

Please stand by for real-time captions. >> Adapter knee and what come to the A Time Machine for Federal Information using Web Archive content government information reference for my name is Ashley and I'm an outreach librarian and elaborate services and content management unit here at GPO before I get started I will walk you through a few housekeeping reminders if you have any question or comment on the presentation please feel free to chat them a slope that in the bottom right-hand corner of your screen like Ali Kelly is going to keep track of all the questions that come in and at the end of the presentation I will try to respond to each of them. We are recording today's session and I will email the link to the recording to you and the slide to everyone who registered for this Webinar and I will also be sending a certificate of participation is the email that you used to register for today's weather if anyone needs additional certificates to get multiple people watch Webinar with you please email if the LP@reachaGPO.gov and include the title childhood Webinar of babes and dress of those needed certificates if you need to zoom in on a slide subjoin you you can click on the full screen button in the bottom left side of your screen to exit the full screen mode mouse over the blue bar top your screen so expand and click on the blue return but to get back at the default view.

Finally at the end of the session my colleague Kelly is going to be sharing a Webinar satisfaction survey with you they will let you know what the survey is available in the URL will appear the Chat Box you would very much appreciate your feedback after the session is through the.

A quick overview of what we will talk about in the Webinar on going to go give you a quick overview of what the appeal Pete Web Archive it will talk about when to use Web Archive content and your reference work and show you some examples and give you a quick demo body as it will talk about some the drawbacks of using Web Archive content and we will talk about high lobbies can take advantage of the web content that is being saved.

VFD LT Web Archive we get a lot of questions about what it is I I do know when Web Archive was even until fairly recently so I'm at the beginning of my learning curve. What I found is at a Web Archive is essentially captures website content and that includes things like the social media and file repositories you can find on it and more it captures that content in its native interface so as much as possible the Web Archive are is trying to bring observed the layout in the site navigation of the websites there I kiting and Web Archive content capture content once or multiple times depending on the purpose the appeal to Web Archive is an example of a Web Archive in our case it is the collection of archived websites that are hosted on archive it in the collections capture our catalog in the GPO's catalog if you decide government publications are the CGP via PLP Web Archive is set up to take point in time captures or snapshots of websites that we've identified point in time captures have the adapter working around Lake front which happens when agency change the layout of the website you can work around click fraud because in the Web Archive you can retrieve content from over captures of the website before the content was lost

and overtime the captured snapshots of the website can be compared to each other to see how changes were made to the agency's website.

That types of files insight capture via PLP Web Archive innertube multi media files like virtual interactive of Capitol Hill that's made by the architect of the Capitol in I have fun with that website if you want to look at it I recommend it. We also pulldown files in HTML HTML 5 PDF Java flash media database content MP4 and also image files and more. Within the appeal be well Web Archive you see content posted on Facebook Flickr Github contrast to be tumblr Twitter instant value to all sorts of accident more and we have definitely beyond static federal government documents and reports in the appeal we moved into a narrow web the content that it agency post on social media is within the scope of public information suitable for the F DLP and I give you a heads up that some of the files to be more difficult than others to capture and archive in this is not unique to GPO's Web Archive project all Web Archive projects take the same as difficulty of tools like Facebook keep up in the security or flag which requires the right version at the time of being captured which is constantly being updated the security of software version issues in the site navigation challenges can make it difficult to capture and archive the content so you may find a successful capture of an agency page and went capture MMX capture it may not have successfully been captured into security or the software on that side changes between the captures the

So how does DPO GPO capture federal websites I will quickly give you an overview of the tools that we use and a nutshell we use a tool called heretics and that's accurately is to do the actual quality of the webpages we actually managed those hair tricks in the cold archive it that matters captures in the storage of that Web Archive content the overtime the website is periodically recalled in in a content is stored in its index so you can do full text searches if you're interested in the nitty-gritty details of a product like this entails you can watch and archive Webinar that with them back in 2014 and the Webinar the fall LSM work was described one caveat the Webinar is several years old at this point in time and our workflow has evolved a bit since then but you can still get a pretty good sense of what the project looks like along with some of the more technical details the Webinar is entitled Web Archive in for the F PLP and you can see the link Webinar archive entry point here -- back LOC and staff have to decide what content to capture and what content to skip long story short we cannot capture everything that the federal government puts online there is just too much content out there we have to throw out some sort of parameters because there's just too much in our case all of the websites that we capture must be within scope of the F PLP Target content that is not be distributed in the print format and we Target content that is less likely to be catalogued by anyone to basically we are going after at risk material. We try to avoid duplication of effort with others to show 10 what content that's art available in tools like a very own FD says [Indiscernible] LS CM is working on collection of elements for the appeal Pete right now we are analyzing what is important to you all in the community and we work with our cataloging staff to get ideas of what's been requested for cataloging the type of the libraries and visit them on-site and we are actively looking at Google analytics for apt DLP

Web Archive content to see what users are actually looking at this is information guides us and decide what to capture the SLP Web Archive and I will give you examples a very easy targets for the F PLP Web Archive at the CTOC Y3 sites which are commissioned committees and independent agencies these websites are often not captured by other Web Archive initiatives and their content is often not distributed in print enough because user temporary commissions and typically choose to disseminate their information online nowadays feel capture on calls nonstandard government sites jointly manage sites like benefits.gov does seem to be a mercurial for other entities to work with because multiple agencies are contributing to the website's content so we can sometimes take that on the >> Be put in place for an opportunity for you to participate in the future of the DLP Web Archive please nominate websites that you want to be captured in the F PLP Web Archive there are three ways you can nominate content you can report it to document discovered through his DPL and also send an email to F PLP WebArchive@GPO.gov and while a mad. The quick reminder of what not to nominate please do not nominate content that is already NFC is [Indiscernible] nominate content that is already in somebody else's Web Archive that is duplicative effort -- back in order to avoid duplicating what others are already capturing here some of the other entities that we do collaborate with any Web Archive in community they include the national Library of medicine the Library of Congress national archives and records administration and the Smithsonian Institute those four on the top or participants in the federal Web Archive in Working Group below them is a list of some organizations in the general Web Archive in community that GPO sometimes work with in that sometimes includes archive it archive it is a too always to manage this process and it has a listserv in the community pages for other partners to submit questions and so on. We are also a member of the Society of American archivists that we participate in the Web Archive in Roundtable and we are members of the international Internet preservation Consortium.

Here's an answer to a question we get asked a lot about when we talk about the F DLP Web Archive GPO does not capture congressional committee websites in that is because that is already captured in their congressional web harvest and that is done at the November congressional term so every two years it is also captured in the term Web Archive and that's done at the end of every four years we'll be referring to these two projects in little while again.

The F DLP Web Archive is our pride and join today to have 132 collections and archive it we have over 9.2 TB of data and just for little bit of perspective the Library of Congress has over 290 TB that's in fiscal year 2016 in that fiscal year 2016 allotment included 25 TB just what elections alone so we are a little bit jealous and I digress in the F DLP Web Archive when we pull down an agency or commissions website we invariably end up pulling down a lot of leaks at the website itself leads to an arcade within our 132 collections we have over 70 million URLs to bibliographic recapture this content we have 152 catalog Reppert and catalog and government publications of the CGP with links that include the archive that link archived at length -- and talk about GPO makes web content discoverable by you all in

libraries and by the public we make the FDLP web tie a discoverable in two ways the first way is to the cattle of government publications like I does mention in the second way is through the archive it interface itself so go to more detail -- back in the CGP there are several ways to discover collections that are in the FDLP Web Archive you can do a basic search in the catalog of US government publications and see here that I did a search for a web art and typed in a keyword -- and type in advance or by typing in keywords and the like but down below be sure to use the drop-down box by catalogs and select the FDLP Web Archive -- >> If you want to browse the collection that are in the FDLP Web Archive there's a direct link to the FDLP Web Archive on the CGP homepage and from there you can click on especially to view all titles. >> Here on top we have an example of what a search result for the record in the FDLP of archive will retrieve on top you can see I added search in the FDLP archive for the keyword Holocaust and note that the subject number has the word Internet and in that how you can tell that the catalog record goes to OS it will be web car Web Archive collection below that in the catalog record we see two pearls that are hyperlinked the first Pro will take you to the calendar page for the captures that were taken for that collections websites homepage the calendar page is where you can select which Snapchat or dates you want to visit their webpage in the second prolixity collection page of the best wish takes you to the archive at page and a seed list is a list of websites that are captured for each collection so you'll will see the collection homepage see the face but see the tumblers seed and YouTube seed and so on and we went to the seedless and the CGP record because we heard from different users that they prefer different entry points and to the archive content we enable rapid

Just appointed out diamond catalog records me from the URL to the original website that was taken at the time the website was called or pulled into the archive so you'll see that URL@time of capture and URL -- and we have done recently a starter linkage whether federal agencies Web Archive's for example the national Institute of nursing research this content is captured in NIH Web Archive which also happens to be using archive it the important thing to remember when you are searching for Web Archive content in the CGP or in your own OPEC is at your searches searching for words found in the catalog record it is not searching the full text of the content that has been indexed and distort and archive it to Sirsi full text of the sites content you will have to search in archive it

Which leads us to the second way that GPO is making FDLP web Ted content accessible discoverable by you and the public as I said you can search directly the archive it interface itself on the homepage of archive and you just search for GPO FDLP in the collecting organization bikes to pull up the FDLP Web Archive in the collections captured in it again a collection is like a website and its associated sub if I could choose my plate got job they twittered this precise that are associated with a a few get to the FDLP Web Archive page you can use the search box to find the collection in the FDLP Web Archive when you click on into and archive site you are able to search within a websites text that is been indexed and you can also search for URLs if you happen to have them and I will join example of the short link to recap we regards to

discoverability you can do is certainly CGP to find the collection and you can do a search and archive it to find the collection and after finding the collection if you want to have the ability to Sirsi full text of the web content.

I have on the next two slides some ideas about how you can tap into the FDLP Web Archive content when you are on the reference desk is this is my brainstorm things that I wish I had known when I was used to sit on a reference desk where 10 years ago that sort of thing in no particular order you can use FDLP Web Archive content when you cannot find what you know used to be on an agency website so I have the same problem at home where my husband and I disagree about something so I set out to prove that I am right by going online only sometimes I cannot find what I know used to be there like give up no I know I am right so I do what any self-respecting librarian would do I plug my suspected URL into the way back machine and I track down what I know used to be there and now with my newfound knowledge of the FDLP Web Archive another web archiving projects that are going on, if what I'm looking for is federal information I check one or two of the Web Archive's that I know of and depending on what I bring I'm looking for I can usually find what I'm looking for peer I know I am right because my husband is wrong and I can usually prove it to

You can also use FDLP Web Archive content when the government shuts down or when he website goes down this is a tip for next September, if the federal government is not funded sometimes the agency websites are not considered essential and they can go dark we saw this before in the past big Web Archive content that is typically hosted content so it continues on even if the sponsoring agency is shut down. You can use Web Archive content when an agency or commission ceases to exist but just still funding references to his bodywork think about it if the Internet had been around during the Great Depression wouldn't we all love to have had a Web Archive snapshot of the Works progress administration or its successor agency the works project administration to I missed both works projects and Works progress administration yes some of what the publications of the WPA have been digitized so we have access to those various and various online corrections but try to imagine what the WPA would've put on their website if they had had one back then the agency was constantly studied had time to Congress and focused on the successes and the work they did their website would have been the publications like post images of the artwork that was commissioned and information about the construction projects going on think of the rich content that everyone would have had access to if they had had a website back then think what would've been captured for everyone not just for the depository library's run then and rapid tempo popular material know everyone has access to Web Archive content whether you profile for it or not I'm not saying that the FDLP Web Archive don't have WPA collections and it is not for obvious reasons the WPA didn't have a website but the point I'm trying to make is that you will come across references to dead agencies or commissions in your work and if the commission or agency of new enough you may be able to track down oaths of office never before printed material in a Web Archive. You can also use Web Archive content when a researcher is studying a point in time you can see what's been captured

across-the-board or in this case across the federal government at a particular point in time to say you want to look at the content found posted in an election year and since coincidentally it's a new form of research that macro examination of a body of content from various entities as opposed to the traditional Michael examination of individual works of an individual entity so people are taking a broad look at a large swath of content and you can also use Web Archive content when a researcher finds a reference to something likely to be on a website but only for a short period of time so I will be showing you an example in a minute for content related to the Olympics.

Some more scenarios here where you might use Web Archive content you can use Web Archive content when a researcher wants to find all iterations of something or track the evolution of something over time and again I will show you an example of this shortly. You can use Web Archive content if a researcher is trying to gauge how would agency collaborative with the public or engaged in e-services so if you want to know what agencies Excel average in the public check out their social media presence and see upper back in time that disability reaches you can use Web Archive content one summary is researching a recurring federal topic federal information topics are constantly appearing and reappearing under different guises and it is interesting to compare the variations for example the electoral college is search like that will be a lot of log entries on federal agency websites on the highly anticipated 20 search like that will be a lot of log entries on federal agency websites on the highly anticipated 2012 election and as more websites are scheduled for the recapture we will start to see content in the Web Archive's related to the 2016 election.

You can use Web Archive content when searching for older titles older tutorials and older guides for example to search on Google right now for a Federal Register tutorial you will discover some outstanding short YouTube videos that explain the regulatory process and the existing Federal Register in the Code of Federal Regulations publications they are archived in the Web Archive so if we ever get off the existing Federal Register and CFR system how we use the Federal Register in the CFR resources are going to be preserved for later research. Take it one step further though what if you're looking for a quick tutorial for Section 508 guidance circa 2015 you are out of luck on Google. Google is only gradually due to current guidance which makes sense from the agency's perspective that they would only lead to current guidance but if you want to find federal agency guidance on best practices for making PDFs or power points and websites 508 accessible for over platforms you may be better off finding that information in a Web Archive that tutorial may be out of date for practical purposes and so it would not be available on a normal Google search but it may be just what you are looking for in an archive. You can use Web Archive content when the content was not picked up in the FDLP and other were set content is fugitive and I have an example of this towards the end of the program and you can also use Web Archive content when the current temperature you're looking for is likely to be overwhelmed by naughty facial or secondary resources any normal Internet search and I have an example of this as well --

Here we have an example of something how many people recognize this -- choose my plate.gov -- how many people remember it is processor the my pyramid -- how many people remember this the big thing -- how many people remember this -- this is actually from World War II here is an image of the USDA's nutrition guide that was trying to ensure that people had access to and knowledgeable was perceived to be a nutritious diet on wartime rations we looked at four iterations of the same publication this is a quick question for you -- how many of you all have had your daily recommended dose of butter in fortified margarine -- but you will be need to get more in my point here though is obviously over time the recommended foods and quantities that the USDA has recommended have changed drastically. From the researcher's perspective some researchers are going to pause that the food industry lobby is a you a day to change your recommendations suitable for their particular food industry other researchers on the other hand they will posit the opposite they will say that the growing field of research backs up the changes that have been made to the daily food intake recommendation in order to see the changes over time you need access to the various iterations of the federal government detention got what are your library never received the paper version what if you were rated the older version because she ran a poster space what if you deem the old version superseded and you ditched it we will go to the FDLP Web Archive and see how one contract on older versions of the title.

There was there with me when I sure my desktop -- here we have the CGP this is the catalog of the US government publications I am going to do a Web Archive CERT by clicking on here FDLP Web Archive and I am going to search for two -- choose my plate -- and and see the biblical graphic record for it enter that record -- in here we see choose my plate.gov have been archived in the FDLP Web Archive and there's two links the two pearls that take you out to it and I will actually click delete go to the archive interface and you can see how to find his collection in archive it as well.

Here is archive it it is archive
it.org and I will limit first to the FDLP and type FDLP --

Here is all of our content. When you are doing a search in archive it, here's the search box and whatever is in this greenish gray box is what you're searching so if you search down here right now we're searching for the collections -- so let me search for choose my plate -- search -- NACA Flickr group and I see a choose my plate so in this interface I can see that it was captured 48 times between August date 2012 into ninth 2017 I'm just one click on the number of times it was captured and this is the calendar page so how do I click on that first throw and the CDP record this is where I would've ended up so you can see all these various times captured it broken up by year I will click on the oldest calendar page the first capture we did in 2012.

English here we go -- this is a quick example of what captures you can see here that the website looks like what it did back in 2012 non-not everything in here's going to work like it did back then but GPL staff do try to customize and configure later call for the site's leg as much content as possible and to retain as much of the functionality as

possible with that the original site had an so often things I found that don't work necessarily are things like the search boxes so that is a heads up for you to let me go under here for just one example printing materials and/or learn I can click on graphic resources, and here I can see some of these images these are essentially what we are distributed to the FDLP those images and posters and stuff when I'm trying to illustrate is that over time as the website has multiple snapshots taken and captured this is going to help researchers tracked down changes made to the USDA's nutrition guide along with old other supplementary information that accompanies a chart itself such as narrative tutorials lesson plans and as we get into more recent years you're starting to see things like app sediment captured and go back to my slides --

The question is can we identify all the various iterations of the dietary guidelines going back to World War II and the answer is no not at this time but over time this content is going to accumulate in the Web Archive and it will deceptive very versions that have been produced and as a bonus you may have profiled to receive this content and paper through the FDLP in the F kept it in paper you may have also weeded it after five years or maybe you superseded it and only kept the most recent version but anyway you look at it what the profile to receive his content or not by virtue of it being in the FDLP Web Archive you have permanent access to every version that is captured --

I will go into another example of how Web Archive content may find its way into your library. We capture the US Holocaust Memorial Museum's website with very few publications in print that gets distributed through the FDLP unfortunately we started archiving the in the FDLP Web Archive backend 2012 and that includes capturing the Museum special exhibits the newsfeeds and event information and more. So let's imagine that you have a researcher who is reading an older news article about a traveling special exhibit on Nazis in the Olympics traveling exhibit has an online component to it as well, the light traveling exhibit in the news articles longer works -- so how do you track it down? Step one try a Google search for the URL or keywords in the news article for the purpose of this let's assume that Google search does not work, step two, as aced savvy researcher you probably know that you can copy the URL from that news article into the way back machine and see if that content was pulled down by somebody on the Internet archive maybe it was maybe it was and I don't really know in this case. Step three if that doesn't work as a backup you can investigate the capture of the website itself has been captured in a Web Archive and I can tell you right now it is not so you put on your God dots reference him the online and terrible in exhibit relates to the 1936 Olympics you can envision that the exhibit was likely only be on the Holocaust Museum's website in the summer of 20 16th which coincides with the 2016 Summer Olympics which is also when the news article would likely have been written so you know you need to look at the Museum's website for the summer of 2016 there will likely not be available in the 2017 summer if the traveling exhibit is over and there's a chance it wasn't available on the website in 2015 so you really know that you need to look at the website at that particular point in time to

Let me go back and share my desktop --

In flip back over to the FDLP Web Archive actually I will do it in archive it and let me go back to the archive page and I am on archive it here's the greenback's green gray box tells you I'm looking in the FDLP Web Archive I will type Holocaust Museum and helps if you can spell it correctly --

Here's the Holocaust Museum entry point, and I will go ahead -- we can see here this is the seedless for the particular collection and you can tell here it says it's US Holocaust Memorial Museum with you if I search a here searching the full text of the content that was pulled down -- you can see the blocks for captured the collections page was captured and dilutions list rebounds and media site news memory project donate Facebook newspaper Pinterest story by Twitter and more Facebook and Flickr is a gram and here is the homepage it was captured 15 times I click on that 15 times to get to that calendar page, I know I'm looking for something and the summer of 2016 so I click on the June 28 capture -- in 2016 this is what the website looks like and I scroll down on liking for a traveling exhibition and looking at Museum exhibits and over here I'm looking for a traveling exhibition -- have to use the site navigation they had at that time -- in the summer of 2016, this was still available the Nazis in the Olympics when I click into this traveling exhibition -- here we have it you can see here an image of the actual traveling part the tangible put the traveling around the country and here is the online exists -- exhibition >>

We got explore the online some videos we can view, can you find these videos online right now and yes actually can you can find them on YouTube they are available on a couple different streaming services but can you find these videos pulled together into an online exhibit along with the transcription of that video -- as well as a little bit of context about that video -- you probably cannot see all of this information pulled together unless the agency's website still happens to have the content available so that is something that the Web Archive is really really well is it preserves the content and still retains all of the files that are associated with the in connection with each other so here's another tip for you especially if you find yourself working with a lot of videophiles if you cannot get a video to Shireen stream or download if you're in a Web Archive try another in the net browser because some of these embedded videos require flash or other versions of viewing software I found that there are times where I can get things to work in one Internet browser but cannot get them to work in another so it's a little tip. >> Let's take a step back and look at this again I've just showed you two examples of searching for Web Archive content you may have noticed that both methods required me to be familiar with the setup I was a Marquette like record or the archive at site itself but what about your users are they going to have the skill set that they need to navigate the information that they are looking for us take a closer look at how typical user might try to find a Web Archive collection in the catalog or in an OPEC record web two options what the catalog record if you're in the CGP you can browse the FDLP web archive content as I showed you the authors typically the websites agency and so you can do it officered and you can also do subject that is to find archive content as well so for example I've listed here the

subject headings for the United States Holocaust Memorial Museum that we just went over and you can see the subject headings listed.

Those are excellent subject headings for the website but they do not shed much light on the breath of the content that you're going to find in the Web Archive itself fortunately for this research example where we were just try to track down a traveling exhibit with an online component to it if I'd done an OPEC search for Holocaust exhibitions it would have retrieved the click on log record would put it directly to the Web Archive collection because those keywords were captured in a marked 520 summary note steadily though if I had searched on Holocaust exhibitions cellular or Holocaust exhibit it will not have revealed the correct catalog record so the user will have to know how to use wildcard characters in their search and searching for any digital or online content in the catalog it is difficult due to the limits of tibial graphically describing and extensive and varied online resource there is just no getting around it. Web Archive content faces this difficulty this challenge so you can capture the essence of the website and subject headings in a few notes field you cannot possibly list all of the possible keywords that a person might use in an attempt to discover the content that was found in the Web Archive websites are just too diverse for that. >> The other access point that users have discovered Web Archive content is the archivist interface itself was user experience as we just saw is not the most elegant of interfaces to use archive it is a difficult for researchers or for libraries to use for reference purposes as one might expect the ventures such as this archive is really focused on the infrastructure to support the content which makes sense but on the high note archive it is very responsive to feedback from its content partners in this leads me to my next point which is an interesting work over onto the user experience problem I would like to show you an example of a Web Archive collection that has a user interface built into it to ensure that Web Archive content is discoverable by users and they do this by building their own user interface. >> Hears that example here we see that an institution is printed a separate portal or user interface to their archive content this example is a Columbia University human rights Web Archive the locally hosted in locally created portal is on the left side and BE Standard archive it interface is on the right you can see that their archive and interface looks very similar to ours. Each locally created portal or site incorporates different features like the full text searching or not very filters to use to narrow down your search results and not subject headings and this case I have highlighted on the screen the people can browse the human rights content by originating website by URL by subject places and languages those value-added search and browse features are added by the collection creative in this case Columbia University not by archive and I recommend that you take a look at the port to see the functionality of them and I'm showing you this because I want to highlight that there is potential here to further develop the FDLP Web Archive in a way that you and your research can more easily tap into forgetting their.

And here is one reference to and I highly recommend that you explore other Web Archive projects Web Archive projects are currently siloed you can see just four of many of them listed here and there is no

better search to search across them at all at least not yet and remember that Web Archive it is a new science and the direction that it is going in is a work in progress there are challenges as a Web Archive in community is still exploring new models that new technologies and practices in the meantime just gaining temerity with them can open up new horizons for you and your reference work for example if you are looking for web content on those summer's Olympics that content has been captured by the eye IPC or the international Internet preservation Consortium members can find by searching within archive it that is the one there on the left, you won't necessarily find that content in the other Web Archive's. A little more related to federal information you can do archive it in the hundreds of collections that are posted there while that itself is a silo you can search across the various Web Archive collections hosted by it the Library of Congress has a whole slew of Web Archive projects on their digital collections made some of which are becoming captured meaning it is archive more than one some of which are captured only once. The end the term web archive captures a wide swath across the federal government every four years and the Congressional web harvest done by NARIC captures congressional websites every two years and we will talk more about the latter two next to be aware that there is some overlap in the content captured amongst these various groups and but again the content is not accessible in any one meta-search so you have the them or at least the aware of them in order to make use of them and one other thing to be aware of the FDLP Web Archive has a high level quality assurance or QA work done on it more so than many of the other sites this means that your odds of being able to playback videos click on website navigation menus open up a social media content and otherwise move around on a website as if it were real and not a capture and to do that successfully are greater here than on other way of Web Archive size do not perform Q&A on their captures no I'm not saying that our captures are perfect to work every time capturing ever-changing websites their social media sites and other entities with their continuous menu changes in their security enhancements that's a challenge for any Web Archive collar tackle I'm just saying that we are very proud of our QA here at GPO to

Again here's one excellent web archive silo you should explore the and determine Web Archive goal is to capture agency website legislative executive judicial and capture congressional member websites including their social media and you've got a do it before the new administration comes in the national archive and record administration or NARA did a back and 2004 and in every captures and sent the partners have very these every four years captures have become very valuable information about the federal government around those time periods I highly recommend taking a look at the site because reference leverage should remember there's numerous snapshots and federal governments website set up and take up here than you might think content you might think life has been captured in one of these projects dashed back here's another excellent site opened to explore that's United States Congressional Web Archive is produced by the library of congress their goal is to capture member websites the House of Representatives and Senate to Ginny with the end of the 107th Congress and they also capture community websites on our web monthly basis the lab library, uses polling technology that hair? I mentioned earlier but it stores are content instead of using it

on archive it be stored in a baguette fire and reply that content through a local installation of the way back machine this means that they are user interface is very different from the archive it interface it is good to explore to get familiar with them all put in a quick plug library congress is public policy Web Archive which has quite a bit of federal information I hope my brief overview some of the silos that exist help you out when you are on the reference desk or when you're trying to pull together course guides on your subjects and things like that in my case when you are in an argument with your husband and you know you are right, here's an example of how you can tap tap into those siloed Web Archive projects if you are interested in national security the Leica clock classified information WikiLeaks and Edward's noted in general you would want to find this unclassified version of the report put out by the House permanent select committee on intelligence this September 2016 report disclosed to the public the extent of damage done by Snowden and a review of the bridge what can publicly be disclosed about the man who committed the lake and where national security needs improvement. This report was never ordered to be printed by the committee and as such was never put in the serial set of the committee print and because of that it was never voted on empties with governmental furthermore the committee's website is up and catalogued in the CGP this report were at one point in time of fugitive from the FDLP although it as has been reported as a fugitive and been catalogued since then it is on the committee's website at present while it is likely to be available for the foreseeable future like all web content it is not guaranteed to be there permanently on that committee's website however both the indeterminate Web Archive in the Congressional Web Archive have caught this website the past and hopefully it was captured again as part of the 2016 and we won't know for sure for at least the Congressional Web Archive because he library congress has a one-year embargo period and that won't be over until September and they have done to their production in the cataloging work and all that sort of thing and we can only hope for current and future researchers at the same and its reports are captured again and remain discoverable long after this report is old news as far as the Internet is concerned.

Was change gears a bit I mentioned earlier that when you're feeling a reference question you can use Google in if that doesn't you then use the way back machine and so however that said even if the topic that your research and is still fairly recent like within the last for five years there are times when Web Archive content proves to be a reference short cut dealing -- I set a Web Archive search can be a reference shortcut even for fairly recent advanced and major events that generated a lot of controversy to explain that though I have to try fails to why Google has limitations and why it isn't always the Blessed placed to find slightly over government information so let's compare what can be found in Google version is a Web Archive search or example is on September 11, 2012 for Americans including two diplomats were killed in Benghazi Libya the question is what was the State Department's official response -- we Google Department of State response diplomat death Benghazi site:.gov and what I add that site:.gov because if you don't you will get overloaded with a little over 4.6 million unofficial news accounts conspiracy theory sites and more you pretty much have to

limit it to official websites in order to find anything that's even remotely official and I did a quick shortcut by limiting all my search results to those that came from aGovernor URL domain -- and limiting a resort result to the.gov domain limits results to 3580 official government channels which you can see is snapshot of your I can tell you now that even playing around with your keywords you're not going to yield search results that are much better or much different from this take a look though what you see are several archived press releases and numerous after incident investigation reports most of the search results are static PDF and I ask you was that the extent of the State Department response to this tragedy and the answer is no pure no pure expect check results from term Web Archive the Web Archive then collaboratively by several partners including the Library of Congress the Internet archive University of North Texas US government publishing office of the California Digital Library I can do a simple search in it for the Department of State and limit my result to those from 2012 captures from the counter paid what you see here, I can further refine my time period by clicking on the 2012 and seen all the captures from 2012 on the calendar page and I select the September 15 capture -- bingo here's the homepage on September 15 bingo here's the homepage on September 15, 2012 this is something you will not find on Google you will not find the State Department homepage taken on September 15, 2012 which was four days after the tragic event why can't you find the Web Archive content in a Google search simple Web Archive content is usually in deliberately blocked from search engines indexing robots think about how confusing it would be if you're doing a normal Google search and archived content enough in the search results you would find that very annoying that when we are looking at this archived content right off the bat we can tell from the State Department's homepage that the agency scrambled to pull together press releases images and more on this major event unfortunately though you can also tell that the center of the page had something that was not successfully pull down in the Web Archive looking at it in Internet Explorer I cannot even tell what is missing when I look at it in Chrome I can see there was a flow played video those richly there remember to play around with your Internet browsers as everyone has different versions of software loaded for different browsers and you may have better luck using one browser over another one.

More indeterminate web results State Department blog death note and no I did not make that name up a really good to call it note it's a blog post that honors and recognize that the loss of lives in the guys in Benghazi or be honest with you GPL has catalogued in Perl to the dip note blog so you can technically scope of the blog by title in the CGP or rollback you can click on the Perl click on Regent curled nontraffic at an Libya and you can guesstimate were in the page navigation webpage number something from September 2012 is likely to be in and have fun sifting through the all of them so can you find the correct blog on the live Internet into a catalog search you can do this find this blog titled in web, you can find that it is active and online but there is nothing in the catalog to direct you to this particular blog post and if you are online would you navigate to all the steps I just told you about in order to see if something was produced on the Benghazi incident in this particular blog I tell you now that there is no search

bar on the current blog page to help you find relevant content facts so this is an example where the Web Archive yields relevant information fast because it was taken at a time of great importance.

Back to our in the terminal archive search because we limited our Google search to site:.gov we missed the State Department's tumbler page and that tumbler page link to so several YouTube videos on the State Department state video channel so now here we can see the Web Archive that a video on our topic exist but as you noticed the video was not successfully captured knowing that it exists though we can redo a Google search to see if the video is still available and online since it was not successfully captured in 2012 we know that the agencies State Department we know the keyword is Benghazi which also we limited time period most important for us we know the channel or the medium that the content was put out on and this case the State Department official YouTube channel is called state video and must be honest at the very outset you could've done a search in Google for State Department YouTube and sifted through the thousands of videos that are available you could have stumbled on the official State Department YouTube channel and you could have sifted through these thousands of videos on that official channel because you can't filter within a YouTube channel by date or subject although you could do a keyword search within a specific YouTube channel if you can find that elusive search box you could also did a more refined search by Googling YouTube a State Department quotes in Benghazi guess what you will get over 1 million YouTube search results and none of the actual YouTube hits at the top of the search results are going to be the correct video there are a lot of news videos in conspiracy theories on video and they dominate the search results so I ask you how would you have known that a video even existed on an official State Department YouTube channel if you didn't already know that one existed ignorable search my to fail to look at this video now knowing what you know from the Web Archive search because you found that there's something that was on this topic, you can simply go to YouTube and search for Benghazi state video and it will return the missing video fairly high up in search results because it is still out there and on the Internet on the State Department's state video YouTube channel what you've done here is combined a site-specific search with the official government channel and the keyword you zeroed in on it quickly because you add more information on hand so let's recap by searching in the terminal web archive by virtue of taking a snapshot of the website a couple days after the event you can easily stumble across relevant content quickly and even if some parts of the content were not successfully captured in the Web Archive you still have information on at your fingertips that you can use to track it down on the live Internet so I hope that I've been able to illustrate how we straight Google search can sometimes be difficult to work with even if the content is not that old and even if the content is still accessible on existing websites remember that the older the web content gets more likely it is to be buried in such a way that you cannot easily find it if you usefully need more information and order to Target the content they are looking for by searching and Web Archive at a point in time when the content was more prominent or when it was fresher you can quickly zoom into major events like what the State Department's overall response was to the tragic event of sup

September 11, 2012 -- back I want you to think about this or strategy logic that we just went through this is not a new strategy for us because we are used to thinking about facets of government information were used to thinking about was the agency was a time. Tina was a likely medium or channel or the resourcing where likely to find something in wears an information resource likely to be found and so on this is a standard reference approach to this topic per one the information is clearly online and hopefully in Web Archive we won't go out of business anytime soon because government information experts are still needed because we can apply our expertise to track down a list of content that is buried and often overshadowed by unofficial or secondary resources in theory we have job security there's just too much stuff out there that people need help sifting through. >> That wraps up our Webinar here. I hope I have been able to shed some light on how the FDLP Web Archive and how other federal Web Archive projects are essentially a time machine that can take you back to the particular point in time and I hope that you have a good or a better sense at least of how these tie machines add up together to help you in your library. It is truly an interesting collaborative project to be involved in building tie machines that is do I have any questions?

Of view what schedule drive the web capture and nutrition my plate example earlier calls were fairly infrequent later on it was every couple of weeks. >>

I've Julie with our Web Archive Web Archive go with what drives it captures and especially different seesawed the time you did see was that we were able to get more resources about two or three years ago and before there were a lot more infrequent you'll see our Web Archive there's a lot more infrequent crawling earlier today because really didn't have the staff resources for dedicated crawling and for keeping up the green calling we were developing the collection so once we got in some contractual help we were able to concentrate the recurring crawl and making sure that we were updating existing collection that we had so and we do regular frequency of calls and we have been pretty good at it the last three years

Archive hair tricks free open-source?

Hair tricks is Internet archive privacy tool and it is open source and you can get that through like source splurges sites like that and I want to mention there's an open way back that is in development for a while and you can get more information on that through the eye IPC website and in on the mailing list for that and archive it is Internet archive description site and it's sort of an out-of-the-box service and it also comes with support but also another advantage of using archive it that we like is it uses other was other than just hair tricks they have a tool called Ambra also they are working on another tool right now that we've been actually assisting and testing called Rosler so we get to use other tools that they have an addition to the hair tricks to which the archive subscription.

What about mental that's a way to search across all Web Archive

Momento is a great tool and it is one that I played with four people who are unaware that you can download from extension to use Momento and choose a date in your browser to search so it is cool but one thing I will say I have had it on before without realizing it from time to time and then when my not getting search results or something I realize I'm searching 1999 with a Chrome extension was on but it's a great tool as well to use for searching and doing research absolutely.

I believe this was in regard to the Benghazi video example with this page have had the original YouTube URL? For the video?

I would say yes. And if you look at the URL, there is the Internet.org URL at the top but then you will also see the original URL is within the URL way back URL as well.

No other questions at this time we will give folks a couple minutes but I placed a link to the Webinar satisfaction survey in the Chat Box so please take a few moments and complete that for us and any less questions?

Any last questions? >> All right we are not seeing more questions so we will go ahead and sign off and thank you everyone for your participation we appreciate it.

[Event Concluded]