THE EVERGREEN PROJECT

ANONYMOUS IN THE FOREST: PERSONALLY IDENTIFYING INFORMATON IN EVERGREEN

11:AM – 1:00 PM

MAY 24, 2021

CAPTIONING PROVIDED BY:

CAPTIONACCESS

contact@captionaccess.com

[www.captionaccess.com](http://www.captionaccess.com)

>> We live in an age where people's data is commoditized. Facebook is looking at potentially billion-dollar losses because of Apple cutting them off. I think it's important that we frame a discussion about patient data, and saves not just patron data, on it. We are given it to provide services to them, and that's all. We don't have an inherent right to it. I think it's an important concept to keep in mind as we move forward we live in a world where discovery owns information about people. How much privacy are we willing to give that consumer? I think we have to flip that and say to what purposes are patrons allowing us to use their information?

So what are we going to talk about in the course of our next hour and 49 minutes? There it is privacy landscape. There is a lot happening at there after many years of governments being slow to catch on to the need to create legislation, to manage privacy for the citizens, a lot has happened, and has been for a few years now. So we will talk about that inherently impact you.

We are going to talk about where personally identifying information is stored and how it is accessed. We want to talk about risks associated with that storage and retrieval. And options to mitigate that risk.

So this isn't just going to be about doom and gloom, were also going to talk about how can we make things better. How can we minimize risk? So in terms of the order of things, we have 4 basic sections. First we're going to talk about how little it takes to be PII. We'll talk about the legal landscape, we are going to think about things everyone should do. I think we might get a couple of laughs in there and hopefully some good discussion. And then we are going to talk about some of those techno things I mentioned particular to Evergreen. Where things are stored and where you should be looking with Evergreen. So what is PII? Information that can be used to dismiss or trace an individual's identity either alone when combined with other personal identifying information that is linked or linkable to a specific individual. It's a broad definition things are pretty obviously PII. Social Security numbers. Email spirit names. All kinds of things can be PII. And I think it's a good thing probably to have a healthy paranoia and ask yourself is something PII? As we discussed, it is easier to be PII -- so is data about someone or something that can be combined with other available information about somebody.

Going to point out a couple resources for people that I think are interesting reading. One is Latonya Sweeney's working paper, from Carnegie Mellon University, called simple demographics often identify people uniquely. Enter is an interesting item from the Journal of Official statistics, finding a needle in a haystack or identifying anonymous census records.

>> I just want to double check that the slides all these awesome resources are going to be available after the fact so we can read these articles and things.

>> Yes. I will be -- after this is done, because Google slides should persist forever, but just in case they don't, I will print a PDF of the slides and should have these links still working. They will be uploaded to the Evergreen website.

>> Spectacular.

>> Absolutely.

What these papers did, and Sweeney use this article as part of her work, is they looked at census records from the 1980 and Sweeney used 1990 records, at the looked at places with relatively low population counts. Some were urban, some were rural. Low population counts, let's take. Minimal data. So in the census record it says this person identified as an African-American of age 90 plus. Then they were able to find them. Because when you have values, this is basic statistics, that are uncommon or you can accept, takes very little to find them. This is especially a problem in our rural service areas.

Anybody who has ever done outreach work, bookmobile work, anything like that, knows that there are rural service areas sometimes with maybe a handful of library patrons. And very little data can be used to find them. And attached information to the activities. In my mind, this is a problem.

So is this a real danger? My answer is yes. Looking back at Sweeney's look at the 1990 census data, she looked at five digit ZIP Codes, gender and date of birth, and found individuals. Not one, not two, she was able to do it pretty reliably. And she didn't have to restrict yourself to rural areas. It became very easy in rural areas, but even in urban areas it worked. So you can eat as well as the year of birth. Low population areas can be most at risk and rural populations are very high risk, but this is not a unique problem to rural areas. Happens in cities, too.

Especially even in high population areas where some is uncommon. What say you go to a locale where a certain ethnic group is uncommon or a certain age group is uncommon because it may be mostly younger people that move into that area. And suddenly, the census can be used to identify them. For the data in the library.

So why do we go overall this? A lot of people may think I'm nitpicking here. Nitpicking everywhere. I felt like this show received by virtual of the virtue of weights of results, a necessary -- but I don't think I take it personally.

I don't think we are nitpicking because we need to keep data that we don't need out of the system. The argument I hear from people is what if we need it one day? And my answer to that is hoarding is bad. Hoarding books is bad. That's why we have reading policies. I will tell you my reference in an adult collection base, I love to read that, neck and take that report and stuff went fast. So I'm anti-hoarding in most regards.

My personal book collection is maybe a different story. I think you can toss out the old stuff. If you're not a government repository with an agreement to hold onto this 50-year-old government pamphlet, you can probably cost them and we don't need those close circulation records.

What do people think? Is the audience here mostly pro weeding of data?

>> I'm moderating, but I would say I am pro weeding of everything. I am the same way -- I was doing a collection analysis and getting my copies per capita ratios in line, and data is one of those things, and everybody knows that if you have had a computer with some type of storage on it, the photos that are on the old computers, the documents that kind of hang there when the hard drive fails.

>> You have to be careful with stuff. My recent experience with old untracked data, I was cleaning out some old boxes in the garage and found some cameras. I found some old photos on memory cards. That is a happy discovery of untracked data. But if you allow things like patriot records or circulation information to show up on a report, and we will talk about that later --

>> And that's the thing, too. And Bradley has a point, walking a tightrope between privacy and keeping enough for robust reports and statistics. To spot trends and things. It's important to keep in mind, Jonathan makes a good point, it depends on the data.

>> It definitely goes back to the point I made earlier about there's no one answer for everybody. I will give you an example from my work. I frequently migrate libraries that have some sort of a statistical category at least we Evergreen would call, with logistical information. That is a topic, I usually encourage library so let me not migrate here however, it worked with a library in the not-too-distant past who said we received certain grants. And we don't like it, but the grant committee in their grant for us requires gender information.

Okay. They have made a legitimate choice. They have thought about it and thought about getting rid of it. They determined they really do need it. And so I'm not going to say that they are doing something wrong. That's all anybody can do is look at something and say do we really need this? Not hypothetically, not maybe if the stars align and the seas rise and that kind of stuff. But they really do need it right now.

This is just have, we do some digital repository stuff. They often are really excited about it though so is a tough sell for your books. Journalists use your books all the time. What is the value? You have to look at that in your collection to make an evaluation for you. Obviously, you have probably done that already and decided it isn't worth the risk. But I can see another library saying we have one journalist that uses our collection, and it is worth it to us. Others watching a documentary on Netflix about the Son of Sam murders. And there were some interesting things discovered based thanks to your book information. Jessica said that she doesn't win the argument often. That's part of what it says feel like Don Quixote. If you hold the banner a lasting patron privacy, and people around you have different viewpoints than yours, then they can be fess rating. I think it's important that we keep that up.

>> And I would say, too, there is in the spectrum, I think of most things as a spectrum, where we have everything is anonymous and/or private, and we have the wild west of everything over there and everybody gets to dig through whatever that pile of stuff is.

And us in the middle, as we are not just weeding, but also curating that information and deciding this is stuff I probably am never going to use, but I can't say for sure is ongoing to put it in this place with these kinds of protections. This is seven probably going to access more frequently, but it still needs to be really sequestered from people unless they have certain permissions and all that.

The yearbooks.

>> Let's get into the legal landscape. I'm one of those strange people that find flaws fascinating. I actually enjoy reading about the laws passed in countries. And I intentionally in going to this non-US centric. We will see why shortly. Even though a lot of us are from the US and Canada, I think it's important to look outside our borders.

And I think by the end of it you may be thinking of Shakespeare's line, kill all the lawyers. The character that says this, is that Richard III? He was not a nice person. This is not meant as something that we should probably empathize with too much. But all the liabilities after can pile up and make you think maybe we would be better off in a world without lawyers. I don't think we would but I can understand the temptation.

So the danger of PII in the system is real. We are going to talk about these laws that are being passed. I believe these are baselines, not goals. If you come into it looking at them saying, we are compliant with the law, we don't have to do anything else, that is a factual statement. You don't have to, but you should stop and ask yourself if you should.

Now, you have to balance things that you may look at some things they leave really should do this task. But we don't have the resources and the threat of this information identifying somebody or this getting disclosed is really low. So you may make a decision to not follow up. But you should stop and ask yourself if those sins are true. So we are going to start with Canada. This website has some information on Canada's legislation involving privacy. They had a privacy Act that is fairly competent of but only applies to federal institutions. Like the United States, Canada is broken up with provinces and territories that have distinct governments. And the government like in the states, have a fair amount of autonomy. And they have laws. However there is a federal, I'm not sure if I should attempt to pronounce this, but PIPEDA, Personal Information Protection and Electronic Documents Act, it is federal. Now,, those provincial and territorial governors bodies may have additional laws. This is going to be a difference with the US.

They have local laws, and they do have a federal law on top. The federal law has a number of exemptions. But you need to know those if you are operating in Canada. Or providing information to Canadian citizens. And I don't have off the top of my head if it is extraterritorial or not but I know if you're providing services to a library in Canada, this definitely applies.

Now, the European Union has the GDPR, the general data protection regulation. This has been kind of a global standard. It's been around for a good few years now. This is the link to it. And one of the things I want to stress about it is that is extraterritorial. That means that if your services target people in the EU, you don't have to be supplying services to an EU library. You don't have to be in the you. But if you have a user base in the EU for some reason, this applies to you.

Now, how granular is that? Ultimately that's up to the courts. But could you potentially have a scenario where a Parisian exchange student comes over to the states, states for a couple years, gets a local library card, and keeps that card for certain services and goes back to Europe and the court would find that your practices fall under the GDPR? Absolutely. It is evolving. It has not remained static. I don't think that particular scenario is particularly likely. The courts have been so far pretty reasonable in their interpretations. But it is explicitly extraterritorial. So if a European citizen uses your services, you do in theory fall under this.

A citizen of the European Union I should say, not necessarily all of them.

China. This is a fascinating one. This is still in draft. The personal information protection law. It is modeled partly on the liturgy that it be that there are. It does have some different to speak. It is more vague. The GDPR is very specific. The PIPL is more vague. Any Chinese citizen who uses your services, according to the Chinese law, has the PIPL, keeping their privacy on your system and you have to conform to those standards.

Again, he single citizen who comes by and get a library card, I doubt you're going to face a lawsuit in Chinese court. But in theory, you could.

>> I have a question about this. And this is just a thought question. So say something happens and this information does make it into a court. And somehow, your database of information comes under a subpoena request. I guess the thought in my mind is, it's not so much a question, but this could potentially end up compromising the privacy of others.

>> Something I have not touched on so far is the ability of these courts to enforce their rulings. It is a big step for a Chinese court to say, you, the state of Indiana, give us a copy of your Evergreen data in plaintext. They have no way to enforce that. They would have to come to the United States and ask the United States federal government under whatever agreements that exist between us to have the US government do that.

>> Absolutely.

>> And federal courts are going to be extremely unlikely to provide wide reaching overstep of that kind of task two and foreign entity.

>> It is China. Any of these laws, for we kind of screwed close to the edge, kind of cutting corners or whatever, do potentially open us up to that kind of what if scenario.

>> A lot at this depends on the sociopolitical landscape. Right now, I find it very hard to imagine that a European court would ask for, say, come to the state of Indiana and say find us all the information you have about every European citizen in your system. That scope is probably reasonable, especially if a lawsuit has been filed. I don't think the US courts are going to find it very high priority to push that along to the state of Indiana. However, the sociopolitical landscape altars, and privacy concerns become much higher between the two countries, and more mechanisms come into existence to freely move that information back and forth, that could change.

>> The other thing, too, is that while it may not even get to the point of here's this text file for these things, but litigation is not cheap. And so even if it doesn't end up compromising patent data at some point in some respect, it could compromise the integrity of the organization from a budgetary standpoint.

It does sound a little chicken little, but at the same point, these things have happened to individuals and organizations.

>> Sure. We live in a landscape now where there entities out there that have a very different profile then libraries generally do. Things like app makers, game makers, who are making decisions to for example, not sell their software in a person that part of the world, because they don't want to deal with those lots. Because not only is litigation expensive, but the form of litigation can be extra nearly challenging. For example, if China were to come after me for some reason, you would need a Chinese law expert. You would need to hire lawyers in China to fight that case. Which provides a whole other level of complication.

So again, the threat to libraries, extraterritorial information, I think is incredibly low right now. But is it a landscape that you should be keeping an eye on just in case things change in the future? Yes. It is something to be aware of. And could I see a court saying hey, we are a court in Belgium and this guy has filed a request to get information from the state of Indiana for his daughter's records while she was a student in the US. Is that likely to be enforced by the US government right now? No. 20 years from now? Maybe.

>> And we have seen GDPR effective development. We do have organizations that fall under that, though it's exit territorial, they are in the territory, so this impacts their day to day. So they would not be able to use those features as we have been talking about, because it requires essentially the opting in for every patron to use a certain part of that to make the future work. So we went away from that to avoid developing in a way that would bump up against the GDPR.

>> And we have to be very aware of these things. Currently, our major user base is North America. But we do have users around the world, and we do want to be reasonably aware, I say reasonably aware because I can't invest the time to become an expert on India's privacy laws, I can't invest the time to keep up with every ruling in the Chinese courts of the indications of the PIPL. But we should be aware of these things.

Let's talk about the US a little bit. There are a few federal laws and regulations that apply to federal entities. Congress. But otherwise we don't really have any federal standards for privacy. Each state and local government has been told to make their own. And they have or lack thereof.

One that is worth pointing out is the CCPA, the California Consumer Privacy Act. It has not been finalized yet, but a lot of states are looking at it as a template to make their own state laws. So I do think it's worth looking at now being able to say in the future, okay, here's this law that affects me. Was a model on this? There will be enough small points that it's worth spending some time on.

One thing about the CCPA is it is a little different from the GDPR. They both are basically framed, CCPA and PIPL, these are things you cannot do with patron information. Do any deposit moment? I see some people having audio problems.

>> I will send a message.

>> I will keep going then. The California consumer privacy Act is different in that it defines things as a series of rights rather than what entities cannot do with information. Citizens have a right to these. It is pretty vague and is intentionally vague so it doesn't get outdated quickly. And we will have a funny list of things I have become outdated and privacy document in a little bit.

And it does have many exemptions. This is one of the things you're going to want to look at and all of these privacy laws is where there are exemptions. Some of them have exemptions for nonprofits, exemptions or organizations of certain size, exemptions for all kinds of things. For example, the CCPA, it really does kind of put the screws to very large entities like Google and Facebook. But if you are running a small business that is not generating millions of dollars in revenue every year, you are probably exempt from others laws. And is a good chance libraries may be, too.

However, it's also extraterritorial. Like a lot because we have been looking at, it says if you are a California citizen, we don't care what server has your information in the world, our laws apply. Is the audio working out at this point?

>> I hear your dog in the background.

>> My dogs are very unhappy probably a score on the deck or something.

>> It's working here, and I see that it is for some people. You will have to translate Brandon's comment. (laughing).

>> I'm not going to get into theories about the US entities, but -- at this point, you might think it's time to put your head down and just like, what can I do? All of these entities are setting up laws all over the world that I may have to comply with, although actual risk is probably really low in a lot of instances. And the answer is you need to be proactive.

My experience with this, I was working at a county library in South Carolina, and South Carolina, the library systems are -- rather than municipalities. We found out after the fact that the local county government had created new worlds that all the County entities had to adhere to regarding privacy. And when I found this out, I asked, what is the implementation date of this? And they said it was months ago. And I said, did you provide notice? We had workshops. Everybody was invited. Can you show me the list? They left the library out.

And as I talked to libraries, this is a common refrain here. They are often left out of these discussions. So you need to be proactive. You need to find out what is going on. You need to talk to your attorneys if the library has some on retainer, and make sure that you are following the rules. What you have to do.

Now, we move on to section 3. Things everyone should do. We are about 45 minutes in. People still feeling good? Does anybody need a break at this point? Before this next section?

>> I would say go ahead and do the next section and break after that.

>> We will see how we stand on time.

Best practice is to only expose the information needed for a specific task. This is something I wanted to call out for people to weigh in. What function do people think occurs in a library that violates this principle that you should only expose what you need for a specific task. You do it over and over every day.

>> I will say what I think for me it's definitely the checkout. Because of opens a whole patron account. And Martha says the same. Checkout.

>> I agree. The circulation desk. This is the horror show right here. And going back to that -- privacy concerns are the environmental waste of our modern world. Why do we expose so much extra information? We do it because we don't know what patrons are going to come up with.

We do it because we find ourselves in these workflows where people expect to come up to the circulation desk and have anybody walk up and be able to handle as many problems as possible for them.

And if they find somebody who can't handle something like voiding a fine or something, they tend to get upset.

>> Absolutely.

>> I presume most of us have worked the circulation desk at some point.

>> And we say checkout, and that's what mostly people are doing, but they are like, while I'm here I can also renew all these things and I just got a new cell phone number.

>> This is the environmental waste of our circulation desks. In theory, it would be nice to break everything up, but people expect it. So that person who comes up to your desk to check a book out, in theory that person should only be checking books out. You should not have access to anything that does not require checking books out.

But is that practical? For most of us it's not. I will talk more about that as we go on. So step 1 of all this. But the first thing you really should do? Have it policy. Make sure it conforms with whatever the roles of your governing bodies are, nestle governments, state governments. Have a policy. And where do you go to start making a privacy policy? You can start with ALA. This page that I am looking here, you can also search for it by searching ALA privacy policy. You will find the link easily. It will give your link for the most part it's pretty good information.

I'm going to tell people, don't just create your policy and let it sit in a file drawer or an aside. Need to have it be part of your technology plan. One of the things that is happening the states and I'm assuming other countries have similar approaches, when they fund your technology, they don't say give us your technology plan for and keep it forever. They want to plan for the future. And they -- I say make privacy policies part of the technology plan because even there is an -- there is an institutional concern, it helps ensure that you will revise it periodically. Because you want to have revising technology plans because technology changes. In many of those changes will also impact privacy concerns.

So that's just a simple way to help give yourself a little bit of security.

So it's harder to forget.

I do want to say that there is a lot of information on that page, there are two areas that I think the ALA has good input on. And even if you're ready have a privacy policy and maybe worth looking at these and going back to your policy and saying, is there anything additional assume you did do this as part of the process to begin with. One is what to audit. The second is questions to ask. Because all of this is about questions. We have talked about maybe not keeping this information for gender. Works for most libraries. Most people say it depends on the data. All of this is true. The only think that helps us determine how long to keep them or what to, I think the ALA is pretty good for that.

But there is some bad there. And I considered glossing over this without mentioning it, but I had to get a laugh when I was reading this. So I decided to share. It is an underscore, respect my statement about making your privacy policy part of your technology plan because you have to revise it. The ALA page mentions specific technologies as a privacy concerns. They list them specific as -- specifically as emerging technologies. This underscores how that page probably needs to be revised. So emergency that -- emerging technologies include smart phones that this was published in 2014, by the way. We can probably debate if smart phones were an emerging technology in 2014. RFC, which has been around since I 14 entities has increased. Social networks. Definitely not an emerging technology. It does have legitimate information that we should be concerned about and people post things on the Facebook pages, things like that. That is potential information you want to track.

This one hurt a bit. I'm building up drama, imagine a drumroll. Interactive OPACs. I want to say interactive OPACs were not cutting-edge in 2014. But this is something people legitimately think about. This one is my favorite. Software. I just want to let that sink in. They listed in 2014 software as in the wording technology.

>> I just can't, Rogan.

>> They actually called them apps, but before you say hold on, you can make it an argument that apps were emerging in 2014. Let's look at how they defined it. This is in the document. A piece of software or program typically small, that can be used on a computer, smart phone, or tablet. So software that can be used on a computer, smart phone, or tablet.

>> They used the actual what it is -- it defines itself.

>> Examples they did are Pinterest, Foursquare and others.

>> They use Foursquare in 2014?

>> It was around.

>> They say emerging.

>> Yes. So if you go to that page to see the emerging technologies you cringe. Don't discount the value of the rest of the document. I do think it has some good information, especially if you are creating a privacy policy.

Step 2. Things that we shall be doing. You had your policy. Now, you need to create a procedure for handling disclosures. Disclosures are a real danger. There can be good disclosures. We are going to use your information to improve services. This could be safe or a third party. We are using this new message that then we want to share information about you with this messaging app and keep you up-to-date on things that are happening in the library. That can be a good form of disclosure. Or we are going to aggregate your information to provide reports to funding agencies to let them know where we need to build new branches or we are going to use demographic information about to show that we need to build the Spanish or Vietnamese collection.

So these are good forms of disclosure. Using information and improving things. Some people think disclosure means the sky is falling, something horrible will happen. Disclosure also needs to include those things that you are using the information for legitimately to improve services.

The bad. This is what people think about data breaches and things like that. And graduations. The patron circulation records are now available on black-market websites where they can also get e-mails of celebrities. The circulation records are thought to be a lot cheaper.

This doesn't happen a whole lot, but it does happen. I know of libraries that have been hit by ransomware. And we do know for a fact or not, but the ransomware may have accessed privileged information patrons and records. So that is something to consider.

And in the context of this presentation, we are talking about Evergreen data, but this can include e-mails and all kinds of things.

And then there is the ugly form of disclosure. I'm using this as an example because this has actually happened at a library I worked at. Circulation staff member was related to a patron and then told that patron's spouse that the person was checking out books on divorce. This isn't a technological breach, this is a circulation staff member had access information they did not need for that transaction. So that's not a security breach because the library made a reasonable decision for their operations, had to make a decision as to -- it is more of a moral breach. Ethical breach.

So after having a policy and a procedure for handling disclosures. You should make it part of staff develop. I think you need to teach staff to handle information with care. To make them understand that they are stewards of patron information to use with care. And to teach them, sure, go into a patron's record to help them. Don't go digging for anything you don't need. And it is important that they hold patron information as a responsibility.

Step 4, ensure that you audit and review periodically. I know I probably sound like I'm just beating this drum on this review periodically, but I so freely see library security policies that somebody made years and years ago and have never been looked at again. I saw one not long ago where they specifically mentioned technological concerns of a previously ILS agents. They had since then migrated to a different solution and they wanted to know -- they wanted questions answered about the ILS in these element of their security policy. They'd apply. These are two specific to your old system. You need to update. I see some folks in Chapter mentioned in some other privacy concerns related to the checkout desk. Been there. And I would tell you that in my time of being a branch manager, that there were many times I had very angry parents talking to me. And I was like, you know what? You want to know what your child checked out, ask them. They are a teenager and you do not have automatic access to their account.

That was true for our library at 13 and under, we would supply parents with information about the children's account. 14 and up? No.

>> And that was in your policy, correct?

>> Yes.

>> That is an invaluable thing for your staff. Otherwise it becomes a confrontation at the circulation desk.

>> And that was based on, we had a policy of I misspoke it was 13 out. Our policy was if you are not a teenager gets, so 12 and under, parents had to sign for a card for you. They were your guardian, and they were essentially the holder of the account. But our policy had been 13 and up you can sign for your own card. So we created a privacy policy that said you have your own car, that is your information and no one else has a right to it. They had a guardian when the card was created, you remove that when it's that they become 13, and it becomes their account.

>> So that becomes a question about fiscal responsibility.

>> That is a difficult area -- that was difficult because we -- that's a quagmire.

>> It is, and I don't except to go into conversation about it or anything like that other than to say that that really does exist because I personally am somebody that will guard a juvenile or adolescent's information to my dying breath. And there is this bill that need to get paid. And it's got an itemized list. And that kid is not responsible for the monetary aspect of it. Someone said they get a card, they take financial responsibility. What we do? That is not the conversation for the day but it's worth thinking about and talking about because it does come into this category of stewarding information and how we handle that.

>> And deciding retention of information, we will talk about retention and a little bit, but when you talk about fiscal element, that adds a whole other level that has to be taken into the policy.

Section 4 is Evergreen itself. I did not do a practice session of this, but we are exactly on time. Do people want to take a five-minute break?

>> We are getting ready to get into the database and staff to see -- are stored, is that correct? My recommendation would be a five-minute break so we can switch a little bit.

>> Let's start again at 1:06. I am going to do the same. I will step away for just a moment.

(Break taken.)

>> Evergreen itself. Nuts and bolts. The greatest risk again, the circulation desk. Circulation information, patron information. This is our big hole. This is the black hole that success in any way. In terms of risk. But the truth is Evergreen is not unique in this regard. Every ILS has the same problem. People expect to come up to the circulation desk and expect people to be able to do everything. One place that every has an advantages we have more granular permissions. This is a screenshot of the permission administration screen, but let's talk specifically about taking advantage of that. This is going to be a project period be honest. Anytime you start talking about adjusting circulation workflows and mapping out new permissions, possibly creating profile groups, you're going to have to start creating flowcharts. You're going to have to do spreadsheets. Maybe poll stakeholders from circulation, and is going to be a whole project. That's just a reality.

And you may partly through go, this is a going to work. We are going to have to give it up. It's not perfect, but that's the reality we live in. At least a privacy issue that you know the dangers and consequences of and are prepared to deal with, but you can't take active steps to mitigate is better than one that you are unaware of.

So if you look at something and see that is not practical to make changes, that's still better than just ignoring it.

But if -- sometimes you can find solutions. Let's look a little more granular late. These are some of the permissions for seeing information about patrons that are relevant. View hold, view user finds, view user transactions, view user. That is kind of a huge hole, because we have this on a relatively few number of screens where I could see a world where we break address information out and have a separate permission for that. Where we break other kinds of information such as birth name and birth date and stuff like that out. All of that would require significant work. None of that is trivial.

As a community, as we start looking at this, perhaps we would want to. What is the down side? Is this doable? Are we nitpicking? Is it practical to actually have one staff member checkbooks out another handle finds and another update addresses? And that's a good question for everybody here. Has anybody here tried to implement something like that? I do know of one library that implement or something like that, or they have simple checkouts, library aids doing that and could not view the rest of the account. If somebody needed something like an address change they had to go back to the main desk. That did not last very long, I will admit. They had a number of workflow problems and ended up having to abandon it. But I thought it was a noble try. Has anybody here done that?

>> We have circulation permission groups that are set up to allow member libraries to potentially do that. There has not been a lot of pickup for it for exactly the same reasons that you describe from my -- my experience plays into it, but also my observation in the years since I have been at a public service desk. There really needs to be an overall look at those workflows beforehand. Because it doesn't only affect those limited circulation staff, it affects everybody else. So there needs to be some understanding so there needs to be a balance. That load is going to land somewhere, and does it make sensory patrons and staff. That's kind of high-level planning as far as how people move through the library.

>> There is a simple barrier of resources sometimes.

>> Absolutely.

>> One library I worked at, it was almost possible, the problem was to do it anyway that made sense with patrons in the space, which wasn't aesthetics kind of thing, it was if patrons are going back-and-forth the runways, people can't see the children's workroom problem.

>> I need staff to accommodate that.

>> That was the other part. We needed essentially another desk that would take up space that we didn't have and we needed at least one person at the desk, and many times oil had two people. So during that time, we couldn't afford for anybody to call in sick, leave, go to the restroom, anything, which means that we would have to essentially double circulation staff.

>> And Chris, point out that permissions are a lot to deal with. Jessica says unfortunately that means they need access to all the information, unlimited access to items. In collections. But they do try to limit what catalogers can do and use the secondary permission groups. That's basically what we have been able to do with Evergreen Indiana as well, as we can limit our catalogers, they don't really care about it anyway, will some I guess probably do, but --

>> Do library catalogers need that information?

>> Sometimes we add those ad hoc and sometimes as a secondary permission group or even as part of our cataloging permission group.

>> We also come back to this idea of it's not one-size-fits-all. Nobody has done this to this point, with an authoritative best practices document for permissions. It would be interesting to do one, but there would have to be really big, bold letters, a caveat that this may not work for you.

Somebody mentioned in chat about small library catalog also work works as a circulation desk. He could split up that into different accounts, does that work -- is that worthwhile?

So let's get into the actual storage of stuff. We primarily store information about patrons are in a schema pattern. Notice we are familiar with database speak. A schema you can think of as a folder. With a lot of spreadsheets. Tables do have data on them. They look a lot like spreadsheets. We will have a column for something like first given name, second given name, family name. And so on. We are going to talk about these and what is stored in them.

So actor.usr is a big one. That is the one that has probably the single highest concentration of patent data in it. It includes usernames, family names, their birth dates, these days preferred names, surnames, email addresses, family members. It goes on and on. All this stuff is personal and in the file. A lot of it is very obvious. Names exist for the very purpose of a terrifying bus. E-mails are that by definition unique. So the risk of this data being publicly available to people who shouldn't have it is fairly obvious.

There are a few fields that are less obvious. I want to spend a moment looking at these. One of these is the ident values. These are the identification values. They don't have to be used. Originally there was a conception in Evergreen that there should be some sort of validating a piece of information that should be in there. But you'd be surprised what I found in there. I have found Social Security numbers in there. Driver's licenses aren't as bad as Social Security numbers, but they are still something that we should ask ourselves, why are we seeing this. You would not believe the stuff I found. And what function does it actually serve? If you're storing a driver's license number, do you actually have resources or workflow to call up the DMV to validate the names and addresses by driver's license number?

>> We have a lot of legacy stuff, too. That migrated in, in some cases.

>> Another field. Photo URL. I don't know how many people are using these. Historically this has been available in the database, but not available for staff to edit there is a pat out there, I'm going to make this available for staff to edit permissions. I know because I read it.

Photo URLs can be very identifying.

And alert messages. Free text fields are scary. Alert messages have larger been superseded by hooks and messages now. And are harder to get to in the modern OPAC. Sometimes that just means you have personally identifying information in there that is still accessible.

So it's something to think about.

What can you do about all this? Evaluate what you store and how. You have to approach that from the standpoint of staff training. What gets stored? Is it okay for a staff member to say in a note, this patron is really stinky and I hate him. And by the way, he lives at blah, blah, blah. There are at least one and maybe more that you don't want in there. So that is a component of staff training saying, what you want to have in those fields. And then you need to ask yourself these questions. Do you need birth dates? Incidentally, birth date seem to be this piece of data that libraries instinctively gravitate to. But do we actually need it? Do we doing frame services off it? Maybe so.

Statistics you say? Do you need the actual birth date? How about just the year they were born? Or the decade? How granular do you need?

I do tell people that digital archaeology is a good thing to start practicing. Stuart inch chat said the year of birth. If all you need is the decade, January 1, start of the decade. Ask yourself how granular you need.

Jonathan mentions in chat that you need to put a date again, but you can make update. And I think there are some options for that. But we can look. I don't think you have to set the birth date as required. I'm pretty sure that's an option. Somebody else in chat can mention that. You might know better off the top of your head than I do.

But you might not click birth dates and you are in a governing group that says on a social level, your organization doesn't need them. Putting January 1, 1900 for all of them. If you need the year, January 1, and the year. There are ways to make this work.

Tammy mentioned that she doesn't have the birth date as required. And Bradley mentions a fake 1900 birth year. I can tell when I migrate data and I'm migrating birth dates, they often don't exist. I may be migrating them into a system where they require them, so they will come up as a bad record. And I often use January 1, 1900. It's easy to look at and say that is probably a fake one.

It's good to do some digital archaeology to see what is happened in the system and even those clutter for a good reason at some point, do you still want it? And stuff, frankly, sometimes it comes from multiple IALs. I have done migration from ivories that have had five, and have mirrored stuff that the library had no idea was in there.

>> Discussion time. I thought this would be fun. What have you seen any library system, doesn't have to be your ILS, it could be some other form of digital archive, newspaper clippings, database, that you have had to remove and clean out? What scares you that might be hiding in records?

>> Social Security numbers.

>> Social skills security numbers.

>> Jessica has that as well.

>> And where they don't conform to standard expectations of format, you have to really dig for them. Stuart says comments fields. Limit text fields. This is not on-topic, but if you do anything in text fields and you put line breaks in, you are a bad person. And you are making life harder for someone like me in the future.

>> John says cleaning up Social Security number is in driver's license numbers. I don't know that it's illegal in Indiana yet, but we are highly encouraging all of our libraries to remove those as well. And just straight outstripped the Social Security numbers.

All codes for invisible spaces.

>> We will be getting to that any second. So we are all good now if we clean this out? No, there are other things. Some people are surprised when I say you need to look at barcodes. Barcodes can be identifying in some cases especially if they are associated with usernames. If you have a username of Jane Doe and that's used as a barcode, guess what? Addresses. There is very little more identifying than address.

And so much more. So all this is stuff that you need to look at. Most libraries probably do have a legitimate need for addresses. But again, ask yourself.

Statistical categories. This is something that everybody talks about. Statistical categories can be filled with personal information. They are just dangerous. But you might need them. Very commonly I see patron types, gender, I mentioned this earlier, age groups, school districts, which I misspelled district period clean that up, municipalities, and these are all things that could be identifying. And they are dangerous. I don't call them bad, and I don't call in bed for a reason. There are things you need to look at and consider.

But the same things can be good sometimes. Let's look at that. Patron types and gender, we have talked about. But age groups, we talked earlier about patrons, and if you turn off requiring birth dates, just keep them as the driven statistical categories. School districts. Maybe you don't need addresses for them, but you need school districts for students. And they contract out that way. Or municipality. Maybe the area of the city or the neighborhood is accurate for your needs.

I would tell you, I am highly doubtful about the value of addresses. I cannot begin to tell you the number of times I have looked at a list of patrons with fines and they have not lived there for years. They are not at that phone number. And if we are going to find them, it requires an investigation pick we are more likely to find an electronic phone book to find and then a library. So I since we question the value of some of the things.

User activity. A lot of people aren't familiar with the user activity table. But you can define activity types, our pre-existing default once, and patients will link them with dates. So you can see things that they have done.

More free text fields, we went through these briefly, messages and notes. They are a danger.

Family relationships. I know we are supposed to live in the age of linked data, and everything been linked to everything is great. But their privacy waivers that says this patient can create a transaction for another patron. We have the ability to link families and groups, all of this is potentially identifying given a data point of linking patients together.

They are things we need to watch out for.

All right. We are back to the frustration level. We are done, right? What do people think?

>> There's always something else.

>> There is more. There actually is. The action schema, tracks what people do rather than who they are in theory. We will see some exceptions. The two big ones are holds and circulations. Going back to PII, it's information about who you are and what you do. Circulations deals with this and people may have legitimate reasons for keeping it private.

But there are other things you may not suspect like curbside. We now have a curbside table in the system that will store information about how a staff member can identify you when they come out to bring you, your staff. So people very reasonably put in things like I'm driving a white hatchback. Here is the urine make. Here's my driver's license -- I mean, my license plate number. This kind of stuff.

How many people thought about curbside data as a potential privacy problem? Circulation history. That has indications. Surveys. Surveys are heavily used in Evergreen, but they list. And who knows within them? Whatever the survey was created for. That certainly can be a concern and something that needs to be evaluated. Should it be used answer to be retained?

So is there a lesson to all this? Yes. There is such a thing as healthy paranoia. Look at everything. And all of this goes back to some of the same questions that ALA proposes designing a privacy policy. I'm not super Pro ALA myself. I have many problems with the organization, but I do think that is a useful starting points. And it does communicate the message that this is about asking questions.

What can you do about all this stuff? The good news is it's pretty straightforward. One, staff education, which we talked about. Two, create a policy for removing inactive patrons. And do this anyway you want. These patients patrons haven't had a holders circulation in years. Now be careful of, perhaps there are only using electronic resources, so you may want to create a way to track that it be validated. Our create a statistical category for them. Get reports from your service like overdrive, use a two-step to statistical categories. So be careful that even if they don't have transactions, they may be active patrons and other ways.

Using authentication activity. That was mentioned in the chat.

But they don't all you fines or they don't owe you much, you can remove them. Anybody who has not been active in five years and old less than $10. I'm probably more generous than that, but I'm pretty skeptical about having to find that person.

And then age circulations and holds. When we are looking at active patrons, this is something you have to do one at a time. It's easy for support staff to do that on the back end if you can give them a clear picture of what they need to do. At age circulations and holds is something that is supported in Evergreen. You can set it up to wipe out this information historically once the closed-end holds removed. And that something that people should consider doing. It does keep statistical information.

Review all content for removal. If you're not going to delete the active pay -- inactive patrons, look at what you don't need. That patron has been there for five years. What's the statistical likelihood that their phone number will still be good? What about 10 years? 15 years? Yet chosen not to remove patrons but can you remove the information? These are good questions to ask. I will be curious to answer Jessica's question in chat. I don't know if anyone would want to use the information on my slides, but they will be posted so people can use them if they wish.

I am going to mention while we are talking about aging transactions, they can be don't have that the anonymous nature removed. And this is the link. There's some back-and-forth on this book. I would love for people to comment on it. I want beat the drum too much this year here. I have an opinion on this book. So people can see what I think there. I would really like to get this fixed. And basically it provides an option for no longer keeping postcode and birth date if you set that as an argument setting. By default right now we do. I think that is potentially an overreach and disagreement the but is that whether or not the default should be turned on or vice versa. In a perfect world it should be opt in, not kept by default. Some systems, where they need that for statistics, like Indiana, I would hate for somebody to miss that when doing an upgrade and have them lose critical information they need for their statistics.

You are free to discuss this.

Community wise. We have that patron purge function. It is really critical to patient privacy. And libraries are purging inactive patrons. It's really critical that it works very thoroughly. Obviously we have talked about a lot of places with the in action schema that information can be potentially stored. So I think it would be nice at some point to do a review of the purge patron function and make sure there are no bugs in terms of missing things.

For example, I don't know off the top of my head, does it wipe the informational information from curbside? I don't know. Those would be good things to check.

I think we should also consider tools for removing all data points. We have these great utility script for aging transactions and I think that's great. Wouldn't it be great to also have similar tools to say, I'm going to pick on curbside again, but these curbside appointments are done. They are not coming back. We are going to remove unnecessary information. I did a full audit for those.

Now, we are going to move on to the single biggest potential threat to privacy in the system. And that is what is in front of you. Reports.

Reports are terrifying. I thought about putting a bracket in here as an in between slide, and everything I thought of that I thought was scary enough to represent reports, I thought the thought might genuinely disturb people watching the presentation.

>> I don't know all those unnamed folders. They are pretty terrifying to me.

>> I washed those out.

>> It was like, how do you find anything? That make more sense.

>> There is that agent with users I didn't feel it necessary to add into the presentation.

>> You made them anonymous.

>> Yes. So reports are terrifying. I thought about putting a slide here of clowns and spiders. So why are reports so scary? They are scary because they are a fire hose. You can do a report that says give me all the phone numbers and all the addresses and all the names of every patron in the system. And just dump it.

Give me other circulations. Give me other holds. You probably want to break that into multiple reports, but we can do it.

And managing that are just these permissions. That's it. Run reports, share report folder, view report output and create report template. It's not very granular. If a user needs to see the patrons and be able to run reports on them, for a bookmobile, they can also get all the information for everyone and everything in the system. If they need to run a circ report, they can also run a sophistical category, and address. So that is a big opening that anybody can freely walk through. Hello darkness, my old friend.

>> We all have that playing in our head now.

>> That is actually my favorite Simon and Garfunkel song.

So what can you do? Reports are open security risks, and you can restrict them to those who most need it. That goes back to staff education. You cannot remind people that have reports permission that they have been given a big trust and they need to be careful. They need to not say patron reports on thumb drives and use them in CSV format. If they don't, they might be subject to the CIA that lost encrypted laptops.

But as always, just like in the CIA, and I don't think the CIA gets compared to libraries much, but one of the holes is that they are not bad actors, their people doing their jobs, acting in good faith, and making mistakes and probably having more access than they should.

What can we do about that? We can create policies don't take this information out of the library, don't send e-mails, not sending information in e-mails is a really good start. Once you send out an email, it's on some server somewhere for a long time.

If you have Google, Gmail as your library email, and you send an attachment, you can probably never get rid of it. So think about how you story report.

Communitywide, I'm going to say upfront this is my personal opinion, it's just my thought, I was out to see us look at options for stripping reports by content per user. Now, this would be a big change it's not going to be easy..It won’t be quick. It would take a lot of thinking before we even start talking about it. But I think it is something that we should think about. It is a very big picture thing.

Finally, we are done. We've covered everything not quite. If you want to imagine the -- playing now, there are some other things you should consider. Outputs and backups. Action triggers. Action triggers allow us to do things like generate XML data for notices. Will any data notices. That is living in an Action Trigger. This person checked this book out on this date. It's due on this date.

That is living in a whole other place. It's not directly connected to a user. Report output. It exists on a server, and they exist on somebody's hard drive, or other drive. We have auditor tables and Evergreen. If your auditing for example, you may have not only information about that user, but every change that is happened to their account. That's something you need to think about.

Data backups. You may be running up occasion, in which case you can probably consider all the versions for all intents and purposes, but there's a good chance that you have some sort of actual backup somewhere. Perhaps even on a separate server. If you have governance rules and you Ashley have cold storage, you may have that. And it should be reflected in our policy that they will be removed at some point. And patrons have been removed and things like that.

And in off-line transactions. This patron was created. Check the book out, and it ends up on a file that should be cleaned out.

So finally, we are done, right? That's everything. We have missed nothing. Sad trombone again. This is my last one. The OPAC. Remember when we talked about interactive OPACs as an emerging technology? It is illegitimate point that OPACs contain patron information. And often we create the problem by installing things like Google tracking. I tell you, if they do that is done, Google has that information. We often need analytics, and I really do think library's could should consider using something like Matomo which is very easy to use and Evergreen. Requires an external service. But as of 3.6, using it with Evergreen is as simple as turning on a few settings. I can tell you I have a cable modem at home, and my I SP has not changed in years.

And that brings us with 15 minutes to go, our time for general discussion questions, Ruth, feel free to add people in for voice chat. The people don't have any discussion points, thank you for having me.

>> Chris has a question about cookies and caching data. He says what about it?

>> It's bad. But we need to use them for web apps to be functional. But we shouldn't be too aggressive about it. It is a difficult point, though, when we friendly, we have taken this thing that was meant to use static documents, and have built it into an -- information provider. We may have to decide delivered some necessary evils there are times.

>> There is another comment here from Jonathan going back to the outputs, remember to delete your old reports. I will tell you I was the guilty one about this because I was sure I was going to go back in there and look at those things for trends and things. I don't think I had patron data once, but I could very easily have been this person. Jessica says I'm sure by e-mailing reports. Never thought about that. I can't imagine transitioning away from that.

>> It depends on what reports. Not all information is PII. If your e-mailing stuff about collections and shelving location and status, that's not hurting anybody. I don't consider information about copies to be sensitive. But if it's a report about patrons, if it's cumulative numbers, if there is anything identifying, that goes on a server that I require access to and identify all this --

>> I want to get back to Bradley here. I think that gets a little in-depth for a second. So using encryption for sensitive reports? Yes. Bradley says using a job to delete old reports is handy. And plain text SIP.

>> I did not plan on talking about SIP today for my presentation tomorrow actually talks -- touches this issue. SIP is a major security issue.

>> So that is going to lead into this thing from Bradley. Thoughts on third-party vendors? This is something we are dealing with hard-core at Evergreen Indiana right now.

>> This touches on a lot of topics I will talk about tomorrow in my presentation, how to connect to external services. So I will fill them in a little bit. I would say SIP2 is like the dinosaurs. It is ancient, unwieldy, probably bad. It deserves to die out, both still occasionally find some dinosaur alive or in the deep trenches of the ocean. We just find it hard to get rid of them. And SIP is like that. I do recommended people that it is the lingua franca of the library world. If your vendor doesn't support and gifted tunneling, they should.

>> Pressure them. Let's talk about an API. But not right now.

>> It would be nice for SIP to be more widely used, but it's not. I will tell you that equinox, does provide supportive tunneling for SIP for anyone who asks.

>> Let me see if there's anything I missed. Also meaning the usernames to clowns and spiders. I fully support that.

>> Third fourth party vendors, discovery records, discovery records I don't generally consider very worrisome. And I will talk about that tomorrow. Authentication gets into SIP2. I'm going to say come to my talk tomorrow.

>> That's legit.

>> I didn't consider these two, Metairie of each other, but I guess they are.

>> They are definitely complement repaired one leads to the other, obviously.

>> I will be talking tomorrow about a number of options for connecting to Evergreen, the full information.

>> That is tomorrow. Do you have offhand what time that is? I'm looking right at it. It's at 1:00 Eastern time, connecting Evergreen to outside services. So that's 10:00 Pacific time.

>> It should be 10:00 Pacific time, I believe. It sort of depends on what you mean by API.

>> Specifically that think called patriotic API.

>> A small spoiler, there is the III -- I took Chris' comments in a different direction. There is a difference between what API technically means and how people tend to casually use it.

When we are talking to a vendor, a lot of times, hate to say it --

>> Depends on the vendor.

>> You are often talking to salespeople who are repeating the things, and they themselves don't know what they are saying.

>> They just want all your data and they want all of it now.

>> Then it somebody else's problem.

>> Are you going to talk about this?

>> I'm going to do it in about 50 minutes.

>> Breads is also vendors, we need local access to everything. Because we do.

>> Tomorrow I will talk briefly about database access.

>> Thank you. Obviously, this is an issue near and dear to my heart. I could go on. Does anybody else have any questions? You have him for exactly seven more minutes.

>> Seven minutes. I will sort of sell this to people who might be interested the present the presentation tomorrow on external services is not a presentation for developers. It is not a super technical presentation Pickett is a presentation for essentially two groups of librarians. Either I'm the librarian who has the technology in my library and they had me talking to other people or I'm a systems librarian. So I'm supporting 30 different things. No, I don't have time to become a developer for Evergreen, even though I might want to. These are the audiences for the presentation tomorrow.

>> Can I expand on your audience for that? I'm thinking specifically of another group and I'm going to use the term broad services but librarians who believe that their administrators can do everything. And I do think that this conversation tomorrow can help provide some context for what is technically possible and what is ethical going to happen. And also the kind of in between for that and provides some context for having those conversations with the vendors who use a lot of technical speak to basically come into they know what they are talking about, so you should do what they say. They are super confident with words. So what you're going to talk about will provide some vocabulary context as well.

>> And I will tell you there are certain of two things that proper me to think about doing that presentation. One is, I was working with one of our libraries who would setting up an external service and the external service kept pushing back that Evergreen didn't change things. Every needs to do this everything need to do that. And I had to have a conversation with the library and say look, all Evergreen is doing is saying whether or not that patron exists and the password. All the stuff that's not working at their stuff. We have no control over that. And I'm not going to say more than that.

The other was, another interaction, where the vendor kind of said well, just have them give us our API box. And this was a salesperson saying that. Just give us the API. And I had to explain to the librarian, they don't mean what they are saying. So we will talk about all that tomorrow.

>> Cool. Thank you. Last call. 4 minutes.

>> And I think we have coming up after me, ASCII.

>> It is from Lynn from Evergreen Indiana and Blake.

>> So that's going to be a must-have sort of session.

>> And Chris is going to be over in track one. And he is doing making the most of Evergreen reports.

>> Chris has done this many times. He's very good at it. He knows it very well. So for those interested in reports, you should definitely check that out.

>> Should we provide them context on what that means? Awesome. Thank you, I really enjoyed what you brought. I hope others did as well.

>> Everybody have a good day and have a good conference.

>> Let the next group start.