

>> [Captioner standing by]

>> Good afternoon. Welcome to the last program of the day. Becoming a digital content contributor. I'm David Walls and the reservation library for GPO and with me as my colleague Jessica and another guest speaker for our program today is Kerman Tanker and I will lead her bio real quick. With the government documents librarian at the Washington state library which serves as the regional for both Washington and Alaska. She initiated the library digital content and contributor partnership and serves as a library's primary point person on the project. What we are going to talk about is I will give a little background on where digital content contributor partnerships come from and why and Jessica will talk about some of the technical issues and how content moves from the partnership and info and Carmen is going to talk about how it's like to work with us. All right. A number of libraries in the FDOT do have the infrastructure to digitize and to do this regularly and the important part is they do this for their own digital collection development they're not doing this necessarily for us and that part of the partnership is where they elect to share that content with us so we can look at it and analyze it and see if it fits what would be good content to serve as a surrogate for the original document and to get that content with the info. So why digitize the pictures of the pages of books is to really create greater access. The access point is bringing these old paper documents into modern research and discovery methods so they can be more easily mined for the content and putting it on the digital repository creates digital circuits for some of the fragile originals and also protect the historic value of some of the more fragile publications. So becoming a good digital content contributor allows the library secretary to the national collection of government information and government publications and for these libraries that are choosing to digitally image their publications for their own use, this local effort gets the national context and the publications that they digitally must join others in becoming the national digital collection. All this content even though they may choose to preserve it on their own repositories and their own organizations it now goes on government info as a national repository.

>> Thank you. I'm Jessica and in the digital preservation library at GPO and am one of the people that you will interact with if you're trying to send us content to go into govinfo and you need help troubleshooting and setting up an account on your site and all those things I'm the one you talk to. And I will be going through some of the information about how to go about establishing a partnership with GPO for contributing digital content and what the process looks like and some of the questions and some of the common questions and answers that come up throughout the partnership process and even after the partnership has been established. So one thing about the slides and presenting today they might have the most utility once you go back to your institution. If you're going thinking about -- full of the slides and use them because they can be used kind of in tandem with all the guidelines that we provide on FDLT.gov. This is going to look like a lot of text and this is a diagram of what that partnership establishment looks like and is going to start with you contacting GPO and identifying what set of content you're interested in sending to us and this can be as few as one publication or as many as thousands but identified some set of information just call us and contact us and email us and say that you are interested in learning more about how you might be able to send that to us so once we have the initial conversation we're going to talk to you about what methods you use to digitize the content and it might be something that's already been digitized so we will ask you about what happened then or it might be content that you will digitize in the future and will try to encourage you to use the guidelines to do that. We will ask you about the availability of any metadata associated with the content and we will ask you if you have catalogued it yourself, if there's records for it and the CDP already and he might have metadata that's not easily structured and you can use spreadsheets and things like that and anything about it that could be helpful to us and we will ask you questions about it then and we will also talk about how your content might fit into the current collection development priorities and sometimes libraries and institutions have collections that are very pertinent to their own collection and their own

geography and they have these topics are subject areas and that very unique and exciting and it's content that will be interesting getting but it might not be of higher priority than some of the other things that have been identified but we will talk about that when we have this initial conversation. After we have this conversation we're going to refer to you to our guidelines documents and we have a set of digitize Asian guidelines and the links on the site you can cook directly out of it want to go home and look at these on your own personal time and we also have another set of guidelines that I will talk to more extensively after for how to package content once it has been digitized and sent to us and in theory we've had our first conversation and you refer to the digitization guidelines and you digitized something in and will ask you to send us a sample and that's a relatively quick process and we assess that the file formats are what we asked them to be for digitization standards and we ask that there's a certain resolution that is clear and legible and particularly things that we digitize several years ago we want to make sure that we can have a quality character recognition on that concept and we want to make sure there's text that's always legible and there's not too much skewer smudging and that all of that works out to we look through all those things and get back to you and say yes, that meets our standards and if everything is good to go up to that point that's when a MOA going to be established and that will document all the conversations that have taken place so far and what you're going to and the role that this will play that will be the official establishment of the partnership and at that point you can start sending us things and we are going to hope that you can send it to a and complaint is that it's sending a package for indoor packaging guidance documents and this will tell you the structure and the structure and what's in a package and essentially we are going to be hoping for a JPEG 2000 digital image file we are looking for PDF file and we're hoping that you can send us some Mark XML ideally we also hope that you have attached some image to go with the file, that's a complete package. Different library partners can do different varying degrees of that completeness depending on the resources you have available but that's the gold standard and you send it to us like that it can be in govinfo in as quickly as one or two weeks at this current point and here is another slide that can tell we were to go and find those guidelines and here is image of the package structure that I was referring to and it's relatively straightforward and there are some common questions that have come up now that we have a couple partners out there and some things we have found out through trial and error a little bit and one of these things is that for a lot of institutions out there when you have digitization project at your local institution whether it's got docs or not most digitization enterprises are going to be digitizing things into tiff and that's definitely a preservation standard at GPO we ask for a JPEG 2000 simply because JPEG files are smaller in size and for a lot of institutions you're going to end up having to convert the tiff you originally make them into JPEG 2000 and for a lot of institutions you do that in Photoshop and we found that in many versions the Photoshop when you go into the default settings and say convert to JPEG 2000 it's actually going to convert it to something that has a file extension .gpf is a subset of JPEG and it's a very easy fix you will go and change that default setting and tell it to do does JP 2 instead what you have to do that you cannot just save the file and then the right click on it and backspace that GPS and change it to dot JP 2. That's one thing that has come up and that was kind of in the weeds of it but you will run into if you using Photoshop to convert tiff. Another question that we get commonly is but if I don't have catalog for this material. As I mentioned before we need a Mark XML file at this point in time to adjust the content and that is a file that you can provide your is one that can be made at GPO and it does not impede your ability to create the partnership with us if you're not cataloguing locally but it impacts the timeline because at some point in time your GPO needs to make their Mark XML so that you know that it will get slower if we are the ones that have to add that to the queue of things to catalog which is fine. Other questions that come up as I mentioned is a question that we get it was only have one publication. Totally fine we can do a MOA for just one and it will be that one time and you can also do an MOA for 1000 of things something we'll talk to you about in terms of resources available and things like that. And then how quickly it's going to appear on govinfo all depends on what document is a

MOA . The less I wonder what the GPO has to do the quicker it goes in and the more we have to do the slower it goes in. Simple as that. And Carmen talk more about her experience working with Washington state library resources and for the sake of you going back home and wanting some visualization about what the files look like and what of the technical things are for packages and sending packages to GPO you can refer to the Washington state library content as a good example and this is what the content ultimately looks like once it's been ingested and has the title and the little thumbnail image and has all of the metadata associated with it and this is what the packages look like before we ingest it it's a series of folders and this is going to be the top level package folder and they are using a file naming convention that we put in our guidelines document and then this is the folders inside those folders and these are the files in Faldo inside those folders and these particular other JPEG 2000 files we will have one JPEG 2000 per page of the document and this is what their Mark XML looks like in this is a review of what it looks like once he gets interested because providing the Mark XML there's going to be a 2:#and we also include the GPO partnership information at the bottom so it will have a message that says this content was digitized and is publicly available through the partnership and we will have the library featured there and more questions. One thing that comes up a lot once you have gone back home and he looked at to the guidelines documentation for packaging you will find that we have recommendation in there for file naming and this recommendation is just that. It's the guideline and its recommendation. If you're ready digitize a lot of things and they already have filenames feel free to email us and say will lease filenames work? What we need is something that is unique and is not repeatable and makes your content somewhat understandable so if you're sending us hundreds and hundreds of them I'm not going to mix things up or get confused about what's what. But if you have not ready digitize and you would like a file naming convention to use feel free to refer to the document for a recommendation. Any other questions that you have for us you can always contact GPO and will troubleshoot and try to help and get advice and recommendation and same goes for the radio forms that you might point us in this metadata or Mark or anything like that and we try to make this whole process as flexible as possible and after that I will turn over to Carmen.

>> I am Carmen and work at the Washington state library and I will talk about what IRIX variant has been with digital content and my goal is to explain [Indiscernible - low volume] and what the day-to-day contributions look like. Ever since I started working with federal documents thought it would be really great if we could be partners but having some big new project was not really feasible given that our staff time and resources are strained and I've been reading up on it and found out that when we get ILR request to circulation request for documents [Indiscernible - muffled] are older and we will digitize them and at the time we are exploring them on the local server and linking them and we get more use but still fairly limited. And you might think well if you're already digitizing why not spread that more widely and share that. We have found there's a whole host of other benefits to being a partnership as well and the primary one is that wider access rather than just focusing on building the collection within our region we are contributing to that wider [Indiscernible - low volume] and it fits our strategic goal and one of our strategic goals for the whole state library is to build and maintain a federal repository program enhance access and participation by stakeholders. I think it fits within that and also our role as regional for both Washington and Alaska if we have a piece of microfilm or [Indiscernible] that cannot travel to Alaska but Alaska library needs that being able to digitize and they authority here is a great way to provide access to that collection we are at wired to do. It is a feasible way and it's an efficient way to [Indiscernible - muffled] to the greater community and its vegetation on the man so if we know that one person is interested in this particular document there's a good chance other people will be as well. We are passing on the responsibilities for storing and maintaining files to govinfo which is far better capable of properly doing that then our local storage at R1 library. And there is some positive recognition. We had a press release put out by GPO and the partnership was official and we featured on the FDLR homepage and we got an email internally from someone up in administration acknowledging this switch was really great

not just to bring recognition to this particular thing but the FDLR program in general within the library from somebody above the library level.

>> Once we know we wanted to do it how did we start? The first step was making sure I was not just having a crazy idea that I was just going out on my own and talking to other people about saying I think this will be really reasonable about this first time and effort and would this fit into our policies and goals things like that. Once I got an okay at the library level I sent an email off to [Indiscernible] and said I have this idea and I don't know if you do it this way but can we do it this way? We had a few conference calls and we talked about what the GPO requirements are and what the process would look like to become a partnership and what the requirements are for the actual digitize station. We send off some sample files to show we can digitize to the right standards and GPO thought about it and said sure, even if it's just a couple of publications we can do it. Once we got the informal okay we had a draft memorandum of agreement to get the formal okay in our contract team decided it was a contract and needed there okay so we went through that process assigned on both sides and we were an official GPO partner so we got the credentials to do as FTP to send the files to GPO I learned the process to properly name and packaged files and I wrote up internal procedures both for scanning and packaging and submitting that we could refer back to along the way and then I make sure everyone who is going to be involved in scanning and packaging understood procedures and where they were and things like that and we submitted our first publication. So an overview of the process. The first step is to make sure it is not already been digitized which is we have not always done that. It was once we digitize something not necessarily an important thing to remember is that it might not be in govinfo but it might be and one of the other GPO trusted partners like a federal website so if it's already available there we don't need to add it to govinfo. You want to scan the documents with the appropriate standards so I will go into more detail about each of them so you scan the file and you packaged files and you send them off and tickets up into govinfo it's pretty straightforward. So for are guidelines there is a few things that we need to keep in mind outside of what we are already doing in all of our digitizing it's straight and basic things like that any to make sure you have both print and phish so checking for print even though the catalog says we only have phish. 400 DPI in color and JPEG so these are settings for the scanner it's not in the extra time to scan to hire DPI or anything like that so it's not an extra work then we are already doing. The JPEG 2000 can be created after the fact so our scanning software for the machine we use does not make JPEG 2000 so we this do the tiff and convert it. So once you scan the files I always pull up my documentation the last one I submitted reminds me of how it's supposed to look and we convert tiff both to PDF and JPEG 2000 and we create the mark that XML file and the way we do that is our catalog will let you download Mark.txt and we convert that to Mark that XML using Mark edit and you name the files and put them in the proper structure which it looks complicated from the diagram but it's actually simpler once you get used to one is it PDF and wanted JPEG and one is a Mark. We send an email saying we're going to send this publication and then resend the publication and then shortly after you receive the link back to govinfo with your new publication up. We keep internal statistics that I update at this point and we keep track of how many as well as what they are and if we were to be able to do a larger scale digitization project I think it will be interesting data to know what we have already been digitizing at request to inform what we might be doing more in the future. So guidance along the way from GPO. While we are looking into doing this at the beginning I was referencing the transects template MOA template is a less the guidance as far as what will be required and what we would have to do and I referenced frequently the packaging submission guidelines. The folks at GPO have been very helpful no matter how silly my question is, there always happy to answer and do not have a digitization background. I did not know that JPEG 2000 and to be a different file for every page so I learned a lot in there always very helpful in clearing up misunderstandings. And as a conclusion I think it's a really high leverage way to be using your digitization. If you are already digitizing at any level, going a couple of extra steps to get into govinfo will really access -- we started in January and we have done seven so far

this year's is not a huge number but that means you can always fit in the time to do a scanning and do the packaging and submitting. Our collection is that it is not fully catalog and once again more and more catalog it will get more requests and it will digitize more. I also think that now that we have the structure in place if we wanted to do a focus digitize Asian project we are more likely to be able to start small if we can do a couple per month we are ready have the system in place to get it up and on govinfo have that infrastructure we don't have to provide that on site and things like that. If you think about it I think this morning it was 1100 libraries in the FDLP, if everyone did one per month we would have quite a large and growing collection on govinfo from the community digitized project. If you have any questions and you want to know what software or hardware we use, anything like that, here is my info and I have business cards you can grab just let me know. >> [Applause] >> All right. Do we have any questions?

>> I have one from the virtual world. Carmen, are you digitizing all the different type of microform? Are you also looking at the basis such as whether the Fillmore phish is cellulose or polyester base? >> For this year only doing phish because I believe we only have federal documents on phish and we digitize whatever we have. We don't really pay attention to what time what type of phish it is. Because we are only digitizing specifically things that have been requested rather than doing any large-scale like preservation focused digitize station sites whatever we happen to have that title in we are digital lysing. >> I am curious for Carmen and to know about the other partners. Are they Washington state library continuing to host the digital publications that you submitted so like, people can go through your catalog or are you linking to what is in govinfo our other partners keeping up their own versions of it? We have three series that we have had up for a long time. Really want to get them in govinfo but I'm curious if other partners are keeping a version on their own platform. >> As to the Washington state library we do not keep her own copy. We have limited space we are using for other things that we have digitized like our state publications and other things within our library and we do have catalog records to the archives everything that's available is eventually going to be everything will be in the CDP and there will be an archived record that we will have linking to govinfo for this publication. Suzanne I would know more about the partnership coordinator but I do believe that for instance Boise State University they do not host their files, which is part of the reason why they established a partnership with us in the first place but another example is University of Florida, they heavily host all their files and also send them to us. >> I think it just depends on the library's own infrastructure and how much investment they have had in developing a local repository and if that's what they have chosen to do, great. And if they don't want to do that we are happy to host them on govinfo.

>> Indiana state library. I just had a follow-up question since Washington state library is not hosting it, how long do you keep the original files that you scan and send to GPO? Have you, like, discarded any of them or are you guys so small you don't have any internal storage issues? >> I think as soon as it's officially up on govinfo and I have that link it is safe to discarded on our" because we know they have all the files. Practically I just leave the last one until I am starting on the next one.

>> Suzanne, North Dakota State University and I'm pretty excited about this I confess. I was wondering even if you find it on a federal site, they don't always have the same quality. Is finding it on a federal site just a no go for your own digitization or do you go like, we can probably do better? >> The official answer [Indiscernible - low volume]

>> Do you want me to take that? It depends. There is a lot of stuff on federal websites that goes back to the early days of digital imaging where the equipment was not as good, public expectation of what digital content was supposed to look like were not as high as now. We looked at older materials and from libraries that have digitized things early on and we have turned people down but his prayer. It is a rare. >> If you were to perhaps find something and propose re-digitizing it to make it better, we are

open to talk about it. I don't think it has come up officially in the context of establishing a partnership yet but it could.

>> It certainly could. If you feel like there is information that online that I have seen a lot of older PDF publications that are not [Indiscernible] very well you can search within the document and you're not finding things are seeing on the screen, that's something that you want to choose to do to do that as a service for the greater good of the FDLP we would certainly put that content on govinfo.

>> One other question with that. How are you handling accessibility especially documents that might have grass or something that would not necessarily be working well with a screen reader or similar technology. >> Grass and things like that we do our best to work with people to digitally image does if it's anything that has content that's relational, we don't pull that. We don't really run into that.

>> We have run into it? Okay. I'm just not aware. >> [Indiscernible - speaker too far from the microphone]

>> We OCR as best we can, as best they can. >> Chris Brown, University of Denver. We acquired a library of Washington DC the interstate commerce commission library and there is many publications in their published by GPO which were never included in [Indiscernible]. If we were to digitize some of that I assign X number two attached it was never assigned a Sue Doc number. Would you consider assigning a official Sue Doc number and we put it in the record?

>> The will to jump in there?

>> As long as the publications are within scope of the repository program, some agencies produce material that are not in scope, as long as the material is in school actual fugitive it was if it was not assign a number at the time we can.

>> That's a great question. If you have an idea like that you can just reach out to us by email or phone call and say hey, can you set aside some time and chat about this we are always happy to do that and if you have questions about your project your planning and you would like to digitally image something anyone to share that content with us we would be happy to talk at the beginning of the project to help anywhere we can.

>> Arlene from the Oregon state library. You talk at one point about collection priorities but I don't think I actually heard a detailed like

, is there something that you can provide us a little bit more detail about your priorities? I guess one of the reasons why I've always been wondering about this is what comes up through ILR requests can be incredibly random so it kind of mystifies me in terms of what offered but it's really not a priority for you so if you could talk more about that aspect of it.

>> Collection development library we have a plan you can look at online that talks about sort of our collection priorities and things like that for developing collections on govinfo and we have a plan to digital image going forward that gives you an idea of what direction you're headed in if you wanted to help us digitally image some of that anyone to digitize it that's great. Some of the random this I had a question about for Carmen and maybe anybody else if let's say somebody through ILR has a volume publication and he wanted digitize that would you ever consider digitizing the rest of that [Indiscernible - low volume] >> If we were to digitize the rest of it is will probably depend on what the topic is and if it relates to our collection development in terms of our broader collection and digital collections so we have a number of different things we are always recommending on digitizing with very limited staff so we decide if it's a priority like [Indiscernible - low volume] Washington and the Pacific Northwest we can add it to the digitizing queue or a national thing like a 70 volume set [Indiscernible - low volume]

>> In the context of partnerships when we have conversations about as it sits within GPO collection plan and priorities mostly it's most a concern if you want to send us lots and lots and lots of stuff. Simply because perhaps you have several thousand documents you want to send and some of them can be

broken down into chunks and some of those chunks might be more appropriate or more in line with our development [Indiscernible] but if you're doing the random one off, there's so much less resources to in just that it does not impact the priority debate is much. If that makes sense.

>> Currently, yes. >> Anyone else online? >> Not seeing any other questions, thank you for coming to this session. >> [Applause] >> [Event concluded]

>>