Transportation Librarians Roundtable-20240125_140435-Meeting Recording

January 25, 2024, 7:04PM

56m 16s

Bob Cullen   0:03

The red light?

Yep, there it is.

OK.

OK, great.

So as I already mentioned, and of course I've already have highlighted this in announcements enough to the day, our speaker for today is Peyton Tvrdy.

And let me start off with some biographical information about Peyton.

Peyton Tvrdy is a data management slash data curation fellow at the National Transportation Library and Peyton received her MLIS from the University of Illinois, Urbana Champagne.

Bob Cullen   0:42

Please feel free to correct me if I'm messing up on any of those pronunciations, but she graduated from there in the spring of last year. Peyton is now very happily, and I'm gonna emphasize happily, living in Brooklyn, NY.

Bob Cullen   1:01

I can say that cause I have some relatives on my dad's side of the family who are from that part of the world, and Peyton is indeed in Brooklyn right now.

Don't let that cactus behind her fool you.

Bob Cullen   1:14

Peyton's presentation today is legacy data rescue and TIUS/VIUS data and I think that all of you are in for quite an informative and interesting presentation throughout a good part of the hour today.

So let me now hand the microphone over to Peyton.
The floor is all yours.

Tvrdy, Peyton CTR (OST) 1:38
Awesome.
Thank you so much.
I'm just gonna go ahead and share my screen so we get started.
OK, we should be good.
Does everyone see my presentation?

Bob Cullen 1:50
I see it.

Tvrdy, Peyton CTR (OST) 1:51
Awesome.
Thank you so much.
So hello everyone.
My name is Peyton Tvrdy, and today I'm gonna be discussing how I was able to rescue hundreds of legacy data tables trapped inside PDFs and how you could do the same at your institution. But before diving in, I wanted to show you all a road map of today's presentation. After today's presentation, you will know what legacy data is what Abby Fine reader is why this work is important. How was able to successfully complete this project and what this project even is and I'll show you how easy it is for you to do it yourself at your own collections.

So what exactly do I mean by legacy data? I have defined legacy data as any data that is outdated, either due to its system, it's format, surrounding technology needed to access it or it's relevance. While this data is outdated, that does not mean that it isn't important or shouldn't be preserved. This data could have significant historical, cultural, or other uses that could help researchers today. Additionally, it is good from a public access standpoint to have as much data possible available, even if it's out of date and could seem unimportant.

Tvrdy, Peyton CTR (OST) 3:02

Some examples of legacy data include old printed reports and data tables, which is the category of my projects, fell under old graphs and source data trapped on legacy formats such as floppy disk raw data that can only be that was only collected on print copies, or an old database that may not be accessible with current database software.

Tvrdy, Peyton CTR (OST) 3:22

So for example, if you have a Microsoft Access database from the 90s, that's not going to be compatible and transferable to modern programs such as my SQL Workbench or Couch base.
So for the purposes of this presentation, just think of legacy data as anything that is not available or accessible to the public in a digital form that you think should be.
So now you're thinking, Oh my goodness, that sounds so great.
Because I have so many documents and files I would like to make accessible that aren't, but I do have a little bit of bad news making this legacy data accessible is not always an easy task.
This process can take a lot of time and knowledge to complete successfully.
However, if you have PDF documents, I have the perfect tool for you for a PDF documents.
This is where Abby fine Reader PDF software steps in Abby fine reader.
PDF Software is an optical character recognition program that could edit PDFs, embed them with text layers to make sure that they are machine readable, and export the contents into other formats.
If that sounds complicated and scary, don't worry.
It's super simple, but how does this actually help you in the public?
You serve well.
For one, it makes PDF searchable and accessible when they are image only PDFs, you can't select, you can't copy and you can't paste the PDF text.
This new text layer makes it so you can copy and paste instead of hand transcribe everything.
This also means that the PDF can be read by accessibility software.
This is not only good for your users, but good for you as a software.
Sorry as this PDF will now be section 508 compliant accessibility user friendly, so we

will also with the software you can also extract different parts of a PDF such as data tables into tabular formats.

So instead of a table being stuck inside of a PDF, it can now be its owned it separate Excel workbook or CSV version open to all the tools and functionality that tabular formats provides, such as tables and charts.

It can also do light editing.

That's not normally possible in a PDF viewer, such as fixing slight skewing issues of PDF pages.

Now it's not a miracle worker, but it can help if you have a poor quality scan and it makes your PDF look a bit more professional in a bit easier to read.

Lastly, is significantly reduces the effort needed to get these results, hand transcribing documents or using free OCR technology is not nearly as efficient as as software and 100% human effort will always pretty much have mistakes.

Stop.

Why is data accessibility important to legacy data if the data is old and not relevant, who cares?

Well, doing this process improves the quality and reliability of legacy data, and if someone today was interested in this report, they would have a more accurate analysis of the data and the document leading to better future use.

Tvrdy, Peyton CTR (OST)   6:09

Additionally, new insights and comparisons can be gained by improving these sets of legacy data.

For example, in my project, if you would like to compare Alabama's results for a specific variable across the years, not only can you now do that easily by extracting emerging the Excel files, but you can make fancy graphs and charts of this data.

Effects will be creating a whole new use and result in this project more than 60 years later.

When making the data transferable and flexible, it also unlocks a whole new world of use in application for modern researchers looking to get insight on old data and continuity and change over time.

Lastly, and maximizes the value of that research output.

While that data is old, it's still be useful to somebody.

That data became a lot less impactful and useful when it was an unusable format.

In my case, an image only PDF now that value of that research is renewed beyond its initial impact and it can still be used and expanded upon, which is not possible with historical data and previous storms.

It also makes them money spent to obtain that research.

Even more worth it so.

Now I'm going to move on to Section 2.

We'll discuss the project I work on, it's outcome and how I got these results.

So the truck inventory in use survey TIUS, later known as the Vehicle Inventory and use Survey VIUS, is one of the surveys included in the Census of Transportation program from the Census Bureau's primary purpose is to collect and publish data on the physical and operational characteristics of the nation's truck resources.

The ties survey was completed in the years 1963, 1967, 1972, 1977, 1982, 1987, and 1992 essentially every five years.

In 1997, the surveys name was changed the virus, and that survey was completed under that new name in 97, 2002 and most recently in 2001, 2021.

Excuse me when it was revived, the goal of these surveys was to collect transportation data not collected by other transportation surveys.

These surveys were released initially and advanced reports, and then later released as a bound volume of each of the 50 states, the District of Columbia, 9 geographic divisions in the United States, and the whole United States for each survey year.

When the census made this publication available online as PDFs, they divided the full report into state summaries, and each of these summaries had the cover, the front part, and the appendices as the Bureau of Transportation Statistics in 2019 began planning for the return of the 2021 Vehicle Inventory New Survey, the National Transportation Libraries Data Services team, was tasked with locating sharing the historic digital data tables and reports from the previous via surveys.

The TTL data services team was able to locate and provide the digital files of legacy bias and tie its data beginning with 1977.

However, the 1963,1967, and 1972 truck inventory and use surveys needed to be dressed as well.

Unfortunately for the data service team, those data tables for trapped in the PDF scans of the original 50 year old print documents, we needed to find a way to make these scans accessible to the public.

We found that our solution was using Abby Finereader software to edit the scans, embed that correct text layer, extract the data tables into a usable tabular format for

the 1963, 1967, and 1972 surveys I have created over 156 PDFs, one for each of the 50 states, one for the District of Columbia and one for the United States 52 a total for each year.

Tvrdy, Peyton CTR (OST) 9:40

Each year yielded a different number of data tables for each of the 52 PDFs, but they added together to make a grand total of 1387 new and modern data tables able to be used in analyzed by the public.

And you can go on rows of P and see them today.

OK, the app you find reader is PDF software with the central part and making the 1963, 1967, and 1972 Tyus reports and data tables successful.

Machine readable and usable.

The process of optical character recognition exponentially speed up the process of fixing these reports and their data.

The software does an excellent job of scanning the characters and overall is very high accuracy.

However, due to artifacts and the scan skews of the text, the type font used and other issues, the software is not 100% independent and still requires human review and action when the software is unsure of a line of text or if a word is misspelled, it highlights that passage for human review in blue.

The reviewer can then fix this text and then move on to the next inaccuracy.

The software not only works with paragraphs attacks, but can also recognize where complex structures such as images and most importantly, tables and forms.

When recognizing tables, Abby fine Reader PDF software can distinguish column headers and row text from the cells of the data.

By straightening the pages before recognition.

This allows the software to better recognize Rosa cells, even for pages that were originally skewed.

And there's no lines between rows and columns as was the case with the tie surveys, as the Times reports from 1963, 67, and 72 are all legacy reports set with mechanical type bounded books and roughly scanned to image only PDFs.

The Abby Fine Reader PDF software does not recognize these complex textual structures such you're tables with 100% accuracy.

Some common issues include not being able to distinguish rows apart, the inability

to read certain numbers such as one and four correctly due to the font misreading.
What is text and what is part of the table and ability to put row lines between rows
that this reading null values as row lines and not recognizing them as values at all are
simply just reading large misreading large sections of text.
This description is nice and very, very wordy, but I would like to actually move on to
showing you how the software works.
So you can actually see how easy it is.
So we're gonna move on to the how it's made part of the sausage factory.
So think of this as more of an interactive cooking show and that I will be like your
Guy Fieri going to flavor town except it's PDF's into Abby Town, which is. That's fine.
OK, so hopefully you still see the left side of my screen.

Tvrdy, Peyton CTR (OST)   12:23
I'm gonna pull up.
This is the 1977 TIUS survey for the state of Alabama, and this doesn't look too bad.
But unfortunately for this year, it's some of the worst skewing I've ever seen on all of
these pages.
And Abby is great, but it's not going to fix all this, and it's not going to read all this
very, very well.
And I'm going to show you a little bit more.
It's kind of a mess.
And then you go to some of these tables.
I mean, what?
What?
What can you even say about that?
That is gross.
So there is a tool where you can correct orientation of all pages and you can see just
kind of do what's best, but it's not actually going to fix.
Everything.
Let's see.
Yep, so it didn't exactly know what to do.
The next step would be to recognize a verify in the OCR editor, but because that
takes a while, I've already done that on the side and as you can see here, Abby
Finereader recognizes 3 different forms.

They also do background picture.

Now that's new, but essentially you have your text elements.

You have your picture elements and you have your table elements and Abby doesn't.

OK, job of trying to recognize what essentially is picture what is body of text which is green and then eventually when we get down to tables they can recognize some tables or a very well.

But for this I'm gonna start with a very simple title page just so you can see how this software works.

So I'm going to drag this picture because that is a picture.

I'm also going to add this as text and then you have to recognize that element and it's stinking hard and it skipped a few words.

So unfortunately for us.

We have to.

Add this text back in, but it's pretty simple as you can see it kind of does focuses it works like a text box like it would in PowerPoint.

Don't worry if it doesn't all fit in the text box.

It doesn't actually show up in the final product.

Going to remove this from pictures and I'm going to add this text down here.

Nor the scramble text.

But you can see it actually does a pretty good job when it knows what it's looking at.

It does a pretty good job and just to show you what an error looks like, let's go over here.

You can see that it's not sure if that space is red clerically, but as you can see here, it is.

So to tell what Peter?

Hey, that's OK.

You can highlight that right click and mark this.

Verified you have to do the same thing for peoples last names because it's not gonna recognize it.

You can also add terms and dictionaries if it is important, and then you can see there's misspelling.

Here I still have my caps locks on, so look at that another page down and this is all fine and dandy.

The text is pretty straightforward.

It's gonna have lots of different errors in formatting for things like this, and we can sit

here and do this all day, but that would take forever, so I'm not going to do that because this part is pretty simple.

What I really want to show you guys are the data tapes, right?

And you're looking at this, you're saying ohh wow. That's not too bad.

This is this is OK, but unfortunately there are some structural issues with this table.

They decided starting in 1967 to do some combined rows.

So you can do different things with the table functionality such as merging these two cells so that would recognize us as one.

Additionally, I actually break up tables like this, so I'm actually going to delete all of this and we're going to make this a table element, and we're gonna make this a table element.

Over here, we're going to recognize both.

Ignore the false text is my version is a poor version now.

Here we go and you can see the tables are now separate.

And there's some artifacting down here that's incorrect do to the.

What do you call those?

You called those brackets.

OK, so the first thing I'm going to do is I'm actually going to wipe out all of these empty spaces and ignore that you can't actually see the text.

If you have a better version ID you can.

But first got married, removing all this empty text and it has to think all of those periods that don't need to be there are gonna get removed and then we can actually look at the table.

Sometimes there's also double periods, so I'm just getting rid of those depending on how it recognized.

All right, so this isn't bad, but do you spot anything potentially wrong?

You notice how there's no line between these two elements?

That's OK, 'cause.

We can go back and fix that for the computer.

So when you click on the table, there is a bunch of different buttons that you have.

You can delete columns.

You can add.

You can add column lines.

You can add row lines and you can even reanalyze the table structure if you think you have the structure correctly, but we need to go ahead and we need to add a row

line, so we'll do that there.

And I believe there's another.

There's two down here, so we'll make sure to add them here.

Here we're going to merge these back two together because sometimes you're changes don't always save between recognizing over and over again.

Yeah.

So another unfortunate part of the program is that if you recognize it will add back those periods because you can't tell the computer never to add strings of periods like that.

But you can get rid of most like that.

So another problem which I talked about a little bit is that it doesn't recognize null values very well.

So you can see here we have 4 null values.

It recognize two of them, so I'll tip my hat to that.

But it didn't actually put them in all the places.

So you do have to watch out for that.

I noticed that there's space is missing here.

Ohh and you can control shift.

You can control to mark things as verified really quickly, so there's more null values here.

I'm gonna add the same null values everywhere so that it looks pretty uniform.

And then we can go on to cleaning up these numbers down here.

We can't tell it to not recognize brackets.

Umm, but we can just get rid of those.

OK, so now we're getting somewhere, right?

So you'll notice that there's a superscript here, and the reason I'm not actually gonna put this in the table is because when you export to excel, you can't have when you have a number format in Excel, you can't have superscripts or else it's a text format and you can't make bar charts and tables and that kind of things.

If it's in a text format, so we have to politely ignore superscripts and then we're going to clean up here, we're going to mark this Texas verified.

Tvrdy, Peyton CTR (OST)   19:42

And this isn't too bad, so that took a little bit of time, but this is one side of the table.

All done, ready to go.

I'm gonna export this page even though we're not done as a Excel sheet.

So you can actually see what it looks like.

It looks a lot better in the viewer than it.

What it actually will look like, unfortunately.

And I have to receive it because it didn't like that, which is often a problem with this program.

Let's try one more time.

OK, just do that.

And I'm going to open it over here just so you can see what a fresh export looks like, OK.

But here is the table I just created.

You can see that it keeps everything from that page, which I'll probably delete and it doesn't quite look that nice.

Tvrdy, Peyton CTR (OST) 20:44

It looks a lot nicer in the viewer, and Abby fine reader, but essentially you have to go in and you have to make everything a number format and space it correctly on the sheet so that it looks correct.

You have to mess with the little the line, the margins and the line to make sure that the borders are correct to make sure everything's a number, and if they're superscripts and footnotes at the bottom, they were actually put at the top and we have to actually add those superscripts back in because Excel doesn't like those.

So instead of doing the entire process, I'm going to show you this is what a table might look like after it's done and it's clean.

So this is essentially where we start and this is what it looks like in the end, after you've made everything times new Roman font, you've made everything size 12 font and you've cleaned up all the stray characters and you have just the data in.

This data is a number format and if I wanted to make a bar chart I could make a bar chart so the next thing I will show you is how this deals with forms.

So at the end of every single project or every single survey, there is the actual survey at the end of it.

And if I show you what it looks like, uh, when Abby recognizes it, just on their own, with no human intervention.

To make this adorable, but as you can see, it's kind of a mess.

There's it's recognizing picture elements down here where they don't exist.

It's making things boxes text.

This is a table.

It's adding more pictures here.

There's also more tables over here you can see is not actually recognizing the text variable well, and when you look over here, it's equally as messy.

So not only are there a bunch of weird stray characters, especially when it comes to things like checkboxes, but the structure is not quite right.

There's pictures here instead of actual boxes of text, and you can see how much cleanup work you have to do.

So Abby is not great at forms, but when you get them correct, eventually with a lot, a lot of effort takes me about 30 minutes per page.

You can make them look really nice.

So not only is this structurally because it is a table the same as the form, so it functions the exact same and it can be recognized in the exact same with the same order, but all the checkboxes are correct, all the text is correct.

You can see everything looks the exact same all the way through, so it is not only correct and the text in that it has the right thing, no straight characters or anything, but actually has the check boxes in each box of the form is an independent box.

Let's takes a lot of time, but that's what it looks like when it's all complete.

And I will now also show you what a project looks like when it's all complete.

Let's see.

Let me pull up, but a pull up 1960 Tennessee, 1967 Tennessee.

Hopefully it lets me in a reasonable manner.

Nope, it doesn't want to.

Let's try.

Let's stick with Alabama.

OK.

There's a bit of a slow program, but.

She's coming together.

So eventually OK so.

Have to add this back in, but essentially this is what one of these reports or one of these uh surveys looks like when it's correctly recognized all the way through.

So you can see how the left looks exactly the same to the right.

Slipper.

But you have all of your picture elements.

You have all your text elements recognized, everything is correct and let me pull up a data table.

So you can see what it looks like when it's completely cleaned like this.

Alright, so on the left, see all the numbers there.

All the texts there and look at that.

So this can be this is the machine readable side on the right you can see how perfect it is.

It's everything structurally sound.

All the text is correct and it makes for a better export into Excel and slightly less cleaning at the end.

So because this is a workshop and a tutorial, I'm actually going to turn it back to you guys if I can pull up.

I teams.

There you go.

And I would like to ask open it up more to a Q&A now because it's more it's better for me to actually know what you guys want to see and what I can show you as opposed to me just showing random features cuz I'm not actually a salesperson for this.

I make no Commission off this product.

Bob Cullen   25:43
OK.

Tvrdy, Peyton CTR (OST)   25:43
Ah.

Bob Cullen   25:44
That, that, that sounds good.

I can go ahead and call on the people I know.

We already have a couple of shows of hand.

I assume that's what you're envisioning for this part of the hour, right?

Tvrdy, Peyton CTR (OST)   25:52

Awesome.

Yeah, it's a bit more of a workshop because it's a bit easier to show the software.

Bob Cullen 25:58

OK.

Truth in advertising.

I that's how I characterize it in the announcement.

Bob Cullen 26:04

So thanks for not making a liar out of me.

Great.

Let me go ahead and call.

I know we.

Like I said, we have at least two shows of hand here and I'm going to.

Call first of all on Michael Nugent.

Go ahead, Sir.

Michael L Nugent 26:20

Alright.

Thank you, Bob, and thank you, Peyton.

I appreciate you showing this.

This is it's basically magic which I find very interesting.

Tvrdy, Peyton CTR (OST) 26:26

Yes.

Michael L Nugent 26:28

I'm going to assume you're able to export a finished product as a PDF.

Tvrdy, Peyton CTR (OST) 26:34

Ohh yes, of course.

Sorry I didn't show that, but essentially if you go to the top you can explore it as a searchable PDF document and what that looks like in the end is that I'll actually pull

up both versions of this Alabama report so you can see the difference.

What's my computer gets its act together?

It doesn't like that I'm running several instances of Abby at the same time.

Uh, which is why it's being slow, cause Abby is kind of a beefy program.

I'm not going to lie to you.

So this is the original 1963 Alabama PDF.

This is taken straight off the Census website.

You can see look you can't copy any text.

Look, I'm gonna control F, and I'm gonna look through the letter C, and it's gonna tell me that there's no letter C because all this is an image only PDF, right?

This isn't searchable the data tables.

If you really want to use them, you go down here and you have to zoom in and you have to copy and paste all these into Excel and then you could start your work instead.

This is what it looks like when it's done and it's gonna be hard to show you this cause it looks literally the exact same because you're not changing how the PDF looks except the skew a little bit.

If there's skew, but I'm gonna type in this first sentence for the 1963 blah blah blah blah blah.

Oh, look at that.

You can actually search.

Oh, look, you can find it again.

You can find it a whole bunch of times because they say that phrase a lot.

I can take this text, I can copy it, I can highlight it, I can go all the way down to these.

Yeah, the table of contents.

Everything about this I if I didn't want to use those excel sheets for whatever reason, I could take all of this and I could copy and paste it somewhere else and then it could be read by accessibility software.

So this PDF was a flat ugly image and now it is interactable.

Prequel, right?

Oh, no, I, well, I don't wanna make those changes, but there you.

Michael L Nugent   28:38

I think you briefly just touched upon it, but one of my biggest concerns is taking an older document and digitizing it in 2024 and then having to apply to it.

All of the section 508 accessibility guidelines, if I if I have a super complex table which I do when you work through the Abbey, does that include if you if you ran your accessibility check on this, is it going to like these tables or is it going to it does?

Tvrdy, Peyton CTR (OST)   28:50
Yes.

Tvrdy, Peyton CTR (OST)   28:54
Umm.
Yes.
Yes, it's all good.

Michael L Nugent   29:11
Ohhh my gosh.

Tvrdy, Peyton CTR (OST)   29:12
Yeah, now I'm not gonna make any, like, huge promises cause you say you have a big data table when the more complicated the data table is, it's gonna take a lot more human effort and you might spend some time really moving.
Like I can go back and show you here like you might spend more time.
Like, really moving these lines re recognizing the table, adding lines over and over again, but it is possible with enough effort.

Michael L Nugent   29:37
OK.

Bob Cullen   29:38
OK.

Michael L Nugent   29:38
Thank you very much.
I appreciate your time.

Tvrdy, Peyton CTR (OST) 29:40
Yeah.

Bob Cullen 29:41
Wonderful.
Thank.
Thank you, Michael.
And I'm Peyton.
You definitely got the attention of the Michaels.
Next we have Michael Molina with a question or comment or both.
Go ahead, Michael.
You're on mute.

Molina, Michael J. 29:59
And.

Bob Cullen 30:00
Or maybe not OK.

Molina, Michael J. 30:03
Great presentation page and this was so cool.
I kind of wanna run out and get a copy just to try it out.
This seems like a lot of functionality.
I was wondering though, when you look at tables and change it to optical character recognition, does it work well with the text to speech software?
Does it change the reading order at all?
Have you found it?
Gets along and the reading and the flow, especially with the tables, is consistent with what you'd find with the remediated PDF.

Tvrdy, Peyton CTR (OST) 30:29
Yes.
Yeah.
So it gets a little bit trickier.

You might have to go back until like Adobe or something and adjust like the reading order, especially when like for example I have like the pages like up here that like have it's being slow again but like double columns and it might get a little confused on where the text starts and where it ends.

Especially when it comes to this next page when I have charts instead.

I didn't really delve too much into this because it's a bit simple on the Abbey side you just have pictures, but this for example is a lot harder to make accessible in this form.

What I did instead is that there's pictures, and then there's alt text to make it more of a description as opposed to like a reading out of every single thing.

Especially because this is actually just all or pictures here like dump trucks and I think factories and bosses.

So that part it gets a little tricky and it depends on how much effort you would like to put in to making sure that everything is 100% accurate, but it can be done using Acrobat.

Molina, Michael J.   31:30

Cool.

Thank you.

Bob Cullen   31:33

OK, wait.

Thank you.

Thank you, Michael.

Any other, I don't say a show of hands just yet, but again, please just do the electronic show of hands.

If you do have a questions or comments for Peyton, I see Roslyn has has something to share.

Go ahead, Roslyn.

Alleman, Rosalyn (OST)   31:50

Yeah.

You mentioned putting things into times new Roman is that is. Is that usually what is that like the standard? Because it seems like a sensor, a font would be more is more accessible and easier to read.

Tvrdy, Peyton CTR (OST)   32:07

That's a fair argument.

That's just what I did with these data tables.

Alleman, Rosalyn (OST)   32:10

OK.

Tvrdy, Peyton CTR (OST)   32:12

That's what, like Jesse and Leighton and I agreed on.

When we transform these, we'd had a set of rules about like how the borders looked and all this kind of stuff, why we did it.

Alleman, Rosalyn (OST)   32:15

OK.

Gotcha.

Tvrdy, Peyton CTR (OST)   32:22

But it was more of just whatever's readable.

Alleman, Rosalyn (OST)   32:26

Gotcha.

Tvrdy, Peyton CTR (OST)   32:26

You can do whatever you want.

Alleman, Rosalyn (OST)   32:27

Ohh, I'm so happy to see all this stuff in a format that's actually usable because this is like a really hot item for people.

You know who need who are in this industry and I'm sure there's a lot of people who really happy to see it being made available again.

So thanks for all your hard work, yeah.

Tvrdy, Peyton CTR (OST)   32:44

What makes me feel happy?

This took many, many months to do so.
I'm glad that hopefully somebody is using it out there.

Alleman, Rosalyn (OST) 32:51
Definitely.

Bob Cullen 32:54
Great.
Thanks.
Thank you.
The Rosalyn and again, just don't hesitate to just raise your hand if there's anything you wanna ask Peyton.
Umm, excellent presentation.
I'll take a I'll take advantage of what seems to be just a, A, some, some reason space here.
Now, I'm excellent presentation, Peyton and I, one of the signs of an effective presentation in my mind is when they're when there are plenty of questions, not because you don't address everything adequately, but because you really, umm, I get a stumped.

Bob Cullen 33:31
I thinking about a lot and wondering about things.
You really them open up that level of curiosity and then I think your presentation has certainly been that so far in this hour and any number of questions I'm gonna ask you all of them, maybe I'll just pick out one here.

Tvrdy, Peyton CTR (OST) 33:42
Thank you.

Bob Cullen 33:49
First of all, just instead of trying to address Davies software in its entirety, I'll just pick out of a certain part because you had highlighted it early on in your presentation.
Just about what?

What? What?

Abby can and cannot do at least struggles to do you had you had cited how it the matter of distinguishing columns rose and so forth.

I know I wrote something else down, but that part of part of your presentation, we highlighted that, that that's a challenge for heavy software and that's something that has to be like you said that the human eye and human intervention really have to come into play.

It's definitely not a passive process.

I think that helped underscore for me and in terms of that example with the with the columns and everything, and in terms of.

We're working to make them as functional and accessible as possible.

Did you find that?

Did you find that fairly easy to do?

Was it in an intuitive process or something?

You could pretty much were able to figure out what we can get done it in a quicker fashion.

And I say that very mindful that unlike me, you, you are indeed very tech savvy.

But was the Davy.

Did Abby allow you to.

To grasp that pretty quickly and.

If not, why?

I mean, in other words, what was the process a lot easier maybe after the initial timer too?

Tvrdy, Peyton CTR (OST)   35:28
Umm yeah, I would definitely.

I definitely don't want to come across like an Abbey salesman cause I again, I make no Commission off any of this.

Bob Cullen   35:34
So yeah.

Tvrdy, Peyton CTR (OST)   35:35
It's a really good program, but it does have a ton of issues.

You can see like when I was doing the find and replace for lots of things like all of my

characters are broken and they show as this weird like wing Dings of an A or I just can't see text at all.

Bob Cullen   35:46
You.

Tvrdy, Peyton CTR (OST)   35:49
I know what I'm looking at because I've done this for hours and hours and hours, but there's lots of bugs in the Abbey software and I've spent a lot of time with the customer service team emailing the back and forth is being like hey, this is broken.

Bob Cullen   35:56
Yeah.

Tvrdy, Peyton CTR (OST)   36:01
Hey this is broken, but as far as like describing like how like my learning curve. Essentially what you're asking with this kind of stuff, the first PDF I did, which would be 1963 Alabama, that one, I wanna say it took me about 2 1/2 hours to go all the way from beginning to end.
Now with the towards the end of the project, and now that I'm starting 1977, hopefully next week.
Uh, it takes me about 45 minutes.
So you do get better with time.
You do learn more.
There's certain things that you can do, like for example, I use a lot of hotkeys like when I talked about control taking instead of right clicking to recognize again or right clicking to mark as verified.
There's lots of different hotkeys you can add.
I add one control Alt Queue because I merge cells a lot and you can just kind of get into a workflow like that.
Just merging a bunch of cells, you do have to get a really good Eagle eye is the program does really mess up.
One of the biggest problems I've had is with where the lines go, and for a lot of tables upon their first recognition it puts lines in the middle of words and which means you have to sit there and hand adjust every single line, not even a centimeter

down, just like the tiniest little pixels down and then 10 minutes later you could recognize the table and you hope it's a lot better.

But it definitely takes a lot of effort.

It takes a lot of knowing when you just need to erase the table and start all over again because the computer isn't doing it at even like even a little bit.

And I will actually pull up.

One of my I'm gonna try and zoom into my poster just so I could show you like some of what because I didn't actually, I do have these photos on my computer, but they're just not very good quality photos.

Tvrdy, Peyton CTR (OST)   37:53

So, like I'm gonna pull this over here, but one of the biggest problems I had is this problem here.

Here it actually recognized like, hey, we're not really sure if this is 700.

You should probably go look at this and it's not it's 100, but both the time completely.

Those numbers would pass and I would just have to know.

Oh, any time I see you 100.

I have to go back and really make sure it's 100 and make sure the 700 because it would never be a 700 is not there.

That happens a lot with fours.

It happens a lot with the letter L.

You just kind of pick these things up over time.

It's definitely a learning curve, but it's worth it for the end product that you get.

Bob Cullen   38:31

Great.

Bob Cullen   38:32

That.

Thank you very much.

And I do want to emphasize what?

What?

What you said during your presentation just now in addressing my question that you, you, you are not it was here to, to promote Abby software and that kind of thing.

And I think I think you are demonstrating that very effectively.

I'm just called you, Abby, demonstrating that very effectively, Peyton.

Because you've made it clear about all the advantages and quite a few of the strengths of Abby software, but you also don't hesitate to let us know where, where, where there might be room for improvement, where there are bugs that she stated and that kind of thing.

In other words, you do.

You're doing the very balanced job.

It's not Abby.

Software might not be the be all and end all it.

It's clearly it's proven to be helpful in a lot of ways.

Wait for your purposes, but.

But I do appreciate the balance you're bringing into that it I I guess for what it's worth, I just wanted to assure you that you're coming across that way.

Now if I turn on the TV later and see you on an Ivy software commercial, I'll have to take that back.

Tvrdy, Peyton CTR (OST)   39:37

You'll see me on HSN, just marketing software.

No, I hopefully not.

Bob Cullen   39:40

That's that Home Shopping Network.

Tvrdy, Peyton CTR (OST)   39:42

Yes.

Bob Cullen   39:44

OK, great.

I did have a question, however I am gladly going to see my time to Michael Newton.

Go ahead, Michael.

Michael L Nugent   39:56

Peyton, you just mentioned a few minutes ago about you contacting Abby.
Fine reader customer service to ask them some questions and perhaps solicit a little bit of additional help outside of contacting customer service directly.
Do they have other types of documentation or help such as tutorial videos available?

Tvrdy, Peyton CTR (OST)   40:18

They too, and they also have like essentially like forum post.
I wanna come for and post like what would you call it?
Like you know, when you're are running into a problem with Microsoft and that you like a type of question and then like the little Microsoft representative, like those kind of like, I guess that would be a forum, I don't know.
Essentially, one of those they do have that and they do have some documentation, but here's where the not salesperson thing kicks in.
Again, they're documentation isn't always correct with every update, and so that was one of the biggest problems that I ran into and this, uh, there is support out there, but there was a couple of times when like for example, what was it when I was working on the forms?
So I had this problem where so the last like I don't know the appendix of the report is the same for every single state for that year.

Tvrdy, Peyton CTR (OST)   41:08

So I was trying to save time and I did figure out how to do it eventually by just doing that once and then just adding it to the end of the PDF.
I ended up having to do that with Adobe and not with Abby because when it would do that it the whole program would break and it would give me warnings.
And if you like, hey, we just can't even recognize this page.
We don't even know what page 57 is like.
We it would just do this and it would completely shut down and I had to go to customer service and they're like, no, that's so weird.
And that's kind of where it left off.
So there is support, but there's definitely there's going to be some things that aren't ever fixed and you're going to have to find some work arounds, especially because they've moved on to Abby 16, which is a really buggy mess.
Like I use 16 for a couple weeks and then I went back to 15 because sixteen was

impossible to use.

So keep that in mind.

They do have a support team.

They are working out bugs.

Will they work out every bug? No.

Bob Cullen   42:01

Yep.

Hey.

Michael L Nugent   42:04

Alright, thank you very much.

Appreciate it.

Tvrdy, Peyton CTR (OST) 42:06

Yeah.

Bob Cullen   42:07

Great.

I thank you again, Michael.

Much appreciated.

I'd like to call next on Connor Campbell with a question.

Go ahead, Connor.

Campbell, Connor   42:17

I Payton.

I'm just curious.

Tvrdy, Peyton CTR (OST)   42:18

Hi.

Campbell, Connor   42:21

Eh, we do a lot of converting CAD to PDF.

I was just wondering if Abby supports that.

You know when we convert via AutoCAD or even Acrobat, we deal with a lot of

flattening issues or it might take, you know six hours for 100 sheets of building drawings or anything is is it quicker with Abby, does it less flattening issues or?

Tvrdy, Peyton CTR (OST)   42:33
Umm.
It does have flighting issues and I wouldn't say it would be quicker from what you're describing.
For example, my computer at 16 gigabytes of RAM.

Tvrdy, Peyton CTR (OST)   42:52
I really need like 32 to get this program working because you can see it's kind of breaking now that I have like two or three windows of Abby open and it gets even worse than I'm trying to make complex things or what my documents pretty big or my computer would stall like when I was doing those form appendices at the end.

Campbell, Connor   43:00
Uh, yeah.

Tvrdy, Peyton CTR (OST)   43:11
The only way to do it is to add lines for every single one and then add column so every single one and then merge the rows a bunch of times, merge cells over and over again and that would break my computer all the time.

Campbell, Connor   43:19
Umm.

Tvrdy, Peyton CTR (OST)   43:23
I would essentially have to merge a cell, get up, walk away, go get a snack, come back, and then maybe in 5 minutes that sell might be merged.
So I wouldn't say it's really good about handling that.
It takes up a lot of RAM, but it's worth a shot.

Campbell, Connor   43:37
Gotcha.

Yeah, we're at the try.

Thank you.

Appreciate it.

Bob Cullen   43:43

Right saying.

Thank you, Connor.

Again, we, we still have some time left in the hour.

So any other questions or comments?

And so far the questions have been wonderful.

So please don't hesitate to raise your hand there if there's anything you wanna ask Peyton or share with her.

Look again, as always, I have a number of questions prepared I'll ask you something.

That's more of a non-technical inquiry but I think you're I'm not to put added pressure on.

Yeah, I think you're more than up to the task of addressing it.

Alright, now you wanna hear it?

Umm early on in your presentation you really gave.

Tvrdy, Peyton CTR (OST)   44:27

Yeah.

Bob Cullen   44:32

What I thought was the best.

I wouldn't say defense, but certainly the best.

Umm.

Explanation or reasoning for legacy data why do we need it?

I mean there there's almost too often beyond librarians and historians and some other select group of individuals, there tends to be maybe this out of sight, out of mind thing.

You know that that's ancient history.

The numbers and everything, the statistics you know, you know a lot of people might not even be aware of it or any anything like that.

Other people might say, well, we don't, we don't really need to worry about that.

That's in the rearview mirror, but as you point out in your legacy data, Bill has it

significance and it still something that, that, that can be sought over and can be very, very important too and you know and then and in other words I think you did a good job outlining the need for having that that accessibility.

I think if I recall correctly, even did kind of an ROI.

Argument if you will.

There that that, you know that there was money in.

If I'm phrasing that correctly here, there was money involved and coming up with that published information in the 1st place.

I've documented information in the 1st place you know, and that how that helps.

Just to justify that too.

As far as legacy data that goes though, I don't know if you'd wanna expand further on the explanation or.

Or maybe I can phrase it this way?

Maybe a little better with.

Was there any light bulb over the head moment when you realized that that legacy data really it really is something that that that's worth preserving and pursuing?

What was it like in terms of maybe during the process that we've been doing with Abby software, but even before that suspect he probably had something about legacy data that resonated with you.

I don't know if you'd like to share anything about that.

Tvrdy, Peyton CTR (OST) 46:26

Yeah.

Before this I was the archivist at the Phoenix Art Museum, and I had to fight tooth and nail every single day to preserve legacy like formats.

For example, I had a bunch of floppy disks that had one of a kind word doc.

Forget The **.

Was it invented yet at the end of that file format, we've already talking the original like 1996 word documents, and I had to fight a lot with administration like over and over again and be like, hey, look, I know it's inconvenient.

Yeah, I know someone printed this out in 2005 and I have a paper copy right here.

But hey, this 1996 word Doc is the original version and I need to preserve that.

I need to make that accessible.

It needs to be digitally available for people because that's how you get access out

there.

So I've kind of always been passionate about older technology and then also making old things usable, especially when it comes into the digital sphere.

I don't know if that answers your question, but hopefully it did.

Bob Cullen 47:26

It does, it does and then it certainly clarifies what but what my starring assumption had been anyway.

That that wasn't like David handed you a folder one day and say, hey, work on this that you came into.

Bob Cullen 47:40

NTL already with that experience and also with what is clearly your strong sense, your strong feeling that you know legacy data is important.

I mean I, you know those kinds of items do need to be to be made more readily accessible.

So you did address my question there.

So thank you for that.

Tvrdy, Peyton CTR (OST) 48:05

Yeah.

Thank you.

Bob Cullen 48:06

Yeah.

And your reference of Phoenix, AZ did just jump started my memory here that I assume that's probably the significance of having a cactus there in Brooklyn, right?

Tvrdy, Peyton CTR (OST) 48:16

Yeah.

Yeah, yeah.

That's why I got that. I actually have more like cactus things now that I'm away from Arizona, I'd have like a cactus Christmas tree kind of light up ceramic thing.

Bob Cullen   48:27

Ohh.

Tvrdy, Peyton CTR (OST)   48:29

I have a lot of cactus things now, even though I don't live in a desert anymore.

Bob Cullen   48:33

OK.

Well, you still have a bit of the desert there said where at least reminders of it.

Tvrdy, Peyton CTR (OST)   48:36

Yeah, exactly.

Bob Cullen   48:38

Perfect.

And any other questions?

Yeah.

Questions or comments for Peyton?

You know, feel free to do a show of hands.

We got a few more minutes left in the hour, not that we have to take up the whole hour, but, you know, you still have the luxury of a little time.

If there's anything you'd like to share, just in terms of, I'll, I'll go ahead and throw out something.

Umm I'm you and I think you're addressing this in one or the other responses you've given.

I mean, I asked about the tables, the columns, but you'd also made reference to both in your presentation and during the Q&A just a few minutes ago about the images and the text.

And forgive me if this is something you kind of addressed anyway, but.

Is that a?

Is that a persistent problem or has that gotten better over time, just in terms of that?

Because I could see that how that would be caused for major headaches there.

If it's a constant thing it would you characterize it that way?

Tvrdy, Peyton CTR (OST) 49:55

Yes, it no, I will say with at the start of the year, it's always rougher because the computer gets more confused because it hasn't seen it before, but towards the end of the year is starting to recognize patterns a bit better and it gets easier with time.

Bob Cullen 50:04

Yeah.

Tvrdy, Peyton CTR (OST) 50:10

Or maybe as I get more comfortable organizing that year.
But I would say that text problems table problems, they're all quite persistent.
You just get better at fighting them overtime.

Bob Cullen 50:23

OK, great.
So that that, that, that, that definitely answers my question.
That had me curious too, cause this things like that that never really occurred to me, but sure there would be some, maybe some, you know, issues they're trying to trying to distinguish between, you know images and text.
So thank you for clarifying.

Tvrdy, Peyton CTR (OST) 50:41

Yeah, it's a big time sink.
I wouldn't invest in like the software if you're already stretched thin.
If you already have way too much on your plate, which is most librarians all the time, maybe not to this cause it's very expensive and it takes a lot of time and it's a skill set that you build up the better and better you get it using the software.

Bob Cullen 50:53

Yeah.

Tvrdy, Peyton CTR (OST) 51:00

So maybe an intern?

Bob Cullen 51:03

There we go.

Tvrdy, Peyton CTR (OST) 51:03

That's how I started.

Bob Cullen 51:05

That's there we go.

Exactly.

You know, give an intern something meaningful to do as well that that that that that that's great appreciate that we are bumping up a little further against the our here.

Tvrdy, Peyton CTR (OST) 51:15

Yeah.

Bob Cullen 51:20

I don't know if anybody else has any questions for Peyton.

Again, Peyton.

Wonderful presentation and the workshop portion.

Which, of course, pretty much has blossomed into the Q and a portion of the TLR as well.

Every everything has been just great and I do have to say I paid, Mama said this at the start of the hour.

But I'll, I'll, I'll say it now.

I almost definitely appreciate your patience and your persistence and I'm coming on board here today as a TLR speaker, I know for several months now we we've been trying to.

I get you scheduled for this.

So again, again, you, you, you are definitely a model of patience there, which probably is also why you're good at do doing what you're doing now with legacy data and TIUS wise data.

Tvrdy, Peyton CTR (OST) 52:10

Ohh thank you so much for having me.

I really appreciate getting to be able to talk about this.

I felt so bad in November when I wasn't even close.

I still had the entirety of 1972 to go, and I'm like, I can't show up to TLR being like.

Yeah.

And I'm like maybe halfway done with this project.

I don't know.

Like I'd rather come forward with at least a more substantial package you'd like.

I have the experience.

I've done this for hours and hours and hours now, so thank you.

Thank you for being so patient and for having me despite all the scheduling.

Bob Cullen 52:41

Ohh, I'd give you.

You're welcome.

And then the pleasure has been all ours and.

I'm glad you got beyond 1972 there and the other challenges she had had to deal with there and.

And so, Peyton, I assume I or at least I hope that that if anybody has any questions that might pop up in their minds later today are that they don't quite wanna know wanna phrase now I assume that's OK for people to check back with you and follow up with you about that.

Tvrdy, Peyton CTR (OST) 53:13

Yes, you can totally email me.

You can hit me up on teams you can find by LinkedIn.

Tvrdy, Peyton CTR (OST) 53:19

You can do lots and lots of things by last name is very distinct, so as long as you remember most of the letters you can pretty much find me anywhere.

Bob Cullen 53:26

Alright, wonderful.

Bob Cullen 53:28

So everybody can keep that in mind.

And I'm glad you cited email teams and LinkedIn as ways to get in contact with you.

You're not challenging people to brave the traffic in Brooklyn to find you so.

Appreciate that.

Alright, well, again, again, many.

Many thanks, Peyton, and many thanks to all of you who took the time to be here today to take part in this presentation.

I hope you learned something about it.

Big thanks as well to those of you who participated in the Q&A portion, but I hope that everyone really got a lot out of this.

I know I certainly did.

And I was just doing a big a check here.

Bob Cullen 54:13

Well, you certainly brought in quite the quite a large number of people that they paid and I should highlight this that even Rachel Cole is here and Rachel in the previous hour had had had been given her own presentation for the SLA Transportation Community showcase.

So thank you, Rachel, for taking the time to be here so quickly.

But what can I say, Peyton, you are quite the draw today, so.

Tvrdy, Peyton CTR (OST) 54:39

That's awesome.

Tvrdy, Peyton CTR (OST) 54:40

Thank you so much for having me.

Bob Cullen 54:42

Oh, you're welcome.

Bob Cullen 54:43

And I again, David and I really appreciate all of you taking part in this.

The first NTKN sponsored TLR of 2024, the next TLR, as I think I've noted at least a

couple of recent announcements, is gonna be taking place on Thursday, February 15th.

Bob Cullen 55:03

Real quickly, I do appreciate everybody having the flexibility to that to gather together today rather than the 2nd Thursday of the month just because of that was when TRB was taking place.
So thank you for your flexibility being here today since it is the 4th Thursday of Thursday, we're gonna have the next month's TLR on the 3rd Thursday or Thursday just to give a little bit more leeway there.
Some breathing space, if you will, between now and when we all gather together again, but it will be Thursday, February 15th.
It will be our annual Open discussion forum so.
Sometimes what I'll say when we have a first rate speaker like Peyton and I'll say the next TLR speaker is gonna have a tough act to follow.
I guess it's safe to say since it isn't open discussion forum, all of us will have a tough act to follow because we'll all be talking that day.
So a good number of us at least anyway though, stay tuned for further details about that upcoming TLR and also for other TLR's beyond that as well.

Bob Cullen 56:10

And I hope that everyone has a great rest of the day and a great rest of the week.