Please stand by for realtime captions. [Captioner is on hold, waiting for event to begin.]Good afternoon and welcome to the FDLP webinar series digital imaging 11, we are here with technical support if you are having any problems. I would also like to introduce our presenters, David walls, who is here for the digital tangible federal government information for the government management program at the US government managing office, since 2012. Our next presenter is Kerri Thompson, the head of the digital library department, has managed libraries for the past nine years, in that time she has worked with imaging specialists. Web staff to determine scanning rates as well as partnering with the archive, for the [Indiscernible-speaker away from microphone] and other projects, for the preservation's for the website, she worked in systems and web development in my prayers, she focused on research data management and research preservation, M ILS was received from Maryland Park 2004. Before we get started we will get started with our housekeeping. Which is pretty typical for those of you who have seen, but I need to go through it never the less, we will keep track of all questions coming in, at the end of the presentation, David and Kerry will respond to each.

We are recording today's recording, -- We are recording today's webinar, along with the slide deck, you can find this all on the FDLPGovernor Academy website.

We will also send you information to the email address that you registered with, if you need additional information for people who are watching, please email FDL P.gov, email all others watching, so that we can get a certificate.

If you are having technical information -- At the end of the session we will share a webinar satisfaction survey, we will let you know when it will appear in the Chat Box, the URL will appear.

I am done with my spiel I will hand over the microphone to David. You can please take it from here, thanks.

Thank you Jamie.

This is David Walsh from GPL -- DPL.

We have a number of requests to provide, thinking back to your college days, the philosophical introduction to a topic, this will be a philosophical introduction. Kerry will talk more about specific projects she has done, leading things from more general to the specific.

Stepping way back. One of the main things we do as librarians, we acquire, manage and preserve and provide access to the information assets, information assets relied upon upon -- Relied upon by customers coming to us.

We create simply just another format that we provide access to. This process of digital imaging tension of -- Tangible content, connecting users, with the latest best practices related and available to us. We manage and protect, preserve for our community of users.

This community of users has changed a lot even from the earlier days of digital imaging, no one in the photo of all users, has a wire connecting any devices to anything, the devices are mobile.

He will notice at the top of the banner, at the slide, access to the government information is post on the go, and on the show.

Why digitize? Why start with paper and convert it to digital content? Some of this we already know, administrators lassies questions, where are you going to spend money on digitizing that stuff?

We know paper locations are specific people can only use one thing at a time. We also know that a lot of the catalog records we have prior to 1976, they are not as complete as we would like them to be. You cannot start in the middle of a document and search, and search content across another document. The tension will publications of space are overwhelming, we hear, and also between people that are used to searching online versus searching paper collections, digitization's brings these paper collections up to date with modern research and discovery methods. In the case of tangible collections and microform, these are increasingly becoming obsolete.

Turning from access to preservation, why digitize to preserve? Older content is frequently fragile due to a certain decay, paper content is formattable to fire, flood, a whole host of disasters people have experienced, in my case it was finding a source valuable to my research. Discovering that someone had been to lazy to photocopying it and ripped it out.

-- Therefore things are less accessible due to preservation's and concerns, there is no real way to provide backup content, if a disaster occurs. The fact that the collections are overwhelming space, has put a real pressure with libraries to come up with storage solutions, which create out of sight out of mind, where content is stored, and we forget about the conditions, in the case of microform. Obsolete formats reef choir -- Require updating to new formats. For preservation.

Be prepared for the administrative questions, why are you digitizing the stuff? The selection for digitalization should fit into your mission statement, you provide access to the content provider users with this content, those users are increasingly using electronically mobile devices. The content should fit within your content plans and priorities, you can also use usage statistics to show the collection is being used. To show this with the tangible content, where more access is needed and can only be achieved through digitization. You may have user driven selection, through networking that your library users, where users say this collection will be good if I attempt access it home or over a mobile device if you have your own sense of experience with collections that will not have [Indiscernible] because they are not checked out but by your own experience there is a lot of used by this particular collection, you also throw your prayer connection -- Professional connection, which there are a few libraries that exist, this content may be rare, and there may be collection need to digitize that content.

Planning for digitization. Be able to justify, and why you will digitize, can you digitize it, explore the copyright issues. Most is not governed by copyright, there may be portions that do follow within the copyright. How do you get support to your digital organization and get this from vendor? What preservation formats do you want to use, where access derivatives do you need for your users? Where will you store your content? How will you host that content to provide access? How will your users find the content? How do you share the careful progress in the work you have me doing all of this planning with your administrative stakeholders?

Talking about copyright if you have questions about fair use, these two sites I highly recommend as understanding the copyright law and what the bounds of fair use can be for ditching -- Digitizing.

The digital imaging is always doing in source delight my advice is to not forget, everyone who works for you already has a job, people think they will buy a scanner, bring it into an existing program and use existing staffing to digitize the whole stuff, but frequently it will break down in terms of the amount of time that people can devote to those projects. It takes a lot of careful planning to set the digital workspace, to train staff on new equipment, to provide the maintenance, calibration and the upgrading on that equipment it will be required, to develop some sense of input. If you are opening this to other staff in libraries and departments in the library, perhaps other libraries within your network, you name, and have some kind of idea of what you do in terms of output, and tracking cost, it can't just be the cost of the scanner, or the labor. To do it right, talk to your business manager, they will probably talk about overhead costs, that is the cost of doing business within the framework of the library within the infrastructure of the library, those costs have to be added to the total cost of doing a project, consistent imaging work in-house.

If you choose to outsource, keep in mind that the vendor's job is to stay up with the latest technology. That his business. The vendors would only do what the contract that you develop with them tells them to do. They can't work off your expectations, or your unspoken expectations. You have to be very careful when developing contract for services, that truly reflects what your outcomes are. The best way to ensure that there is appropriate communication between the lender and the outcomes you desire, is to test those publications and outcomes in a pile or a project, where you take a small sample of content, the contract, a small sample of that content, have it scanned, evaluate, and sit down to talk about anything that does not quite match the expectations or what you felt the contract actually said, you can revise the contract to move forward in stages like that.

Frequently we talk about digitization, that has best practices, or policies, or guidelines things like that, in terms of actually looking at your internal best practices, I think this set of best practices is probably the best condensed way to go about looking at content to be preserved. That is, when you sit down to look at the item, tangible publication you want to digitize, you want to look and the basic feel and look, you do not want to digitize in a way that obscures content or dementias of the size of what it looks like. Some people put a actual ruler to show the page size of the source publication. It sounds like a basic thing, but a lot of commercial digitization projects did not follow this. That is the source, the publication must contain all the information intended by

the publisher. The information from the source item should be completely captured in the digitization process, some of the complaints about Google imaging, they did not take the time to unfold foldout's. They did not take the time to deal with pages that have been pulled out in the process. You can't have dirty pages or pages with a lot of scribbling or underlining that will mess up the OCR software. The optical character recognition software allows you to search text within the document it is critical force for discovery and access.

Without an excessive page curvature, which looks tacky on screen, but more importantly this can diminish the accuracy of the OCR software, if you cannot get it where you need to get it in terms of the page curvature within accessible limits you may have to find a copy that you can sacrifice by cutting the publication. Into single pages of scanning them in individually. Finally the chosen file formats for both preservation and access, have to conform to establish best practices, which we will talk about in a moment, and the access, the content you are using. In the last I have current preservation formats, on the right I have a slightly incomplete list of access formats. As you go down the list on the left, both PNG, portable network graphics can be accessed formats as well.

I neglected to put PNG on the list I see. Increasingly I believe five way compliance, for access for disabled folks, publications like Daisy, which is a top talking books format will be increasingly in demand, the people can hear the publications read to them as they connect to the files format.

So within the Federal depository Library community, and looking at the partnerships and the people digitizing content. We have request for a partnership guidance for providing digital content for info, the document I know is too small to read on the left, it is findable if you search, pull up Google and search for guidance for contributing guidance to FD cyst info -- Assist info. The URL would have filled half the page, so we had to move it to a better location. If you search for the title you will find it on Google.

This title is a multipage document that talks about best practice. It encourages libraries within the program, if they are setting up scanning operations, or working with commercial vendors outside or in-house, to contribute to these goals that we are increasing public access, and providing there gets of copies of records, and best protecting pensionable originals that may be in a fragile state. Finally we are contributing to a national collection to federal government publications, and we are currently seeking trustworthy depository status for that digital repository, this past year, we finished working on self assessment, and we are going to the acquisition process, to find an external acquisition person to work with us. We follow the basic heart of this specifications and best practices, of what we do, what is called the federal agency guidelines initiatives, these are frequently miss discussed as standards when they are actual guidelines, this is a group of federal agencies, of which GPO is a member, we sit down to talk about the best this and best thinking. To digitally image publications, and you can find this on the site, which is a combined many different views and needs of these federal agencies.

With each agency, what they will have to do is develop their own internal best part of this, to fit their content and how that will be to Michie -- Will be digitally imaged. Then stored to their users.

We don't talk about metadata at all, we will talk about this in upcoming sessions, the holder of a lot of guidelines, for various metadata, we are all familiar with MARC 21, many of the good sources for the data, is the original catalog, in terms of the metadata you need for these projects we are starting with a very good MARC record, and then eventually XML -- Where to put your content is something that you can work on before you digitally image the first document, that is what content and access systems will you store it in, and provide access from? This is not intended to be an endorsement of any particular one, we use FDsys.gov info, there are many digital forms out there it is important to do work for the system that works best for your library. Putting in that content. That is nothing more than desktop PC, until you end up with a lot of content that you need to preserve.

Here is some more information sources for your knowledge and training, this listserv is a wonderful place to work. To see some of the different conversations going back-and-forth as people start to discuss problems and needs about the various digitally process projects that they have going on, the Northeast projects that they have going on, the NE. Conservation Ctr., provides an ongoing training webinar, you can join, for a fee, I believe it is relatively not -- Relatively modest, -- They provide training, and much of this preservation issues we are talking about -- Here's the link to the digit in guidelines #the digitally -- The guidelines.

This is minimal digitization capitalize Asians, which are important to look at as well.

Assuming that your outcome is success. If you do all careful planning and thinking, if you take it slow, and carefully, the success will be a successful digitization project. To support the library's mission and your preservation priorities. You will have an enhanced user experience for the users that rely on your information product, this will be realistic goals, you will not set a bar too high for yourself. The careful planning and documentation of the knowledge that you gain will provide increased knowledge that you can provide in the future, and you will be providing tangible content for things that will need to be remaining intentional content by creating digital circuits, you will be collecting the tangible form to a new community of users, that seek that information to be able to use on their mobile devices while on the goal -- The go.

Thank you very much. We will move on to our next speaker.

Great. Hello everybody. Let me just switchover, this Iscariot this was sodium library, -- Smithsonian library. I will briefly talk about working with vendors for your collections, I will only point out the key things to look out for, when you are outsourcing your digitization project. Although we have a robust program at the Smithsonian Library. , there are things we have chosen to work with over the years, this allows us to be more flexible on how we allocate our time, allowing us to focus on the things that we do best, particularly the material that is too fragile to leave in the institution or have other people handle it. Without using vendors we would never munched -- Launched into match this mass digitization, we had to develop workflows, enabled us to ramp up our digitization, since then we have adapted our workflows to use in-house. Not only did we increase our capacity by outsourcing, but also internal capacity as well.

Going outside for digitization enabled us to digitize formats we were not used to digitizing, like audio. The main reason that people outsource, we get grants to digitize, it is unfortunate. Like many of you, most of our digitization is still project-based and grant funded. The good news, project-based grant funded things are ideal to use mentors for, assuming that you can catch your contract before your grant design is up.

Now how much work you will need to do upfront before you contract a vendor, it is dependent on the sky -- Size and scope along with the restrictions. I will not talk about contracts that is not my area, but the roles will be a little different from an organization to organization. You will start out with the description of what your goals are, if it is a long-term open-ended project such as the agreement we have with Internet archive here, you might want to periodically renew the contract. You definitely will want to request for proposal, or quote process. Find several vendors and get several quotes to do the Garris -- Vigorous vending. If it is possible to talk face-to-face with potential vendors this can be helpful to supplement any information they give you". -- In the quote. Sitting down them -- With them is a big part of working with these vendors and what the capacity is, how'd the is there -- How deep is there match.

You may want to do RFP regardless on the content size.

This proposal should focus on describing the material you're trying to digitize, what is important about it what is unique about it, what are the characteristics are and what you hope to get from the digitized version. Those RFPs can include who your intended audience is and who it will be. If it is a more straightforward project. You know how it is going to be done. You can do a statement of work. If it is a large project, always request samples from potential vendors". We can read and understand the statement of work, can we meet them?

Even if your project is small, and you do not need the statement of work you can draft a one pager, before you can do that, you really need to know, just what you want. You have to articulate and be specific you have standards and deliverables, working with archives, most archives digitization vendors know what FADGI is, and it can provide a convenient shorthand in discussing guidelines for digitization. Your project might not, to me it is all sorts of guidelines, you should probably work with the Bender -- Vendor and what they need to give you. Especially if you have a fixed budget. You might not be able to get everything within your budget. Think upfront. What is absolutely mandatory, in terms of your goals for digitization, and what would be nice to have? And then the statement of work in terms of mandatory elements and option elements, this statement of work enables a vendor to give you one or two quotes to help keep you within your budget.

When you write up your details of statement in your work, be specific about anything if possible, specify whatever your file naming conventions are ORA metadata you expect to use, as well as what derivatives you will want to get back. All of these imaging guidelines should be in there to 20. Should -- There imaging guidelines should be in there too.

To want blink pages scanned? -- Blank Pages scanned?

You should lay out any expectations you have, turnaround times, and communications expectations, if you have a deadline, or if you expect them to respond to questions that you have within a couple of days, put that in the statement of work, make sure that you know they can meet those expectations.

If the vendor is creating extra data for you for instance, if you are doing it transcription project, or rekeying, providing extra market for you, Sarah -- You should spell out very clearly how your data should come back to you, ensure that the data fits into your current workflow environment. You do not want to create a whole separate workflow, to get your creative material into your repository or your catalog, or website. Don't forget to spell out expectations like, for example if you are doing a rekeying project, typical error rate you should demand, would be no less than 300 of a person. If you think of -- No less than 300% -- If you are thinking of errors.

Depending on the size of the project, you may want to receive your deliverables in batches, even if it is not huge, you may want to just get a may -- You may want to get one sample back.

Invoices and manifests are not just good for it billing and tracking, they are really good for keeping straight what has been reviewed, and what hasn't, as well as providing a record of what has been sent to the vendor and what has been received.

Obviously these are electronic manifests spreadsheets.

We have definitely found, if you have them. Identifiers like our coats they are really useful in digitization, a couple years back we involved a collection, some have been bar-coded, some had not. You can bet having the vendor barcode the bits that were not done yet was involved in that project. The vendor is starting out with data from you. Such as an inventory list, make sure everyone will understand how the data will be provided before the digitization starts, this is a good way to get the technical contact with the vendor. Everyone is on the same page in terms of how the data will flow back and forth. Mostly you will be coordinating with the project manager, or if you are unlucky sales manager, it never hurts to try the email address from the technical contract person, or a metadata specialist, added that project manager, sometimes when we had complex projects, when questions or problems arise from the person we need to talk to was the engineer or the metadata specialist. Having this content, and their contest -- Content information helps.

If there are some things you may not want to put RFQ, but definitely look out for. This can happen in any face-to-face conversation with the vendor, if you are transferring material off fight -- Off-site, provide proof of insurance and how these collections will be secured while you are working on the project. Don't forget to ask how they train the people who will handle your collections, depending on the value, you may also want to see resumes of key staff. If your project is large, you want to get in this a better sense of what turnover is like, do they have the capacity to train or retrain digitization staff to keep up with large incoming projects, as well as your projection targets?

Last but definitely not least, you will want to have a vendor do quality insurance on their work. However ideally, you should also allocate staff time, to allocate work they have done, ideally it

is 100% review, for most of us it is not realistic. The good news. We found doing review on a statistical sampling, has been pretty indicative of the job, this slide shows the table we use, which is based on a standard for manufacturing, to help us know how many items we review out of each week's worth of material.

We also defined early on in the project, what a major error is, which would require rescan or every key, versus a minor error, sometimes they can add up to a major error, and we have insisted at least in one project, that if any shipment, fails review, in terms of how many major errors that gives, we will return the entire shipment and have the vendor do their own quality assurance, and rescan if necessary.

So there are many advantages with working with vendors, even though sometimes it seems managing your relationship with them, moving things back and forth is almost as much work as doing things in-house, it really is not. Consider your capacity by working with vendors, and hope some of this is helpful or a jumping off point for questions. Thank you for GPL for hosting. Thank you for listening.

We will take questions now anyone who participates from the audience, if you would like to ask the presenters question, feel free to chat at the bottom box at the screen, while you are presenting, I will submit a couple of questions. The first one for David. What about blank pages, wet -- Should they be obtained in the original Sarah get clicks -- In the original? --

It depends on whether you are trying to require a true facsimile, if you are trying to create a complete circuit -- Surrogate, then you should probably go ahead and scanned the plate pages before you get to the actual content, they are counted bibliographical he in the description, why that would matter is authentication, sometimes to determine which version or versions were published in printed Turco if you just -- Printed. If you just scan on the page, you may be missing some small attributions, on those pages while maybe not entirely blame, but they have the actual content that could be valuable in the future.

There was a second question?

Second question is, there is a couple of places where the spec DIV urges from the FADGI spec. -- Where the surrogates diverge from the FADGI.

You do have to make the FADGI guidelines real for your situation, any time that we put out partnership guidance like this. We do a lot of bidding entrapped form, both to the Council and our stakeholders, and our community things go up on the web for content. We do individual receiving of emails and content. The guidance documentation as is, we were actually discovering for a lot of people distributing content it was easier for their scanning operations to meet these standards, then it with great skill management. -- Grayscale management.

The color most people have is easier for people to produce than trying to manage the grayscale issues. Again for the community feedback we have received, it was easier for them to essentially deliver the uncompressed JPEG 2000 format then it was to do various compression algorithms. A lot of conversations with our community of content providers and partners and stakeholders, a

lot of discussion over the past year and a half, of rolling this out, as to what they can provide. The guidance documentation is a living document, if this is changing, as we get content in, working with our partners if things change, we can change some of the guidance documentation. The guidance documentation is within FADGI to provide a framework of what is largely achievable for content going into FDsys/govinfo We have another question. What other things do you have an mapping, and what solutions do you have to make this workflow smoother?

That we have seen almost everything imaginable, that maybe something too Kerry can comment on, we were trying to provide a basic pathway as far as metadata development, that would be a mark to MARC XML transforms, a we will see this mix as well in the guidance documentation, some people cannot provide much metadata cop --, -- Whether they outsourced to another vendor or in-house, the amount of metadata they have is part of the project, if this is something they have done in the past, we work with them with whatever they created, if we can help them plan, we prefer FADGI -- The MARC XML.

We match it to either MARC XML, or the mods that we don't have a lot in a pinch, worst-case scenario. Back to Dublin core, then back up to mods, but we do not usually have to do that.

As many of you have noted, I apologize for this, when I went into the webinar, I goofed and I totally apologize my fault, this is Ashlee by the way, if you need to chat a question, chat to all panelists, everyone can see them chat, -- The chat -- I don't know if you see the questions I have, I have three more questions.

I have one more.

Go ahead and answer that one.

Incoming students in digitization newbies, what is your advice for someone just starting out in the field with what you wish to have known going into the beginning?

Personally for me I graduated from a library science program in 1995. I have had to burn the midnight oil, and trial by fire to learn about this after the fact. Digital imaging was not really a part of the library experience we were microfilming, we were doing facsimiles, book scanning at the time, where the library I was working which was photocopying onto paper, those photocopiers became scanners. The vendors would say what do you want done with this digital content come that prompted heavy investment and learning curves for me, taken a lot of classes, school for scanning at the time. Completely immersing yourself I'm sure Kerry can provide a more up-to-date perspective of her experience and she has come into the field relatively later, much later than I did.

I came in luckily for me I inherited an existing program, and I just try to refine and spruce it up a bit. I never had to actually do reformatting, folks have done that for me. But the things I keep learning all the time. All about the metadata. You will never do it just once. There was a third thing. About metadata. -- Yeah -- Most of the stuff we have done, works great now. I think I have tried to pave the way, to having things scanned more than once, even though we like a special collection, of we only handle it once. In 15 years, we will be able to do something much

better than what we can do now. This is pretty good. I always assume there is room for improvement.

I have a question from Carolyn for Kerry, can you give an example of a rekeying example that the Smith those in -- Smithsonian Library. libraries have done? -- Smithsonian Library. has done?

This is a standard in the botany world, that lays out basically a biography of all the botanists, which includes bibliographies, all the issues they have described. We have had this work, and we scanned it, from the scan we actually outsourced, because working with the OCR, it was not enough. This title was a giant database in vapor form, we wanted to get out the more database elements of that, so we had a vendor rekey it. Formatting the page heading, indicating that different parts of the page, when it was a biography entry, things like that. That was a pretty big project. I just managed to digitization part of it, and it was extremely detailed. I cannot even remember how many volumes in the series. For that series may be 4?

We sent specifications and have them sent samples, review the samples, making sure that they understood what we were recessive fine, -- We would change the statement of work ever so slightly, and it was a successful project, it was still on her old loves, on the website, you can see it is all marked up in XML.

I am interested in statistics and the proportion of content that you need to do for adequate QA, are there any other sources ice should -- Sources, I should consult for that?

We just use that one standard. It seems to work pretty well for us. That isn't to say, that we haven't after-the-fact found errors. Using the guidelines, and making up what you wanted to call a shipment.

Wheeze to do this in terms of actual shipments, now we do it in whatever we can scan in one week on shipment it is very indicative of the quality you are getting.

If you get more than a couple of errors, something is off, and you will need to check it.

This is David. Of all of the projects into choosing work. I have done in outsourcing over the years, it is human error in turning pages. I don't know if Kerry has experience that as well? Unless the document has been cut into individual pages, human beings turn pages get pages get off track things like that. We adopt 10% rules of the sample, then have the thresholds, the circuit breakers built into the contract as Kerry was talking about, a certain number not attainable.

Part of this has to be the calibration work, in house that you do to your scanning equipment to make sure that it is delivering what the specification on the screen says they are, the vendor that sold you the machine can provide calibration assistance. For a lot of people the quality assessment is basically giving all of that calibration work to ensure that you are going to get good quality out of it, and then check and balance of the sample, a random sample, or whatever you feel you need to ensure that the scanner is delivering what you are expecting. In terms of the actual content, that it has not been skipped. There are programs also, that we did not get into.

That do quality assessment on quality formats, to ensure that the resolution is what you expect it to be.

Great, I am not seeing any more questions. We just pushed out the webinar Satisfaction Survey, we would appreciate if you fill that out for us, helps us with the programming here at GPO, we want you to know the second part of the preservation serious it was just announced right now, on FDLP.gov, I will send out the link, the title is planning for managing and storing digital collections, Wednesday, April 12 managing and storing digital collections, Wednesday, April 12, 2017. We hope to see you there. We want to thank our presented topics, and the presenters for their topics thank you very much.

[Event Concluded]

Actions | Save as Text | Save as HTML