

## Benefits and Challenges of PDF Migration

Please stand by for realtime captions.

---

We will be getting started in five minutes. This is Laura with the GPL. A last audio check for you. We will be getting started in about two minutes. Hello everyone. Good afternoon and welcome to the Academy webinar. Benefits and challenges of PDF migration. I name is Laura Flynt. Ashley is on tech-support. Presenters are Jessica and Anna. I will walk you through a few housekeeping reminders. If you have questions or comment from the presentation fit please feel free to chat them in the chat box. We will have our Q&A session at the end of the presentation. We are recording today's session. We will email a link to the recording and the slides to everyone who registered for this webinar. We will send you a certificate of participation. Using the email you used to register for today's webinar. If anyone needs additional certificates because multiple people watch the webinar with you please email. Include the title of today's webinar along with the names and email addresses of those who need certificates. At the end of the session we will share a satisfaction survey with you. It will let you know when the survey is available. We appreciate your feedback after this session is through today. I would like to introduce our presenters and give you background. Jessica [ indiscernible ] is the digital library and it GPL. With a specialization in data care ration and digital preservation. Since 2015 Jessica has supported GPO's trustworthy digital certification. And providing standards-based stewardship of publications at GPO. A scholar communication and discovery services librarian at the Federal Reserve Bank of St. Louis. Since graduating from the University at Illinois where she stuttered studied digital preservation her work focuses on the scholarly lifecycle. Including pre-print publishing and information retrieval. Thank you for presenting with the Academy. I will hand the microphone over to Jessica.

---

Thank you Laura. This is Jessica. I am going to go right into our presentation. Today's topic is benefits and challenges of PDF migration. To start I will talk broadly about what is migration? What am I referring to? When we talk about migration we are talking about a response to the risk of file format. Within the preservation community [ indiscernible ] has always been considered to be one of the great threats to successful preservation. It's one of the most common ways that content fails to become accessible or usable well into the future. As consumers, we know that is a devil different file format. It can contain functionality that alter significant challenges to our content itself and how we preserve it. New file formats might supersede existing file formats. It also may be extinct as new commercial industries. Any been open first communities. Failed to support the development of tools that allow us to use that content. Or make available or usable or understandable in any way. Migration, which is sometimes simply called data migration or it might also be file format migration. It's one strategy that is used by digital preservation by institutions. And involves converting data to another format source. Due to these changes in technology over time. For those of you that work in digital preservation or institutions, a chart on the screen is very familiar to you. It's been around since 2013. To make everyone aware it has been referenced throughout the digital preservation community for the last seven years. It is referenced so frequently as to if

assume it is just a general activity we do as digital preservation. It is an activity that all institutions will actively plan for, as part of their responsibility as digital preservation stewards. In 2013 this is where the concept of migration became so ubiquitous within our community. The national digital storage alliance has levels of digital preservation. It is maturity model still used till this day. Really the only thing to point out here on this screen is when it comes to how institutions are preserving file formats specifically. There is level 1 and level 2 and level 3 and level 4. Level 4 is considered to be that state where the institution is doing the most it can do to preserve content responsibility. In this level. Saying the institutions will format migration. Or similar activities as needed. These are considered to be sort of the goal of what our institutions should be trying to do. When do we do it? When is there an appropriate time to migrate. When should we implement into business procedures? The preservation best practices really encourages institutions to develop a robust file format monitoring strategy. This is a process of proactively characterizing all of the content you have in your repository. Knowing with the significant properties of that content are, and what needs to be retained over time. From that information you would then develop a monitoring plan, or a monitoring strategy. You may call it something else in your institution. You will essentially have some sort of proactive plan of how you are going to [ indiscernible ] about what is happening with the file formats. So you know is this file format likely to become obsolete in the next five or 10 years? Is it already obsolete? Some of this might include developing communication plans are to medication strategies around how you communicate to your administration or your users. How certain types of content may or may not have supporting technologies to actually use it in the most ideal state. This provide some examples that are out there. Of what you can refer to when you putting together a monitoring plan. This can be the [ indiscernible ] Congress sustainability digital file format registry. And the national archives also has a very robust digital preservation framework now that has just recently been released. Both of these are good resources when they break down every individual file format type that is there. Another really good resource is the digital preservation collations watch report. These come out infrequently. They are still a good resource to subscribe and stay informed about how different file formats are changing over time and what the different levels or risks are across each of them. To relate this specifically to the work that's going on at GPO and some of the project teams I am a part of , for those of you that might not already know on December 28 of 2018 [ indiscernible ] made history by becoming the first organization United States and the second organization in the world to achieve ISO certification for our [ indiscernible ]. The isil standard is considered to be the highest level of achievement for any digital preservation institution. It's a process that includes a third-party audit by an accredited body that assesses all components of the institutions digital reservation processes. This is across all areas of organizational infrastructure. Digital object management, and also infrastructure and security risk management. At the end of our audit I certifying body provided recommendations for how to continue our success. Going into future audits. At that time, in 2018 two of the recommendations that came from our auditor included that we work with our content producers to improve the standardization of [ indiscernible ]. Submission information packages. So the content we are receiving -- I think somebody else is not muted. Okay, sorry. They recommend we work with our producer so there is more [ indiscernible ] of the concept coming to our repository. That can be centralization in a variety of ways. That means that GPL has more of a relationship with how the

content is actually created at the start. So that we have no more control over how it is being preserved. Later in our own workflows. That would be an ideal goal that we have for an opportunity we could look into. The other suggestion that they made is that even though our current preservation system works great for our current holdings we do need to prepare for new publishing paradigms going into the future. That will be the case for any digital repository institution anywhere. Inherent nature of working with digital anything in this era. Things change over time. Specifically with the changing perspective of what is or is not a publication. And working with the federal government. And the various types of information that are made available to the public. That GPO may or may not and just going into the future. While neither of these recommendations explicitly NHTSA needs that GPO develop a migration strategy, both of these recommendations do implicitly reflect this awareness that GPOs is determined by our capability to standardize content within our collections. Which allows us to provide more control over any possible threat. And as well as our opportunity to proactively monitor future developments in software capabilities as it relates to file formats and future publishing paradigms. From this GPO has actually established this group called the risk review board. This is an internal working committee that was established that documents all high-level risks for our repository, and each of those risks are documented input into a registry, we then develop a specific handling plan. And mitigation strategy. Specific to each of those risks. Of those risks we have documented is the risk of file format obsolescence. Specifically what would be the greatest risk would be some sort of obsolescence relating to PDF. The large majority of our digital collection is PDF content because we have documented file format obsolescence is a potential risk, something that is definitely in our registry and we want to create a handling plan for it, for GPO this means we might have to look into migration as a potential solution. It seems to be the appropriate route to go. Specifically because our content is so heavily PDF content. We thought it would be best to start to scope this specifically to the PDF file format. And think about specifically what would it look like to develop a pilot project potentially. To explore PDF file format migration, and if we were to do that we focus in on it says all other PDF formats, these are PDFs that are not PDF a. These are versions of 1.0 to 1.57. They are deemed endangered. They faced technical challenges to preservation or responsibility for [ indiscernible ] poorly understood or were the responsible agencies are poorly equipped to meet preservation needs. It made it seem as though there is something inherently more presentable about PDFa and more preferable about PDFa. In that any PDFs that are not PDFa are inherently less presentable or inherently more difficult to be preserved, or bringing these other sort of risks into considerations deemed by the digital preservation community at large and the international preservation community. It seemed quite logical from this. Also a lot of the other literature that exists out there that PDFa would be the file format we would want to move all of our PDFs to in an ideal state. This is an assumption we are working with at this point. Just to reiterate there are other reasons why this seems to be logical. This is also consistent with all of the recommendations and guidelines that are currently being provided by many federal institutions. The national archives and members of Congress and [ indiscernible ] and Fendi and even GPO. We all currently have language that exists. In some form or another. Instructing other libraries and partner institutions that when possible you should create your content in PDFa. When possible you should strive to preserve your content in PDFa. This is ubiquitous recommendations across all of our major library institutions at this point in time. After this it

does seem as if a full endorsement of PDF/A is appropriate. It might not be that simple. As we found out. This is where I will let everyone know I am not a PDF technical expert. From what we do know it helps to revisit what exactly is a PDF. If we talk about PDF versus PDF/A, PDF is a container to several components of data. When combined they can experience similar to interacting with print media. These components, if you think about a PDF file and the document it made up of all of the smaller parts. These parts can include line art, text, images and metadata and embedded objects and text. Color scheme is an different object types. All these different components reference one another. There is dictionaries and occasionally these interactive components like JavaScript. They are generated almost as layers. They are compressed out the maroon thickly. This compression allows a PDF reading software to make them available to the users. When you're interacting with them from a user perspective you're not seeing the various layers. For most people if you are using screen reading different technology that's a different experience. There are different components of the PDF. It is all enabled through semantic and coding. That semantic encoding allows us to experience these parts as if they were not layered or if they were not parts. It is all enabled do three different tagging and protocols as well. It is all of those different parts. As you might imagine. That create dependencies on other proprietary technologies or standards. To make a PDF successful. Dependencies could be dictionaries or iso-standards for colors. It could be security data or digital technology. There is all of these other things that exist outside of the PDF itself. That the encoding of a PDF makes reference to. Every time there is this back and forth referencing that creates more inherent dependencies. Potentially those are things that could be a risk to digital preservation. You may not be able to control those other dependencies. Those dependencies might change over time. A basic PDF, a basic PDF that is not PDF/A, it allows for more flexibility when it comes to those dependencies. That allows potential for those flexibilities and and dependencies to exist. The PDF/A is more strict. That is why PDF/A supposedly has the benefits that it has for preservation. It is more strict and does not allow for that many dependencies. PDF/A formats are intended for long-term preservation. Their intended to represent, to preserve a static visual appearance over time that is independent of any of the tools and systems that are used for creating or restoring or rendering files. There's more constraint from eight PDF/A standard. The intent is for PDF/A to maximize device independence. It is to be self containing. And self describing and self-documenting. All of these self things. So you are not referencing other things but the self-contained document. This is achieved through a set of restrictions. The restrictions can include the requirement for all fonts to be embedded. It does not allow for JavaScript. It might not allow for video and audio content. It has a specific compression scheme. It might not allow for color spaces that are not embedded. It also might not allow for certain types of inscription. These are examples of things that PDF/A really has a light of control over. In order to make it PDF/A standard embedded. These are also all of the things that specifically allow PDF/A to really be about maintaining that visual image. That is what it is trying to do. We felt as if PDF/A would be the appropriate file format for us to do a migration pilot with. We were starting from this point. We were learning about the differences between PDF in PDF/A. We really wanted to see are there other institutions that have done something like this? It is nice that I get to present this presentation with Anna. Her master thesis work was a large component of information that we referenced in our own research going about this topic. Anna is more familiar with a large-scale migration sort of tool pilot that we were envisioning for

ourselves. We found that these restrictions of PDF/A provide this long-term reassurance of fidelity and integrity of content. There's a certain and controlled risk that when you take a PDF and you convert to PDF/A, in order to have it conform to the PDF/A standard you are potentially going to be disabling certain features of that original PDF. And that transformation process. That might impact information or significant properties in the PDF that could get lost. They could get altered, or they are always different things within the PDF that any of those components could be changed once you try to migrate PDF/A. You can look individually by PDF. And make sure you are not losing parts of that visual integrity that you really wanted to maintain. When we were looking at some of the work that Anna did, when we try to find other institutions that have done things like this we found other institutions are not choosing to do mass migration. At least not with PDF. We did find that the library was an institution that outright concluded in their own assessment that wholesale migration of PDF content to PDF/A is unwise. Not a good idea. This did make us pause even more. Definitely brought us to this whole question again about is migration really the solution? This is where GPO is sort of it right now in our own research phase of this migration process. It is ongoing. We do not have a clear-cut answer or solution specific to PDF just yet. We are grateful that we recently received the opportunity to become allies of the member of the PDF Association. The archives are also members of the PDF Association. This is an opportunity that will allow us to speak directly with industry experts at the PDF Association. It will allow us to have access to active technical working groups that contribute to the PDF standards. These technical experts can provide us more detailed information about validation software for PDF file formats. And help us better understand the very technical individual components of PDF. From this opportunity we are realizing that there might be paths going forward that are more beneficial or more practical than doing migration at this time. For instance, as you are talking to industry experts we learned that implementing PDF validation software such as one called vera PDF might be a better solution now than what our time would be spent trying to do a migration pilot. There is a technology that provides a report of why a PDF can or cannot be migrated. If it were to be migrated what components of it would need to be altered? And would potentially impact properties in the PDF that might have implications for being rendered for its authenticity. This is a technology supported by the PDF Association. It might give us a way to better characterize our PDF content. In order to better develop registries of what features are or are not within our PDFs. From there we can create more specific risk handling plans. Or determine exactly how risky migration may or may not be across the heterogeneous nature of approximately 7 million PDFs that we have in our repository. On a philosophical level the research endeavor brought us back to our initial hesitations. About the DPC bit list. That question of our PDFs that are not PDF/A inherently more endangered? On a philosophical level we actually think it might be possible to argue that isn't exactly the case. Or at the very least this is not an assumption best to approach the topic of migration from. If you have a large collection of PDF content and you migrate to PDF/A, that migration process itself might create so much risk through the needed changes to those PDFs that it does not outweigh any benefits of the migration and itself. It does not make, if you're ultimately having to have the most presentable collection of migration might not be the solution. From what we are learning it seems as though at the end of the day you want to preserve the content. You want to preserve the user experience. You want to make sure that you have to always understand your content.

More than anything else. If you do not understand the individual components of this PDF file you cannot really make a well informed judgment call on just how risky it is to convert from PDF to PDF/A. From that it does not seem to be so clear cut as simply saying anything that isn't PDF is inherently less preserve a bowl. This slide contains all of the main resources that have really informed our project group. And some of the conclusion our team has come to on this. I am more than happy to answer the questions at the end of our presentation. For now I will transfer over to Anna

---

Thank you Jessica. Hello everyone. Good afternoon and thank you for joining us today. In this portion of the webinar I will expand upon several components and considerations specific to the PDF/A for a PDF archival file format. The real question, do you really want PDF/A? We will investigate this question in the context of different institution types. We will consider the risks, specifically some cases of loss when migrating to PDF/A. Also, we will investigate why PDF/A should not be considered a silver bullet for ensuring the sustainability of electronic documents. As Jessica noted in her presentation. PDF/A is a variation of PDF. Regulated by the international organization of standardization in the [ indiscernible ] 005 standard is an electronic document file format for long-term preservation. It's a variation of PDF. What is exactly the difference? PDF/A differs from the standard PDF as well as other file formats. Its intent to mitigate issues by embedding information needed to render content. For example, when I opened this presentation on Windows the slides I drafted on a Mac operating system. The thought specific to my Mac operating system were lost. They were not installed on my Windows operating system. As a PDF/A the font information would have been stored in the file. The following are rendered as expected even on a different operating system. We know that PDF are excellent containers for aggregating content. That is why we use them, right? Let's think of PDF/A is the ultimate container for consistently rendering the content. Like the preferred thoughts I had. In the file on my Mac operating system. In 2005 the iso-005 standard was first released. This version expresses how a PDF version, 1.4, might be created in a more sustainable or archival format. That is with PDF/A 1. With this version came two levels of conformance. Conformance level a and B. A fulfills all the iso-00 five requirements. Notably information about how the file is structured. Level B captures information about the file visually appears. Note that while the conformance level A uses the term accessible, that does not equate accessibility. Such as accessibility for screen readers. There is another variation of PDF called the universal accessibility to actually get sent to accessibility for users who need a screen reader. Back to level B, as I mentioned it captures information about how the file appears visually. Digitized files often contain obstacle character recognition. Which doesn't conform to the file structure requirement of come performance level A. As such digitized son content is best suited for basic performance. When working on a newspaper project among the most common issues I encountered in the PDF file structures with the absences of linearization. Which improves loading up the many layers of the PDF file in a web browser. So that they OCR doesn't render first. Also, it improves and requires PDF tagging which establishes a logical reading order of content. In the case of digitized newspapers the embedded OCR text blocks are not always ordered correctly. PDF tagging is impossible there either. On January 29, 2007 Adobe announced its intent to release the full specification of PDF to AIM or the Association for information and image management. For the purpose of publication by the international

organization for standardization. In July 2008 the PDF version 1.7 was published as a digital standard in ISO 3200 1. Of years later the release of PDF two. With that release performance level U for Unicode was introduced. Level U is almost the same as B inquires the [ indiscernible ] of embedded fonts. Here we have PDF a one and two and all the comport performance levels now on 2 PDF a three. Like PDF a two it supports all levels of performance. The disputed difference is that PDF/A three allows embedding of any file type. This flexibility is almost exquisite in some cases which we will look at in a bit. Under criticism are notable risks of archive quality. It is illustrated in a [ indiscernible ] report. The report acknowledges that the file format maybe is for certain situations that require manipulation of files. For general long-term preservation. The use of PDF/A one or two which do not allow embedding of any file type. Those are suggested. The underlying difference between the ISO standards is to accommodate different versions of PDF. Version 1.44 PDF/A one , while PDF a three circumstance dashed by embedding any type of file thus introducing risks associated with those file types. It is not in a safe contained PDF a world. On the previous slide you may recall that -- adding sustainable quality to the new PDF version. A new version of the new version of PDF. Not extended version to the PDF format described and other ISO [ indiscernible ] versions. Enough of the details. For more information on PDF/A I recommend viewing the PDF associations PDF/A in a nutshell guide. Which you can see in the resource list. On to more interesting content , being cases of loss. Interestingly enough to me at least. As referenced in the NASD report there are risks to migrating to PDF/A no matter the standard version. Earlier just the highlighted elements for bid and in PDF/A such as JavaScript. And certainly file types, excuse me font types. Here I will highlight three cases in which the rules are problematic. First up is font embedding. A colleague who knows ancient Greek much better than I do. That is anyone who even knows the faintest commands of language. The colleague noted even the slightest change in placement is a critical risk. Illustrated in the slide. Between the last letters of the third line actually better represent the meeting when created as a PDF/A with Adobe acrobat D.C. The performance tools. It is creating more meaningful representation than in the source file. That is good. This happen without our knowledge. It had to be visually assessed. It happened without the author's knowledge. The author's intentions it could prove detrimental in some cases. If the die critic was lost altogether the die critic has structural meaning to the artifact this script was written on. Without the structural information a lot of content is lost. Just with that one that we lost to conforming to PDF/A. Here we have identified a significant property for student certification that uses a [ indiscernible ] script. I.e. the font embedded from the source file that did not confirm conform to PDF/A. Onto image and tribulation. And a master thesis on PDF/A the author states that several PDF/A conformant software removed interpolation from embedded images. Conforming to the ISO -- specification. Which states if and image dictionary contains the interpolate key its value shall be false. The example and a slide, all other embedded features including image width, height and color space and compression remain the same after migration. The embedded features remain the same the pixel array was not consistent across migrations. As you can see here on this slide. On the top left the source file image and the other three images are post migration. These last two cases were identified in electronic theses and dissertations that have been deposited in an institutional repository. These cases would not have been result of mass migration but rather the result of recommended practices for creation of these and dissertations. The content creators would have been responsible for any changes

to the file. We will look at the key to recommended practices within research institutions a little bit later. On to our final case. The archaeological has published notable research on their own investigation of PDF/A's for cultural heritage data. These publications are rich with CAD files in GIS data and reading models and other complex digital objects. Which would generally not be allowed in PDF/A, but are allowed in PDF/A version 3. In a report on preserving the great literature explosion PDF/A and digital explosion [ indiscernible ] limitations of PDF/A for heritage data . the report we looked at earlier acknowledges cases in which embedding of any file type would be useful for capturing their original content. Despite the issues in the other concerns in the PDF/A three. More evidence states that the appended or embedded content will lack the appropriate metadata that can provide important contextual information about complex data streams, that can assist in the assessment of significant properties, and can aid in the development of digital preservation strategy. Still the authors remain positive acknowledging that perhaps onerous familiarity of the complex embedded thoughts or embedded formats along the digital presentation community makes the task for enhancing [ indiscernible ] impossible. Among limitations in both preservation and access with a focus on preservation of look and feel rather than the actual data, even PDF/A three files which allow embedded objects A difficult to extract and reprocess digitally if the objective in preserving and disseminating a piece of research is to capture the data in PDF a makes the data difficult for a machine to use the PDF/A immediately becomes inadequate for any use before any consideration of digital preservation challenges and risks. We learned about some potential cases where PDF/A would be problematic. Let's look at PDF/A within a context of institutions. In these cases represented two institutional perspectives. Depository institutions and educational institutions. The latter which has the capacity to support the ongoing research lifecycle. The first at-bat being repository institutions such as the GPO or any of the with an 1100 depository libraries in the United States. These context digital preservation actions are subject to the notions of what you have is what you get and what you get is what you have. These institutions might have deposit guidelines and format policies or other documentation that mitigates potential risks. Ultimately indeed what you get is what you have. Unlike a research institution depository libraries and archives do not have the opportunity to work with creators. Research institutions are invested in the lifecycle and works. This perspective of her research institution is that [ indiscernible ] content creators will not always listen. Nevertheless in the case studies I referenced two student theses. University of Virginia and [ indiscernible ] are two among many research institutions which required the deposit of ISO files. They publish helpful guides for students when recording these files. They have to make retroactive modifications when transforming their files to a PDF a. They may not make thoroughly considered decisions in making those transformations when there fixing a deposit deadline. These guidelines are not supportive of the creation process. Remember creation is the entirety of a research lifecycle. Not just saving a single file. Let's think about research institutions. In order to receive grant funding projects must detail the [ indiscernible ] of their data. Creation and preservation exists within the data lifecycle. In order to preserve the data more easily users must be conscientious about how they create the data. The same could be said about electronic documents with embedded content. Like that we saw from the archaeology data service. Creators have the opportunity to learn about sustainable content creation practices and information stewards. Such as have the opportunity to guide them through the lifecycle creation process. As part of the digital

preservation project the collaboration between Oxford and Cambridge they developed and delivered educational content on digital preservation. Of the publish web outlets introduction to the preservation guide. For students among the workshops here on how to format your dissertation libraries and research departments could support those users in how to create sustainable content in the form of a publication. Okay, I have taken you through way through much information on PDF/A and some potential institutional perspectives. We are aware of the different versions of PDF/A. Those dangerous in the requirements to assess our content and understand the content to mitigate those dangers. What is next? Unfortunately I do not have a solution for you nor will anyone just as PDF/A isn't your silver bullet. Helpfully what you can take away from this plethora of information on PDF/A 's information related to PDF/A and research institutions. And actual files in the real world. Rather Jen just saying PDF/A is good for this particular version of PDF/A is good. It's important to have institutional policies for preservation requirement. To understand your content as Jessica noted earlier. And have a way to identify risks such as automated risks assessment. Rather than just dumping everything from PDF to PDF/A or whatever file to PDF/A in a total migration. Jessica Mensch and the PDF is a great tool for pushing through a PDF/A document to identify what pieces and what components exist within the structure of the file. And from there you can say these particular components such as maybe a pot file, we cannot preserve a CAD file for whatever reason. You can say CAD files are problematic and we can have the embedded in our PDF. That needs to go through a workflow. Maybe you can have JavaScript. That's not problematic for your own workflows. That is something you would allow. Having an understanding of your collections and understanding what your institution specifically considers a risk and what is on your no list and why it is a no. Versus what are exceptions that PDF/A does not actually allow. These checks would help you mitigate preservation risks without causing more preservation issues by imposing a mass migration from PDF to PDF/A. Having an implementation of long-term recommendations for electronic content is important. If it is PDF/A that is fine. If it is not PDF/A is not the silver bullet. It's one solution and maybe it is not even that. I am going to stop there and say thank you. Here are some resources. Also the standards that I referenced. I suppose we will open it up for questions now. Thank you.

---

Thank you Jessica. I have not seen any questions come in yet. Now is the time everyone. Please go ahead and text your questions in the chat box. In the lower right-hand corner of your screen. We do have time for questions. Why you guys are thinking and possibly typing I will ask Ashley to send out the links to the webinar satisfaction survey. She has done it instantaneously. If you have to go please fill out the survey before you have to go. We will give it another minute or two for any more questions that come in. Last chance for questions everyone. To your knowledge how have the UVA and [ indiscernible ] guide [ indiscernible ] creation of PDF?

---

I simply pull the University of Virginia. Guides. I have to double check my references. They require PDF/A Fort Deposit. There are some research institutions which require PDF/A for electronic theses and dissertations. With that said if you are creating a PDF/A in Adobe Acrobat you may not actually be creating a PDF/A. There may be issues that Adobe Acrobat does not capture. In its preflight. Or validation. That is where tools like PDF come in handy for identifying, or validating files as PDF/A. I am not sure about the University of Virginia. To the [

indiscernible ] guide, it is on the greater digital preservation and is a general introduction to internal preservation. I do know some of my research reference has impacted how the [ indiscernible ] library for their institutional repository electronic the season dissertation [ indiscernible ], how they create PDF/A from Word documents. I hope that answered your question.

---

Okay. A follow-up on [ indiscernible ] would need to be part of the student workflow?

---

If an institution is requiring PDF/A but something isn't actually a PDF/A I think there PDF will be hopeful for ensuring that. If you are requiring a particular file type and somebody puts a file extension that is not accurate to the actual content of the file, what is the point of requiring that file format if it is not actually what it is. I would think so. Thank you Catherine.

---

Does anybody else have questions? Thanks for sharing the links in the chat everybody. Will the recording be made available? Yes. We recorded it and will emailed everyone who registered for the webinar. A copy of the slides and a link to the recording. A PDF copy, but I do not know which kind. When you watch the recording the chat will appear. You will have the chat transcript in the recording. Thank you Ashley. She will add the chat log to the archives. Then you don't have to watch the recording, you can download it. The developing your ETE, is that from Virginia Tech and not UVA? Or are we talking about another document?

---

Let's see, I don't remember the last name. Meghan [ indiscernible ] mentioned that the University of Virginia attribution list is incorrect. It should be Virginia Tech. The slide should be updated to acknowledge that. Set my apologies on that typo.

---

No problem. I'm sure we can fix slide deck before we put it up. Any other questions anyone? It looks like we are good. I want to say a big thank you to Jessica and Anna for a great webinar. It's fascinating to learn about experts talk about their topics. We hope everyone will check out past webinars and join us for future webinars. Again, please take the satisfaction survey. Thanks for coming today.

---

[ Event Concluded ]