congress which is an effort to determine agency web publishing practices out of what offices does that information come. We know that this has changed this to better help us understand what the agencies are doing are doing with digital content. The board digital content. So that we might better learn and to be able to make connections with them. One of the actions that makes federal agencies aware of the FDL PNV tooting their content. When we know where we need to go, then we can better work with them before it becomes [Indiscernible]. Select just a quick follow-up. Right now when things are reported to lost dock in that group is working anemone catalogued them they are born digital which Jeanette -- teams talked about. The are creating a Perl and that means there is a copy be made in the permanent as well.

Yes.

Okay. Thank you.

This is Lori Hall. I just wanted to tell you a little bit more because we are entering into an interagency agreement that with the Federal research division at the library of congress for them to do a study, case studies. It is a two-year -- about two years. A year and a half investigation of federal agency practices on publishing. So you can get a sense of how much the government is publishing or how much the agencies are publishing to help us figure how many documents we are not getting and to talk to them a little bit about their strategy. I suspect that probably by the spring we will have something to report on that study. A

I think that we will. And the final part of the agreement is for them to come next October to report on this. The final analysis and report will be here next year.

It is one of the big unknowns. We usually just that at the end of the trough and get things. Findings. You guys find things. We decided we would see if we can get to the agencies themselves and find out what is going on.

Kathy Rice. Also be declared the cataloging system like an open beta system. I did not hear about that yet I'm not sure if that -- where the is or if this is the right format for that?

Could you say that again?

I was wondering where the discussion is having to do with data for the catalog system whether it is a frame or another system I'm curious to hear how that may evolve.

Also mentioned the LDS update yesterday. We've got some requests from our community asking us where we are with the framing data. We also had the development of this frame. Cytomel whether all of you have heard of this frame it is really initiative from library for Congress that aims to develop metadata standard for the postmarketing environments. Based on the successful and limitation of RDA. We wanted to do our metadata for the library community for the library committee. So far our Labrie staff and the SEM we have been following the development

through webinars and attending conferences. The most recent we participated was in May was an optical meeting and Library of Congress. We actually participated in the moving away and participated in the virtual [Indiscernible]. They showed us the big frame editor and how to create and use the editor to create work level records and the manifestation level record. And we have shared that with our staff. Actually with the big frame conference over in Richmond Virginia that our staff will be participating in. Stay tuned.

Anthony

GPO chief products and systems and just to add to the comments, they are in the process from a systems perspective working on the next generation solution which we have effectively named LSF library services system. We are in the process of gathering the functional requirements needed and a lot of the functional requirements will come from the technical service areas whether we need to be looking at how to accommodate in this new system environment semantic capabilities and so forth. But that is in the works right now. We are very close to having a document. If you have a model, a conceptual model to begin building out this new system requirement. I hope that helps give you a sense of where things are in the process.

Yes. I also wanted to add in terms while we also have the perfect example. So there is a capability of linking publication in FDsys so if you want to look at different versions they are available. So if you wanted to find different earth -- versions they are all linked together. Thank you.

Rebecca Owens. Thank you for this presentation and for helping us understand all of the planning and processes that go along with it doing all of this amazing work. My question is actually stepping back a little bit and looking forward to the future, and how confident are you that the level of funding will remain so that this kind of work and actually take place or even increase? Are there changes in terms of communicating all of this amazing work to the appropriate funding bodies in Congress, what kind of advocacy do you need to do something. Just basically looking at that big picture Moving Forward

I must say that GPO is very lucky to have a library community that is very supportive of our efforts and often go up to the Hill and voice their support for what we are doing not only for appropriations, but when we are building FDsys and wanted to have an online catalog. Through the end library associations we gave triptych support for advocacy. As far as predicting what might happen with the budget, I don't have a crystal ball. But GPO has been very fortunate to get the funding that we have asked for, is not at that level, not far below it. As we will see what the election brings.

Join me and taking all of these great participants. This has been a great program.

[Applause]

[Event Concluded]

Actions