

LPC Forum Transcript

File: LPF03_0305_2pm.mp3

Total Time:

00:00:11-00:00:56

...This panel is on, it is titled Infrastructure Alignment, and the main purpose of this planary session is to discuss interesting developments in library publishing infrastructure and potential ways of aligning such infrastructure. Our panelists today were chosen for their involvement in major exemplar projects. They can speak to infrastructure developments. Let me introduce them...my name is Martin Halbert. I am Dean of Libraries at the University of North Texas.

00:00:56-00:01:44

Our presenters today are Curtis Fletcher, who was the Program Manager for The Alliance of Networking Visual Culture, which produces the Scalar Online resource. He has previously worked at Cambridge University Press and USC in many different capacities that bear on the publishing arena and has a lot of perspective to share on these topics of infrastructure. Since 2010, Jeremy Morse has served as Director of Publishing Technology for Michigan Publishing, a division of the University of Michigan library. This continues and expands on his work as a programmer analyst in the scholarly publishing office which he joined in 2005. Before that, Jeremy served as Digital Preservation Specialist and Head of the Reserves Department at Northwestern.

00:01:44-00:02:20

Dave Ruddy is Director of Scholarly Communication Services at Cornell University's library. His unit is responsible for user support and operational management of several library repositories, including Project Euclid, the archive, eCommons (Cornell's Institutional Repository), and the Cornell University Archival Repository, and Digital Preservation System. He has been involved with Project Euclid since its inception in 1999 and is the Cornell lead on that effort. He holds an MA/MS and Ph.D all from the University of Michigan.

00:02:20-00:02:50

I was asked as a way to get us started in this session to frame the discussion with issues from the perspective of the dean in provisioning infrastructure development and comment on some of these examples of leveraging infrastructure that we've seen and as a dean, I find encouraging at our institution and others. So just very briefly to speak to that: I'll give you a couple of quick points.

00:02:50-00:03:27

The obvious one is that infrastructure is expensive to develop, and to the extent possible, should be leveraged for other purposes, especially as we get into this library publishing arena. One of the things that we Deans look for is ways of extending and leveraging the existing investments in our infrastructure and seeing what new applications they can be used for. I'll give you just a small example of that that I think echoes many of the comments we heard from a lot of the presenters this morning.

00:03:27-00:04:09

We developed at UNT, over the years, a lot of infrastructure around our various Digital Library efforts; our Portal to Texas History has millions of items in it. (It is) a collaborative project that goes on in the state of Texas between 240 different libraries, museums and historical associations...it is part of the Indian P effort on the national level. It is a large scale infrastructure that they regularly digitize a hundred thousand pages a month out of historic newspaper content...it also runs the back end for all of our digital libraries infrastructure that you see the front end for here

00:04:09-00:04:40

When we got interested in this notion of library publishing, we are one of the 20 institutions that were mentioned this morning where the university press reports to the library, and I got very interested in this question of “Well, okay, what can we do to develop some synergistic efforts between the Digital Library unit in the UNT libraries and the press?” We looked for ways of leveraging that big investment, long-term development curve from the DL operation.

00:04:40-00:05:13

One of the collections that you see in the Digital Library at UNT is the UNT Press Collection. It has been a budding collaboration that my press director Ron Chrisman was at first a little nervous about. He was like, “Well what are you asking us to do in this idea? You aren’t asking us to give away our stuff?” I said, “No, no, no, no. Let’s just honestly look for any and all possible ways of collaborating between different units and the campus.”

00:05:13-00:05:52

Once he kind of got over that gun-shy moment of “What is he being asked for?”, then he got together with Mark Phillips, our head of our Digital Library division, and really did some very interesting experiments. They started with taking items out of the background of the press archive, the back list, and things that were no longer in print, and putting them up in the infrastructure. As you can imagine, with doing 100,000 pages a month, this was sort of noise, honestly, in that big work flow. They were able to very quickly do 77 books that they mounted up in this new collection.

00:05:52-00:06:28

What really got Ron interested, motivated to look at this as a possibility, was when the DL folks, at his request, developed this little small thing here that makes a big difference. That is the “Get a Copy” service- you know, “Buy a Copy” of this thing. You click on that, and you instantly get presented with a range of opportunities to buy a copy of that item out of the backlist. Ron suddenly started seeing revenue generation out of these items that had been out of print for a very long time.

00:06:28-00:07:20

That quickly made a convert of him and we now have mounted, in very short order, this collaboration has only been going on about a year, we’ve mounted more than 250 items out of, not only the back list, but the stuff that was still in print because what he was seeing is the stuff that had not sold in a while, because the UNT Digital Libraries does a lot with discovery

mechanisms and Google and other things, suddenly, as also was mentioned this morning, there was a whole new mechanism for their materials that hadn't sold in a while to be discovered and become discoverable in a whole new way. That quickly got their attention. They put up all these volumes, these are some examples, as well as journals that are published out of the UNT Press, like Thera and soon the Journal of Schenkerian Studies.

00:07:20-00:07:49

I was very encouraged by that as someone who (is a dean). Deans are often very unhappy about when there are hard silos that arrive in our organizations, and we like to see those leveraging of investments and this was honestly no additional investment on the part of the Digital Library unit and it has already had a significant impact in the Press and shows a lot of promise for work to come.

00:07:49-00:09:00

With those as some framing thoughts, I think we have a stimulating panel of presenters here today that are going to tell us a bit about quite a significant range of different infrastructures that have been developed for library publishing initiatives. With that...I will turn it over and we will have presentations by each of the presenters and if anyone has a burning question, as Charles said, we can do it, but otherwise we are going to reserve some time for a Q&A at the end of the session for all of the questions. (Inaudible shuffling and instructions)

00:09:00-00:09:36

What I want to do today is talk a bit about Scalar, go over the anatomy so to speak of Scalar projects, Scalar books, and then talk about who is currently using Scalar by way of showcasing what it is that Scalar is really built to do, what it is good at. At the end, (I will) briefly talk a bit about the bigger picture about the Alliance for Networking Visual Culture who produce Scalar and its overall mission.

00:09:36-00:10:09

Scalar is, on the one hand, an authoring and publishing platform which is built to bolster the technical and infrastructural connections between digital archives and publishing systems, academic publishing systems. You see at the top, these are some of the affiliated archives that we have that are partners in the alliance and a few are just larger, online repositories that we built technical bridges with.

00:10:09-00:10:46

On the other hand, Scalar is...an authoring platform that is built to encourage media-rich, multi-modal on linear scholarly productions. As such, the other section of our partners are academic presses who are now starting to embrace those new modes of scholarly production. Scalar is sort of situated in between.

00:10:46-00:11:40

(Gesturing to screen) This is what Scalar looks like. These are the front pages of a couple of books. These authors have sort of taken advantage of the built-in templates, although I think the bottom one they've customized a bit. You can do customized CSS for individual pages and for the book overall. Again, the top one is just a built-in template with a background image. The

bottom one has used a good amount of CSS customized look. There is a lot of flexibility there...again, these are the landing pages of these Scalar book projects.

00:11:40-00:12:28

Scalar books have a couple of main elements. There's the pages; there's the media; there's the annotations, which are easily done to any media within a Scalar book and author interface. There are ways of structuring the overall content, which are paths. On a Scalar page, you import media and reference it. Let me back up a second, I should say that all of the archives you saw before, both the affiliated and unaffiliated archives, all their material can be directly imported into a Scalar book when you are in the media section of the dash board. The main part of the user interface is a WYSIWYG editor, but there's a dashboard that is a backbone that is a lot like WordPress in that sense, or many other platforms.

00:12:28-00:12:58

Media is referenced just like a lot of other platforms. The interesting thing that Scalar does is once you've referenced the media, you can then dictate a number of views. This is the text-emphasis view (shown in presentation). You can also do a media view. You can do a split view. You can cover up the media altogether, and you can set these as default or any individual pages.

00:12:58-00:13:38

Once you've referenced the media, then you have all of these options of different media views. The last view is a visualization, these built-in visualizations that we have. I'll get to that in just a minute. Again, the media can be imported directly from any of our affiliated archives. They can be uploaded, anything from a URL, and annotations can easily be made. For instance, with video annotations, you can just stop and start the play head and make annotations (and) comment on those annotations quite easily in Scalar.

00:13:38-00:14:30

There are a number of ways in which authors end up organizing their content in Scalar books and one of the unique features that we have that we call "The Paths." Paths are linear sequences of pages, much like chapters in a book, or sections within a chapter of a book, but one of the interesting things that you can do is you can confront your readers with a myriad of paths, giving them options. A lot of our authors will have separate paths for different languages, separate paths for specialized readers and layman readers. Some authors will take all the media within their book and make it as one path; something like that is really easy to do. You go and you create a path and you go to the media tab and you just drag and drop all your media.

00:14:30-00:15:08

One of the things Scalar does well is you have a sort of, I guess you could call, flat ontology. Everything that is created has its own page. For instance, in WordPress, if you have a piece of media in the media library, you are going to have to put that into the page and make it viewable. Anytime you import anything or create anything in Scalar, it becomes a page in itself. For instance, even reader comments. If an author wanted to create a path of just reader comments, they would create a path and just drag and drop all those comments. Everything in Scalar can relate to everything else.

00:15:08-00:15:44

(Shown in presentation) Here you are showing that not only can you have multiple paths, but you can also have paths intersecting. Maybe a more concrete way of saying that is any given page or any element in a Scalar project can live on multiple paths. We'll see a little bit later that there is a way for, in Scalar, in the reader interface, it cues up and...becomes a navigation item where the reader has the option of switching paths when they are on the individual page.

00:15:44-00:16:24

Tags are another way to organize the material, and again we are going to see some visualizations later. Tag visualization can be really helpful to get readers to explore the material in a Scalar project. Tagging in Scalar is a bit different than what we are used to. (In) platforms like WordPress, the tagging is essentially database tagging. In Scalar, tagging is less a noun; it is more of a verb. You are establishing a relationship between different elements. Instead of saying that all my blog posts that have to do with "1966 World Series" are here, you are saying "This item and this item relate to each other in some way."

00:16:24-00:16:51

If you have, let's say you have an annotation of a video of Fantasia, and you'll also have an image of it somewhere else in the project. You can tag that media, that piece of that annotation, to the video and to the image so that when you are on the page for that annotation, you know that these other things relate to it.

00:16:51-00:17:20

(Shown in presentation) Here is sort of a depiction of that flat ontology; these things all equal each other. I can tag a path with a piece of media, if it relates. I can tag a page with another tag if it relates. What you get are paths that contain all these different elements, and tags, in which any number of things are tagged with any number of other elements.

00:17:20-00:18:01

The other unique feature is these built-in visualizations. These are generated; the author doesn't have to do anything. In this case, this is, I think, (a book) by Mark Marino, a book on border codes (shown in presentation). He's created all these elements in a Scalar project, it's all color coded. There are pages and there's media, and there might be a couple of annotations, but there are no tags there, I see. He's put them into separate paths, and if he wants, he can make this...these visualizations work in two ways. They can be used a navigational items. You can set any page to be this default visualization. It is also a good way for scholars to get a sense of what their overall project looks like.

00:18:01-00:18:31

This is the thing about Scalar: there is no built in architecture. There is no built in metaphorical narrative. There is no "If you are writing on a blog, it is chronologically ordered. If you are using a website template, it is order by the navigation items." In Scalar, you put all this content in and then as a scholar, you decide how to organize it and how to relate it. That's really the tough job.

00:18:31-00:19:15

There's an index visualization. Again, you hover over...it's color coded again so the tags are red and the paths are blue and the pages are beige. The intensity of the color has to do with how many elements, how many other elements, in the Scalar project it relates to. This tag, there (are) maybe 119 items...tagged here. I should've had a visualization, but if you roll over one of these blue...paths, it will show you all of the elements that are on that path in a number of ways (shown in presentation). This is a good way of getting a global sense of all the elements in a Scalar page, or Scalar project, and how they relate. This is the sort of thing that Scalar is really good at.

00:19:15-00:19:56

This is the tag visualization. Again, color coded- pages, tags, media. You can set an individual page to be this tag visualization and what happens is just that page will come up and when you click it...again, I should say these are all navigationally enabled so you can click on anything within any of these visualizations and go to that page. Here, imagine that all of these are gone except for the one that is highlighted, and when you click on it, the two that it relates to will pop out, and then you can click on each one and everything that it is related to will pop out. It's a good non-hierarchical, non-linear way of getting your readers to explore your material.

00:19:56-00:20:17

This is the radial visualization. Not a lot of people use it. It is a little sy-fi-y. It's a bit harder, I think. It doesn't show it here, but when you click on any one of these blocks, it will expand and show all the different things it relates to. In this case, that tag (shown in presentation).

00:20:17-00:20:58

(The element of) multiple authors: we have professors right now who have 50 students using one book. Again, they've each established their own path and then they end up using each other's material. If someone creates a page, someone creates a piece of media, someone creates an annotation of a piece of media, the other student working on the other path can import that in, just drag it in. When the reader is on that page, they will be cued up and see that they are at a kind of cross roads that they can explore that page within all these different contexts within these different paths created by these different authors.

00:20:58-00:21:27

Comments (are here)...Again, the comments live on the bottom of the page but they also live as a page in and of itself within the overall Scalar framework. Scalar Open Source can be found at GitHub (and) can be installed on anyone's server anywhere.

00:21:27-00:21:59

So what are people using Scalar for? In the most basic way they are using it, they are using it as a way to get material you can't get in the print format, namely media, moving media, into larger projects. In this case, Matthew Delmont with the help of the University of California Press came out with the companion for his book. We get a lot of these.

00:21:59-00:22:54

Another way people are using it is as a way to ask the reader to explore material in a way which isn't necessarily dictated by the author. This is Virginia Koon, her filmic text, and this is the main landing page where the reader is confronted with options. In this case, she has decided to divide it up between scholarly material that relates to a subject, and pedagogical material that relates to a subject. Again, there are a couple of pages here that live on different paths. In this case, we are on a page that is on Digital Pedagogy. When you scroll over it, you will see "Oh look. This is page two of 14 on this path. It's also page three of 16 on this other path." Now we can switch contexts.

00:22:54-00:23:23

This is tag visualization (shown in presentation). College Art Association, I think, came out with a book on the publication of The Art Bulletin, if I remember correctly. They've tagged all the material. They've tagged it by year, they've tagged it by theme and this is how all that material relates.

00:23:23-00:24:18

A few museums have come up with ways to create sort of virtual experiences of exhibitions- in this case, it was the Bernini sculpted in clay. The last way to utilize Scalar is API. A few authors have decided to create their own interfaces. In this case, this was built in Flash a couple of years ago when Flash still existed. That user interface sits on top of Scalar's organizational architecture so they've tagged and created paths with Scalar material and then that's what organizes the material in this front end. We've had, and are having, more and more people use the API for just that purpose.

00:24:18-00:22:55

Let me just talk about the...Alliance for just a minute. The Alliance is, The Alliance for Networking Visual Culture, is headquartered at the University of Southern California, but it exists as a set of partnerships between archives, libraries, humanities centers and academic presses. On the one hand, Scalar is there to fuel and shape these new genres of media-rich, multi-modal, non-linear, scholarly production. One of the ways we do that is we work closely on the ground with people at our humanities centers. We have a sort of direct pipeline from scholars on the ground. That feedback goes directly to the development team and that's how we build Scalar.

00:22:55-00:26:05

On the other hand, the Alliance is committed to investigating the "best practices" for closing the gap between visual archival digital material and scholarly publication to strategize about work flow between carefully curated digital material, visual material, scholars who are writing, authoring, researching and arguing and academic publication.

00:26:05-00:26:54

While the Alliance is obviously very interested in libraries using Scalar for publishing journals or monographs or other works, we are also particularly interested in libraries and archives who have digitized collections of visual material and are interested in getting it into the hands of humanities scholars. I have to say, the fact that one of our main missions is to close that gap

between archival material and academic publishing, we're very excited about the fact that libraries are moving from one to the other.

00:26:54-00:28:37

I'll end there. You can visit us at scalar.usc.edu. You can contact us... if you are interested in any way, we do training for Scalar. We are happy to do individualized training. We do training at set times every semester, but if you have a staff and you want to get going on Scalar, feel free to contact me and we can set something up. That's it. (Applause, inaudible audience questions)

00:28:37-00:29:58

What gets imported from the archives is not the media itself, it should be said. What gets imported into Scalar is all the meta data and the reference to the media. We do encourage our authors, or we warn our authors, about YouTube who, occasionally by fiat, change the parameters and suddenly the URL that you were referencing will disappear or change. The meta data, if it is updated, won't change. It will stay in Scalar. We encourage our users to make sure that whatever URL they are using, (it) is going to be stable. Obviously this is an issue because you want publishers to move into this realm, and publishers are very worried about sustainability, very worried about reliability because buying a book and putting it on a shelf is not the same thing as publishing or curating a publication which is not on their servers and references media that is also not on their servers. It's an issue, and one of the ways of getting around it at this point is to just make sure that everything – the URLs you are using- are stable or hope that they are going to be stable.

00:29:58-00:30:06

Audience question: Is this a template or are you hosting this material on the server?

00:30:06-00:30:43

We are hosting all of it except for the media that you reference. You can upload media to our server, but we have a very tight limit because Scalar isn't really meant to be a repository. It's a 2 mega bite limit, which is very low. There are pictures on my phone that are bigger than 2 mega bites. All the content is hosted on our servers but the media is not. (*Announcement to audience about questions later*)

00:30:43-00:31:19

Second speaker: First of all, I'd like to apologize to Curtis for all the times I've called it "Scalar." That's the mid-western in me coming out. I'm here to talk about iMPact, a set of tools that we are developing at the University of Michigan to build a publishing platform on top of HathiTrust. It's being developed by Michigan Publishing, along with the digital library production service, which (is) the development team primarily behind HathiTrust technology.

00:31:19-00:32:36

First I want to give a little context, though, about the story of Michigan Publishing building on top of existing library infrastructures. In the 1990s, Michigan, along with many other places, were doing these sort of library collection reformatting projects- scanning stuff and putting it online. These were all one-off silos of content, each its own web app, basically. Not only does that quickly become unscalable, but unmanageable just because the practice of both the

encoding, the text behind the scenes and doing the coding of the application, kept evolving as we kept learning. Back porting your improvements to older versions became more and more costly. We didn't want to do that anymore so figuring out what all those things had in common, building a common platform to host all that and then migrating most of that content into the new platform, which was DLXS, developed by Digital Library (Publishing) Service, as I mentioned before.

00:32:36-00:33:01

This was great. It was so great. We made the code open source so others could use it, and some did. Even though the code was free, there were still pretty great costs to running your own instance of DLXS. You had to have the infrastructure; you had to have the expertise to get it up and running, it was not turn-key by any means.

00:33:01-00:33:42

None the less, we had our own instance at Michigan, so when the Scholarly Publishing Office formed in 2001 with Maria Bonn at the helm, we needed a platform for our experiments in library publishing. We went with DLXS, did some repurposing of it. (We) took a little closer look at DLXS; it's pretty obvious when you see it, you might be familiar with this interface, that it is designed with library collections in mind. I think it was John Unsworth who once called it "an interface only a librarian could love."

00:33:42-00:34:33

You can see it is designed with page scans, with OCR text for the search engine, (and) physical volumes being presented as a library collection. We just pushed the software a little bit further to make it (a) fully encoded text, article-based publishing, born-digital publication and gave the interface a face lift so it has a little more of a publisher-friendly view. Here's the Journal of Electronic Publishing (shown in presentation), one journal that we happen to edit at Michigan. It is the only journal in our journals program; our journal program is headed by Rebecca Welzenbach.

00:34:33-00:35:10

Normally, our arrangement is to partner with the journal editor. We do their production, we production-alize their manuscript once they've given it to us and it's gone through all the editor (process), and we (do) the digital distribution in our DLXS instance. This is all pretty good. It's showing its age, it is 13 years old in our use. All the components are very tightly coupled so you can't use just the search and do a different display very easily at all.

00:35:10-00:35:34

We are obviously concerned with preservation. This does kind of "preservation lite" where the content doesn't need to be exported from the system; it is already sort of separate, outside the platform, but is really just sitting on a file system that we try to take care of. It's not very rigorous in terms of preservation.

00:35:34-00:36:31

We started thinking about "Well what is our next step of evolution going to be for our publishing platform?" Along comes HathiTrust, developed at the University of Michigan so we see opportunity here. It's sort of history repeating. Once again, like early DLXS, it's reformatted

library collections, page scans, (and) OCR, but with iron-clad preservation. It's (a) track-certified trusted repository, one of only four such repositories in the world. It not only has very strong requirements for content normalization to be in a preserve-able state before it can even be ingested, but then both the content and the access mechanisms are redundant across geographically dispersed locations- at present, Michigan and Indiana.

00:36:31-00:37:02

It's also going to be part of the DPN Network so we are even further building redundancy to the availability of the content. HathiTrust is all about preservation. It is all about keeping content as open as legally possible, which is not always so open, and available forever. (Those are) basically the core principles of HathiTrust. We like those principles so we want to build on top of those.

00:37:02-00:38:04

Here we are back to iMPact. What is iMPact? iMPact is modular platform, so you can try to pull the pieces apart better so you can just use the parts that are useful to your work flow; if you have a better tool, you should be able to use that instead, right? It is, however, tightly coupled with the HathiTrust repository. This is sort of a core design principle of the system; it is the core design principle. They designed everything around this principle, and that is that access must be through the copy, meaning you can't just publish in your own place and then deposit a copy into an archive somewhere. Our stance is that, over time, inevitably, the resources to do that archiving won't be maintained and the two versions will fall out of sync as the access copy gets refreshed, renewed, updated (and) the preservation copy with wither or become outdated. Access will always be through the preservation copy, that's sort of the principle here.

00:38:04-00:38:43

We are designing this for open access journals. Open access is, again, a core principle. Journals is just the first problem we are going to solve- monographs and other types we will have to deal with in the future iterations. It's meant to be, if not a built-in end-to-end system, something that you can make part of an end-to-end system. If we are successful, we think that eventually, an end-to-end environment that uses iMPact would exist. We are working toward that.

00:38:43-00:39:27

Here's an overview diagram of the environment (shown in presentation). Our rounded box here are things that actually exist, they are in production or development. Starting with the orange box, which is HathiTrust as you know it today, with some added functionality, which is being developed as we speak, unless they are taking a break, to not only serve up page images but fully encoded JAX XML; we are going with the JAX.dtd journal archival text set. If you don't know, is the NISO standardized version of the NLN.dtd.

00:39:27-00:40:05

Displaying and indexing and providing that kind of format in search, and also some extensions to the collection feature in HathiTrust to use that for general presentation. I will show you that in a moment. That blue box there is the prepper, which would be "prepper" because it helps you prepare your content. Because of HathiTrust's standards for preservation, the bar for ingest is

actually quite high. We want to make tools to lower that bar considerably and I'll talk about that in a minute.

00:40:05-00:41:04

Sort of the more amorphous part to the left of that is the green box which would be your peer review or editorial workflow. The prepper takes over once you have the manuscript that you want to publish. We want to be able to interface with tools that manage that workflow for you. Here is a mock up, parts of this are actually functional, of a JAX article in HathiTrust. It is repurposing the page turner application. Instead of turning one page now, it is just a long scroll. It's using XSLT and CSS style sheets that are available on GitHub. If you want to make use of these to display your own JAX, you can, but you can also make your own improvements, too. That's great.

00:41:04-00:41:48

There is some branding here for the Journal, and ways to navigate to the collection view. Because of the strict preservation standards, like I said, building a submission information package to submit to HathiTrust can be a little difficult. You have to know METS. You not only have to know METS, you have to know our profile of METS and our encoding practices for the XML content of the article so it renders correctly in our environment.

00:41:48-00:42:26

Here is how "prepper" helps you build that set- first of all, just drill down on itself a little bit. It is using the Baggett Standard. The METS is the wrapper that holds all the meta data. If you are not familiar with METS, the key part is actually the struct(ural) map, which describes the relationship between all these other parts, the JAX for the article, and then any kind of embedded media you might have, or supplemented material that might be available for download but not necessarily displayed as part of the article.

00:42:36-00:43:11

You could build one of these yourself. People do, but "prepper" will help you do that. How will it help you do that? Here's a screen shot of what we have in development. So far, it's not very pretty but that's not the part we are concentrating on at the moment. The idea is that a journal enters into an agreement with HathiTrust, asserting that they have the rights to publish their content and that HathiTrust can keep their content open forever. The journal would also enter in a relationship with Michigan Publishing, which would host and administer this "prepper" instance.

00:43:11-00:43:47

Once your journal is set up, you can come in here with your login (and) create vimes, create issues, and mostly importantly, upload your manuscript. Right now what we are supporting .docx to JAX conversion. We have our own tool that we've been developing called Norm because it normalizes your content. Basically, what you do is you take a .docx template that we'd provide, you use paragraph styles to identify the semantic elements of your manuscript, and then Norm can take over from there and render that into JAX.

00:43:47-00:44:17

Here is where you would include meta data that might not be in the manuscript for various reasons, maybe because it just came through peer review and you don't want to include this stuff in the peer review copy. It extracts the media from the .docx. It gives you an opportunity here to upload higher resolution versions for preservation. We want those if you have them.

00:44:17-00:44:59

(It) previews the rendering. You can see, this is exactly the rendering engine that HathiTrust would use, so if it doesn't render correctly, you can re-upload your .docx file. (It) pre-validates it and sends it off to HathiTrust for ingest. It also would automatically add the article to the special permanent collection in HathiTrust for the journal, which would serve as the main presentation of the journal in HathiTrust. The content records and journal article would be discoverable through the catalog and through large-scale search.

00:44:59-00:45:25

Then, if this wasn't the ultimate presentation in your eyes, you can always pull the content out of the data API. The entire package is here so you could render it in some way that is more useful for you or somebody could build a platform that did that- the more sexier publication view that HathiTrust may not be interested in supporting.

00:45:25-00:46:02

We are developing it based on our use cases, the workflows that we've been using with DLXS and have been developing over the past decade-plus. We are very much interested in opening it up to other parties once we've worked out the kinks in those workflows. We would start out just using it as a replacement for our own workflow doing the production as we do it now.

00:46:02-00:46:27

Typically, like I said, the manuscript is handed off to us and we do to the production, but as tools get better and we get more confident that they are ready, we can open it up to the journals to do some of that production work themselves- get ourselves out of the way of the process (and) get faster turn-around time for them in they are interested in being that hands-on.

00:46:27-00:47:01

Past that, (we want to open) it up to journals that aren't necessarily Michigan Publishing journals. That could take a few different forms. Ultimately, there could be separate hosted instances of "prepper" elsewhere where you would administer yourselves. There would have to be some sort of relationship set up. I think our "prepper" would still have the privileged access to the HathiTrust backend so there would some kind of nodes talking to nodes over a network, but we can visualize what that world might look like, I hope.

00:47:01-00:47:40

Beyond that, like I eluded to earlier, support for peer review and editorial workflow, particularly, I think, and OJS plug-in that would help existing OJS publications. Depending on how much work you've done, they can live on, but when it gets ready to publish, HathiTrust ingests it. You have your preservation solution kind of built in to OJS. Any number of other interfaces like that are imaginable.

00:47:40-00:48:15

Title management: I think this is an opportunity for shared infrastructure that I really want to start thinking about. I don't know about you, and I'd really like to hear about it, but we're not happy with our title management solution. It's quite costly, not only in the terms of dollars, but in the work it takes to maintain it. I think shared title management infrastructure is something I'd like to talk to you more if you'd like to talk to me.

00:48:15-00:49:02

That's it. Any questions? (*Applause, indiscernible shifting of speakers*)

00:49:02-00:49:49

Third speaker: It is great to be here. Thank you very much. I was talking with Maria Bonn earlier and we were reminiscing about trying to put together sessions at DLF a decade ago where it was hard to find speakers outside of what we called the "usual suspects." It is really rewarding to see so many people here and the level of interest and engagement in this topic. It's great. It's also great to be here. I'm from Kansas City. I don't know if you've ever met someone from Kansas City- watch it with the Kansas City jokes. Not that I haven't heard them all.

00:49:49-00:50:28

This presentation may well be a little different than the others in the sense that I'm not really here talking about software that you're going to use. I'll explain why that is. I'm really more here to talk about some of the experiences that we've had over the last 14 years as we've built and maintained the infrastructure to support this project, project Euclid.

00:50:28-

As I've been thinking about these issues, sustainability and scale keep kind of jumping out as real drivers, as real central issues. That's why I mention them here, and I'll touch on them as we go.