# PDF Pit Stops in Library Platforms: An Essential Feature for Blind Users

Presentation at *Access*

Conference hosted by Carleton University, Ottawa, Ontario

October 12 to 14, 2022

## Presenters

Ashley Shaw ([shaw2610@mylaurier.ca](mailto:shaw2610@mylaurier.ca))

Mark Weiler ([mweiler@wlu.ca](file:///C:\Users\mweil\Documents\mweiler@wlu.ca))

Bart Kawula ([b.kawula@utoronto.ca](file:///C:\Users\mweil\Documents\b.kawula@utoronto.ca))

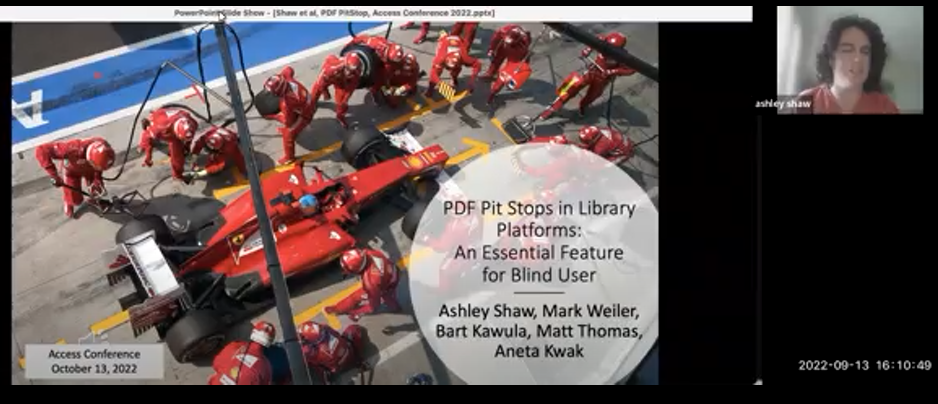
Matt Thomas ([mthomas@wlu.ca](file:///C:\Users\mweil\Documents\mthomas@wlu.ca))

Aneta Kwak ([aneta.kwak@utoronto.ca](mailto:aneta.kwak@utoronto.ca))

October 13, 2022



## Introduction: Ashley Shaw



[Video clip](https://youtu.be/0uC-arfXCmQ?t=8)

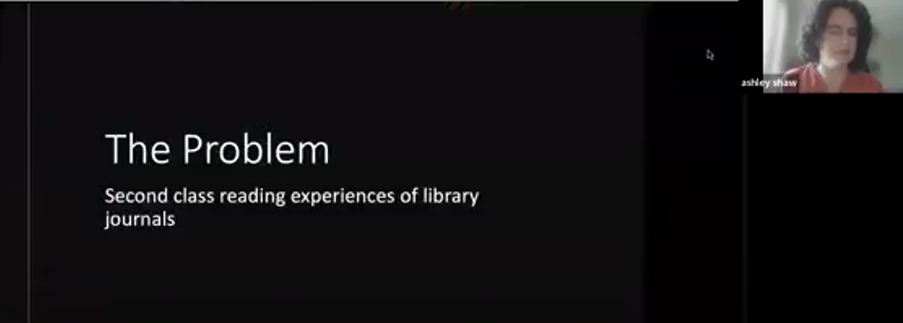
Good afternoon. everyone. My name is Ashley Shaw and I'm a master student in the Community Psychology program at Wilfrid Laurier University, and welcome to our presentation entitled PDF.

Pit stops in library platforms and essential feature for blind users.

With me today are Mark Weiler, Bart Kawula Matt Thomas, and Aneta Kwak

And to assist those who are blind this afternoon you'll hear a click that marks the transition between slides, as you heard just there.

## The Problem: Ashley Shaw



[Video clip](https://youtu.be/0uC-arfXCmQ?t=48)

So I love libraries, and I love doing research. But every time I open new journal articles, I feel a sense of trepidation about what I’m going to get.

Many journal articles are only available in PDF format. and as a blind researcher using a screen reader, the PDF Format presents me with a number of accessibility barriers. The creators of PDFs typically do not verify accessibility using a screen reader, which means the first people who experience the PDF.

Using the screen reader are usually end users like myself. I tend to waste a lot of time during searches, simply checking all available copies of an article.

To figure out which one's the most accessible. I have to do this, since there has essentially been no proofreading of the layer of text.

My screen reader has access to. In many PDFs, the layer of text accessible to screen readers is full of spelling errors, missing spaces between words.

Has unclear reading order, contains no descriptions for images and lacks identifiable headings, links, tables, footnotes, references, and page numbers.

Barriers like these create a second-class reading experience for library patrons who use screen readers. Instead of focusing on comprehension or making connections between content and concepts, We're distracted by missing or incorrect content. We can't skim through long articles when properly formatted headings are not used. And statistic and scientific symbols are often not available or are incorrectly represented, which deprives us of the same opportunities to interpret data as our peers who are not print disabled.

We also can't be certain that the text we're quoting directly is actually accurate, and cannot create proper citations without access to page numbers.

The good news is that it is entirely possible to generate content that is fully accessible to screen reader users.

This is known as born accessible content, a term you may have heard before. However, retrofitting or remediating existing files that are inaccessible is time and resource intensive, especially in the case of PDFs. For my master's thesis I am conducting a scoping review, and we have needed to design a remediation workflow for the 215 articles eligible for full text screening. Inaccessible journal articles have resulted in additional time and financial costs throughout the project. So I have to wait for accessibility remediation before screening full texts and pay for things like extra file storage. As a graduate student I've learned scoping review protocols from my mentors, but have needed to spend a lot of time and energy helping design a workflow just so I can read the same journal articles anyone else can read without a second thought. we're going to discuss the components of this workflow with you today. Given what I just described, the fact that publishers and vendors continue to provide inaccessible copies of content is frustrating and disheartening. The fact that I as an end user am expected to identify inaccessible content and request remediation indicates that the system is designed to offload barriers such as these onto patrons. Continuously requesting assistance with remediation requires a good deal of emotional and explicit labor on the part of patrons with disabilities, and this labor is not required of patrons without disabilities. There are so many accessibility barriers with library reading materials that I and many others prioritize which we request help with while giving up on others and living with them.

I'm incredibly grateful to the individuals who have chosen to accompany me on this journey, and most of them are presenting with me here today.

Together we are solution focused, innovative and creative, and this collaborative focus on solutions is what I call for from others within library systems.

## Designing a PDF Pit Stop: Mark Weiler

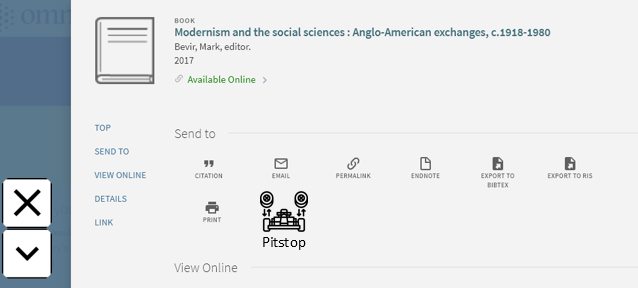


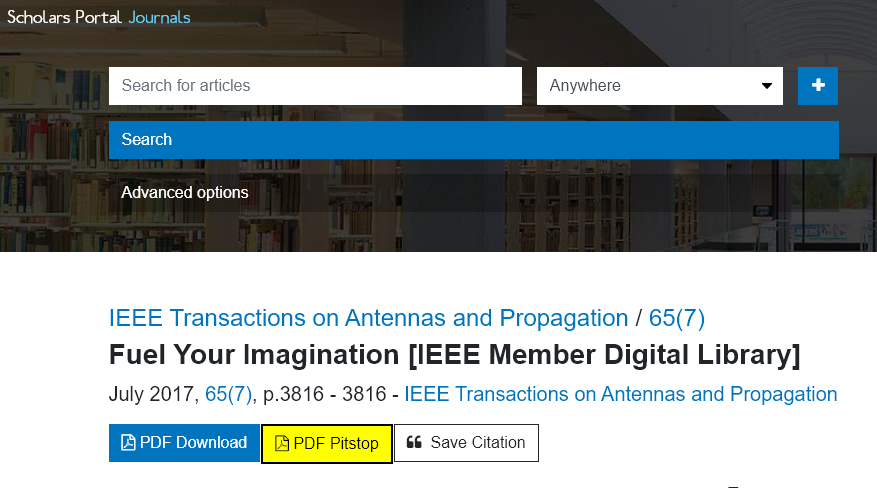
[Video clip](https://youtu.be/0uC-arfXCmQ?t=283)

When Ashley met with me to get ahead of the obstacles on our path I became aware of the problems Libraries were presenting to blind screen reader users. The quality of journal articles and their discovery were just 2 of them.

Although we talked about the solutions for her, Ashley also wanted to think beyond any band-aids and address root problems, so others won't have to go through what she is going through.

So we imagined the PDF pitstop integrated into discovery layers, as these images show.





Next to the download PDF link is a pitstop link, which sends the article to a pitstop where someone finds an acceptable version of it or transforms it into whatever format the print disabled reader needs. Our design has assumptions. First, a powerful creative force in the world is toppling dominant structures to make the gifts of disability be part of our lives.

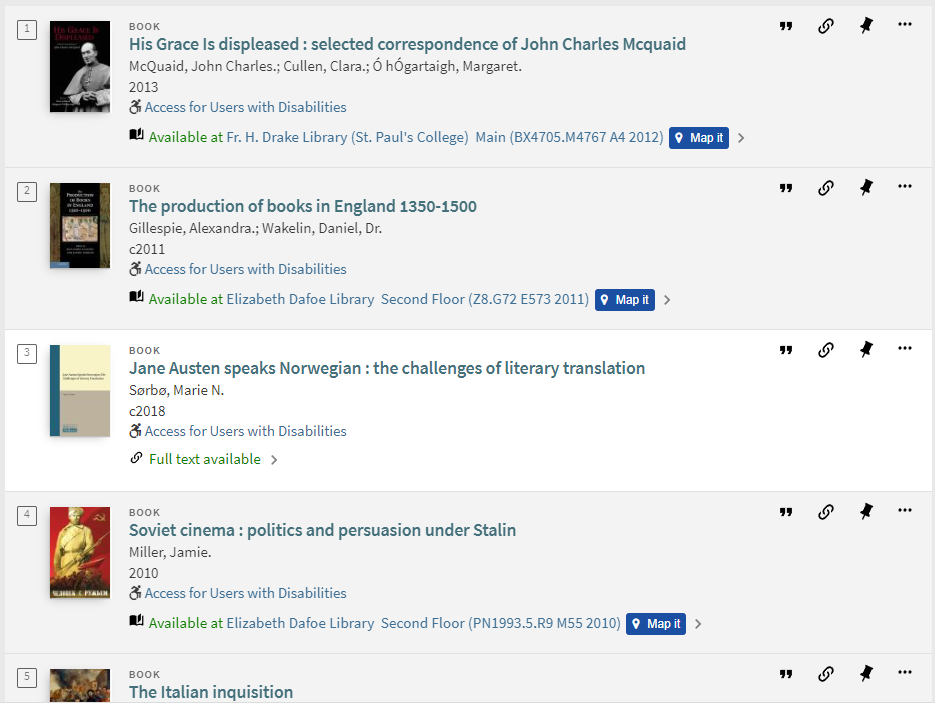
Second, the legacy of disabled activists have been clearing the path for us. Third along these paths resources have been left for us. And finally, we will meet others, starting at different places but walking in the same direction.

This presentation illustrates that others have adapted the routines and joined us, offering their talents to stop the unfairness.

Matt Thomas at Laurier University, Bart Kawula, Scholars Portal, and Aneta Kwak, at New College, at the University of Toronto.

And from around the world others are joining our network, and you can too. We've also found examples of libraries doing something similar, as these images show.

McMaster University, Osgoode Hall Law School, and the University of Manitoba have integrated a “request accessible copy” link into their library platforms. If you know more, please tell us.

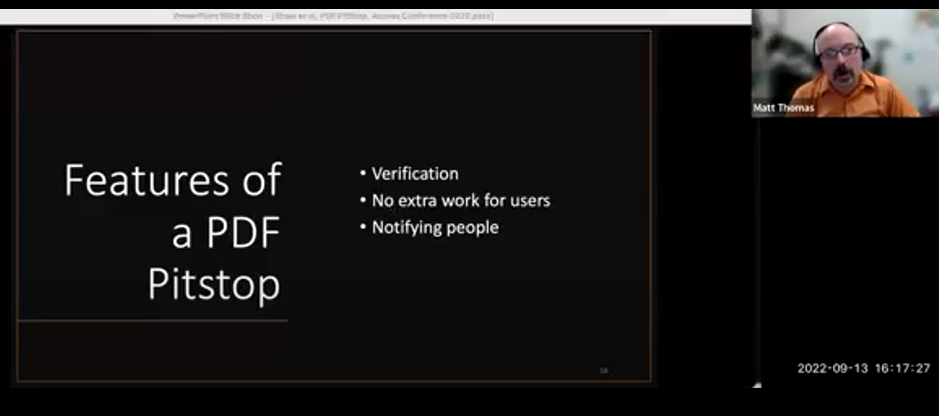


Library Record showing Request Accessible Copy
A library record from McMaster University's old library catalogue which shows a clickable link on the right hand side which reads "Request Accessible Copy"



But adding a "request accessible copy" link in a discovery layer or library platform would risk being a band-aid to a deeper wound that the legacy of inaccessible journal publishing, procurement, and discovery has created.

## Features of a PDF Pitstop: Matt Thomas



[Video clip](https://youtu.be/0uC-arfXCmQ?t=403)

There are 3 main features of our PDF pitstop.

One: verification. It is crucial to verify the quality of PDF documents before they're provided to a blind patron.

Verification of the quality of these documents is done for cited users by almost the entire public and provision process.

But we know that there are common, and often in surmountable barriers for blind users in these documents, like the ones mentioned by Ashley earlier that need to be at least identified, if not resolved.

Number 2. No extra work for users. Blind patrons already face multiple barriers and, therefore, extra work when finding, collecting, reading, and organizing their research material.

A key goal of our PDF pit stop is to remove as much additional work on their card as we can.

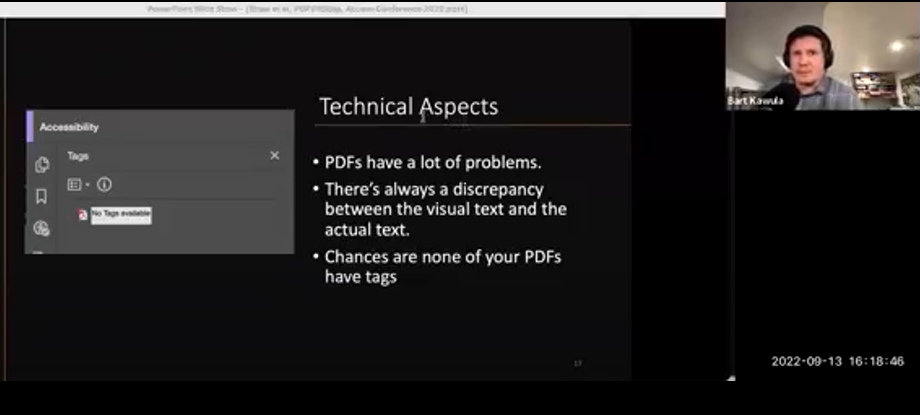
Number 3: notifying people. It's important to ensure that not only is this work done, but that more than just the people directly involved or informed about the issues and the resolutions.

For example, if Laurier, is encountering a problem with inaccessible content in one of our resources, other institutions in OCUL who also have that resource should be notified.

Their patrons are almost all certainly going to encounter the same problem as ours have; or if there's work being done supporting a blind student, a certain subject area, the subject or liaison librarian should know so that they can provide additional assistance, if necessary.

And since we're usually paying for this content, those in acquisitions in the library should know that this work is having to be done on top of what we're paying. It's an additional cost that should be considered.

## Technical Aspects: Bart Kawula



[Video clip](https://youtu.be/0uC-arfXCmQ?t=492)

In addition to managing the workflows associated with remediation, request a large part, and perhaps the biggest bottleneck of the pit stop is file format authoring.

This means transforming PDFs into other formats based on a patrons needs.

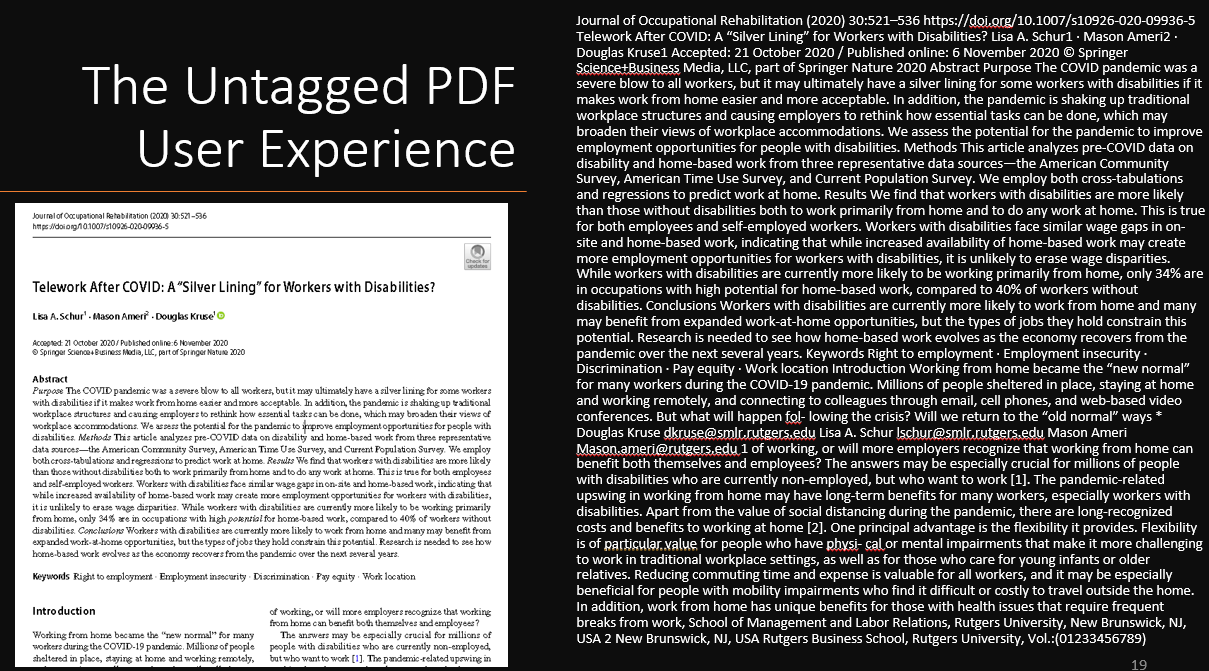
In many cases, a properly tagged PDF document is sufficient, but unfortunately, the vast majority of PDFs are not tagged.

PDF is a visual format created for print media, and so it can be difficult to offer the level of accessibility provided by a WCAG conforming web document.

This is why HTML and Epub are rapidly emerging, as the de facto, accessible formats.

Markup tags like headings, lists, tables, and landmarks were specifically designed with the assistive technology in mind and there's a greater level of certainty that what's shown on the screen is what will be read by a screen reader.

On this slide is an image of an article in each HTML.



With clearly defined headings which someone using assistive technology can use to quickly navigate or jump through the document without having to use any extra effort to figure out what's being read.

For fully sighted readers, visual cues or design patterns are what give text structure as shown by the image of the PDF on the left.

But without proper tagging, the same text is experienced like the solid block of text on the right.

Sadly, this is the default experience of almost all PDF based library materials by people using screen readers.

And this creates a lot of unnecessary cognitive load not experienced by fully cited users.

Based on our own work thus far, the only way to properly tag a PDF document is to do it manually, and given the amount of content being published, it's easy to see why most publishers are referring people to Html as the accessible copy.

The graph on this slide shows the percentage of full text Html available for all else of your content in Scholars Portal Journals.

Image on the slide. Articles with full text html or xml across all Elsevier collections in Scholars Portal Journals. The key point is that the percentage changed from about 10% in 2007 to almost 95% in 2022.

Data available at: https://nextcloud.scholarsportal.info/s/g7idKfR6rCEEoKe


It was below 10% in 2006 and jumped to about 80% by 2008.

Today, it's about 95%. Nevertheless only 20% of our total collections and Scholars portal journals has full text Html. and reflects the larger publishing landscape. Hence the need for a PDF remediation plan.

Also publisher provided HTML and Epub are not necessarily accessible.

There are often issues with missing alt texts for figures and missing page numbers, but correcting us are a lot easier than modifying a PDF and having html version gets us almost to the finish line.

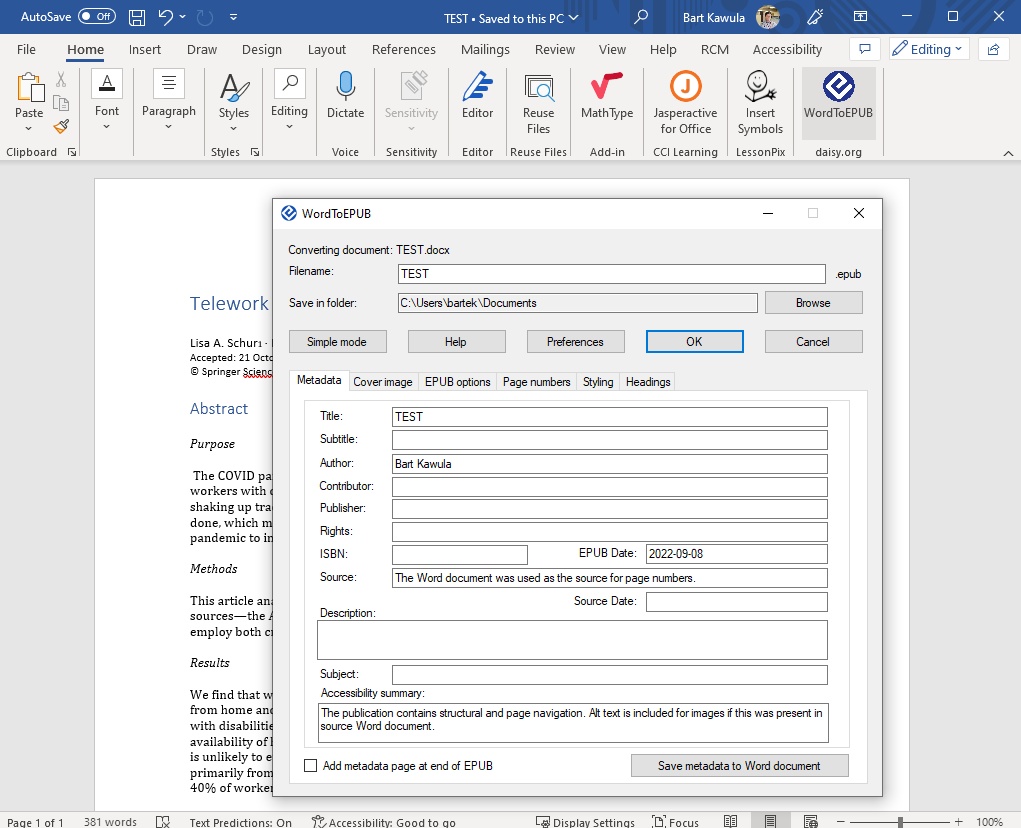
But what do we do when the untagged PDF files all we have?

If a vendor can't provide an alternate format in a timely manner, we need to either tag the existing PDF using adobe acrobat pro, orin cases where PDF tagging doesn't mean a person's needs rebuild the whole document from scratch

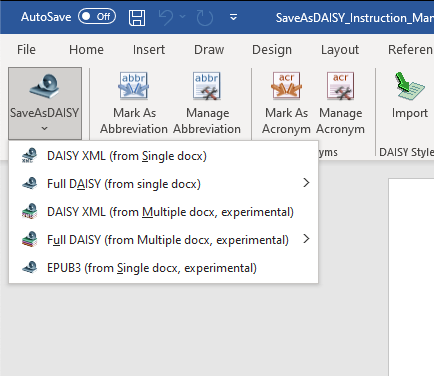
We're still figuring out the best way to do this, and we've evaluated quite a few authoring tools in the process.

At the moment, we've mostly settled on Microsoft word as our primary authoring environment.

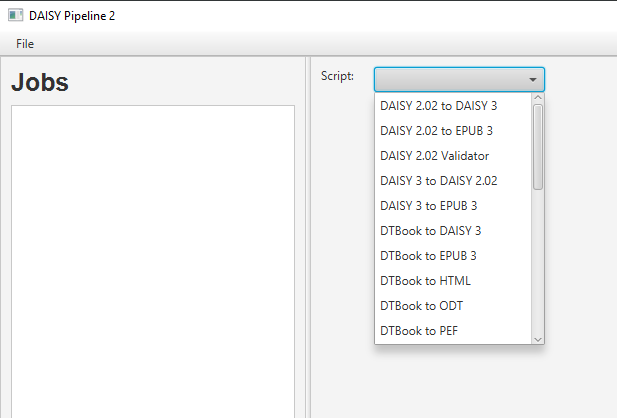
It offers the most features for creating an accessible document that can be exported and converted into formats like Html, EPUB, or Daisy. The Daisy consortium, who officially maintains the Daisy standard also maintains a variety of tools for authoring. Both daisy files and EPUBs directly in word. This first mage on the shows the WordToEPUB extension inside of Word.



The next image shows the save as daisy extension.



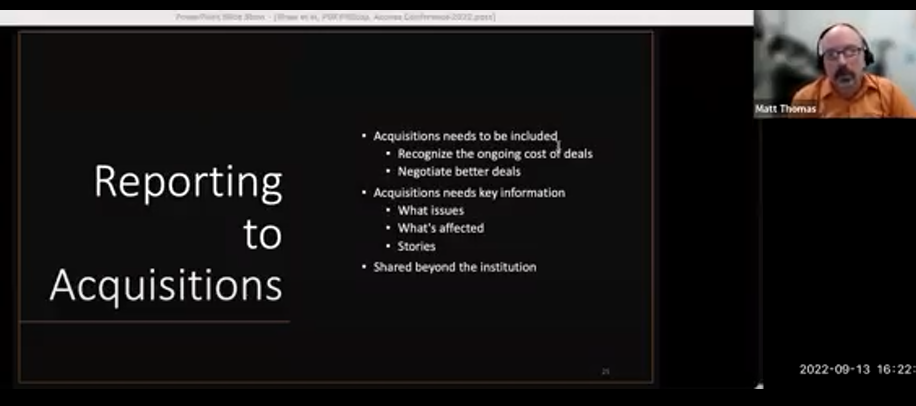
Once a Word document is exported to either format, it can be converted to other accessible formats, such as a Braille Ready File through the Daisy pipeline tool.



So to summarize. We're traveling down well-worn paths. But we still need to.

We need to find 2 in our methods and workflows to make this part of the pit stop as fast as possible.

## Reporting to Acquisitions: Matt Thomas



[Video clip](https://youtu.be/0uC-arfXCmQ?t=723)

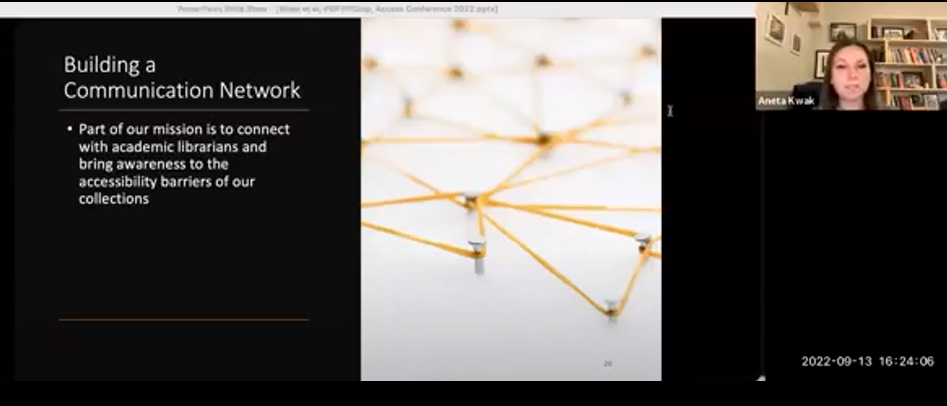
As previously mentioned, it's important to include those responsible for acquisitions and collections somewhere in this process. Not only is it important for us to recognize that the work that must be done to make a acquired material accessible to our patrons is an ongoing additional cost of those materials, but if these issues are never shared with those negotiating deciding on products new or renewing the can't be considered in that process, and it may be difficult to ever see any improvement. Although more information is usually better it's not necessary for every little detail to be shared. Decision makers need to know the overall nature and scope of the problem and the solution.

For example, what kinds of accessibility issues are appearing most often, and how are they being resolved or avoided?

What resources are platforms are being affected the most, and therefore for which deals are we paying good money only to have to fix before handing over to some of our most already overburdened patrons? And this data is good but minds aren't changed by numbers and lists alone, but through stories.

So someone understanding of the specific issues that blind patrons face, if, or rather, when this work isn't done = issues like the ones that actually shared with us earlier. This is important at the local institution level, could even be more powerful and impactful if shared between institutions, so that we can all include these concerns that are negotiations and decision making and compiled and used in broader consortium negotiations, where more people and more money is on the line.

## Building a Communication Network: Aneta Kwak



[Video clip](https://youtu.be/0uC-arfXCmQ?t=818)

In addition to reporting, to acquisitions librarians, we also need to make sure librarians and library staff who are on the front lines and engaging with our end users are aware of the issues some of our patients experience when accessing library materials.

Part of the PDF Pit Stop mission is to build a community to help tackle the issue of accessibility barriers in Academia.

We want to connect with and bring together subject librarians from all disciplines to create a network of communication.

We want to notify librarians in this network each time a journal article from their subject area arrives in the pit stop

The idea of this notification is to continually bring awareness to the issue, so that it does not fall into the back burner, but is instead front and center in how we approach our work from our perspective. The more people who are aware of the accessibility issues with our materials, the more chance that libraries will be able to catch these accessibility issues before they reach our end users. We want to minimize and eventually eradicate situations like the one Ashley described, where our blind patrons have to spend their valuable time locating accessible copies of journal articles.

To do this we need to work together to bring an end to the barriers of experienced by our users. In addition to notifications, we are also meeting with subject librarians. At this time we have met with a group of 3 Canadian liaison librarians, and informed them of the work we are doing, because we received a number of requests from their subject areas.

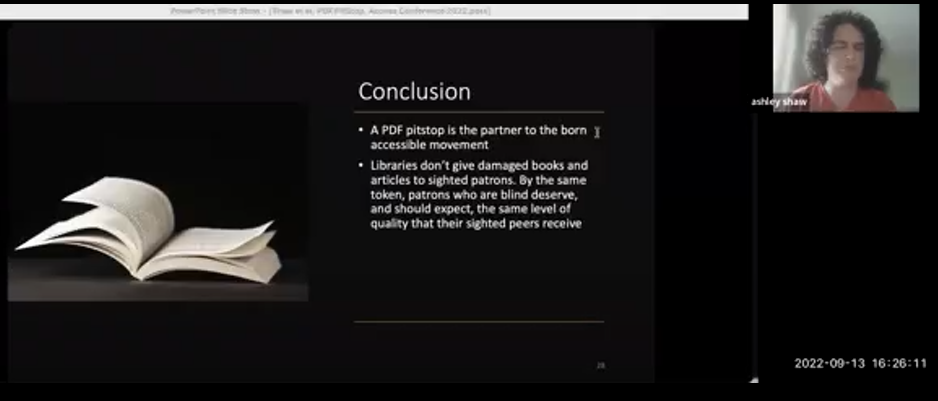
Awareness is one step, but we also want to expand our team, to have more people available to help.

We hope to add more members to our pick crew to be available to test, review, remediate, spread awareness, or to assist in other capacities as needed.

If you want to get involved with our pitcrew we're stay connected with updates about the PDF pit stop please fill this form and select how you are able to participate.

The link is: <https://bit.ly/PDF-Pitstop-Network>

## Conclusion: Ashley Shaw



[Video clip](https://youtu.be/0uC-arfXCmQ?t=940)

A PDF pitstop is the partner to the born accessible movement. It recognizes the inevitable need for document remediation to address the legacy of born in accessible journal articles.

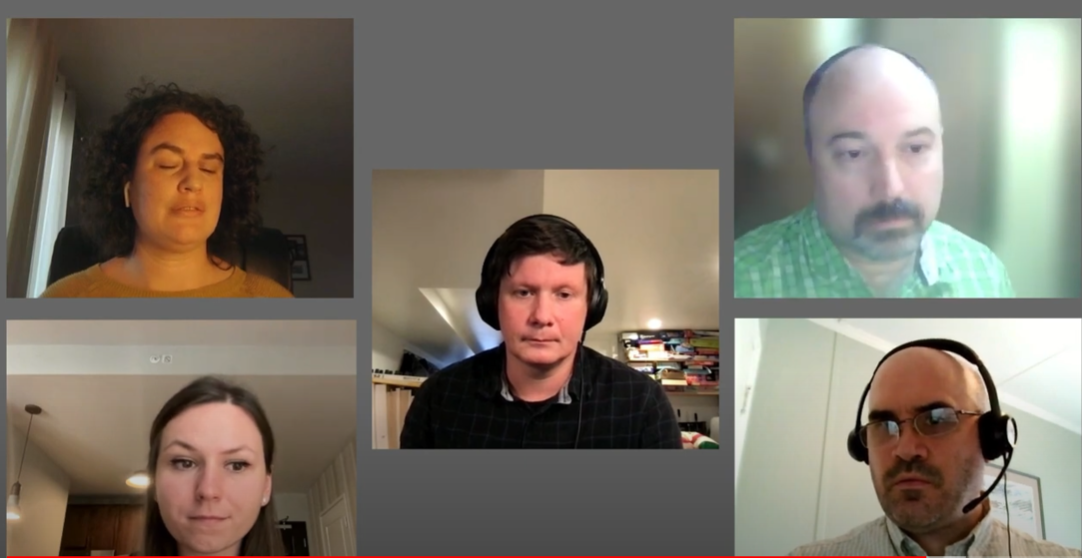
Libraries don't give damaged books and articles to sighted patrons. By the same token patrons who are blind deserve, and should expect the same level of quality that their sighted peers receive.

**I have a message for librarians:**  accessibility is part of everyone's job. If you do not have a disability right now, if the word accessibility isn't in your job title, you're still urged to join the movement toward making information truly accessible to everyone. Each of you has some degree of influence skill and resource you can bring to the table, so consider this a call to action.

**I also have a message for blind listeners:** The library is a place for you. Whether you are a student, community member, researcher, or scholar, you have the same right to access librarians, library resources and library materials that your sighted peers do. You may access all of these things in different ways from the typical but curiosity and innovation are highly valuable skill sets that I encourage all of us to embrace. In this era and increasing universal design, I encourage us to make a home for ourselves in any and all libraries.

## Questions and Answers





[Video clip](https://youtu.be/0uC-arfXCmQ?t=1037)

**Moderator**

Okay, do we have any questions? Someone is asking can you please show the link to the network again?

**Mark:**

Yeah we can do that. Is there a way we can put it into the chat? I'll see if I can.

**Moderator:**

If you can send it to Joel then the organizers will put it in the chat. Thank you so much for that amazing presentation. And it's great seeing how that resonates with our keynote speaker's speech earlier.

I'm not seeing any more questions coming in so thank you again. Oh one more. Wow, okay. What can large content vendors such as EBSCO do to help this?

**Mark Weiler:**

That's a good question. I think a priority is to really listen deeply to end users. You have to believe.

One of the things that we've learned in this process is to some extent blind screen reader users have given up on libraries. We don't listen well and we think we understand what's going on. And the more we work from that position, the less credibility we have. So I think that's a key important step. I don't know Ashley, if you want to comment?

**Ashley Shaw:**

I think the other thing I would add to that, Mark, is take users seriously when we tell you that something isn't accessible. We know when things aren't accessible and we have so many configurations and variables to consider that we have a high degree of expertise in our own situation. And I've had a number of kind of, you know, low-scale arguments with vendors who insist that an item is accessible when it doesn't meet basic accessibility criteria.

So really just try to familiarise yourselves with what kind of the very basic standards for accessibility are, depending on the content you're producing.

**Mark Weiler:**

Yeah and if I could add one thing to that: just because something is WCAG-compliant does not mean that it is free from disadvantage. Content can have disadvantages and still meet web content (standards) and we as librarians are called to respond to the disadvantage.

So a very good example of this is page numbers. It's very difficult to get page numbers. Very difficult maybe even impossible to get page numbers from journal articles with a screen reader just because of the way things are structured.

That's not in WCAG - Web Content Accessibility Guidelines. But that is a very significant disadvantage and so we have to respond beyond those narrow interpretation of the guidelines.

So that's what I would encourage vendors to do.

**Aneta Kwak:**

If I may also add that vendors work closely with publishers to ensure that the content they're receiving is also accessible so that our patrons and the end users are not stuck having to experience these barriers and then request remediation. There's lots that we can do before this content gets shared with our end users to make it accessible.

And so if vendors have a chance to work with publishers to go through and make sure that these materials are accessible before they're shared with libraries then that makes it a communal work that we're doing to improve this environment as opposed to all the work being on certain groups.