# DAISY Webinar: Publishing, Accessibility, W3C Standards. Where Are We Going?

Date: 6/3/20

The recording and further details from this webinar are available at:

<https://daisy.org/news-events/articles/future-publishing-standards-w/>

This is a Captioned transcript provided by CIDI to facilitate communication accessibility and is not a verbatim record of the session.

This is the second of two webinars examining the revolution of born accessible digital publications. Our first session was entitled “Publishing, accessibility, W3C standards. Where are we and how did we get here?”, which reflected on the EPUB journey to become the world’s most popular and most accessible digital publishing standard.

This session will discuss where we currently are and some of the challenges we face, summarize what is planned in the near future, and gaze into the crystal ball of publishing standards to explore where we might be headed longer term.

I will be back after the presentations to moderate the questions, but at this point I’ll hand over to our panelists, who will introduce themselves and address the question: where are we going?

>> Tzviya: Hi, I'm Tzviya Siegman from Wiley. I chair the publishing work group at the W3C. I'm the lead of the publishing activity which is a group of all of the publishing groups within the W3C.

>> Dave: I'm Dave Cramer. I've been complaining about EPUB for about 25 years. I helped write the fix lay out speck. I was part of the community group of 3.2. I hope to be the chair of the forthcoming EPUB 3 working group.

>> Wendy: I aspire to be one of the chairs of the EPUB 3 working group. I'm the editor of the audiobook specification. So today we are going to be talking about a number of things that are coming up. First we will give an introduction of how the standards groups work and EPUB. We will talk about the EPUB 3 working group. EPUB accessibility. DPUB ARIA and WCAG. One of the first things we want to talk about is how publishing at the World Wide Web consortium works. It's a collective of a number of different groups. The main groups now are the publishing working group which is responsible for the publish and the publishing business group and [inaudible]. They also are in charge of EPUB check fund raising. Two additional groups are the proposed EPUB working group which we will talk about. And the EPUB 3 community group which we are in the process of folding into the community group. And the one group that oversees is the publishing steering committee which is a group of all of the chairs and invited advisors. We come together and discuss the strategy and then do outreach. If you ever wonder how does an idea make it through the publishing activity, ideas are recommended to be filtered up to the community group. The community group is a great venue to look at ideas and if they are possible or room for specification, if they need to be sent to another working group like CSS and talk about the business specifications. That's how it all works.

Earlier this year the publishing activity hosted a survey of the publishing industry to understand how EPUB was being used. As we look for the strategy for the future we want to understand what is going on in the industry, how are people using EPUB, do people still want to use EPUB, is it doing what it needs to do? We learned EPUB is the format of choice. EPUB 3 was overwhelmingly the majority format. Publishers like the format but need it to do more. We include things like modernizing CSS, clarifying support for Java script, support for web browser and conforms via tests. A lot of the complaints around EPUB are around interoperability. Reading systems need to know what they are trying to implement. There's a lot of misunderstandings about EPUB. We hope improvement to the speck and communication through webinars like this and other activities will help fix that.

>> Dave: EPUB 3 seems pretty good to me. Why does it need to change? Why do we need an EPUB 3 working group? In the previous webinar in this series, my former [inaudible] described EPUB 3.2 as future proof. The future is hard to predict as has become clear in the last 77 days. Luke is correct in that EPUB was designed to evolve with the under lying web technologies like HTML and CSS. Yes, EPUB 3.2 is the best ever buzz there's always room for improvement. Many features are not supported by existing reading systems, reading system rewrite our files for good and bad reasons. There are hundreds of reading system with variable operability.

So we want EPUB to be better but now we have a new tool at our disposal. The W3C process. The W3C has a long history of creating high quality specks. 25 years of collective experience and wisdom in creating specks has been distilled into the process. There are two fundamental concepts. First, you must prove that every single feature in a specification is capable of being implemented. How do you do that? You write tests and show that at least two independent implementations pass. A recipe that looks good on paper may be terrible. You have to try it out. Second, the web and by extension e-books are for everyone. W3C has the concept of horizontal review where every speck is scrutinized to show it is accessible for all and every language and writing system and secured the privacy of its users. EPUB has always focused on accessibility and EPUB 3 was driven by the need to expand support for many of the world's languages but there's always room for improvement.

There are also political and administrative advantageous of going through the full W3C process. Governments or standards organizations are happier with such specks. EPUB exists for the readers not for standards bodies or governments. We hope to incorporate some suggestions from the community. We must be careful with EPUB. We don't want to break millions of existing books and cause trouble for millions of existing readers. So every valid 3.2 file will remain valid to whatever new version we create. We are call it EPUB 3.X until we name it.

So what sort of work will we work on in the working group? We will be working on testing, identifying all of the testable features in the speck, writing tests, and running the tests in many different reading systems. We hope to clarify how reading system are supposed to present EPUBs. It's likely that this is one significant cause of our interoperability problems. Of course we hope for a few new features. Scripting is confusing and most reading systems are not willing to implement it. We have been asked for full bleed images. So we hope to include those without drastically changing how people create and consume EPUBs.

>> Tzviya: Thank you. Additionally in the EPUB 3 working group we will take a look at a revision, improvement of the accessibility specification. The current version of EPUB accessibility is a note. The scope of the work of the specification is to make the EPUB accessibility 1.X a formal W3C recommendation. There's a lot of work going on right now to make sure that the EPUB accessibility specification meets the EU directive for e-books. Along with that we are looking for new features as long as they are compliant with the EU directive and WCAG 2.X. we will talk about WCAG in a few slides. We will look at fixed lay out accessibility and make sure to reference specific versions of WCAG.

We are also working to make sure that we are providing clarifications and reducing ambiguity. There's a lot of information in the specification that's for e-books and we want it to be clear information and not just a bunch of metadata. We are hoping to have this out by May 2021. There's a lot of work going on in that group right now.

Additionally in the accessibility task force there's work ongoing with the metadata cross walk. There's collaboration with other organizations including IFLA. So publishers can include the same information in multiple languages. Screama.org, onix. The work is ongoing and there's a great deal of work happening and I'm sure the people working on that would be happy to share that information. We want to convey information about accessibility to as many people as possible.

Next topic is DPUB ARIA. Some of you may be familiar with ARIA, accessibility rich Internet applications. It was to make HTML and other markup languages more accessible. Functionally I'm going to read a paragraph here. ARIA roles states and properties are analogous... [Reading from PowerPoint].

What that is saying is that if you use ARIA correctly it significantly improves the experience for people using screen readers. If you use it incorrectly, it can have horrible effects on those using screen readers. The purpose of ARIA is to override some of the semantics. Digitally publishing extension to ARIA we added roles of the specification that are specific to publishing and they map to the roles in the main ARIA specification. So you can say this paragraph is supposed to convey the information that this is a chapter and then it will convey to the Assistive Technology that this chapter has a structural role in the HTML. It's important to make sure to use it correctly. You can take a look at the link here ARIA authoring best practices to make sure to use ARIA correctly. If you are not sure how to use it, take a look at the daisy accessibility knowledge base. If you are not sure how to use it, avoid using it completely. You could cause more problems for your users than solving.

Coming soon we will release a small update in DPUB ARIA. There are small bug fixes and depreciation and some people are been logging issues with dock and note. And dock and biblio entry. Hopefully, this will not cause problems in your systems because you should have been getting validation errors with this all along.

Outside of the publishing world there's a lot of work happening in the accessibility guidelines working group. There are updates to WCAG. WCAG 2.2 is expected in 2020. There are updates related to cognitive and learning disabilities. There's some related to EPUB called fixed reference points known as page numbers. Information about low vision and mobile. Watch out for these. There are significant changes to the work that you do. Hopefully not too significant because hopefully you are already paying attention to these. They are bringing together the world of accessibility that had not been previously been in WCAG.

Coming up in the future is WCAG 3.0 which is the next generation of WCAG. The work for this is beginning. What WCAG 3.0 is incorporating testing and the whole world of technology has harmonized guidance so we can see all of web guidance. This work is going to take several years to incorporate.

>> Wendy: One of the last things we are going to talk about today are two specification that are being worked on. Main is audiobooks. Audiobooks is the specification that we designed to bring the specification to the audiobooks industry. Right now the audiobooks specification is on track to reach full recommendation status in Q3 of this year. Because the audiobooks industry didn't have a specification before this we are trying to fill that gap. We are trying to make sure the specification allows for creation of accessible audiobooks. Content creators will be able to create born accessible audiobooks or audio books that don't start accessible but are easy to make accessible thanks to specification like sync narration.

The synchronize narration note is the successor to smile for content that is built on the web. Synchronize narration is used for J son. A content creator can have line by line word by word paragraph by paragraph linking to a text document.

>> Dave: Finally, we wanted to talk a little bit about some things that might be happening a little further into the future. Wendy was just talking about audiobooks. Those who work with EPUB know that all the ingredients in an EPUB are listed in this XML package file the OPF. That's really not what the web does anymore. The web has drifted away from seeing XML as a primary way of expressing information. It sort of Java script won that battle basically. So much information on the web now appears in the form of JSON. So for a long time we have thought about could we make some sort of e-book format that was based on the technology the web uses today?

EPUB has always been based on web technology but still a significant hurdle to parse the XML files and do all of these crazy things. So we were wondering could we express the content of an e-book in JSON, so it is easier for developers to build. The audiobooks is the first example of that happening. An audiobook consist of MP3 files and metadata. So we are expressing the metadata in a JSON file that we call a publication manifest. It's similar to an EPUB. It lists here is everything. Here's what is in it. Here's the authoring and publishers and the person who read the book. Now the information is expressed in JSON rather than XML.

We hoped that idea would take off for books too, but it hasn't really so far. EPUB is very entrenched in the ecosystem. There's so much investment in EPUB and EPUB reading systems everywhere. So we haven't found a business need for this publication manifest. So we are just trying to figure out where it works for now and that first place is the audiobooks.

A second place this manifest might become useful is for, you know, digital narratives like comics. These are certainly currently possible and done in EPUB, but I think there are people out there who hope for much more expressive and interactive manga which have panel to panel navigation and complex transition affects et cetera. So there's a community group working right now that's taking that publication manifest JSON idea and creating a vocabulary to express all the geometry you have in a complex comic or manga and see if they can start experimenting with that and see if that's something that might at some point get some traction in the world. It's interesting work that's ongoing but we don't have a clear idea where it will go.

On an unrelated note there's also a lot of renewed interest in better CSS both for print and for content that often appears in pages. Most e-books are paginated. CSS has a lot of really interesting new capabilities like flex box and grid and animations and 3D stuff and so on, but those technologies haven't trickled down into e-books yet. There's a renewed energy among the browser vendors themselves. Google and Microsoft are -- and Firefox are talking about improving the support of print defining what happens if you split a CSS grid across pages. How could we actually do foot notes and notes in a more elegant manner? Can we have control over running heads and page numbers in e-books? There's a new community group devoted to some of these ideas and I think that's something that has the potential to help make our e-books a lot better than they are now.

Finally, there's -- Google has been working on this really interesting web packaging idea. One of the powerful things about EPUB is that you can create this file that you can distribute, e-mail to someone, that you can create without having a web server basically. And they are trying to extend this idea to sort of the web in general that you can bundle together a whole bunch of HTML, CSS image files and have it work directly in a browser rather than needing an EPUB reading systems as a middle step. The original motivation for this was for places where the Internet is not very available and there's not a lot of bandwidth and people would collect the webpages of news sites and sell them on USB drives or set up local beacons to rebroadcast web content. But the idea expanded, and it looks like it may be possible to create stand-alone web applications. I have made a self-contained audiobook I can e-mail to you and you can double click on your browser and it would be right there. No reading systems or distributor. It seems like an idea with a lot of potential, but there's also a lot of controversy about it because it changes the web security model in certain ways and has interesting connections to Google's other plans for world domination. And that's what I have to say about that.

>> Wendy: So we have talked about a lot of things in this presentation. We have a list of links here in this slide deck. DAISY will post them on the website. There's the EPUB survey overview and results. A link to the EPUB 3.2 specification. The EPUB working group charter. It is a draft right now and open for feedback if anyone is interested in providing feedback. We have a demo of what an audiobook looks like on the web and an implementation of sync narration. The accessibility metadata crosswalk. A link to DPUB ARIA 1.0 and the ARIA authoring practices. If anyone wants to get in contact with any of us, you can feel free to e-mail any of us at our e-mails or reach us on twitter. We are always happy to talk about anything related to digital publishing or if you want to post a question, do that here. And it's discussion time.

>> Richard: Fantastic. Thank you. This is Richard again. Thank you Tzviya Siegman, Dave and Wendy. If you have questions in your mind and you are watching this webinar and would like to pose them, now is a great time to pop them into the Q&A. We are going to start with some that relate to the reading experience. One is about the thing that was mentioned that is more work being focused on the conformance of the reading systems and specifying what the behaviors should be. The question is haven't we been there before where there were lots of tests around reading systems and that ran out of road and wasn't successful. What will make that successful this time around?

>> Dave: First of all, the EPUB test.org was really focused on the features of the EPUB themselves. Does mathML work in this particular reading systems? Does it support these image types? Does it screw up vertical writing mode? It doesn't address the behavior of reading systems in like do they display the metadata -- how do they display the metadata in your package file? Do they show the book title if you have three authors, how do they list the three authors? How do they handle spine items that are linear equals no. Do they support inline mark up in the navigation document? So we are hoping to clarify in the speck itself how reading systems interact with readers. That's largely something we haven't really even tried to think about much yet. As to the difficulty of EPUBtest.org did get overwhelmed because there are so many reading systems and so many tests. I think that's what worries me most about the working group. We will have to find a better solution than what happened there.

>> Wendy: I would add -- Dave said the important stuff but as someone who has run EPUB tests a lot of the tests came with an element of interpretation which is not always a good thing especially when you have very specific -- you are hoping to have very specific technical requirements. Dave mentioned linear no. The behavior as described in the speck addresses more of the use how linear no should be included and not what a reading systems should do with it. So that's a good example. There's many others. It's one big challenge for the working group is defining that behavior in a way that reading systems and users have a good experience.

>> Tzviya: I will add to somebody who has tested for EPUB test that there are few aspects -- the interpretation aspect was very challenging, but it also was extremely manual. And we are hoping in the working group to automate the testing. As far as what Dave was saying in terms of the language of the specification, we have numerous statements that are "should" statements and when we say a reading systems should do X, often that leaves room for interpretation and they aren't testable statements. We are going to have to make decisions about how to rewrite those statements so they can be interpreted into testable statements and that is one of the largest goals of the working group. It's not just testing itself it's revising to test things.

>> Richard: So there's challenges ahead but it's good to be aware of them.

The next question relates to the reading experience. You talked about the W3C reading specification. How does this relate to playing within EPUB readers?

>> Wendy: Audiobooks specification is not technically written for EPUB reading systems. I know there are some who have an accompanying player. The audiobook specification was designed for dedicated audiobooks experiences. Because we have the added text element, we also wrote it in a way that it could be implemented depending on the reading systems developers choices implemented. Because we are providing a manifest format it's up to the developers to how that gets presented. I have some dream possibilities but it's up to the developers.

>> Richard: Dreams and crystal balls is what we are all about on this webinar. Let's move to books on the web. Dave you talked about web packaging and the desire to have an EPUB like reading experience but actually through the browser. What are your thoughts on reading books on the web and that whole piece? What's your crystal ball or dreams in that space?

>> Dave: I think my crystal ball and dreams contradict one another here. I dream of it happening. From the point of view from accessibility, I think a lot of people know their browsers well and it's hard to deal with a reading system that does something different than the browser for this case. What we haven't figured out is the business model for books on the web. I sometimes imagine -- hope that there's a future of independent books and content that works as web publications in some sense outside of the big 5 commercial EPUB market which I think is likely to continue for a long time. 10 years ago we were about to have EPUB 3 and not that much has changed in the last 10 years.

>> Tzviya: I agree with Dave. Dave's the best person in the world to ask this question. He has created many prototypes of web publications of books in the browser. He calls them toys, but they are great templates. EPUB has its limitations. I work for a scholarly publisher for Wiley. There's things that don't work fantastic with EPUB and I don't think they are going to unless we change the model that makes better use of the web. So web packaging could change everything. We could do things with that model where it basically is just a book on a browser that you could package up using whatever functionality it is. You can have linking into very specific points in a publication and share those links. Citations are crucial. Java script would be part of the way it is on any website. I don't see this changing quickly. I don't want to attempt to fracture the publishing world. I could see this taking off in a different way in scholarly publishing or education publishing.

>> Richard: Speaking of the browser for one more question. Ted asks do we see browsers warming up to using EPUB 3. The edge browser supported it for a while and then the new version dropped it. Do we see native support for the EPUB in the browser as a reading systems at all?

>> Dave: I haven't heard anything. I think Microsoft edge was an interesting case. One reason they were able to do it is that under the hood of the browser they had technology that made doing things like pagination easier than it would have been in chrome or something. I think some of that functionality is now going to come back to chrome which would make it easier to work with the browser to achieve these publication like affects but I doubt the browsers themselves will be making EPUB a native format. I don't see how it fits their plans for world domination.

>> Wendy: I agree with Dave. There's a lot of challenges in the current modern web views with things like pagination. It's not easy to create an EPUB reader. A lot of reading systems developers are looking to create them in web context and there's a lot of work to be done to make a book render the way a user expects to render. Until the browsers make that easier and have a vested interest in it, it's still a huge challenge. Though I think there's a growing interest in browser based readers though it's still coming from the book people and not from the browser people. Though, I think Dave and I agree we would love to have browsers part of EPUB 3 and welcome them all the time. It would be really a big change.

>> Tzviya: I agree. I will say that I expect to see before we see browsers is what we used to see with PDF. Where there was a plug in to read something in the browser. It would not surprise me if we see that for EPUB.

>> Richard: Okay. You mentioned the PDF term so why don't I ask that question. You are speaking with your perspective as publishers, but people read content from a wide variety of services and places including public services, government bodies and so on. So far we see the use of EPUB as a replacement or alternative to PDF. It's kind of slow in taking up in that space. Can you think of reasons for that and do you see that changing in the future?

>> Wendy: I think one of the biggest reasons for that is the ease of creating -- today still it is easier to create a PDF than an EPUB from however an organization is authoring their content. However, in the last 6 months there's been a huge flurry of EPUB tools. For instance DAISY word to EPUB plug in. Google doc now allows you to exported to EPUB. The easier it gets to export an EPUB from whatever format you are using to develop that document -- I think most people aren't -- especially when working in government or industry, you are not writing raw HTML code for your internal documentation or reports. The easier it gets to export the EPUB format and the better those EPUBs are, I think the more proliferation we will see. This is my shout out for the DAISY word to EPUB plug in. It's the best one I have ever used.

>> Tzviya: Especially dealing with government, it's not just about the creation tools which is a huge issue but also the reading tools. I think all governments and certain private companies need to authorize everything the people download. If someone has an EPUB that they can't open is useless. So now with PDF everyone opens it in the browser. With EPUB they need to be authorize something to open it. So we need to work on advocating for pleasant reading experiences.

>> Richard: The different reading systems and user experiences will be subject of a future webinar. Thank you for that suggestion.

Ken and ted ask almost the same question. I'll ask ken's question. Ken mention new media types that will be included in the next version. Have you identified any examples of those yet?

>> Dave: One thing that is come up really a lot is an audio kodak that is optimized for speech. I think people have mentioned opus is a candidate. There's a lot of audio formats out there. I think people seem to have strong opinions about audio codex. As someone who has never had a strong opinion, I don't understand that at all. I think another obvious candidate might be image formats like web P. I haven't heard calls for that yet, but I wouldn't be surprised. I think the number 1 thing is the audio stuff.

>> Richard: Great. Jan ask the question do you plan to separate the 3.X speck into modules that could be supported individually by the different reading systems implementers? This is making them possible to use outside of reading systems. So making it more modular?

>> Dave: I certainly see some of the appeal. HTML went down that road with the modular XHTML and it was a confusing disaster I think. I'm especially sympathetic of what is the interoperable core of EPUB. What is the stuff that works everywhere? Leaving out things like multiple renditions and collections stuff like that. It's hard to know where -- how to modularize things that way. We already split out the package file, the packaging mechanism of OCF et cetera. It would be hard to know where to draw the lines.

>> Richard: Great. I have some accessibility specific questions here. One is a straightforward question. Does EPUB type help with accessibility? Who would like to take that?

>> Tzviya: My son has joined us. He see going to help with this. EPUB type was planned to help with accessibility. The goal was that if we include more information in the tagging than Assistive Technology and other tools would be able to pick up and call inflection. That never materialized. Right now it's a wish and doesn't help anything. I need to talk. If you include it to help with work flows that's fantastic. You can help -- you can include information for your workflow. If you have something like P EPUB equals chapter all that is doing for you is is including information for your workflow. No Assistive Technology is picking it up.

>> Wendy: What Tzviya said.

>> Richard: We have talked about the future and a little bit of crystal ball gazing. What does this future of publishing standards mean for people with print disabilities? How is the future brighter?

>> Wendy: That's a great question. Where we are really trying with the future and so things like refining EPUB and audiobooks and things like that, we are trying to make it easier and easier for content creators to produce better content. EPUB 3 we like to say is accessible by default, but the fact is that a lot of publishers that are producing EPUB 3 are not producing accessible EPUB 3 content despite it's easier to do so with EPUB 3. I think with us making it easier for both reading systems and content creators to produce better content and make it clearer and clearer what are the accessibility requirements, what should a reading systems -- accessibility is one area I would like to refine for reading systems so reading systems understand what they need to do with accessible information like the metadata or things like DPUB ARIA. By doing that it means that people with print disabilities can have a little bit more confidence over time. I want the -- the end goal is complete confidence in what they download. I hope with these changes there will be an increase confidence that any reading systems they use will give them the best reading experience for their needs.

>> Richard: I'm going to ask each of our panelist to make a prediction of something they think will be really different in 10 years’ time. I will give you a moment to think about that. We have a question from Austin: For a person just starting to learn about EPUB publishing, what are the resources that you recommend to help with their development?

>> Wendy: I would recommend read speck. I think there's a lot to learn from it. The EPUB 3.2 is easier to read. Follow e production on twitter. There's no vowels except for the e. The other resource I would check out 100% is anything by Laura Brady. All of her videos have been helpful to myself in learning how to be a better e-book producer.

>> Richard: Any other resources?

>> Dave: The books by mat Garish.

>> Tzviya: I think the DAISY accessibility knowledge base is one of the most accessible tools out there.

>> Richard: I second that. Let's hear from you. You can pick the order you go in. A prediction for 10 years’ time what will be your amazing thing that you for see and watch out for?

>> Tzviya: I will try because my kids are quiet now. I'm hoping that we will incorporate more web technology and enable things like citations and deep linking in e-books.

>> Richard: Who is next?

>> Dave: There will still be EPUB 2s in the marketplace in 10 years.

>> Wendy: Don't say such a thing, Dave.

>> Dave: There will also be web packages.

>> Wendy: That's better. I'm going to be the wild optimist and say that we are going to see way more accessible reading experiences in the marketplace and that's going to be the standard. I'm hoping that it comes from all directions in terms of retailers demanding it and publishers producing it. I hope that audiobooks are produced using the standard accessibly for all users and we just see really great digital experiences for books and audiobooks and content in general.

>> Dave: If I had a wild prediction PDF will be seen as flash is seen now. That would make me happy.

>> Wendy: I will second that.

>> Richard: We will wait to see how good those predictions are. Wendy, will you move us to the next slide?

OK, we’re coming to the end of this session. Tzviya, Dave and Wendy, thank you for sharing great information and insights.

Thank you to everyone who joined us for today’s session. If you missed the first of these two webinars examining the revolution of born accessible digital publications, you can access the recording and slide deck of “Where are we and how did we get here?” on the DAISY webinar webpages.

Coming up in the next few weeks we have some more wonderful topics for you:

On June 10 we’ll be learning about some legislation that will impact on one of the world’s largest marketplaces- “The European Accessibility Act: considerations for the publishing industry and benefits to consumers globally”

On June 17, in response to many requests for this topic, we have “Describing Images in Publications – Guidance, Best Practices and the Promise of Technology”

And on June 24 we have another requested topic, “Metadata in publishing – the hidden information essential for accessibility”

Find out more information at daisy.org/webinars, where you can also sign up to the webinar announcement mailing list to learn about new topics as we add them. If you would like to suggest a subject, or if you are considering presenting a webinar, then please email us at webinars@daisy.org

I hope you will join us again next week. In the meantime, thank you for your time and have a wonderful rest of your day. Goodbye.