**Episode 6: Neil Scully and Adam Snook**

[Intro Music]

**Amanda Ferrante:** Welcome to Long Overdue. My name is Amanda Ferrante and today I'm speaking with Neil Scully, who is OpenAthens' IT Director, and Adam Snook, their product manager. Our conversation will focus on several prevalent concerns amongst libraries at the moment: Data Security and User Privacy. We'll aim to address those concerns and talk about the ways in which OpenAthens enables libraries to manage authentication more securely.

**Adam Snook:** Release... minimal data release... so only releasing the relevant data necessary to gain access to something.

**Amanda Ferrante:** I'll start by speaking with Neil and Adam about those tangible issues present in library authentication, especially affecting the end user experience, that movements like RA21 are working to address.

[music]

**Amanda Ferrante:** What sorts of issues can fraudulent access cause for libraries?

**Adam Snook:** Okay, so I think libraries pretty much have the same issues as any other organization. There's obviously data security, data privacy concerns. Certainly GDPR has brought a lot of that to light.

**Amanda Ferrante:** Do we find that relying on IP access contributes to these concerns at all?

**Adam Snook:** Right, so there's a couple of things. One: people typically going elsewhere. So you're not getting the value that you're putting into the library, I guess, if people are going elsewhere. But also if people within your library are identified as making fraudulent use, then obviously there's fines and reputation, potential reputation damage.

I mean pros and cons of IP. One: the individuals that are making use of the content... from whatever IP it is... they are absolutely a hundred percent anonymous. Which obviously from a user privacy thing that's great, but from an accountability and security perspective it's not so good, namely, I guess, for the publishers. So if peak, I guess, or abnormal use of their website, so too many downloads or whatever. Typically, they'll then block access for the entire organization and in order to either fine you or not fine you, then you need to locate that user to do relevant actions, to find them accountable and whatever that might be... banning them from the library.

But with IP that's incredibly difficult. Whereas, with what we typically deal with around Federated Access. It's a lot easier, I guess, to track that person down, to identify them...the person that may or may not have been misusing things.

[music]

**Amanda Ferrante:** I see. And so how do the RA21 guidelines and recommendations seek to address those types of concerns?

**Adam Snook:** Initially, RA21 was more about the user journey, I think. So with a bad user journey... so trying to access content online... often you'll see Login with an Institutional login or Shibboleth login, OpenAthens login. Again, there's not really the consistencies, people don't know necessarily what to do, and if people are having trouble signing in to things, there's that perception of a barrier, a barrier to access. When barriers are put in place... intentionally or otherwise... that's where people go elsewhere. So potentially that is where the fraudulent activity comes because of the poor user journey that we experience too often online. So I think that is certainly initially where RA21 was coming from: To make that user journey a lot, well, more friendly, I guess. To encourage people to use content in the ways in which they were intended.

**Neil Scully:** It's also about consistency because the way that Federated Access has been implemented across publisher platforms at the moment is often very different from publisher site to publisher site. So RA21 is trying to make that more consistent by creating a set of standards for publishers to adhere to.

**Amanda Ferrante:** And so who informs those guidelines? Are libraries central and considered as a central stakeholder to those guidelines?

**Adam Snook:** Very much so. Yes. RA21... there's a steering committee, lots of people from lots of different organizations. Yes, a lot of publishers, but libraries are involved in that. And there were a few pilots, which, of course, involved library stakeholders and there are still a kind of chance just to feed into the guidelines, the potential standards that might come about and there were a lot of UX workshops, so looking at the user experience side of things. So that covered both publishers and libraries and providers such as, well, the likes of us, I guess.

**Amanda Ferrante:** And I know that in your positions as Product Manager and IT Director, you probably take a lot of tact from the RA21 guidelines. So I'm curious, just at a very high level, how does the OpenAthens service seek to address those same concerns for libraries?

**Adam Snook:** From a user journey consistency approach. So, we've been around for many years. We've seen these problems and we felt that we would like to make the whole thing a bit more consistent. So a couple of years ago we did launch something called Wayfinder and this is, in effect, an organizational search tool or organizational discovery service. So that that will make the act of finding the organization on a publisher site to get sent the relevant login, making that much more consistent. But also things we've been doing for years, but it has certainly been highlighted in RA21, is reemphasizing what Federated Access is all about, and that's minimal release... so minimal data release... so only releasing the relevant data necessary to gain access to something. And then anything for personalization, that's obviously optional, but we've always adhered to those kinds of standards.

**Neil Scully:** We should point out, though, that OpenAthens does provide IP recognition and proxy alongside Federated Access and that's because currently there are a number of publishers who can't support a Federated Access model, so we are not exclusively a Federated solution, currently.

**Amanda Ferrante:** All right, and actually that's a good segue into my next question, which really tried to demystify SAML a little bit because I know that Federated Access is sort of predicated upon the SAML 2.0 protocol. So I'm wondering if you can give us just a rundown of why SAML 2.0 is considered to be more user-friendly and a more secure access method than other common authentication solutions?

**Adam Snook:** From a user friendly perspective, I will just caveat this to say: it's kind of subjective. So there are people out there that prefer the user journey with IP because it's effectively zero touch, but then there's a load of people that prefer the more Federated or SAML approach because of the, I guess, additional features you get around personalization and things like that. But the key thing to, I think, differentiate the two... so IP and SAML, or Federated Access... just want to talk about authentication and authorization. So authentication is... well authenticating a user... so based on a set of credentials, you're verifying that that individual is who they say they are. And then the authorization piece, which typically on the publisher end is about them confirming that okay, you've authenticated this person, you've verified them, we now authorize their access to our content. So with IP, though, in effect there is zero authentication.

You can really embrace remote access with SAML. That's one of the key advantages. So regardless of where you are, what physical location you're in, you can always authenticate the user and validate that they're an individual within this organization. That is, obviously, licensed to access this content.

With that you've got the personalization aspects... the good thing, even if we say personalization...you don't need personal data to make use of personalization. That's one of the things I hear a lot of around conferences. People saying "SAML isn't that secure. You're having to give more data away." On the contrary, you only need a unique user ID, which is called a Targeted ID. Not only is it unique to the individual, but it's unique to the organization they're from and the publisher that they're accessing. So it can't be transferred across publishers or across services. So it's secure in that nature, but it is able to uniquely identify that there is an individual, don't know who that individual is, but you know that it is an individual. So for continual professional development and other personalization, that's key. But also that authentication and accountability piece.

**Amanda Ferrante:** So OpenAthens in its single sign-on can authenticate and personalize the end user using just that anonymous targeted ID attribute?

**Adam Snook:** Yes. The caveat is we provide the ability for publishers to provide personalization services where we have authenticated the individual.

[music]

**Amanda Ferrante:** Now that we've established where some of the more common authentication solutions may not meet the needs of every library, let's get some insight into features of the OpenAthens service that empower libraries to improve the end user experience and protect privacy.

**Amanda Ferrante:** And so I'm curious to know what features are in place in the OpenAthens service that are unique in terms of enforcing security and what features give control to librarians over what data is passed or not passed to providers.

**Neil Scully:** There's two things here that are worth highlighting. It's common for OpenAthens to be connected to another authorized authentication system at the institution end and in that situation OpenAthens only requires a unique identifier to be passed to OpenAthens. More data can be passed into OpenAthens if required. So for example, if you want to be able to run reports based on a set of information about your users, but that's under complete control of the librarian at the customer end. And similarly when OpenAthens then sends data on to publishers, as Adam has explained, only an opaque user ID is sent to the publisher. If there is a requirement to send more data to the publisher, again that can be done within OpenAthens. But that's entirely within the control of the customer.

**Amanda Ferrante:** So it sounds like there are some features in place that help librarians and administrators control what user data is sent where, if at all. And past that we also know that some libraries want to support the ability of patrons to opt out of patron data collection or release. Whose responsibility is it to support that opt out: Libraries as the data owners? Or service providers as data receivers? Or both? And that said, will OpenAthens support a patron opt out in the future?

**Adam Snook:** At present, we do not support opt in or opt out facilities for individuals when managing their data as we are the data processor. We need to provide the means to ensure that organizations are managing data legally and all of that kind of thing. But ultimately it is the responsibility of the organization as the data controller to ensure that relevant consent has been obtained in relevant cases. So if you're sharing their personal data for example. But at present we don't support opt in or opt out. If we were to provide such thing, it would need to be opt in, like from a GDPR perspective. You need to get that consent up front. So, when you're collecting the data or when you're storing the data.

**Amanda Ferrante:** So I want to pivot just a little bit and talk about the other side of the coin, which is the service providers, because they are half of the entities that are using the OpenAthens service in a sense. And so I'm wondering if you could sum up security concerns that also affects service providers? And the example that comes to mind immediately is Sci-Hub.

**Neil Scully:** So publishers rely on libraries to authenticate the users. It's the authentication process that happens that then drives their platforms to unlock access to the content. So the risk from their perspective is if an unauthorized user gains access, they can access and download content they're not entitled to. And if that's then achieved by somebody deliberately doing that on a large scale, and then make it elsewhere for free, then publishers obviously stand to lose significant amounts of money from that.

**Amanda Ferrante:** I see. And so does the OpenAthens service seek to address those concerns for service providers as well?

**Adam Snook:** I would say so. I think, maybe 18 months ago we released a product called Keystone, which is effectively our service provider software for publishers and the like to provide Federated Access. Without wanting to get too technical...SAML is quite complicated and that's basically the thing that underpins Federated Access. So what we've done with Keystone is basically take the complexity of SAML away using more API-like things like OIDC or OpenID Connect. And in doing so, one: It, in theory, is a lot simpler to integrate. Secondly, by providing access via Federated means, you get those additional security gains. The fact that authentication has happened upfront... unlike any publisher that supports IP recognition... you've got the accountability features, but also with Keystone you've got that Wayfinder product that I mentioned, which is the simplifying the organization discovery piece. So whilst we're not directly managing or addressing those security concerns, we're putting things in place to hopefully remove those barriers to entry or those, at least, perceived barriers. So then the user journey is improved. There is less likelihood that they're going to more nefarious sources of the content.

**Neil Scully:** We trying to make it much simpler to deploy than alternative products with service providers. So I mentioned that we continue to support proxy and IP recognition because a number of publishers don't support Federated Access via SAML. With Keystone there's a number of things that it does that other products currently don't do, so it runs entirely in the cloud. There's nothing for the publishers to run on their network and it's based on a protocol called OIDC... OpenID Connect... which is, for the developers working at the publisher end, it's simpler to implement than SAML is, and it's also available across multiple technical platforms. So most common programming languages have an OIDC library that publishers can use to integrate. So by making it simpler, we hope to get, to help make Federated Access more widely adopted.

**Amanda Ferrante:** Do you find that Federated Access, with the seamlessness and the improved user experience, drives increased usage to the content that libraries purchase subscriptions for?

**Adam Snook:** We don't exactly have any evidence either way to back up either argument, whether it reduces or increases usage. But through the things we said in the previous question about helping to improve the user journey by hopefully removing that perceived barrier that would drive more return usage...returners...I don't know the term...people returning to their site. That's certainly the hope. That's why we do what we do. That's why we provide these services to publishers and other service providers in exactly that aim, to help drive usage to their platforms.

[music]

**Amanda Ferrante:** I've worked with many libraries who have a strong preference for Federated or SAML access for all the reasons that we've already laid out. And I'm wondering if you can give us some insight in how libraries can advocate for their service providers to move towards SAML support rather than solely allowing IP or even just username and password access?

**Adam Snook:** I think ultimately it's about trying to promote the benefits, but not only to the library, but also the benefits that it gives the publisher. So with what we were mentioning previously about authentication, the publisher and the library can both be assured the users or individuals, patrons... whatever you want to call them... are being authenticated. That they are valid, verified individuals that should have access to content. From a personalization perspective, anyone that makes use of whatever environment that they're in, that makes use of the continual professional development type things... so if you're in a medical organization, those are the CME credits... personalization is key for that.

Same with any academic or research type organization that will save searches and bookshelves, bookmarks... all those kinds of things... are incredibly useful. With IP, you don't really get that benefit of single sign-on or seamless access that everyone seems to think exists with IP recognition, because if you do make use of personalization... which, granted, not all users do... but if you do, you get automatically recognized by IP. So you get seamless access to the content. But then you've got to register for each publisher that you want to make use of personalization. So then you're volunteering more personal information to many more publishers than if you were using Federated Access. So it's a method of how you could sell that, as it were, to a publisher is they are no longer managing personal data. They don't necessarily need to worry about all those GDPR concerns.

I mean they should, but you know what I mean? Whereas with the Federated approach, you've got that seamless single sign-on aspect because you've got that opaque user ID. So you don't need to register and volunteer even more personal information, it should just be tied to that. Plus the accountability piece.

So if an organization that does make use of IP a lot and they're constantly having all of their access blocked because someone looked like they were misusing the publisher platform. Well, that might be that you had... all of a sudden... the library was packed and you had 1,000 people accessing that publisher. It might not be misuse, but because you don't know how many people are accessing that content, the whole organization had their access blocked. Whereas with Federated Access, publishers can put things in place so that they can just block that individual, so there's a lot less grief and admin on both parts, if that's a concern. And then where Keystone and Wayfinder fit into this is down to that simplicity piece. We're trying to basically remove barriers to... not only content... but to Federated Access. We're trying to simplify how you implement our Keystone product, how you implement Wayfinder so that then you can provide that better user journey to those people that are actually accessing the content: those patrons, those students, those physicians.

[music]

**Amanda Ferrante:** All right, so my next couple of questions will probably require you to look into the future a little bit. I'm curious to know what we might see next in authentications' data security practices, and whether libraries can partner with their it publishers or anybody else to move toward those best practices.

**Neil Scully:** I think in the short term we have to go back to RA21 because I think that that is going to bring about some much needed improvements in the user journey for Federated Access and, much more importantly even, consistency between publishers, and through that there should then be greater adoption of Federated Access and then that should improve security for those sites that adopt it. So I think in the short term that's the most likely thing to happen.

I think further out, slightly longer term, there's likely to be a move towards a passwordless authentication method. So, passwordless authentication is something we already see in many other sectors. It tends to be more secure because what it's doing is it's tying devices to individual accounts, usually using some form of second factor authentication and then it means that users don't have to continually sign in all the time. So it achieves a sort of holy grail, really, of security, but more convenience as well. And then what typically happens is that patterns of user behavior are tracked. And when systems pick up departures from that pattern of user behavior, it triggers an alert that perhaps something suspicious is happening. So I think we will see that.

I think we will also see, user consent will definitely be the mechanism that's used. So the end users control exactly what information is being passed to publishers and they'll be able to decide that on a publisher-by-publisher basis. So at the moment the consent is mostly controlled at the library level and it's whatever decisions are made tend to be across all publishers. But if end users can choose, "I want to release this information to this publisher because I use that personalization features, but when I access content on this site, I don't want to release any information."

**Amanda Ferrante:** So they can basically weigh the pros at the user experience with...whether or not they want to release any of their personal data.

**Neil Scully:** Yes. Exactly. That's right. And then finally I think we will see lighter weight protocols likely to come in, as I mentioned at the start, it's often quite complicated to implement Federated Access. And of course you don't get adoption if things are complicated.

**Amanda Ferrante:** Complicated for the service providers and publishers?

**Neil Scully:** Both parties, yeah. Absolutely. Has to be as simple as possible on both ends. So yes, I think those would be the things I'll flag.

**Amanda Ferrante:** Excellent. And so that actually brings me straight into my next question, which is whether there are any features, new protocols, any sort of adherence to what you see on the horizon for data security practices planned for OpenAthens? Do you plan to adopt any of those features?

**Adam Snook:** Yeah, so I'm certainly keen with the opt in feature and that's namely because it's incredibly difficult to strike that balance, which is incredibly important, between security, privacy, and convenience. So my concern is if we rush in and just do that it will just create yet another barrier or perceived barrier to access. And that's what patrons just really don't like at the moment. And why certainly a lot of them, but by no stretch of the imagination all of them, but why they're going to the likes of Sci-Hub and whatnot. So it's incredibly important, but we need to tread carefully with that. But I am very keen to explore it. We were part of RA21, and we did launch Wayfinder before RA21, but the recommendations, the guidelines, those kinds of things are moving absolutely in the right directions. So I'm definitely keen to ensure that we're still with those, and that's probably about it I think at the moment other than just general security and privacy concerns.

**Neil Scully:** So I did mention a couple of others in that we will almost certainly implement user consent, as I mentioned in the previous question, to give users control over the release of what actually goes to which publishers.

**Adam Snook:** User consent, so opt in opt out, is definitely something we're keen to explore further because it covers all kinds of aspects of the service. There's one: What attributes are we recording within OpenAthens? So clearly need some form of a consent model there where appropriate. Same with... we do have a reporting function that our organization, so our library customers, can make use of to basically get a bit more insight into the value of the library services that they're providing to ensure that things are being and various things like that. And then the third aspect of consent is around what attributes OpenAthens releases to publishers. So what we're storing in general, what we're reporting on, and what we're releasing to publishers. All of those areas are explicitly turned on, or configured, by the library. So we are not doing this without library consent, but we do need to think carefully about how we empower library customers or librarians to ensure that they are handling and therefore we are processing user data legally, correctly.

**Amanda Ferrante:** My last question is really just going to be about where libraries should go if they have questions about this. But before I get there, I actually just want to find out from you both, as the experts, whether there are any other relevant topics or any other questions that I should be asking that maybe I haven't thought of?

**Neil Scully:** As you know, we monitor usage of the service and so we see behavior going out of kilter with certain rules. We then temporarily block users, so the typical example is multi-country misuse. If you see somebody coming from IP address in one country and then coming from an IP address within another country and it's not within a certain time window, then we would pick that up and we'd flag it.

**Amanda Ferrante:** I think that example that you gave actually relates, to my mind, it relates directly to why Federated Access is such a benefit. So, it's sort of peripheral to the idea of if misuse is identified on a particular IP address, a publisher will frequently just shut down access to that entire IP, but rather SAML access allows misuse to be targeted and detected down to an individual user. Isn't that correct?

**Neil Scully:** It is. Yep. Yep. That's right.

**Amanda Ferrante:** Finally, a lot of libraries might not already be thinking about data security and user privacy, but if they do want to broach that topic just for their own edification or for a conversation with their IT departments, where should they go to learn more?

**Adam Snook:** Regardless of where an organization is from, it is likely that GDPR will affect them. RA21 do have a lot of information on their site. And, again, there was that Scholarly Kitchen article about the Five Myths of RA21... I find that quite a good read... so if you want to Google Five Myths of RA21 or Myth-Busting RA21. Otherwise, RA21.org, they, too, have an FAQ.

**Neil Scully:** There's a wealth of material online and there are some really good high quality guides from reputable companies like training companies such as Pluralsight who offer a sort of free guidance and free white papers around both data security and data privacy.

**Amanda Ferrante:** Okay, well that was everything on my list of questions. Any final thoughts that you want to leave us with?

**Neil Scully:** Yeah, so I would say generally that data security and data privacy are incredibly important topics. Any organization that's conducting business on the internet in particular and exchanging information online needs to have those two things, sort of front and center of their considerations and at OpenAthens we certainly do that. We're constantly looking at our practices and our procedures and we're also looking at the features in the product. We need to make sure that ultimately the data we are responsible for, the processing, doesn't fall into the wrong hands and isn't misused in any way.

**Amanda Ferrante:** Thanks for joining us for this brief dive into OpenAthens' security and privacy features and thank you to Neil Scully and Adam Snook for guiding us through common authentication pain points for both libraries and publishers and how OpenAthens' team works each day to improve the user experience and control over the authentication process.

[music]

Thanks for checking out Long Overdue: Libraries and Technology. If you like what you heard, be sure to tune into the next episode. ISBNice talking to you.