DOIs, PURLs, URIs Oh My! An Introduction to Persistent Digital Identifiers - Transcript

Please stand by for realtime captions.

Hello everyone, once again thank you for joining our session today before they get started I just want to remind you if you have questions or comments or technical issues, put those in the chat box in the lower right-hand corner of your screen and I am going to keep track of questions that come in and will relay those to the presenters after they are done with their presentation portion of the session. The session is being recorded and everyone will receive a link to the recording in their email and also be available on of the LP.gov and in the session and there seems to be a little bit of a web ex glitch where you cannot chat to everyone and you can only chat the presenter so if you have a question or a comment, just chat it to the host, panelist and the presenters the entire group of folks and then we will get the questions re late back out to the audience for this session. Some thank you and with that I will hand it over to will straight fellow.

You so much, Kelly and welcome everyone and good afternoon and welcome to the OI PURL URIs all my an introduction to persistent digital identifiers in this program is presented by the DLC PURL working group which is also known as explore the durability of PURL and their alternatives in our speaker today will be James Jacobs from Stanford University and my self from Vanderbilt University and dolphin GPO and here's a list of our working group members and the help put together this presentation as part of our research. And part of our judgment part of our deliverables as a working group. This presentation is a lot of our preliminary findings at not really findings but preliminary information research into persistent i dentifiers and such.

And I want to say thank you to all of the group members for contributing and hope you all enjoy so let us get started.

Before we really go into the actual identifiers we need to have a very basic understanding of electronic document retrieval. What we mean by electronic document retrieval is how is it you go from your computer or at your library catalog to retrieving the government document to or any document for that matter? So we first need to understand that the electronic documents or electronic objects like files and documents are stored and they are stored on file servers and file servers exist inside of computer networks and computer networks made up of multiple different computers, servers, and they all serve different functions. The fileserver is really important because it is a computer, computers, sometimes even just a program on the computer that work within the network and have a specific purpose and that purpose is to store information and store files and then they also do all of their work dealing with making files accessible, so for instance on the back end, on the private end there is the ability to update, at files remove files and on the public end is where the files are actually communicated usually through a Web server or other server to make these document store on its storage available to the user.

When you go to access a document, or object, first thing you need to do is you actually need to dial into the network that you are retrieving the document or the information from.

Now normally you actually are going to -- you type in your URL, and what actually happens is it finds the IP address of that network. So then you most likely interact at this point with the Web server of that network which is the public end and you do not have access to an entire network but it is the part that will communicate and you will then actually go through and send in a request to the network I'm sorry, to the Web server that would send something to the fileserver if you are requesting a file.

The fileserver gets the request, it is then going to search its inventory of the object that has an there in the files, and then it will provide a copy of that object back to the user. Now it will do this in many number of ways and it may allow you to view it and may give you some metadata or let you download, but it will be a copy of that. It may also return error message of the document or item cannot be found. The question why is this important when looking at digital identifiers is as you can imagine there are a lot of networks that store a lot of documents. And finding ways to uniquely identify documents so that you could find the network that they are an and once you get into the network file servers, we really would not have an issue locating files or documents of there were five of them on the server, the downside or reality is they usually have a lot of documents on there so these identifiers are critical in helping to find these documents in pieces, and not only the happy find them in the other goal is to make sure we can continue to retrieve these documents long into the future.

I will turn it over to James who will talk about uniform resource use.

Okay, thank you. Actually I will talk about URI first [Laughter]. Pardon me. The infrastructure of the Internet as Will has gone through, is filled with acronyms that we have all run across but maybe do not have more than a vague understanding of how they work. URI, you are in and URL are the lingua franca of mapping traffic on the Internet and key parts of the domain name system, DNS, which has been in the GPO news recently, we will talk about that later, but DNS is hierarchical and decentralized naming system for computers, services and other sources, resources, connected to the Internet.

In their uniformity a high school student in Des Moines, I want or historian and Seattle Washington get to a hearing at one of the 4.16,.1624,.242, which is of info IP address, or the American memory image on Elsie server somewhere and Alexandra, Virginia. Through this massive network we are able to find, identified, read and harvest websites and online content. The URI or uniform resource identifier is a globally unique string of characters of a specific resource and it can be an ISBN number, web domain or IP address. In other words, URI need not be on the web. URI is a superset made up of you are ends and URLs, dose human readable web addresses that we all know and love.

Now I will go back to you are in, how about that?

You are in on the other hand are universal resource names and they are a kind of URL, I'm sorry, URI that provides a unique name for a digital resource. Unlike URLs which always start with http or HT TPS, or FTP, which is the new way that you are going to access the file on the fileserver, URN do not specify a way to retrieve an instance of the resource to find a resource named by a URN, you have to use some sort of resolver servers. Such as CTS or canonical tech services or DNS.

URN name spaces are required to be registered with the Internet assigned to numbers authority, I ate and and. These resolver services in which our working group is primarily interested because in these resolver service, EPO can provide persistence to the national collection.

I am going to go to persistent identifiers okay, everyone is properly confused now between URN and you are in and you are I, right? But in order to talk about and understand the PURL , those permanent links that we have an our library calendars, that GPO create and so maintain, with to start with an understanding of URIs in order to talk about those persistent identifiers of which PURL is just one example. Our working group as Will has explained has been tasked to explore the world of persistent identifiers in order to make recommendations to GPO on which technologies to leverage over long-term sustainability of the FT LP.

Persistent identifiers are long-lasting references to a document, a file, webpage or another object in addition to digital repository and these identifiers allow for traditional identi fiers like I is PN, ISBN, sorry, orchids or URLs, to be resolvable over time.

Any digital and institutional repositories rely on persistent identifier systems to better manage their content and provide a persistent link for citation services over time.

Some of you may know about permit CC organization that provides permanent links to -- especially to legal documents, but they are a good service so you should check them out permit.cc. The key to persistence is two fold, identifier itself has to stay the same over time. And the object to which it points to or resolve must stay the same over time as well. Regardless of whether that object name or address changes. Persistence is both a matter of good policy and practice and technology. That is what we are interested in with this group.

And now I will hand back over to Will to expand on the idea of permanent links and then we will discuss some of the permit link system that this working group is exploring in order to recommend the best option for maintaining long-term access to the national collection, arcs handles, DOI's and PURL, oh my , will, take it away.

Thank you, James, start with permit links and as you might imagine a permit link is a permanent link or a permanent URL. So karma link has kind of turned out to be sometimes use anonymously with persistent identifier but is probably more accurate to say it is a type of persistent identifier rather than say it is a persistent identifier. And so, one of the things that is really critical about permit link is that it really has a lot to do with the intention of the link,

when you create or purchase your domain or acquire domain, you intend to create your links and you want to change them and you're going to keep them -- your links will continue onwards and so it is a decision you will make upfront that these are going to be think that you will not change so most often these are actually created within a content management system, so if you think of a case where like a newspaper for instance if they generally purchase a domain for their newspaper, as they add on articles they can create Permalink which are permanent links to each one of their articles to make it easier to access each of those individual articles. Now if it is an organization planning on going for a long time, these are a pretty good tool for ensuring long-term access. But once again these are usually done within organization content management system and there is not a registry for Permalink's. Now permit links are also very helpful with search engine optimization and there are a few reasons why so if we look at the characteristics of a Permalink, first of all -- if you look to the right you will notice there in the top there is actually an example of the Permalink. The first thing you should notice is going to be HT TPS and so that is a secure page. That is something that you really want to have an that is the first sign of a good Permalink. The second part is that when we look at the link itself, you want the link to describe what the webpage is so in our example of superintendent of documents public policies I am about 99.999 percent sure that that is not going to take me to a page of Gov statistics but to the super Internet documents public policy page. Another think that is really critical is separating the words with hyphens and the reason being is that these are or when a search engine looks at these they can identify the words themselves and one of the things that people do not normally think of as computers are incredibly stupid, they do exactly what we tell them so if we don't put in the dashes, in the Permalink example here, the computer does not know that they are supposed to be dashes because we did not tell it that they're supposed to be spaces there. And so, and using the practices for creating this, this makes it easier for search engines to find these materials and also able to remember -- so if you frequent this page, it is something users could connect with them to have an idea of what this actually is on that page.

So Permalinks and PURL or forms of persistent identification and they will direct you or redirect you to a rep source. This is where things get a little bit different though . URLs -- I'm sorry a permanent URL is going to have the same domain which is the first part of the web address. So in our Permalink example the www.PLT.gov will actually be the domain part. The slope which comes right after that or kind of the sub pages that go into her link back to the main domain, now generally domains do not change and your Permalink will be stable as long as you have the same domain. However, if you do change it, well it is probably not going to be accessible. These time of links, dryer Permalinks are generally to be used for years, PURL's on the other hand are slightly different payment now one thing to note , PURL's are actually example of permanent link because the object of a PURL is to provide persistent access to a resource and the big difference though is that a PURL is going to have an independent domain and so as we see in our PURL example goes Perl.PLT.gov and that will redirect us to the domain of the Gulf info.gov with this long URL for this item here. Now because this is going to an independent domain, these links and these permanent links are actually tending to last instead of years they are intended to last for decades. And so we will now move on to our next persistent identifier and that will be arcs and I'll hand back over to James.

Thank you, Will .

Yes, so the first permanent identifier system that we will talk about is the arc or the archival resource key and this was developed by the California digital library specifically for digi tal archives and the arc is a persistent actionable naming scheme. An arc links the user to, through things the object itself, its metadata, and the commitment statement that the provider will maintain long-term access to the object. The ARK scream --'s screen produces unique URL as you sleep here on the slide and hope you could see that on the slide and the difference between ARK and other permanent identifier systems is that ARK are generated and maintained by the organization that creates them like a library or digital archive while other systems like DOI and handles rely on a few named mapping authorities like handle.net or DX.DOI.org and that is centrally registered and generate and maintain these unique Permalinks. Working group will be sure to explore the pros and cons of centralized versus decentralized Permalinks systems because there are pros and cons to each. University of North Texas, which is this example here, assigns arcs as permanent identifiers to its digital library. As part of their statement of commitment, UNT, says that no arc should be reassigned, that is once an arc to the object association has been made public, that Association shall be considered unique into the indefinite future. So the arc will not change and hopefully the arc that it is describing to the document will remain persistent as well.

Another persistent identifier system is the handle system. Like DOI's, the handle system is undergirded by a registry of persistent identifier streams, similar to ISBN's. Handles consist of a prefix and you'll see that here on the slide, like 20.1000,/100, which indicates the namespace and the naming authority and a suffix which indicates the local name to yes, directly to the digital resource itself. And so if you look at the handle you can look or just like you look at ISBN you can tell who the publisher of this content is with the handle you can tell who the organization is because that organization is registered at handle.net. The global handle registry is underwritten by and run by the Corporation for national research initiatives in conjunction with the don't know foundation nonprofit organization based in Geneva and I will pass it back to Will to talk about DOI .

I think you have the ball, Will thank you, James all right, DOI, digital object identifiers and coming from handles over two DOI is a perfect transition.

DOI is digital identifier of an object cannot to be confused with identifier of a digital object. Is a go through the next couple of bullet points I will further expand on what is meant by that. Simply put DOI is a series of numbers and punctuation that provide unique identification is for objects. What is really important is they could be objects of any type. The DOI actually uses the framework for persistent identifiers, specifically what we just saw with handles, and they also provide actual unique identification or digital resources. Now I don't want to just harp on digital resources because the OIs can actually be assigned to any entity. You can assign them to a physical entity and you consign them to the digital entity, you can also assign them to an abstract when idea could actually receive a DOI but the important thing to take away is that the

identifier is actually the digital part so it is a digital identifier that identifies an object and now the object can be digital object but it doesn't have to be so that is one of the things it does get a little bit confusing. Now one of the things about the DOI is that you do go through registry process for the DOI and when we start looking -- sorry, when we look at the DOI's, here is kind of what the sample one looks like it is not filled in and it contains both the prefix and the suffix just like with the handle and they are separated with A/P make us look in depth at what a DOI looks like.

So this DOI 10.1000/100 -- 108 to and this is real do I owe what I want to do before I go to display and this one is that many DOI's are actually displayed and resolved as URL so you can actually put in this URL of the DOI.org with the DOI numbers on here, that is actually going to take you directly to this digital object that has been identified.

When we break down the prefix, we see the 10, now this is one thing that will not change with any DOI. All DOI's will start with number 10. And the reason they do that is this is actually the part that designate this as part of the DOI namespace. So this flat out says, hey, this identifier is a DOI and we need to utilize their dictionary. We do not want to use number 82, that is not going to find these items that of the registered. After the first period or full stop, we see the number 1000 so this I guess we could call it a variable actually identifies a registrant. So what I mean by registered I mean this is the entity person or organization that actually registered the identifier. Now this does not mean that they necessarily own the rights to the object that they are identifying. They are just the ones that register the actual identifier itself. And that is what takes up or makes up the first part which is the prefix. Now the suffix is that -- sorry which is after the flash is actually the item ID. And so, when you combine all three of these, we note that it is the DOI namespace, we know who registered it, and then this is the item that they registered, and it gives us a unique yes, it is a way to uniquely identify just about any object that you could think of of whether it is physical, whether it is electronic or whether it does not actually exist and I guess in a real sense but it does an internal sense.

This is the basic overview for you guys and I am happy to yes, X think we will discuss is -- PURL and I will hand it over to Ashley Dalen.

Okay, I have the ball now, let me get to the slide so hello my name is Ashley Dalen outreach librarian had to peel and I work in the federal depository support services unit used to be called outreach and support but now we are if DSS and I'm here today to talk about one type of those Permalinks that GPO is actively using and that is of course PURL, not the one that you see on the right side there but persistent uniform resource locator and GPO began using PURL Peck and 1988 and we continue to use them to this day .

So if you're unfamiliar with GPO PURL you can find them in our to catalog records and this is a screenshot of GPO's catalog the catalog of he was permanent publications and anyone who has downloaded or otherwise copy catalog from our workers will have the PURL and their local records as well .

You can also find PURL embedded in various webpages upped as [Inaudible static] and more and this is just a quick example of the guide and found just happened to be on the JFK assassination and there are several documents that are linked to in the subject guide that if you click on it you'll actually realized they are PURL's. This is a very rough diagram of how PURL's work at least a GPO payment when a user clicks on a PURL their computer or device will go to GPO PURL server and automatically be redirected to wherever the content resides. GPO does maintenance on those pearls and we will change where a Perl Routes a user to when it is needed. In that way, elaborate needs to only catalog the record or link to that Perl once. Permalink remains the same but with the user is seamlessly redirected to it will change to wherever that concert happens to be at that time. So the point been GPO is one doing the maintenance. On the right side there I am included some of the major locations for GPO cataloging staff may choose to redirect a PURL to and includes places like the of info.gov and agency websites just to name a few.

Piggybacking off of that, we have been asked to report in approximation of where GPO is redirecting PURL's to and I will be honest and upfront here this was just my eyeballing a very long list of web domains that we PURL to so bear in mind there was a mistake in my flagging category where we were routing PURL's to and that was my mistake entirely, my own doing there. Overall though I think we could roughly say it appears that about 38% of PURL's and redirect content found on the GPO server, 32% redirect to go in Polk content 12% redirect to the content found on other federal agency websites and 10% redirect content found on digital access partner website and about 7% redirect to a web archive of some sort and that number is growing.

Next question is, when GPO catalogers are cataloging a news resource, how do they choose where to redirect a PURL to? This is just a sampling of some of the decisions that each cataloger is going to have to make when they create these PURL's. Catalogers know that content found on of info is probably the most secure because it is a preservation repository so if they have content that they are cataloging and it is located there the PURL will redirect there and the content being catalog is tricky to harvest on the agency website tends to be stable, they might opt to redirect the PURL to the agency website and this is especially true of the content is updated frequently .

They do this because catalogers want to ensure that the end user is redirected to the latest and greatest edition which may be located on the agency website before it is identified and harvested by other entities.

The content may be pulled down are harvested by GPO staff and the cataloger is going to identify first of the content is a monograph or a serial. Attends to be easier to harvest monograph unless it is a pesky monograph with the online version is broken up in their incipit files for each chapter. That me present a problem from a usability standpoint for the end user so the cataloger is going to take that into consideration when they are choosing whether or not to harvest the material.

Sometimes the cataloger can spot that a special feature cannot be captured when we harvest it which might be under the reason for them to redirect a PURL to the agency website is they do not want to lose access to the special features for the end-user. Another consideration for GPO catalogers is digital access partner is in place. In this case the agencies have signed an agreement with GPO that they will archive content and if they take it down they will give GPO a copy of it so GPO catalogers are not going to harvest content that is already locked into an agreement.

And web archives also used when creating PURL's and GPO catalogers are going to evaluate the completeness of the web archive capture to make sure the content that is described and in the new catalog record was sexually captured in the web harvest. In some cases, the catalog record might actually have multiple PURL so for example one PURL may redirect to the agency website that may for example have three additions to the tile and latest addition being the 2019 addition. A second PURL in the catalog record may redirect to a GPO server where three additions are also available in the latest edition also being the 2019 once a why do we duplicate it? That is what is going on is the cataloger wanted to ensure that the latest addition that was available to the end-users so they created one PURL that went to the agency website that is likely to have the most recent condition and created another PURL to the GPO server and in this case though GPO had caught up with the agency website and had harvested the latest 2019 edition so it is not common to see a link to federal agency content but then also another PURL that goes directly to GPO server.

Okay, GPO catalogers work pretty extensively with digital content and you can see here that since GPO first began using PURL's GPO catalogers have created over 257,000 PURL's and on those PURL's we have done some form of maintenance about 520,000 times so that means that for every PURL we have created it has been updated an average of two times and the number is a bit misleading though so we will talk about why that isn't just a minute.

Up next we will talk about some of the challenges to using PURL's as they relate directly to the work of this depository library working group in the first and foremost challenge that many may recall is what I call the great 2010 PURL server crash were PURL were not able to redirect end-users to content and to address the issue the PURL server is now a hosted site, secure, and back up. We've actually done test rollovers in the past to see if there any glitches when we redirect anything to alternate PURL server and you all did not even notice that it ran so smoothly. That is one challenge I believe has been addressed but I'm superstitious and always knock on Wood.

It is not also salient also when a PURL does not work there could be a handful of reasons. Here is just a running list of some of the more common ones that we encounter.

Agency websites could go down which will obviously impact any PURL that redirects to that agency website and see this happen for example during federal agency shutdowns in the government may be bombproof that if an agency does not have funding to operate, it can be really, really hard for an agency to maintain their websites.

Changing federal agency web domains, for example going from ABC.gov to the alphabet.gov, that will break all the PURL's redirecting to the older domain. Sometimes when that happens where able to go in and fix the PURL's and have a local change but not always. There are handful of agencies that have a reputation for domain changes, no names.

Agencies redesign the websites. They usually we design it to make them better. But any sort of massive file restructuring of a website forces GPO to identify how best reroute PURL's to the contents we location and hopefully we can do that with a global fix as well.

For PURL's the rep the Internet archive, which is a web RK, we've knows the Internet archive does routine maintenance sometimes on their files during normal work days. When this happens, the content is not accessible so if you click on a PURL or running a link checker, and the PURL route to the Internet archive and you get error message, wait a day and see if the PURL's are back up and running again.

Under the problem is that agencies may not be following the proper HCTP protocol for the 404 error codes a link checker be it yours or ours, it may see a page that says no longer available or something. But because upper protocols we are yes, were not being followed the link checker things everything is okay even though we as a human being can edit and see clearly there is no content there and that's a problem.

Interestingly though, most of the problems that we are seeing with PURL today are actually not related to the stability of the content and the content is stable and the PURL server is properly redirecting the under user and the problem is increasing network security. There are a couple of different flavors of network security that I am talking about.

Up first is increasing network security on federal agency websites that make it hard to get to content and the easiest example of course is the Department of Defense and they are always changing or improving their network security so what you used to be able to get to with one click might not require additional steps or you might be blocked from the content altogether and GPO has to make decisions about how to get people to content at that point. Ever PURL RSU to http web address, you should automatically be redirected to it http S address and the federal government was directed to migrate to http S because it is a secure connection however, sometimes owes rollovers to net happen seamlessly with certain agencies.

Content may actually be there and maybe your library needs to install security certificates so if you have a web browser that says this page is on say, you can usually view the content if you know the browser and how to work around it, otherwise you shouldn't so the security certificate on your computer so you computers recognize the status stated.

Okay. When a PURL does not work, -- I apologize I lost track your. The most recent security issue that has popped up is called the DN SSC protocol for domain name system security extensions and it is a new protocol that your IT may be adopting. If your Internet service or your network provider has adopted the DNSSEC protocol, it will not resolve links for you to websites

that are not configured for that security protocol. And that includes retired domains that are unable to validate and confirm what this new security system is requiring. If you look at the chart on the left, UC permanent.access to GPO.gov effort web date.access.GPO.gov and PURL .access.GPO.gov those a retired GPO access domains and they are not able to validate that they meet the security protocol. Bottom line is that PURL's are redirecting directly the content is still housed on those GPO servers, but your network security, if it has adopted the DNSSEC protocol, it will block you and your users from getting to the content.

The new security prompted GPO staff to modify PURL's and we did it in a new way so this is why as said earlier that we are doing various types of maintenance on PURL's over the years, the bulk of the maintenance that GPO staff to on PURL's have simply been changing where PURL's redirect to, right? We change where it redirects to but that PURL link itself remains stable and with this last challenge that arose, we actually went in and changed the PURL link itself to the new domains that are located on the right side of that table .

Those new domains such as permanent.F DLP.gov are active domains and they are configured for the new security protocol. So GPO change the PURL links in our catalog records. Did you all change the PURL links in your catalog records ?

This is not a normal operation that you have had to do before. If you have not refreshed your records, the PURL links in your catalog records are trying to redirect users to the old domains. Help libraries out you can download for free a zip file of over 170,000 marked catalog records and you can find them on the CGP on GitHub and if you're not familiar with it CGP on GitHub it is linked to office F DLP.gov and you just mouse over collection tools and you will see it listed there in the drop-down menu. Also put out doing news alerts about the situation and I have included the links here on this slide.

Sometimes despite all our efforts to ensure that PURL's rap content the content does in fact go dead and this is largely a reflection of how the logging and archiving practices have evolved over the years. Remember that PURL's are designed to route people to the content described in the catalog record. Very concept of archiving the content that was being described in the catalog record actually developed on their own and until 2005, if there was a print and a digital edition of something, permanent copy or less was considered to be the rent copy that actually recited in your depository libraries. So if we had it in print and electronic, we relied on your print copy to be the archival copy. Beginning 2005, when two formats were available print and electronic we still started archiving the digital copy as a backup.

So fast forward a couple of decades, at the content is no longer available it once was, we have to try to find another digital version of it because we don't have it on a GPO server. If we cannot find another version, you get redirected to a PURL history page and the PURL history page will contain the history of where the content originally recited and all the various redirect corrections that may or may not have been made over the years.

There are some other reasons for dead or lost content so some of you may have noticed recently that yes, I will give your bread example here that some of the Federal Reserve content is no longer available to its PURL, and that was a result of changing scope determinations per the Superintendent of documents public policy statement 2019-1 so the content scribe in the catalog record is deemed to not be in the scope of the F DLP, pearls will get redirected to a history page.

Digital content may also be withdrawn at the request of the agency or it may be withdrawn at they were west of Congress let me give you an example of that. The legislative branch of Appropriations Committee, that committee requested that content be removed from Gov info as individual annual files and they were placed those individual annual files with one large appropriations omnibus file. The annual files described in the catalog records no longer exist as annual files anymore. So the resource described in the record is no longer there. The PURL's were redirected to the history page. And GPO subsequently created a new catalog record for this try enormous omnibus appropriations content so the contents there, just in a new record and the other previous records that closed off, if that makes sense.

Finally sometimes we see PURL to go to a history page and the tool described in the record has ceased and I think we will probably see more of this as time goes on so let me give you an example there was a catalog record created for FD scissor federal digital system, that resource is no longer in existence and has been retired in favor of Gov info.gov so they went in for that catalog record the DYsys and they reset the PURL to go to the history page because that resource no longer exist and then they created a catalog record for the info and that PURL obviously correctly leads to gov info. Fun fact for y'all in September 2020, there were approximately 230,000 catalog records with PURL send them and again as I mentioned earlier, many of our to catalog records of more than one PURL and them so that is why there are fewer catalog records and there are of PURL's. 561 records had content that were no longer available so in terms of grand scheme of things, did content accounts for about 00 -- I'm sorry .00 24% of records in the CGP.

Okay, that was a crash course in GPO PURL's and hope I've given you sent to some of the challenges we face when redirecting end-users to digital content in a streamlined and permanent way, and I do want to give you a heads up that the DLC and one working group will be gathering feedback in the future but the challenges your library has with making digital content available so please stay tuned in the future for that.

I have sort of seen the checkbox score by and I'm kind of petrified [Laughter], what kind of questions do we have?

Ashley, I am trying to follow this so other presenters please take a look and make sure I do not miss anything here. Jenny says in early days of web I was at a science library and there was talk of DLI been standardized for journal articles, did that continue?

Now [Laughter].

No, it is not been standardized, there are certain areas that tend to use the OIs more, but no, it is not a standard which is kind of one of the reasons why -- I mean this also explains, I mean GPO uses PURL's for identifying those in different organizations use different once and some actually still use permit links for their material but no, DUIs have not been meet the standard for really in the of the areas of -- I'm sorry, a persistent identification.

If I could just add to that, I was chatting along on that question and there are several agencies who do create the OIs and they tend to be agencies that are members of the data site organization and those tend to be more on the scientific site for example no, or oats FTI EPA and those kind of agencies but as Will noted, the OIs are not being used across the government and one of the issues that someone mentioned was that DOI are costly and they do have a DOI is run by a central organization which uses that cost to rent the DOI service. This has both pro and con because on the con, DOIs cost money. On the Pro, that money goes into central organization which assures the long-term maintenance of the DOI's. And so, you -- which side of the coin you are looking to, to note whether that is a pro or con?

I kind of want to follow-up, since we are talking about the OIs, we had a question that is a true the institution seek to pay to register DOI's for the materials? Yes, they do and you actually need to register a DOI you go through a registry organization and if you go to the DOI handbook, they actually have about 12 or 15 different organizations, that do the registering for those but yes, all of them are going to charge for that. And for the reasons that James had mentioned.

Yes.

To answer another question though, about that the OIs which kind of comes off of this, is also different vendors may identify the same object with different DOIs based upon the path to excess in the registry of those objects, they don't own but provide. I am not sure how the registry -- what duplication crosscheck and they do in the back end of that, but in theory if they are not crosschecking, [Inaudible static] potentially have that happen but I am not too -- personally too familiar with what chicken they are doing. There are multiple registry agencies so that could make it or increase the possibility of it, but I am not 100% sure what they do -- to keep from the duplicating part of it especially if it is two entities that are identified the same resource that is housed on the same server because it would be a unique resource. And so, I am not what their error checking is.

You pretty much addressed this but I just want to make sure there was nothing else you want to say about it, Tom said how prevalent art to OIs and government documents? To have a professor who asked about using to OIs of federal documents and I see USGS uses them a lot.

Yes, and we kind of answered that by saying there are some of the scientific agencies, do participate or to pay for DOIs for some of their content at least, not all of their content. One of the organizations are one of the examples with honesty, USGS, APA, some of the other

scientific once. I also want to note that one of the -- as far as I understand the OIs, the DOI itself can actually point to multiple copies of that content. So for example, if you look at or anybody knows FSR in social science research network, they used the OIs for their services and there are other journals that use the OIs for their services, and those can resolve to the different servers based on your location so you could download the same file, the same journal article or the same report from a server in California versus a server in Washington DC, versus the server in Europe somewhere, depending on how close you are so they try to limit the network traffic or make the traffic faster and so that is one of the positives of using DOIs as post to PURL which is going through a central PURL server at GPO.

If that makes sense?

[Laughter] and if my microphone did not go in and out and hope everybody heard me?

I think we skipped over a question from Lisa Prichard so let me read it out loud, besides clear descriptions and hyphens, are the other standard naming convention practices?

That is a fun thing [Laughter]. No, they are not. That is kind of -- so do standard practices would be dependent upon what your organization is and depending upon what you are doing. So for instance if you have a blog for instance you may want to create Permalinks that indicate certain dates for each of your posts. And that could be part of your naming convention for your Permalinks. But in terms of like a universal standard, there is really not much other than keeping it simple and easy to read, easy to memorize, easy to remember because it is kind of another good -- because they want to be more -- it is kind of a mix. Partially for humans to read, and so a lot of what we see are for human reading and then the computer speak in their own language. So most of the conventions are going to be things that the human can utilize. But there is something else that you can to on the back end of your website if you're using Permalinks is in the HTML you can actually tag your Permalinks to web crawlers can actually go through and identify those as well.

So that is something that would be a good practice behind the scenes. But in terms of that it is really going to come down to your local decisions for your content management and what does make certain organizations [Inaudible - static] different needs for their naming conventions.

As Will mentioned, one of the simplest naming conventions is to not put a space in your file titles, because a space to a Web server is actually a% 20, 20% or something like that, and so it makes it more difficult for a human to read that filename.

So make it simple and make it human readable and make it understandable and do not use lots of capitals and non-capitals and keep everything -- if you're an old UNIX person like me you want to keep everything lowercase which is why if anybody received an email from you you notice my name I never put uppercase J but just say James Jacob with all lowercase is because it is an old UNIX thing. But yes, that really doesn't have anything to do with the PURL system , but it is more of what you should do to name your files.

So you can find them later.

Not sure if you want me to talk over but I do see the next question that I do not think has been addressing want me to go ahead with that?

Yes, please, Ashley.

My checkboxes frozen at the moment so thank you.

Leslie [Laughter]?

From Lisa, different vendors may identify the same object with different DOIs based upon the past the access and their agency registries, of those objects that they don't own provide?

I mean, you could and I don't know why you would not necessarily -- would necessarily want to -- I take that back. Yes, you could do that. Because with the registry you are actually just registering the identifier and you would provide the metadata for accessing whatever it is that you are identifying, so it does not -- just because you get the identifier registered does not mean that you necessarily only item, it is just you are identified enough and that is kind of the best way to think of this so yes, you could do that and that would be something that would be helpful for instance if you have a project that was like a crowd source project, where I could see it in something like PG of documents if he had fugitive documents, you know where you want to ID them and you were not submitting them to GPO for cataloging and you just wanted to ID all of these so you would not have created them but you could go through and register a bunch of these items and still make them accessible and have -- have those identifiers created so you can do that, but once again I am not sure if there would be a lot of cases where you necessarily want to especially because you would have to pay for them for the registry as well.

I am not sure if that would ever happen actually and I have not seen that out in the wild because basically the system of DOI or any persistent identifier is that you control the object to which you are creating the permanent link and if you don't control that object, and I'm not talking about copyright, I am talking about if you don't have the server the file or the object is sitting on, then you cannot create the permanent or the persistent identifier, the DOI to that object and that is why do OIs I mostly used in the journal literature because many journal publishers are DOI registrant so they register the DOI because they have the files, that are under their control. So does not make much sense to create a DOI to point to an object to what you don't have the control of the file.

Does that make sense?

I think this is one of the issues with PURL said some people have been talking about is that you can create a PURL of permanent URL which GPO does but then if that PURL points out to a

live.gov website and the to permanent.access.GPO.gov, org of info.gov, then that permanent of the PURL is or could be called into question.

So that is one of the issues I think our working group will be working through.

And I'm looking at the chat log and I have one question from Jenny Groom but not sure what the context of it when she typed it in was so she may need to provide some clarification although maybe you all can fuse it out, from Jenny, what if we need the earlier version of a document and she is thinking of legal cases?

And this is where something like I mentioned permanent cc is a group that does permanent links, permanent URLs especially to the legal materials and somatic if you have got a permanent cc link to a document it is always going to go to that document and if there is yes, if that document is superseded then in theory that link -- there would be a new Permalinks to the new document, so the permanent cc link would always go to that or to the document to which it originally linked to.

If I --

[Indiscernible - overlapping speakers]

Now we are going down the rabbit hole a bit.

If I could add to that little bit so each one of these digital identifier should be unique to every virgin script to a brand-new identifier and the question is not going to be so much about the identifier itself but it is an issue that James mentioned earlier that if you are not hosting that material or that content, the concern would be -- is that they take it off of your server or was it backed up in another place because you could have the digital identifier but if it is removed from the data server, where it is stored, Hewitt lose access unless it has been backed up somewhere else and that would be really the concern more so than the identifier itself.

Yes, and not -- running [Inaudible static] some [Inaudible - static] GPO does this for several sometimes on their permanent access server they have a splash page an index page of all of the serial items themselves, all of the various issues and that PURL is a link to the title of that serial and when you get to that PURL you see all of the various issues of that serial and serials is a big, huge messy issue that lots of digital archives and libraries have try to deal with serials, and there isn't a good way to deal with them yet.

Just a reminder we have about one minute left and our session.

I think we have one more question that we can get to and it was kind of a unique question from Jenny, two different issues, catalog records and research guides, because Connecticut state libraries migrating to a new digital collection we are trying to link the catalog records rather

than directly to digital item so we do not have to update in both catalog and research guides and is anyone else done that and ate thought it was rather interesting.

We are not doing one or the other but we are utilizing both but -- yes, [Laughter], so we utilize the catalog orders and we link -- either took catalog record and sometimes we will link to use the identifier for bookends.

Jenny, I would add that when the PURL working group puts out I don't know how we will do it then the server you are what we're going to do to assess the challenges you all are facing in your library's with linking to digital content, would you please chime in with that to trying to kill multiple birds with one stone potentially and that is a really good tip though to take into consideration as we think through recommendations for the future.

Hallway, we are at 5:15 on the dot and I'm sorry we did not get to all of the questions and I apologize for that but feel free to reach out to the working group and the contact information is on the -- there's a handout on the [Inaudible static] library we can find the contact information. Other than that I think we are good so Kelly, did you want to wrap up real quick?

Sure, thank you to all of our presenters, wonderful question, for all attendees. Tomorrow at 1230 Eastern time, we start our virtual snacks with council sessions a please join us for that and thank you.

[Event concluded]