Gone Digital? An Introductory Discussion of Issues Related to Digital Government Information – Transcript of audio

Please stand by for realtime captions.

Welcome, everyone. We are just getting set up. We are testing audio and we will be getting started in just about five minutes. Let's see. Laura, do you want to test your audio?

Hello, everyone, and welcome to day two of the spring 2022 depository library virtual Council meeting. Before we get started, just a few reminders. Leaders as your questions and comments to the chat box that is located in the lower right-hand corner of your screen. Please follow along with us on our Twitter screen, GPO is tweeting you should and please use your photos and your updates. With that, we will start with our first program gone digital, and I will pass the controls to Renee Bozeman to get us started.

Hi, everyone. Can you hear me?

Yes we can.

Great. Today, I am presenting with James Jacob from Stanford University, Robbie Sittel from University of North Texas, and will Stringfellow from Vanderbilt University. I will start my video. I realize that didn't go on. We are presenting an introductory discussion of issues related to digital government information. A little bit about the learning outcomes. We hope folks will get out of today's session. We are starting with a real basic definition of digital deposit. I would like to say that some of this was born out of discussion on the listserv during the biannual survey completion time. There was a lot of talk on their about one of the questions related to digital deposit. That prompted this piece of this presentation, just as realizing that there are a lot of new folks in this community. There has been turnover in the community. We would like to make sure that when we are using some of this terminology, that everyone is understanding what we mean by digital deposit. And some of these terms that we are using, coming out of these working groups. So a lot of that is just a learning outcome to define the terminology that's being used. To go along with that, we are going to be talking about the difference between pointing to digital government information, as opposed to collecting or preserving digital government information. Talking about the importance of collecting and preserving this, and how it relates to the national collection. I know a lot of folks, me included, when we hear the term national collection, my mind often goes to the tangible materials that are physically sitting within our library buildings. So we would like to talk about how the infrastructures of the national collection don't necessarily just mean tangible materials, but that it doesn't translate to these new, more digital materials, and that we will need to start shifting some of our talk about the national collection. Finally, we are going to talk about the difference between persistent access and persistent identifiers. I am going to start off and just talk a little bit about national collection. I mentioned that when I think of it, my mind tends to go towards what is tangible. The national collection is defined as all public information products of the U.S. government, regardless of format or medium produced by federal employees, or paid with federal funds. So everything from print and --, which I think of the primary tangible. We have floppy disks and DVDs and other types of media in our collections, all the way to the foreign digital material that we are seeing more of now. It is always a great reminder to just remember that the national collection is a geographically dispersed collection that goes for print and the digital deposit, and that there is not a central repository at GPO in one of their buildings in Washington, D.C. The national collection is all of our collections. Again, accessible to the public at no cost. That goes along with our American informed

volumes. The national collection requires continuous development maintenance and preservation. This is easy to think of in terms of the tangible materials are collection development of tangible, how we maintain and preserve these tangible materials in our buildings. We also need to extend that thinking to the born digital formats, as well. So I think many of us know, with development materials, we have a lot of these links in our -- to maintain. And how we preserve the born digital materials. We will be talking a lot about that today, in terms of deposit. I believe I am going to turn it over to James, now. Oh, sorry.

Do I have the ball yet?

Let me see how to drag it. All right. You should have the ball.

Oh, yeah. Now I do. Excellent. Thank you, Renee. So as Renee said, the national collection is -- let me move this slide forward. The national collection is all public information products of the U.S. government, at no cost. Importantly, geographically dispersed not in a single building in Washington, D.C. And accessible to the public at no cost. We, as FDLP librarians, know after many generations of experience how to care for and feed the geographically dispersed national collection. When we are talking about physical objects and paper and --. In the born digital age, this need for the care and feeding of the national collection hasn't gone away. In fact, the need to curate more digital government information is even greater as the amorphous Internet publishing platform continues to grow input naturally. Only now, there are some different processes and workflows and an increased need for collaboration that need to take place in order to do the work of maintaining the national collection. I don't think we can stress strongly enough that the national collection needs continuous development, maintenance, and preservation. We can no longer simply sit back, receive boxes from GPO, but documents on shelves, and call it a day. Through the work of the PURL and physical depositing working groups, of which I am a member of both groups, as is Robbie and a couple of others we will introduce eventually, it became clear that our charges of the two groups actually overlapped. They were all part and parcel of the care and feeding of the national collection. So today, we would like to talk about the three legged stool of digital deposit, PURL, identifiers, and unreported documents. All key legs to the digital development stool. These three issues support and prop up the national collection, and need to be explored and conduction with each other. So that is what we will be doing today. First, we have a poll for you. We love our polls here. So I will give Kelly a minute to bring up the poll and start the music.

Thank you, James. I just need the ball.

Oh. I need to give you the ball, don't I? You have the ball. I wish I could just actually pass it to you. So take a minute and to this poll. What are ways to expand the national collection? I should have put hot sauce on there as an answer. I hope everybody doesn't just choose E.

Okay. Our results should be displayed.

Excellent. So, it looks like, for the most part, people chose F, which is all of the above, or basically, all of the above. A, B, C, and D. PURL libraries accepting digital deposit of digital files, FDLP libraries accepting metadata from GPO, notifying GPO of agency reported publications, and PURLs linking dispersed digital collections. Excellent. You all passed. And now, if you will pass the ball back to me, Kelly. I will go to the next slide. I've got it. Okay. So let's jump right in with the digital deposit leg. The term digital deposit has been around since the 1990s, if you can believe that. It was first mentioned in an official way in 2005 when it was included as a question on the 2005 survey. But to this day, there is confusion in the FDLP about exactly what we mean when we say digital deposit. So the first charge of the digital deposit

working group was to come up with a definition. What is digital deposit? We had some very intriguing conversations about it, and finally hit on the following definitions of what we are recommending. Digital deposit is the practices, services, and workflows for the collaborative acquisition of born digital and digitized federal government information for the national collection of U.S. government public information. So why is this important for the national collection? In the digital deposit working group, we were at first thinking of digital deposit strictly as the online equivalent to paper deposit. That is, GPO creating digital publications and depositing or distributing files in some manner to FDLP libraries. It soon became clear that deposit, as a clear workflow to development, can flow in various ways thanks to the Internet. It can flow from GPO to libraries in the form of depositing digital files in library institutional repositories and digital archives. This is how lots of copies think things safe part of the equation. With libraries building their designations for the local communities. This can be a plus. We discussed the different technologies that can go out and grab contact, or GPO could push content to libraries based on some sort of selection criteria, or something like that. All to be decided. But digital deposit can also flow from libraries to GPO. This is the libraries undertaking digitized leasing products of historical products and depositing those products with GPO. It could also be libraries doing targeted webinar information. Inc. of this as targeting important publications for inclusion into the digital national collection. It can also flow from agencies to GPO. This is the least applied flow of the content that is under the greatest threat of loss, in my humble opinion, since most agencies nowadays ignore their title 44 requirements at the detriment to the national collection. The greatest amount, by far and away of unreported content in our -- is from the executive branch. According to my analysis, 81% of the 300 terabytes of data that were collected in the 2020 and of term came from the executive branch. That is a whole lot of content. The key phrase, though, in the entire definition, if you think back on it, is collaborative acquisition, which hints at the need for many librarians and digital libraries. Just as the network of libraries has insured reservation of the national collection for the last 200 years, a distributed network of digital archives that is necessary for both preservation and access, digital deposit allowed libraries to build digital collections focus on the specific needs of their communities and patrons. This could be based on format like PDFs or spreadsheets, or GIS, or some other format. It could be agencies. Think -- collections of excellence, which focuses on specific agencies, or document type like maps. Or region. Oregon, Southern Oregon University, has the soda are cut, which is the Pacific Northwest biome in which they collect and give access to. Or any combination thereof. It is flexible on the Internet. So GPO, FDLP libraries and federal government publishers, will need to collaborate in order to curate national collection of digital deposit, which will be one of the critical parts of the next generation, digital national collection infrastructure. That is why the working group recommended that GPO undertake a digital deposit pilot project to map out the workflows of this critical piece of the all-digital FDLP. Next, I am going to pass the ball to Will so that he can talk about PURLs. Hold on a second, Will. I need to find your name in the list. There you are. So, take it away, Will .

Already. Thank you, James. So, PURLs. What is a PURL, I'm sure we have heard PURLs, but we are not talking about the jewelry. Of course, we are talking about libraries in the federal government. There is one thing they both have in common, and that is acronyms. What exactly is a PURL? You have all seen that this stands for persistent uniform resource locator. Or, PURL. You are probably familiar with uniform resource locators. Those are the web addresses that you put in when you are trying to search out a website. So a persistent uniform resource locator, though, as in an additional element. It is one that is meant to persist, or last for a long time. By longtime, we refer to, when discussing digital identifiers and technologies as the current technology and these next technologies. You have to be forward thinking. What does a PURL do? It does two things. The first of which is that it acts as an identifier. What it does is identify a specific document or object. Uniquely, it does that to where it disambiguate one document from anything else, from another document. In theory, it should apply if

you have version 1 of a document and aversion to. Those are two different documents. You should get those identified. Currently, GPO uses PURLs for identifying online content. So first, it helps the library identify the online document. The second part of the PURL is that it acts as an actionable redirect. What that is is the pointing part that was referred to at the very beginning. The PURL points the user to the document. So how does it get there? The very simple explanation is that when you click on a PURL, this takes you to the GPO PURL server. At that point, in the PURL server, the request is then sent to where the document is housed. This is where things get a little tricky. Some documents are housed on servers, a.gov, and it will take you to that document. There is a specific government information that is not duplicated on GPO servers, and may reside on an agency website or a digital portal. The portal will actually seem to take the user to that document that is on a different server. What makes it be persistent uniform locator is that that initial PURL link has underlying -- let me step back before I go into this part. This is one part of the national collection, but this is not digital deposit. So if you are working with PURLs, you may receive a catalog file or catalog record that has the pole in there. Or, if you just have the PURL by itself, you can take that and put it into the web address and put into a document. Notably, when working with PURLs, it allows access to the national collection, but you are not actually receiving a digital file . So there are actually two elements that are related to this. The first is the persistent identifier. So a persistent uniform resource locator is a form of persistent identifier. What that means is that, first of all, it uniquely identifies an object, which the PURL does and labels it. Then it is actionable. Like a PURL, you can put in the web address that takes you where you want to go. Well, it is where you want to go. You want to go to the document. How this works is that with the PURL itself, that is unchanged. So what that means is that your catalog record, the PURL link that you have, once that is assigned to an online digital objects, that should not be changed. However, the residing object itself -- sorry. The one that it is pointing to, let's say it is an agency server. It could be moved to a different server, or the web address or domain name may change for various reasons. If that's the case, normally, you would have to change your URL to get to the document. However, in the case of a persistent identifier like a PURL, what you do is use those pins, which is the identifier for personal identifiers. They have underlying metadata, which is separate from the descriptive metadata of the object that you are going to. The underlying metadata, for instance, is the web address of where the object or document is located. Let's say, for instance, let's move to a different server, and it has a new web address. What you can do is, the GPO can update the underlying metadata of the PURL. What that does is when the user clicks on that PURL link, even though the location of the object is changed, they can still access that because of the redirect changes. The user, even at the --, will usually not ever see that. It goes from server to server, and we still have seamless access. These must be maintained. For instance, if that location or the domain is changed, those two have to be updated or the persistent identifier does not work. This, however, is different from persistent accessing. This is the second part of PURLs and PIDs, and I will use those interchangeably. This persistent access refers to the accessibility of the object that is directed to by the PURL. This object is actually maintained separately from the PURL , or if you are introducing something different. In our case, GPO is using PURLs. In libraries, this is the system we are referring to. The PURL systems can be updated, however, if that digital object is completely removed off of the server , so let's say we have an example where you are utilizing PURL and it is linking you to a digital document on an agency website. If the agency website then goes off-line, or the servers are shut down, you can no longer access that. So the object -- as you can see, we have two different things that have to be maintained, the first of which is the actual object that you are relying on. This is where, for instance, digital deposit becomes relevant because if something is on an agency website that goes down, if you have a copy elsewhere, you can have that redirected to a different copy. Now, one of the things to remember is that in theory, you don't actually have to have a PURL or a PIDs to maintain persistent access. You could actually utilize the URL that the digital object resides on. However, and I will point out, in a practical sense, you still really want to have a persistent identifier or a PURL. The reason

being is that if you are relying on a URL and it is the domain name that has changed, the digital object has moved, and you are not aware of the PURL, that is an object that is much more challenging to find. As we know, in libraries, for instance, if you have something that is mis-shelved or that is not in your catalog, while yes, theoretically, it is still accessible, it is not practically accessible. These two things work together to ensure a long-term accessibility to digital objects. Notably, this is not digital cause because this is utilizing the redirect service of PURLs. At this time, I am going to hand it off to my colleague, Robbie Sittel. Robbie, I am going to pass to the ball.

Thanks, Will. All right. So I am here to talk about unreported publications. As part of the work of the digital deposit working group, we conducted a pilot project that assessed the strategies and time commitments related to discovering and reporting unreported publications. If you are interested in learning more about that work, please review our final report, which was noted yesterday during the updates. For those of you that may not be familiar with unreported documents, GPO has an official definition or description for an unreported publication. That is, unreported publications are public information products that are not discoverable through the government publishing offices catalog of U.S. Government publications. That's a mouthful. Unreported publications do represent a leg of the national collection stool, and provide an example of both digital deposit and the use of PURLs. So here is how it works. Once unreported publications are found and reported by the FDL community or others, GPO reviews the item to determine if it falls within scope of the FDLP and or the catalog indexing program. If deemed to be in scope, the publication is catalogued and made discoverable through the CGP, and as part of the cataloging process. GPO determines if it should be held on one of their local servers. If this determination is made, the publication is harvested and added to one of the GPO servers . This is digital deposit. So it would be the public to GPO pipeline of digital deposit. PURLs are also assigned to these newly reported publications as part of the cataloging process. As we will describe, a PURL provides a unique identifier for the publisher or object pointing to where it resides, whether on an agency website or on a GPO server. While PURLs offered GPO an easy way to update where the PID directs or redirects, to resolve access to the resource, digital deposit aids in ensuring permanence of access by bringing the resource object under the control of GPO and perhaps, one day, a network of trusted digital partners. So that is unreported publications in a nutshell. I am going to pass it, now, to James for wrap up.

All right. Yes. I have the ball. Thank you, Robbie. So, Vernet, Will, Robbie, and I have just touched on the surface of these important, interconnected digital infrastructures of digital deposit, PURLs, and unreported documents. We are hopeful that the work of the PURL and digital deposit working groups can go a long way to informing the outcomes of the digital FDLP task force. The key is to the care and feeding of the national digital collection, including collaborative digital election development by unreported publications and other avenues, distributed preservation and collection building by a digital deposit, and meeting together of various digital archives of government information through technical services, data, and the management of persistent identifiers. The three legged school can ensure that the GPO and FDLP libraries stand firm and work together on the long-term feeding of the happy collection. These workflows will make us all happy monsters. It is up to each of us, though, to guarantee that history is curated, collected, and preserved so that it can be evenly distributed to all. The future is already here, and we as a community need to step up to support the promise and continued maintenance of the national collection. Only you can prevent government information loss. Thank you, and thank you to my co-presenters. And now, we will just go to questions, if anybody has questions.

Thanks, James. We have a couple of comments to start us out. From rich, could digital deposit also be taken on by various -- around the country? FLV C in Florida covering all this date universities and colleges, for example?

Yeah. I can take that. I think I answered it in the chat, but that is a great idea, Rich. Individual libraries might not have the digital infrastructure to do digital deposit, but they may be working in consortia that do have that ability. That would be an amazing thing. I could also see other partners working as digital deposit partners. For example, the Internet archive or trust, or other large digital preservation archives.

I know you saw this one, too, but we have had a couple of folks give a +1 to this one. The question for discussion. As GPO continues to move digital, how do selective continue to fill open access to government documents if the public library doesn't have the ability to do digital deposit, postdocs, and a repository to the online archive, et cetera?

Yeah. I think that is an important question, and probably something that will need to be figured out. As Robbie mentioned, there are other pieces to digital deposit other than receiving digital files and creating digital archives. For example, unreported documents. So this is a huge area that is really difficult to tackle, because many agencies don't work with GPO for whatever reason . They choose not to. So there publications don't get into the national collection, but a group of librarians -- I used to call them fugitive hunters. Fugitive is no longer the term used, it is unreported. It is still the same concept. A group of librarians who were dedicated to this could track on agency publications. I noticed Daniel Cornwall mentioned that the Alaska State library is gearing up to do a more systematic work on unreported documents. So I would like to see lots of librarians working on that, whether it is targeting specific agencies or specific types of information, or even just serendipitously looking at the news and finding those often unnamed reports that are in the news that could be reported as a reported document. Lots of things for libraries to do.

James, if I could add --

Please.

To further go with the question, if you are looking at open access to golf docs, if your library is lacking in resources, that is precisely where PURLs have a great role, because you can receive those catalog records that will point you to those documents, or heck, you can even take the PURLs and form them into --. There are various ways you could put those out work, and that way, you're not infringing on the things that would actually be necessary for hosting the digital files. The bigger issue is that there is just so much information out there that I think one library, or one entity, would really struggle, first of all, to even identify all of this. If you did, the amount of server space and resources that it would take to host all of that. You still wouldn't want in one place, anyway, for security reasons. Lots of copies keep some safe. That is very true. I think with the digital deposit, it kind of works like the ideal library, and the concept where the idea is that one person couldn't possibly afford all of the books they could ever need. But if we collectively work together, we put our resources and fill a library where we could all utilize them. It is the same concept, except we are talking about digital files. The deal is, there is an increasingly large amount of literal information that is produced. When I think about the amount, I start to get sick to my stomach. I think about infinity, because it is just so much. So there are options, but even the ability to redirect or exchange, which is identify some of these documents, as Robbie noted earlier. Those unreported documents are then reported. PURLs are assigned to them. All of a sudden, everyone has a much greater ability to access those, and had, you could even go to the CGP to find those and you

don't even have to take catalog files. I'm a big proponent of libraries getting those catalog files and having more accessibility, but there are more options for those libraries that they have a little bit less resources to still contribute from accessing.

I will also add to that. To think about digital deposit and the lack of many libraries infrastructure to take it on, that we need to look at it as a community in the same way that we look at preservation's and other partnerships with GPO in that a lot of libraries are just not going to have the bandwidth for it, but some are. For the work collectively, there will be pockets of digital deposit across the country the same way that we have partnerships with GPO for tangible materials distributed across the country. So I think we're looking at this, knowing that many libraries are not going to be able to take on the deposit piece of it, but the knowledge of the concept is just going to be help as we move forward.

Okay. From Jim Jacobs. Kudos to GPO for aiming to have all public information products of the U.S. government regardless of format or medium! But, isn't the term national collection misleading since it implies that everything in its scope is in a collection? I would suggest that we limit the use of the word collection for the content that is actually collected, described, securely stored, and preserved for long-term access. Then GPO tried to describe how complete or incomplete the national collection actually is at any given point in time.

Yeah. I think that is definitely important to look at. You can't define something if you don't know what the denominator is. In the equation. As Jim points out, and just so you know, that wasn't me. That was my doppelgänger, Jim Jacobs. There are two of us, for those of you who don't know. Defining the entire scope and then what it is that we have control over, I think, is really important to see how well we are doing and create benchmarks and target information that is not currently in the national collection, but should be. I think that is really important.

Daniel Cornwall says, regarding unreported documents, should we be reporting new documents as soon as we notice them, or give GPO some lead time to catalog them?

I think Robbie answered that in the chat, but it is probably good to let GPO know as soon as you find them , because you don't know whether they are in the pipeline for cataloging, or whether they are not in the pipeline. So GPO can easily resume the form and know whether or not it is in the pipeline. So yes. Report away.

Amy says, I think it would be nice if someone will revisit the scope of what falls within the FDLP and the national collection, and what does not for some of the new coordinators here.

It looks like Cindy has pasted into the chat the scope of the national collection is the corpus of federal public information regardless of format or medium, produced by federal employees are paid for with federal funds. The cost of materials and material privacy are included within the scope of the national collection. That is interesting. I have never seen the declassified piece of that, Cindy. And I wonder about documents that are made public through FOIA requests, if they would fall within the scope of the national collection if they are, indeed, quote unquote, public declassified information.

Okay.

This is Cindy. We have that last statement about declassified documents and some early presentations that we did, and we also included that some discussions that we have had. It is certainly one of the

things that we are looking into as part of the national collection strategic plan we have as an objective in the next couple of few years. Prior to get in the loop of the declassification of documents, so we can figure out how to best bring that content in the national collection.

Have I got some FOIA documents for you.

Yeah. We have looked at FOIA.gov and looked at some of that stuff. Some of it is more defined as records.

Yeah.

As opposed to publications. So there is a line there that we need to investigate when we start looking at FOIA materials.

Okay. From Brian. Are there official guidelines or standards for what can become an unreported document?

I guess a guideline for me, and my panelist pals can jump in on this, as well. Anything that you find that you think is within scope of the FDLP , so if it is a public document that was published by a government entity, then you can check the CGP to see if it has been catalog. If it has not been catalog, then it is an unreported document. And you can send it in.

Pamela asks, how likely is it that some small selected depository will run across one of these fugitive documents?

Yeah. I think, probably, the possibility is high, because there is lots of stuff out there that is published. You might find it on a tweet, you might find it on Facebook, you might find it written about in your local newspaper. These are all places where government publications are mentioned, and many times, the fact that so much of the federal government publishing is unreported, chances are very high that you are going to run into an unreported document. Just as you are walking around, or on the Internet, watching YouTube, whatever.

I am going to add to that that the project that the digital deposit working group conducted, we found that even though we sent most people searching for things, most of what was found was actually found serendipitously. It was not what they intended to find. So you never know what you are going to run across.

Yeah. I think that somebody in the chat mentioned that maybe adopting an agency as a good way to do this. I think that is a really good way. Alaska state library looks like they are looking to do just that. At my library, we have been doing that for probably about 10 years now, where we track on 10 different agencies like NIH and other agencies like EPA. We just go to their publications page every month, or every quarter, and go down the list and check to see if they are in the CGP and send them in. So it doesn't have to be serendipitous. It could be more proactive.

Doris asks, how does GPO address PURLs that are no longer active ?

Hey. This is Lori. I guess I can take that on. We have quite a few staff who actively go try to find publications. The publication we are pointing to the agency site. We will go look for versions. We will

relink the PURL to a stored version. There is a lot of different mitigation ways that GPO tries to ensure that the PURLs link to documents, that we don't get errors. So in the day-to-day maintenance activity, I don't know if that helps or not, but that is how it works.

Yeah. I think that is one of the inner workings of PURLs that a lot of people don't understand. When GPO creates a PURL, when do they create PURL that points to a live agency server, versus collecting a document and storing it in government info or permanent GPO, which I think is permanent. GOP.gov now. Other local servers, wind is a PURL that is pointing to a public .gov URL refer to the stored coffee of that document. So those are all, you know, intricacies of GPO cataloging and technical services workflow.

Which reminds me. Thank you, James. There are quite a few webinars on FDLP Academy go through that process. Ashley Dahlen has done a presentation a couple of conferences ago about how the intricacies of PURLs . Folks might want to look there. Here's Donna Kramer, who is one of our cataloging staff, talking about PURLs, too. So there is a lot of information in there , and at the Academy, about how we do PURLs and catalog PURLs.

From Elizabeth. How are libraries handling digital deposits? Are they dedicating a server, using a commercial product? I don't know the practical aspects of archive and digital materials.

I can jump in real quickly to talk about my library. We do not do this on a large scale, but we are part of the locks U.S. stocks, maybe. James will probably correct me after this and give the correct name.

Know.

And so, we get archival units every so often, and it is only for a certain subset of U.S. stocks that we get in the Congressional record, and a few others. We do have a dedicated server that those go on, with our other locks material. For my library, that is a really manageable amount for us to be able to take on, and that is all I can speak about from our end.

Thanks for mentioning locks U.S. stocks. That is one form of digital deposit, if you will. The one piece of locks that maybe isn't really about digital deposit is that locks is a preservation server, and it's not meant for public access until such time as the content that they are collecting goes down. So right now, locks U.S. stocks is only on contact with info. We harvest all of this content and distribute it out to the locks library who run their servers. In such a catastrophic event that of info would go down, that would trigger the public access of those archived copies.

I would just chime in to say that I don't know if we know of anybody that is actually doing a federal traditional digital deposit type project. Some of us may be doing piecemeal things, or be part of the -- which is a way to preserve federal content access. It is still a struggle. Part of what the digital deposit working group has done is identify libraries that may be potential pilots to see how those workflows might happen. So, fingers crossed.

Yeah. There are libraries who do have government information on their Web servers, or in their digital archives. I have linked to the soda archive in southern Oregon, digital archive. They have lots of content in their digital archive, and government information. I don't know how targeted or holistic they are in collecting, or intaking government information from guv info or from CGP to put into their archive. But there are libraries who do have government information in their digital archives. Our library, when I

report something as unreported, I will frequently but not always collect a local copy of that document. Especially if it is something that is of particular interest to Stanford researchers, or just something hot in the news. I will save a local copy and store it in our Stanford digital repository, and catalog it, along with reporting it to GPO .

Let me just add, I put a link in the chat to a digital deposit working groups final report. Part of that final report includes our survey responses that we got, and one of the questions was about identifying or finding unreported publications. What they do with them after they find them, whether or not they download them or make them publicly accessible. That might be of interest to you. We did find that 50.7% actually saved the unreported publications to their computer's hard drive, and only 12.3% added them to an institutional repository. So that gives you some kind of indication there, but there were a large number of different responses there, if that is of interest to you. Check out the digital deposit working group. Final report.

Okay. Holly says, digital deposit items should be deposited in more than one server, I would think. As the group considered that as a recommendation or issue?

I think this is something that the PURL group is actually addressing, and some of their principal documents. Anybody else wants to chime in?

I know, at the very minimum, we have discussed this. Digital preservation, just like any other thing, you want to have more than one copy. In terms of with the PURL working group, you started -- one of the things that we have run into, also, is in drafting principles for personal identifiers. One of the one big things you want to have is a 121 relation. This personal identifier to really be unchanged and go to one document, and that document is exactly the one that it says it is. So if I access today by the PURL, if I go 20 years down the line, I should be able to access that same document. The question you make it into his, well, if I have a copy 1, and I have a copy 2, those are technically different digital objects. It doesn't seem practical to assign different personal identifiers, but you could make an argument for it because they are in different places. From a practical standpoint, it really wouldn't make sense for the copy to have a different PURL or a different PID . It would make way more practical sense to have your geographically dispersed copies on different servers, just like your --, and if one of the servers goes down, you could redirect to a second copy on another server. My understanding with gov info is that there are multiple copies that GPO has . So, yes. If you are looking at digital deposit, that is precisely what you want. You want multiple locations for this stuff, in case someone's technology updates cause compatibility issues. Or you have an Internet outage in a server area, or even a cyber attack or some ransom where it shuts down one server. It definitely is in the interest, and relive the spirit of the national collection, from a digital standpoint, to have this deposit in those files in multiple locations, preferably geographically dispersed.

Yeah. I think that is one of the strengths of PIDs , not PURLs specifically, the other PID systems like DOI and handles. You could have the PID resolved to multiple copies. DOI's do that now, if you go to journals, an article, and you click on eight DOI for an article, you are not necessarily getting the article from Wiley's central servers. They have servers across the U.S., and it might resolve to a server that is in California, if you are in California, because that is the closest geographically close server that is close to you. And so, ostensibly, the content would be served faster on a server that is closer to you, rather than getting a copy of the article in Belgium, because that will take a little bit more time. Nanoseconds on the Internet, but still more time to serve that out. It is also a load balance, and I think that would be an amazing aspect to build into GPO's future PID system.

This is Cindy. I think one of the reasons that we have a group of the depository library coordinators who are in favor of digital deposit is that you are going to have something deposited in many libraries. If you relate that to what we are doing, historically, and still currently with the tangible materials, we are sending them to a bunch of different libraries. A lot of people have an addition of digital deposit that would do the same thing with digital files going to multiple libraries. There are a lot of intricacies that need to be worked out, how to do that, and the whole preservation question. All of that. But the duplication and the redundancy of having the content on different servers underlies the desire for digital deposit.

We have just a little over a minute left. Bernita asks, will digital libraries be required to accept all the GPO offers?

I think that is a question to each of the regionals. Currently, there are 10 regionals that are working on the locks U.S. docs project, so there are some regionals who are participating in digital collection development, as well as in paper. The regulations aren't written in such a way that regionals will be required to take digital content, but I would hope that there would be some who would be able to and would be interested in doing that.

This is Cindy. Let me add a little bit to that. As James mentioned, since we actually aren't doing this yet, we don't have guidance in place, and we won't until we see what happens and how things will work. Currently, regionals do not have to select all formats, so if receiving digital files is one of those formats, it wouldn't have to necessarily select the digital content. They could thought to have tangible -- if it is, indeed, available as tangible. Those are among the many questions that we will have to look at in the guidance that we provide our depository libraries as things progress.

Unfortunately, we are out of time for this session, but we will capture all of these questions and comments and share them with all of our presenters. We will take a brief break and be back with you at 1:15 for the retro median session. Thank you, everyone.

Thanks, everyone. [The event is on a recess. The session will reconvene at 1:15 P.M. Captioner on standby.]