

Please stand by for realtime captions.

>> Hi everyone, this is Jamie again at GPO doing an audio check, so feel free to adjust your volume.

>> This is Jamie at GPO. We are going to get started in about five minutes.

>> Good afternoon and welcome to our FDLP Academy webinar. My name is Jamie Hayes and with me today I have Ashley Dahlen on tech support in our presenters Jessica, David, and Suzanne. I will walk you through a few housekeeping reminders. This is nothing new if you have seen any of our webinars. If you have any questions or comments on the presentation please feel free to chat them in the chat box located name bottom right-hand corner of your screen. I will keep track of all the questions that command and at the end of the presentation our presenters will respond to each of the questions. We are recording today's session and email a link of the recording and slides to everyone who registered for this webinar. A certificate of participation using the e-mail you used to register for today's email. Additional certificates because multiple people watched the webinar with you please e-mail FDLPoutreach@GPO.gov and include the title of today's webinar along with the names and email addresses of those needing certificates. If you need to zoom in on the slides being shown by the presenter, click on the full screen button on the bottom left-hand side of your screen. Mouse over the blue bar at the top of your screen so it expands and then click on the blue return button to get back to the default view. Finally at the end of the session we will be sharing a webinar satisfaction survey with you and we will let you know when the survey is available and the URL will appear in the chat box. We would very much appreciate your feedback after the session is through today. Please keep in mind to reserve your comments about presentations and value of the webinar for the survey at the end of the presentation. With that, I will hand the microphone over to Suzanne.

>> Thank you Jamie and good afternoon everyone. Thank you for joining us to learn more about the contributor partnership. My name is Suzanne and I am GPO partnership coordinator. Joining me today are David Wells the preservation library and and Jessica team and Howard digital preservation -- an overview of digital content and Jessica and David will talk about items documents we released last fall. Let's get started. In order to better meet the government information needs of current and future users GPO is working to increase the availability of historic content. This is not a goal we can achieve without your help. While GPO has begun to enter take its own digitization effort, and to work with other federal agencies to increase digitize content -- [Indiscernible] the sheer size of historic Federal depository library collections means that we GPO federal agencies and libraries must work together. Before we dive into I want to share brief overview. Priority for inclusion on Gov Info. Content are made by our collection development and collection development working group. And the Gov Info team on the working group. Development priorities outlined in the GPO system of online access question development plan which is publicly available on [FDLP .gov](http://FDLP.gov). The plan is updated annually to reflect that fiscal year priority and the accomplishments of the previous year. At the collection development plan provides guidance for addressing content into the system, collection priorities, and planning resources -- and an example of some of the priorities for GPO produced content such a style manual and monthly catalog. To encourage depository libraries and other and the tuitions to share digitized federal content with Gov Info, we established established a new partnership category in 2017, digital content contributors. They provide digital content to DPO GPO for Gov Info. Contributors provide content on specifications for repository. And GPO provides are going access and preservation for the digital content. The University of Florida was our first digital content contributor. As a result, digitize content for which University of Florida is also a preservation toward will be added to Gov Info. The first content, publications related to the Panama Canal. Hopefully after learning more about digital contact today you will be -- join University of Florida.

>> How does digital contributor content appear in Gov Info? All digital content contributor content is authenticated and here is an example of a publication digitized by the University of Florida with the odd pen dictation seal on it in the lower left-hand corner. We want to be sure our partners receive full

acknowledgment for all of their work. Information on the partnership is in the record and on the Gov Info details page and labeled GPO partnership. The acknowledgment reads this content was digitized by the University of Florida library and is made publicly available through partnership between the US government publishing office and the University of Florida. Like GPO other partnerships becoming a digital content contributor is easy and painless. -- [Indiscernible] planning to undertake a digitization project the first up is a discussion with David, Jessica, Jessica, and I. We want to hear more about the content you have as well as digitization specifications. If you're in the planning stages David is very happy to answer questions and offer guidance. If you have completed digitization the next up will be to submit a small sample of master files. Possibly an access file for us to review. We will double check the files meet the quality standards. GPO working group will review to ensure content is within scope of the FDLP within our collection development priorities. Draft and signed a memorandum of agreement which outlines the responsibilities the partner and GPO. A template of a digital content contributor is available on the partnership page of FDLP .gov. Please keep in mind this is just a template and we are very flexible and changing language to suit the situation at your library. Final step is transfer the digitized content to us. We have an FTP site set up so you can easily set up the content. Once the content is in our hands we will begin -- begin the work. In late August 2017 GPO released guidance documents to assist libraries considering becoming a digital content contributor or any library starting a digitization project. I will turn the mic over to David and he will share information on one of those documents, the guidance on contributing digital content to Gov Info.

>> All right, good afternoon. This is David walls. Preservation library and for GPO. You see in front of you the actual guidance document on how to contribute content to Gov Info. Digitization guidelines can be fun about the partnership and preservation sections of the FDLP .gov webpage. On the preservation page, it is under a lower section down on the page called guidance and best practices. The digital imaging guidelines that we are provided are designed around traditional best practices creating a preservation master and access file in separate file formats. Accompanied by technical and bibliographic metadata. The acceptable specifications in that category are designed to accommodate local institution practices that may still prescribe [Indiscernible] as the master file format. And local practices who may be unable to create the specific PDF. The preferred specifications or what we ask our commercial contract vendors to provide when they digitize for us. Be uncompressed files provide significant space and cost savings over uncompressed TIF for a comparable page image. In addition to the file formats and metadata, there are other practices to follow to ensure that the best page image is captured in the digital imaging process. The decision of whether the source publication will be digitized whole or just bound is a decision that should be made early when planning to digitize the publication. The source publication -- too little margin at the center of the book when it is open and the page curvature may interfere with the clarity and readability of the image and the ultimate goal of the optical character recognition that allows us to search the contents. If you don't want to unbind your library ask FDLP exchange for a copy from another library that can be dispelled. Let us know if you're having trouble finding a source copy and we can assist with finding other copies and the cost of shipping volumes to your library. The issue of SKU is also addressed because the slant of the page makes page images look sloppy but it can also affect the quality of the OCR. The amount of pixels per inch you should capture depends on the size of the print and the image details. For most government publications a range of between 300 to 600 ppi with 400 ppi serving as a sweet spot that offers optimum quality for most text and table information. Pages with small type and graph or line should be imaged at 600 ppi. Some Skinners have settings that go way beyond 600 ppi and instead of getting extra pixels doesn't necessarily produce a better quality image or an image that can be better preserved. The quality that you build into a digital imaging effort begins at the selection of the source publication. Our partners is that they are digitizing publications for their own digital collection development and as digital content contributors they are sharing that content with GPO and contributing to the national collection of

content in Gov Info. While many contributors are experienced in digital imaging there douses about this so we offer this basic advice. If you have space us to do your bibliographic review away from the area. Vacuuming the volumes and dust from volumes in another location will keep dust out of the scanning equipment and reduce the maintenance you need to perform to keep the scanner within calibration. During the review as you look through the book, look for marginal, other underlining or anything that may reduce the user experience or interfere with the quality of the OCR and the search capability. If you need replacement volumes don't hesitate to ask your FDL colleagues or GPO for help.

>> If it is done right the digital imaging should provide a better experience for the user online than they would have using the original source publication. The careful planning that you do at the beginning of your digital imaging work will produce a quality product that you will be proud of and that will enhance the user experience, bringing the source publication up to modern research and discovery methods. Some of the best practice in front are to make sure that you capture all of the content from the original, develop your documentation and digitize specifications. And conduct final chance to evaluate the results. Also make the usual scanner test to ensure that they scanning equipment is working properly. And now I will turn it over to Jessica will talk about packaging and metadata guidelines.

>> I'm Jessica and this presentation somewhat technical and so I think that it is useful for viewers if you have content you are considering packaging and sending it to us and you have content you have already digitized this presentation is probably more beneficial to watch in the future but I am going to go through the slides and the steps for how to put together packages of content and what the metadata requirements are for our packages anyway. As an overview to find the guidelines for how to put together packages you can go to the FDLP .gov website, preservation tab across the top, and then you can use the drop-down menu to provide lines and best practices and there will be a file that you can download that is called packaging guidance and that is the actual link to it there. And they document looks like this document here on the right, and it will also be a step-by-step process of the same information that I am providing here in this presentation. This slide provides a digitalization of what a complete package looks like. This is the package we will use to induct content into the repository. One thing I will reiterate is that this is a really collaborative process between GPO and the partner, so in the ideal perfect situation you will be able to send us a PDF, a JPEG 2000 or test file, complete metadata, a thumbnail file, all of these things and everything will be completely ready to go and if it arrives to GPO looking like this we can very quickly go through and do an assessment to make sure everything looks the way it is supposed to and in just did along with wherever it sits within our scheduled priorities. If by chance you only have a PDF, we encourage you to contact us and talk to us about what possibilities there might be to work with you to complete the package and find the other missing parts. If you don't have metadata or complete metadata we encourage you to talk to us about what the opportunities could be. We do often reuse the bibliographic metadata that is generated from our CGP records. It's possible that this is a collaborative -- or complete package. And this next slide this is one more review of what the files are that go into the package. We ultimately anticipate having a single PDF document of that [Indiscernible], one PDF that contains all of the pages in that title. There is -- in addition to that multiple JPEG 2000 files and multiple test files so you will have one PDF that you will have a JPEG 2000 or a test file file for each of those pages that constitute the entirety of that publication. The reason why we are saying JPEG 2000 is that both of these files are not recognized by the digital preservation community is being high-quality preservation file formats. However TIF files are considerably much larger in size than JPEG 2000 and for that reason at GPO we decided that our preferred file format between those two is JPEG 2000. However if you only have TIF or can only send a stiff we do have processes here at GPO to convert those two JPEG 2000 if we chose to do that for the purposes of saving storage space. The other files to go into these packages include a mark XML file so that would be and XML file version of a mark record that accompanies this content and that could be a mark record, a market XML record that you send us or that we generate from our own system. And then additionally

there is also another XML file that is called the template XML. And I will go into details and further slides about what goes into that document. This is additional metadata that is used in our repository to generate metadata that is reused for certain discovery so it contains additional information about the document relating to the agencies it generates from and may contain information about where digital signature goes on that file for the purposes of authentication. I will go into further details in further slides. Lastly we also hope that we have a thumbnail image that will represent this document pixel dimensions basic JPG file. A GIS file. But that's another type of file that we can also work with you to create here at GPO if necessary. Here is more information about how to create filenames for these files. It's important that within the packages, all all of these files have consistent file naming conventions. And within our guidelines document we recommend that we use the following convention that contains government offer. Title of work, -- designation of work. Volume, a series of some other parked information. And then we realized there is a 30 character limit for filenames. Making abbreviations wherever possible is advantageous. Using CDC, DOT, making the title shortened. This information isn't necessarily going to show up on the user end, but it does help us keep information straightforward and trackable throughout the line of preparation so we can easily look at the file and no what agency produced it, what the title is and so forth. It makes our process go a lot quicker and it maintains a level of consistency because at the very bare minimum, all all of the files do have to have the same or similar file naming conventions and this is another thing we are willing to work with you. If you send us files and we have to rename them here that is something we are willing to do and we have done. This is the ideal. The metadata requirements. This is another iteration of the fact that we anticipate there being a folder within the package. Metadata submitted and it contains both the mark XML and the template XML file. And these files I have pulled out separately from the previous slide because no matter what come at these actually don't follow the other naming conventions. They always have to be named Mark and always have to be named template. I am not going to go to into detail about this page. This is mostly useful for you to go back on and referred to in the future if you have specific questions I am more than happy to answer them within this webinar. But we do require that within those mark XML records, we do need some changes made to it if possible. We do need those XML records to have a full suit up and we also need dates to be written in a year month day format and that is a system requirement. Not necessarily a videographic requirement but for our repository to handle that information cracked, it does need to be in that format and if it is not in that format where willing to work with you, the 245 field must include information about if this is -- if this title is not a monograph. We encourage the use of subfields a BNC for the 264 fields. And the 500 field should have information that says this content was digitized by your library and made publicly available three partnership and this information will show up on the public view and well be recorded so throughout over the years we can trace back where the file came from. And we also encourage you to include the information with your library same. This is information about the template XML file that I was talking about in prior slides. As I said, this is a metadata file that is purely used to generate information within our repository system. And this is another file that we are more than happy to create for you or create a template for you to start out with or just work with you in general because it is a little bit foreign and it does contain information that might not -- you might look at it and truthfully you might not know what the appropriate information is to put in this file. We do within the guidance document we do provide recommended values for these fields. We provide information about which values in these fields are consistent and always the same for every file. As you can see this is and XML record. It is quite brief, it really doesn't contain that much information. It can contain more or less depending on the nature of the information and we explained that in the guidance documentation as well. For example, this file here that I have created, there is a line that is embargo date. I don't have any values in that field in this example but this is where we could perhaps put something like December one of 2018. And that would tell our system that even if we ingest this package today it will not show up to the public until that December day. It might be

something you prefer to use as perhaps the information in the content you are sending us is particularly commemorative. The top 10 displayed publicly on memorial day or the anniversary of a major world war or something like that. This is a field that is totally optional and I wanted to explain. The use of it is there. That we have digital signatures, where our digital signatures placed on the document. That represents our authentication process where it could show up on the top corner or the bottom corner, wherever it goes on the document. That we have information about Everman author. This is required. And depending on if you have provided us a full suit up and a full mark record you might not have to supply it here but if that information isn't in the mark record we do need it here and this information will be publicly displayed and it will also be used for search purposes and required in the preservation any videographic metadata for that object. We do the need you to include information about category and we do provide the different values that can go into this field in the guidance document. They are quite self-explanatory. Executive agency publications, Senate publications and so forth. Relatively straightforward. Content is always the same so it's converted, as I was going to be reformatted digital. We recommend you put the -- this will be a record that we provide you, this is a library system ID that is consistent with the ILS ID and our CGP record. The rest of these things are quite straightforward. Content search is true, a means it will be -- search. This name is something that we might use to group content that is all within somewhat of a subject but for example we have a lot of content that was sent to us from the University of Florida about the Panama Canal. We can call that -- we can call all of that Panama Canal. We always include a right statement and if we want to we can provide subject headings, this rank is our way of providing an additional search term so for example, one document we used to rank for is the -- legislation associated with Obama care. But we can put Obama care here in this rank line and make it so users can search for the word Obama care even if Obama care isn't in the actual title of the document or anywhere within the bibliographic information. Put that alternative name here so it shows up. Let people Google or search for.

>> Now I will go through the rest of the slides and hopefully tie in this information together with an actual example. This is an example of digitized content relating to the Panama Canal that we received from the University of Florida. University of Florida is an example of one of our partners that sent us content that they had packaged and sent us their own records and own files and once we received them here GPO we work to make some changes as necessary. And this is the outcome of that process. Here's a file or package that was ingested from that collection. Branch information, category, government author, all of of this information here and any summary, that shows up as a combination of information pulled from the record and that template. That is all a combination. If you look on the left-hand side you can click on PDFs and open the PDF content. Look at the mods metadata and the mods metadata is specific bibliographic information that has been translated from that additional XML file. Metadata is a civic to your preservation metadata about that object and a lot of that is auto generated. It may or may not link to the CGP record depending on if we have the ILS them are integrated with this content and as you can see the thumbnail image here that says report, that is an example of that thumbnail image I was talking about earlier in the slide deck. You can create here. When I talk about package and folders and whatnot, if you're looking in your desktop this is what that looks like. It is a simple it is as simple as it sounds but there is a folder, this is a folder that can -- naming conventions, and within that folder if you open it up you will see that there is a JPEG 2000 submitted folder, metadata submitted folder, PDF submitted folder and the thumbnails submitted folder. With each of those folders are the files. If you open JPEG 2000 there they are. The one PDF, the one thumbnail, that is what it actually looks like. And here is a real example of the template file that we use for the Panama Canal content. A file that we made here GPO. And it contains information that was digitized by the University of Florida libraries. It contains that rank information on Panama Canal that people can search for by that name. Other information I have mentioned already. And that is that. I realize I went through detailed and technical. If

you have any questions feel free to ask. Any questions about -- David can answer that as well in partnership with Susan.

>> You have one question so far and it is in regards to slide 23. The question was are these submitted by e-mail?

>> We have an FTP folder which is a way that we will provide you login credentials to access the system. Different software systems out there that are free that you can download. Files are one example. And if you are familiar with those types of systems that is one way to submit content. If you're not familiar with that and the things I said sounded familiar, basically SF FTP allows you to log into their computer at work, and they can just simply drag-and-drop files into it and we can -- and they show up here GPO on our end also XML hard drives or thumb drives. And we are flexible about how we receive the content. We can't receive content through e-mail. Mostly because the files would probably be too big and it is also a somewhat security sensitive way to send information.

>> A couple minutes more. If anybody has any questions for our presenters. Feel free to chat them into the chat box. And while we are waiting, I will let you guys know about some of our upcoming webinars. Tomorrow we are having the federal government databases and websites a surprising source of state and local data. That will be tomorrow at two PM Eastern time and then we have on March 15 making science more open efficient and reproducible. Also at 2:00 p.m.. And then on the 20th make your smart phone a depository library at two PM as well. You can sign up for any of those on FDLP.gov. We have another question and the question is make sense, I had had a question about digitizing. What is the reason for not using D SKU software.

>> There is some feeling that some software is okay. It is just a setting on your scanning software that is probably okay. It will essentially rearrange the pixels to comply with that and there is a feeling that if done excessively it can introduce artifact into the file. Ideally, we would like the SKU to be corrected at the Skinner when the item is being scanned physically. A tiny degree, 1 degree plus or minus is fine. But anything larger than that is best not handled because of those issues.

>> We've got another question. Are you targeting your web archiving harvest to fill in gaps and target documents listed in the collection development plan?

>> The question is are you targeting your web archiving harvest to fill in gaps and target documents listed in the collection development plan?

>> The web harvesting we do is within the collection development plan. Gaps I'm not sure that I would use the term gaps. There is a sense that federal publishing right now is consistently largely over the web electronically for the latest so we are attempting to get a handle on harvesting that to deal with the fact that the website content changes so much. But the web content is a publication just as much as any of the others in the past that we are converting from paper to digital. Does that answer your question?

>> We will wait for a response to that but move on to the next question now. Whoever asked that question response back. My apologies if this was covered earlier but I was late. We have digitized some material from microfiche for interlibrary request. To accept those? These are only from print copies.

>> I think it depends on quality. I think if you were scanning them, at above three ppi -- 300 ppi if they were clear enough. For ILO those are put through like and FTP or attached to e-mails and things like that. Those files are compressed by the email protocol and things like that but if you've got the original scans that you did, you can always send us a sample and we would be happy to look at it and let you know.

>> A clarification about the question on web archiving. In other words, signed digital documents rather than digitized documents.

>> Digital content contributors, partnerships are looking at digitized for his store content. We have another program that David is involved in and Joy Bauer here GPO who are looking at web archiving of digital content. But the collection development plan I mentioned is Pacific to digitize content. His store content that we want to ingest into others.

>> A lot of work on their way to web harvest Ford digital content. Mainly we archive it through the tool.

>> Is there a been a month commitment to become a partner in terms of number of publications that are digitized?

>> No, if you digitize one title or your entire collection, we would like to talk to you.

>> Getting a low here in the question so Ashley has pushed out our satisfaction survey. And if you click on that link in the chat box it is to survey monkey and you fill that out, we we would appreciate that. Helps with programming here GPO. I am not any other questions.

>> If you think of questions later or content you have RD digitized you think would be a good fit or if you are starting digitization products and you want to talk please contact us, we would like to talk to an answer questions. Our email is up on the screen. David Jessica and I also the email address or fill out the partnership inquiry form on FDLP.gov and that comes directly to me. We would really like to talk to you.

>> Thank you all, and we will leave this screen shot up for a couple minutes before we actually end. I want to thank all of our presenters and also everyone for participating in this webinar and as Ashley said the slide deck and archives will be made available shortly afterwards. Thank you. [Event Concluded]