

Partnership Opportunities with Internet Archive - Transcript

Please stand by for realtime captions.

Hello everyone. Welcome to our session partnership opportunities with the Internet archive. Just a few housekeeping reminders before we get started. First of all, please check the checkbox for any questions or comments or issues that you have. We will be keeping track of all that and answering any questions at the end of the presentation. Second, know that we are recording the conference, and everyone who registers will receive the sites to the session after the event. Please join me in welcoming our presenters Karl-Rainer Blumenthal and --

Hello, I am Karl-Rainer Blumenthal, if you have any questions just adjusted in the chat. First I want to introduce my colleague Andrea who is on the line as well, Andrea, can you say hello?

We are, hello. [Indiscernible].

Sorry to put you on the spot. Andrea, we can go back to the second half of the program to talk about digital books, digital text, basically, digital bound publications. But before we do that, that would start us off with a general interview an introduction to partnership opportunities with the Internet archive that are more specific to web archiving. Because that something we are generally known for, you may have heard of us before today. So just so we are absolutely clear on terms of reference in meeting, when I say web archiving, what I'm talking about is this. His process of collecting, preserving and ultimately if they are successful, enabling long-term access to things that are native to the web. To this could include publications that are shared via an agency or organization's website, but they can just as quickly beaters websites themselves. But long or social media feeds, there are only as many territorial reasons for doing so as there are web archivists out there in the world, but maybe not as immediate and universal to all of us in an recognizable as that first preservation imperative because among other things, we find that the web is naturally kind of a simple, it is an at risk format as soon as it is being used. Our own data science at the Internet archives have for instance looked across the [Indiscernible] what we have collected and found that their typical webpage will last fewer than 100 days before it comes down entirely off of the web and you are left with one of those wonderful 404 file not found in good luck with your research. Or the content of the page in question has changed so much that it's not really the resource that it used to be. It's not the information resource that was intended to the other end of the hyperlink.

Our colleagues at UCLA actually studied this a couple years ago and found that if you just look for instance at the links contained in the citations in stem journal articles you will find that seven out of 10 of them suffer from what we call in this way reference frauds, that link is either dead or takes you to Somerset that doesn't fit the same purpose of the original citation. So what we do about this? Well librarians and architects come archivists, do what we do best, we have them ourselves in control them and provide kind of an easy point of access where we can.

And it all starts and ends with people. A lot of this process is the near of automation and that does tend to obscure sometimes how much human duration and judgment is going on, I would like to point that out as well.

But speaking kind of a 10,000 foot view looking down a web archiving, you can kind of envision it's about this. A person to person at a computer, that leaves a piece of software called a web crawler comment about callers something that you may have made use of as early as this morning without knowing it because he said the same things as search engines, to travel all around the web and index what is founder so you can turn back to the user in the term search results.

Web archivists use web crawlers as well and to do additional things, when we reach website or group of them have a clear calmly pull the content down and put it in standard archival file format called web archive work file, and put it into our own repository sure.

So if I wanted to for instance collect a copy of the FDLP safest conference I could point a web crawler at it, pull all that content down into a work file, put in storage, so later I can about the file again, and using some special software, just like you would use Adobe reader or some TV a file or Microsoft Word or doc file, use something in the family what we call way back software, to open of these work files and interact with web publications as the original time that they were collected. So this includes all the same information, but as well interactivity, so it has a look and feel of everything you would have expected from a fully functional web [Indiscernible] not just a status or static screenshot.

We have been doing this for about, but as of now, almost 25 years. And in that time, several different unique technologies have been developed to kind of account for one of the parts of this schematic. You have the Internet of, the Internet archives for instance, we have the roster and heretics tools, if you have older collections in the repository or come across them you might find files held in ARC as well as WARC, it's predecessor, and the replay tool for something that's been collected in the web browser.

And already he can see from the amount of options that are starting to build that, this is a lot. He gets to be really well are way too much for anyone human curator to handle they have a rich life and they are doing other things with the library as well.

So thankfully as 10 to 15 years, some services have been developed a penalty things together to put them in a software stack that anyone can use for more of a pushbutton type interface. While the rest happens for the behind-the-scenes. And a lot of these were developed for and I national libraries. Places like the UK and Australia and across Europe. Since we don't have the same kind of legal deposit mandates that they do, this hasn't happened in parallel kind of in the United States, but what we do have is the Internet archive. And the service that we developed to kind of fill that gap and now serve and if you are nurse outside of the national library.

If you have heard of us before, you might be interested to know that the Internet archives is an actual factual will place, in addition to being a website, archive.org, in fact in other times you'd be able to visit us in person, at our headquarters that you see in the San Francisco district, we can even visit the facts. What you see there are servers in place of where all the been printed materials might be, service that holds the books, movies, games are increasingly popular on the platform. But more often than not, I think we're still probably best known in first for the way back machine, and of the well-known uses of the way back [Indiscernible] it's the most largely used web archiving today.

Is close to 500 billion unique web captures in at corpus and a 100 million websites among them. And adding to that at a really good clip. Billion captures of URLs are added per week and moving forward with it. It's a big collection, and you can find a lot of interesting things and it's going back in time, for instance if I wanted to be the GPO website in 1998 or 1999, jump on the Clinton impeachment, I can do that.

However given that we have kind of a portfolio that is the with an breadth of the web, it's a large mandate and we try to capture as much of it as possible, which has its limitation.

Because we do try to program and configure our web crawlers of much of bread, a broad content as possible in the time we get it, that can mean that there isn't the same kind of disability and completeness and cohesion and as a library of cohesiveness you might expect. Because I have the homepage for the GPO in 1989, there's no assurance I have two or three pages on the tree structure into the website. It's no guarantee I'll have that website archived on a predictable page like if I need to find a master

Date and time, that's not possible and ultimately the corporate is large enough now that we can to a fully muffled search and prescriptive for browsing. Basically have to we're going to get there. You have to have a URL in mind and plug it into fine regain two.

With a specific is a concept it's useful as a nexus point for people like journalists, teachers and students, anyone who needs to find primary source material for the web, but it has kind of a natural delimitation where you start thinking about 2005, 2006 we found a critical mass especially University, government, that is to say state and local as well as federal government, NGO type library and archives if you don't have sex what we want and you don't have exact timing that you want for the completeness that we want, but we work with them to develop the dashboard of all the same tools that we use so they can run the levers themselves and create their smaller curated versions of the weight have just what they want inside. In a format that can be portable, that we can move these work files among different repositories and reflections and index everything for description for the search and if you go to the website, they are archived what you will find there is thousands of thousands of curated collections on specific topics and covering specific domains of the web. And that includes the FDLP's top review. There's one operation for instance based out of GPO that collects government agency websites and social media on behalf of of the Trinity for instance, that claimant.gov collection. I think it's screenshot it here like it has been for several years.

That's a different example there. And the kind of platform got off the ground, just up into the this is the upper right charts we like to see these days, closing in like where the 800 partners by the end of the year. And maybe as many as have this in any given time a university-based because of the broad purview of University libraries to cover their institutional records at flow through special collecting areas and anyway their faculty or curiously them and the other half again to make sure federal, state and local governments, nonprofit organizations are for profit organization and so forth.

So dealing now with the corporate the kind of large curated over what 14 years now, it kind of prove that we can collect and preserve things and provide more direct discovery layers and access points through things like metadata and search and the can get to the individual discrete publication, for instance you might be looking for from the web. However at that scale, we're talking about two petabytes now of data that is added to the archive collection and starting to do some interesting things as well through text mining, surface reading, data mining and I confrontational opportunities in materials and the scale and kind of technology between some of our partners developed in parallel. Just as an example here, a archive collection made for the 2016 Chicago architecture biannual. And what I did is I sent it through which processing tool and of Stanford for the names the entities in branch, which is people, places and institutions and file for instance by pulling up the names of the most mentioned architects and designers that allow journalists or really pointing towards any basis in the young designers taken ambitious projects that might also want to brush up on the modernism with the context of what people are talking about.

And such as about the text, because they are dealing with foreign visual materials here, we get all sorts of the things, for instance a recent hack up on with the DC public library, and the collections department. They have sure to us that the collection may have curated the website related to the local [Indiscernible] and we were able to have a way to pull up all the image files from all the websites and webpages in that corpus, so all the flyers, and tape and inserts and put them in a possible kind of zoom and browse clickable format.

Just a different way to kind of envision accessing something that is web-based.

For this scale up we have done some fun things with this kind of concept in the Internet archives, for instance. If you are a digital homesteader from the nineties like I was, you might be familiar with these GeoCities platforms that not everyone a chance to design their own website, and they did things like use a lot of web one point oh aesthetics and created a search engine for those so you can look up a little piece of web history in that way, it's not just for fun, it's actually can a practical useful at the same time I designers have found great use for this work. The New York Times has used it a couple times to illustrate some of the more fun illustrated articles that they have done.

And this is equally applicable to things like .gov and in this case that .mil domains that we have played with. You would be surprised for instance, just quite how many PowerPoint finals you can surface out of the corporate .mil website elected in the way back machine.

So again kind of a silly fun examples, they decided they can like where the next stages of the evolution of this product are, is what you can do with what's in the collection. What would your stakeholders or your types of publications do they had this kind of skilled access to the -- what's already in the corporate or if you have special knowledge, what should be a part of it.

Because the ways going forward for me is making all of this [Indiscernible] actually open access is good. Just a reminder that everything that web archiving partners, and the platform or other direct services to provide Internet archive, that they desire, they can make all the data available openly on the web via documented API side, that includes the actual data that they collected itself in raw files as well as records of where to find everything in the way back machine interfacing. Kennebec in the book indexing things. As well as the full text search index, all the descriptive metadata that they have added for discovery, this cannot be kind of ported and redesigned and remixed, however you feel best serves your stakeholders. In addition to feeling dashboard that all of the archive the partners use to build the collection for the backend.

So just as an example, you see another curated collection on the left from our website. Every archive partner can spit up an access point like this sort of out-of-the-box. All the designers take care of. You just have to collect and we will plug in the parts there for you. But since this is all again openly assessable and API, some partners have decided to put their own portals on the website or include them in union catalogs or federated search, at the kinds of discovery layers where they can afford in the same search index, the metadata even provides direct access points to that original resource in the archive form. So on the right, you have kind of a remixed of the very same collection and theological seminary library. And knowledge effectively was used the three API queries to what they're going to pull in stylized the front end and anything is they don't have to update it from there. They continue collecting and adding their archives and everything that's pulling automatically to the backend. They can save it in real time.

So let's talk about scale again. What is responsible with this kind of customized infrastructure at the scale of something like FDLP. I looked at a map for instance of all of libraries that we can reach with this technology and found those that are already archived as partners. So this is not to say that they are collecting in the area that you might specialize in with the kinds of materials that you might [Indiscernible] too. Just about 10 percent of FDLP libraries have archive accounts, that is a have access to the web archiving dashboard. These could be libraries in your organization's already. We use it for the purposes can kinda provide dashboard of access you have a few ideas of how to put it to use. Or this might be an opportunity for us to talk about hold it collecting area, or for neglecting domain paid

That is really most eager of hearing about in the chat, and while we have the opportunity [Indiscernible] but also afterwards, we have the opportunity to get to talk about what sort of is

not in the collection already that you would like to see there. How we should go about curating it, how you do that if you have access to all the same high-tech tools they need to. And what kind of access and discovery and knowledge building is made possible or even imagine coming from that scale of availability. So I'm going to hang out here. And listen to Andrea in the meantime, that the get a chance to talk about that more at the end of the hour.

Turn it over to you Andrea.

All right thank you Carl. And let me see. We go, my beach ball is in my category here. So thank you to the organizers for inviting us here and thank you for actually filling a blank that has been missing in 2020, and what I mean by that is normally this week of the year is our library leaders forum, and our partner that all of our colleagues so around around the world from around the world to send on seven Cisco we talk about our work and have a big party. And in the spirit of 2020, it's not happening this year. So it's good opportunity to talk to a bunch of librarians and it's filling a place in my heart. So actually I will turn my camera off while I chatter, and so say hello, and I'll be back the comment section and hopefully proof of identity so far.

So I think it's Carl mentioned, and the organizers, I am a digitization program manager for Internet archive. And I have been with the archive since 2006. So lots of work, projects and change during the time. So it's a great opportunity for me to talk today about our digitization work, and I think I will share three main categories today.

And see if I can do this properly. So first I will give an overview of our digitization purposes. Second some highlights and collections of info that we have worked on overtime. And last, some projects that we are currently working on and I think would be of interest to this group.

So if you haven't seen before, this is Brewster, our fearless leader and Carl mentioned as well, we're founded in 1986, archiving the web, which is how we get our name and Internet archive.

And Reese are Brewster has a book that I will read aloud if anyone is filing it. So we begin in 1996 by archiving the Internet itself, the media that was just beginning to grow in use. Like newspapers, not to publish on the web was ephemeral, but unlike newspapers, no I was saving it.

I think Brewster's quote here is very relevant to the web in the work that Carl and his colleagues do. But I think a lot of government information, both print and digital is largely ephemeral despite all the hard work of programs like FDLP.

So this is a headquarters in seven Cisco. It's the other side of the building that Carl is on in that mission and Internet archive has always been universal access to all knowledge. So a lot has changed since 1996. But the web really still can be the highway for equitable access to culture and excess heritage and Internet access to and with Internet access and a computer and more more cell phones.

So preparing for the session, chatting with rooster, he used the phrase universal access to government information. Which was amazing. And I took out of it. So it sort of makes me think government information has been published and disseminated to libraries for centuries. In the modern digital age, when lines don't exist, universal access is harder to obtain. Especially during our strange COVID-19 pandemic year. When he can't jump to fax we need to refer to something when we are all working remotely. So what we do to further this goal? So since we started in 96, the e-books digitization group has grown. I continue to be digital library growing every day. And our text collection currently has 28 million items which I can't believe after all my years, and also almost 5 million of these have been digitized by archives along with hundreds of library institutional partners.

And on average more than 2 million texts are viewed on our site every day.

So how do we get here? I am harnessing the power of the way back machine and going back to December 2005, hopefully some of you remember when designed like this, and our pilot site at the University of Toronto was just getting into production. And many of their first partners and projects were actually digitizing Katie government publications and information. And we had hit 2500 text. I don't know if you can see on the screen, the first year and we were totally ecstatic like high-fives all around. Which was, Schmidt. But now in 2020, we digitized more than 2500 texts every day. And actually trending to 3000 on Sundays.

And so far in 2020, we have digitized six and 50,000 texts and we have 20 locations around the world. Some of them are large centers that have lots of staff, and some of them are single tribes. Prescribed space a minimum by our partners. D typing specialized in important material. In reading rooms all overpaid

In the bulk of our digitization currently is modern text for our open library project.

We have this army of what we call scribe machines. So this image in the background is somewhat of our staff digitizing. And it is, this is what we call our full-screen scribe, and he said at it like a desk. And it's the size of the person, and we can fit a wide variety of material on this machine. So we have this team a very dedicated people, some of you have been with us for years and years that have special experience, and as average, we work at a rate of about 1000 pages per hour. And we at this point of experience, a wide range of formats including finding, [Indiscernible] addition, orange writing covers, when orange painters weren't sell and someone to the health of the library, so we started see everything. But we have really been able to maintain our mandate of nondestructive digitization. So proud to say we did have a pile of findings behind the building. We maintain the integrity of the materials.

Said transitioning a bit, specifics related to government info. I would like to start with highlighting some key partnerships that already made hundreds of thousands of government publications easily accessible. So of course the partners library of Congress and we been working for years we have digitized 166,000 texts on-site at the Library of Congress and it's a very rich collection of ulcers material. Of course the Fed links library, 277,000 texts among the

different libraries. And so many representative. And I always like to highlight our site at the national agricultural library. Not to speak of we digitized over 170,009 of. But many gorgeous plates of flowers and fruit and flammable, one might say for their office, there's really a lot of dutiful materials.

Let's see. And I just wanted to also highlight some of the state and special libraries that have digitized materials as part of their collection. So these may not be libraries that are as top line a government library, the included as part of their work digitizing some government libraries.

And the other party's activities are some locally digitized material or more digital material it's been collected in archived so I want to give a special shout out to that as well. Very important part of our collection.

And is also an ongoing project that we have sort of been filtering into for a long time to capture and tag you discover docs that have been digitized or uploaded elsewhere, sitting at best to [Indiscernible] the government publication. But that definitely has title runs and publications, so hidden within it, but we haven't really had expertise to sift through it. So we finally, I will admit, can be found in the area.

So in this last section, I thought I would review a few key projects that would be interesting and can be implemented another libraries and some new projects at IA.

So we have done a lot of what we call digitization for preservation. So similar to what Carl was discussing as often the top line need. So to focus on the story of digitizing just one title. Which illustrate some of the challenges that we face when we are digitizing and some of the results. So here we have 120-year-old text that as you can see it's well used in flextime to provide a lot of reference. And his maybe the last chance to digitize a title like this before it literally turns to dust. Sometimes these books arrived to us in a bag or envelope and full of chaff in the bottom. So we have several challenges working on the title.

So pitcher really can set thousand words. You can see just some of the challenges we face. So we had findings, turn pages and a term that we use sort of lightly run the office, book dust. See can see in the top right there, the little thanks of paper. It seems relatively benign, but when they are floating around, they can fall and of scared text as we are working.

He can see this is another title that really was suffering. So as methodical as we can be when we are working, the snowflakes and braille pages can really pose a challenge to the imaging process. Especially if this is the last time we will be able to handle this material before goes in to deep storage forever. So really requires, in some cases, a lot of the work in reshooting and he can definitely going to down 1000 pages an hour when we are working on material like this.

So he can see sort of a happy end of the story. Seven images with the final actually put it together. You see the vintage tape that always seems a good idea at the time and it becomes

very brittle and acidic over time. But I think it actually, it turned out well. When you see with the before pictures look like.

So here's another example that I thought this group would be interested in just in the spirit of collaboration. So he also had cases where it's a challenge to even find a complete set our collection of material that's in good condition for vegetation -- for digitization. And what it means is it took a real village of librarians to make this publication complete, to find books and bindings that were conducive for digitization pitching Exodus collection was only [Indiscernible] but it took more than a year actually to source and find interlibrary loans, assess all the issues in order to bring together.

. So there is at least one table like this which I'm sure anyone in the government information area is familiar with printing like this. And the materials had been printed very close to the gutter and nearly every institution had rebounding issues. So even using this item in the reading the people in that reading them had trouble seeing and of course the figure you need is in the gutter. So you're reading level alone digitization.

So what we were able to find one library in the group that had not rebound the copies. And they were in the commission. So we can have these key tableside is a single image, and it took a lot of time to complete, but this is an important set of financial data that is now available for use online.

And it's a sort of, we call it digital creation, when we started ringing these books physically together in the digital world, physically bring them together for digitization and making them available in the online collection.

So I hope to see more of this work. Because some of us, these competitions and information, and initiative. It can be referred to online.

So I also wanted to highlight some unique owners. So the men aspect bring him young university is one of our partners and they have used digitization as part of the practice they have to upload hundreds of maps that include the Geo reference data for public download. And how the Internet archives can help is to derive JPEG and [Indiscernible] from these high-resolution files, and make them available in the browser. And sort of where they can come in in terms of being helpful, is that looking at these images in the browser actually, the resolution is quite good, and if you have the time and the bandwidth to Domino the images, they can do that. But you can refer to it in the browser and that satisfies your needs, you can get on with the rest of your research.

And also take a moment to know here that if you have already digitized material, IA helps to preserve reservation access and let's not do it again. Moving on to the next collection together.

So on to universal access to government information. I have colleagues at that [Indiscernible] project in the have excellent working environment to in compiling the data preservation, it

really gives detailed challenges facing not just US government librarians, but those around the world. And I think we really could have a whole conference one day trying to find a government location is. We could fill a week with those discussions for sure.

And I have digitize thousands of locations, but it really can't miss the chance to talk about microfilm.

Film and fiche as you know the spec microfilm and fiche, recently shocks that general public, but at this group I think. And through a variety partnerships, the archive has digitized thousands of microfilm and fiche with the government info like census and records.

It was really historically a really slow and involved process of preparing real trading images and uploading to archive.org to be derived. And really several years we get really work on microfilm at all. But this year we have revisited and revised our microfilm digitization activities and we are working with microfilm. So new process and booklets we are imaging processing uploading 400,000 images a day. Such as a lot speedier than I remember as a decade ago.

And so far within our collection, we have five US government publications that have been digitization or digitized, and unless is over 260 more to go over the list titles. And a few international titles.

The Fed registers a very popular publication, and we have digitized it from microfilm, again, and this is an issue of effects that we frequently finished.

As you know, we've done a wonderful job organizing and making a Federal Register available online. With excellent metadata. Downloadable by documents and fully searchable. Documents are only available from 1994 onward, which will satisfy folks working in certain areas of research. So as somewhat of a proof of concept would say, let's hope is much more of this type of work, we have linked to the FR website from the IA collection, so if you would like to or need to go back to 1936 free work, you can do that. But if you pass and if I want the fully assessable collection, it's late from our site, and also linked to the way that mission. So if you like to use it that way, then wonderful. And hopefully we want to linked to more curated web collections to create the best research for researchers and for the collection.

So you can see the FR sites are very well crawled. I [Indiscernible] on his calendar view because I said not to worry, FR to handle as far as register. And here is a page from the archive.

So how can the collection help fill in the research blank? The full tech search collection brings you through a set of reference screen the document in the hyperlink that you will see that is generated can be saved and shared with other users. And you can see here, the function can be helpful for giving research are getting research started. If you just are searching for secured terms and or zero collection without knowing where to start, he could be a starting point, and in the current reality, mostly working at home, this can also kind of take a place of writing two

stacks. If you know it somewhere in the Federal Register, we are not sure where, maybe we can be of service that way.

And lastly, I would like to highlight the Internet archives command tool because I think it would be helpful for any of the digital collections have online. And so these Python tools are very well documented, easy to use and, if you use an assessable to anybody with the online our library card to use online for free. You can is a gamut of a choice. So if you have the time and energy and you want every JPEG from a collection, you can do that. We just want a set of fulltext, from the OCR input, the candidate that way. I did a quick test, just a text file from the entire run is about one point three gigabytes download it. So usable on your own machine, it can be used for text research. Now, using a variety of tools. Some of which Carl was showing them the tools that can be used for the first one.

So last word here. Keep digitizing, crawling and curating, but let's not duplicate our we can. My archive experience is something worse than being an account is something worse than being a conference like this and watching a collection being unveiled and you been working on for months. In your office. So you know, let's do our best to keep in touch with who's doing what. And they're also very happy to provide states to upload any digitized materials or digital collecting. If we can help derive those big tips that you will have from even in the vendors, we're happy to do that. And if you have expertise and energy, you can use the.gov info. Please let us know. In my experience I know librarians in this area are very passionate and motivated. So I hope that we can work together in a more full way.

So up in the last picture up of the storage and physical archives. We're putting workflow at the speed and quality we had always hoped. So with good metadata, you know, well organized and efficient, but we try to be as cost-effective as we can. So we are open to partnering on you microfilm projects, so that something that hasn't happened in a while. And including the titles that we have on the physical collection, that we collaborate on metadata and that sort of thing. So I will leave you here with a call to action, and I will post us in the chat. One week from today. We will be hosting a roundtable discussion, which may be the first of many, but Brewster will join us and other members of our team. To sort of continue the conversation. And we want to hear how we can digitize and preserve and make government them are available and better on archive.org. And gap. Thank you all. And I see there are some questions popping by his I was speaking. So I will test it back to questions. They keep it

Hello everyone, we have that 30 minutes for questions. So we did a couple while you are presenting and I will gladly just you see can work on answering those while other people are putting in questions in.

Jessica wants to know regarding the books you digitized, are they in the public domain? And if not, how do you handle copyrights?

So I think for many years, and certainly in a government area, thus the organization was public domain material, but we are doing what we call modern digitization and we're using something

called control digital lending, which I will also relate to that in, we have a mini site that explains the rationale and process that we use to digitize and make available modern texts.

Excellent. And Angela wants to know how did IA obtain the microfiche collection, are they accepting microfiche from libraries and repositories [Indiscernible - low volume] eight

So I think that is exactly how we required it. So a lot of donations. So if you have microfiche or film you are are you need to hone, please in touch with us. Without a lot of donations a time, and we have actively done some purchasing as well. So we do maintain a physical collection of microfilm and affiche. And also have done projects where we digitized film and we gave the back to the holding library rates we have done projects either way.

Great. [Indiscernible] wants know will the Roundtable be specifically focused on government information?

It will. Yes. So I will find my Ear. It's called working with a knife, Internet archive for government library. So it's just for this group.

And Nicole asks, are there any fees or costs associated with the archives as a partner?

Thanks, great question. And answers typically yes. The way it is done is kind of on a level of need basis that is, we take a look at the shared data volume of material that you need to collect and preserve and like that out on the storage racks and kind of figuring out the amount of space you need to rent a time in order to do that. So that can range really widely, but we do estimate that out for new partners so they can plan accordingly.

Great. With about 10 more minutes, so if anyone has any more questions, just go ahead and type them in the chat and I will make sure that they get read.

Melissa, this is Laura, Luke one question from Elizabeth. Have you digitized Michael card format?

That sort of funny, we have taken, I think we have some examples where we just forgot to, but what anyone can do with that, it may be in back of my research. I think we should talk about give it a try and see what we can do.

I hadn't even heard of that one before. [Laughter] if there's a format, we can to take it as a challenge, do we?

I think if I remember quickly, it's a [Indiscernible] card, that would does?

I am honestly not sure myself

Micro print.

I will have to do my research on that microformat. I don't know that we have been a project like that.

Always wants to know are there specific government agencies that you are still interested in getting?

It's the question. I think with the web archive side of things, we are certainly always interested in anyone has a national scope increasingly also trying to decentralize and prevent more partners with local special interest and broad distribution across the country. So that is kind of the more special focus, the more special the focus, the better the opportunity, but the King go a lot of different ways.

I would agree. I think we've had a lot of federal representation, and we are sort of creeping down states and sort of municipal's. We sort of just scratch the surface with usable organizations and they may have all the information that runs our life. [Laughter]

James at the I know you mentioned one of the national libraries, what are the biggest agencies you are working with? Are the date more of the digital collection of resident preservation or is it mainly digitization?

Again on the website of things, just in terms of the size the collecting scope, the larger ones tend to be areas of Health and Human Services. National Library of medicine. A lot of health and medicine basically. Sensibilities. But that doesn't mean that others can't jump on board with FDLP . Sometimes what agencies will do around this time of year, and it semiannual or quadrennial cycle, is contact us for a one-time web crawl in a full conference of capture of their online presence before training teaching administration, or before the end of a administrative term. Rather than like ongoing permanent collecting partnership.

So those are smaller in scale but it can happen throughout the lifecycle.

Thank you. Can anyone excess download all materials or do. Different level of access depending on membership?

I will say for the books and microfilm, five. And for those of you -- you have the same copyright concerns in Canada. If it's in a public domain, you just need a library card, and you can download whatever we digitized.

And as for the way back machine collection, you see at archive.org, that one what you see is what you get. Everything is always public access. The cardiac being they can't is a member the general public looking at it, download actual work files that we hold back in storage. Archive it is a little bit different. But when libraries and archives the tools themselves, they had a curated collection, they had the option to either make that public, in which case the materials also added to the way that machine in the original preservation access, that can be all public, or it

can be held more discreetly. They can be an access point shared just among the team or with the donor or with any smaller group of stakeholders without indexing it for general public access like on our website. You have full access as a partner to all of the backend materials that you would want as well, whether the collection is public or private to others.

Great. Just a quick reminder, we have about four minutes left for questions, so if you have anything you want to know, just type it into the chat box please. James wants to know is your government metadata hopefully Mark format, made available for libraries?

So if we, when we're sitting at the return the system, we always try to connect with the library that has catalog record. And if we can't, [Indiscernible] and we capture that it a major, original record and we make available with the item and other metadata is [Indiscernible] as well. We ask partners, tend to comply, if it's all available for download, sometimes even in the content, if the content is available, the metadata is.

And that again is up to the partner and whether they want to mature to be public or private, but if it is private, the easy

And I think it's a little switch of the dashboard and that adds their collections to be server. So if he can get that either in double core or in Mark.

Thank you.

And [Indiscernible] to send comment that it would be nice if you can also get New Zealand our Australian government resources as well.

I think we, I could be wrong, I think we worked on the digital side with New Zealand, is that right, Carl?

Yeah, and especially in the website, UK has a great web archiving and based on its own, based on the British Aire, but New Zealand and Australia and other callers contacted us for what we call domain calling, that is when you get a full kind of company is a snapshot of their top level dock domain, use that intersection for that, it's a little different than archive partnership. But same tools.

And yes. Specifically in Australia, with able to sometimes at the national Library of Australia, does its own domain level calling in has come but these archive for special collection. On specific topics. COVID obviously is a huge one now, that a lot of federal and especially local governments are using, archive to collect specific collections just on the topic, so they have local control of it. That they can kind of put their own separate services if they want to as well.

And also if you have colleagues working in the UK or New Zealand and the government Aaron hadn't heard of us, please let us know . Absolutely.

It's like a knife in the heart sometimes of people are like Internet archive, what is that? But some people haven't heard of us. So any references would be good.

And if anyone has a during the call today that you would be interested, we have a seamless for the Roundtable next week. And we talk about legal cards, and we can talk about the next week.

And I see this last question. I know it's a lot of notice, and I don't think it be less people have an opportunity.

Thank you so much. A huge thank you to Carl and Andrea for a great program. They you for the audience for participating today.

Next in this room we have uncharted waters, lessons learned by depository libraries during the time of COVID pandemic.

And in the other room is new government enhances, if you want to join the meeting, you have to close out of this first and use the other URL to enter the other meeting room. This can be found on our page.

Are the FDLP homepage or go to click and join on sessions and I'm also going to put the URL there so you can get to it more easily.

And for now, we will take a short rake and pick up again at 2:15 Eastern time. Thank you all.

[Event Concluded]