

Electric Libraryland - Presentation Introduction

Good afternoon, and I hope you're having a wonderful time here at IUG 2025! Thank you for choosing to spend some time with me today. Before we get started, if you want to take notes, I will not run into the audience to stop you. But if you like, you can find everything about this session at the URL on the screen. You'll find links to various sources that I've drawn on for this talk, like, literally an export of my Zotero repo. You'll find further information, videos, music, heck... I've got my talk written out here in front of me so you'll find a transcript too.

Just to make sure you're in the right place, this is Electric Libraryland. If you're my age, you may have had a class in high school called CWP, or Current World Problems. That's kinda what I'm going to be doing here today, talking about some things I've noticed in the world of library technology. Because, folks, there's a lot going on and it's impossible to keep track of everything. Believe me, I know. My goal here is to give you some things to think about, nothing more.

As the title of the talk implies, we're gonna look at AI, copyright, ownership, the varieties of Celtic music across Eastern North... wait, no. Sorry, wrong talk. And the world of tomorrow. Or something like that. But to get some introductions out of the way, Hi, I'm Dan! I'm one of your Members-at-Large on the IUG Steering Committee. I've been working in the library field for almost 30 years now, which yeah, is well more than half of my life. I'm a Polaris ILS Administrator for Library Systems & Services and the author of *Bricks Without Clay: Demystifying Polaris Transactions* which is not for sale on Amazon because who's going to buy that? You'll find the book for free on the aforementioned URL. I've been working with Polaris since it was in beta, from the front desk to the back of the house and database administration. It's been a wild ride and it's the reason I say that "I'm an ILS administrator and that's why I drink." If you listen to library related podcasts you might recognize me from a show called Cyberpunk Librarian and, as things happen, this talk is going to be so cyberpunk flavoured that I should be talking to you while wearing mirrorshades.

I'm... I'm not going to do that.

But on that note, William Gibson, the father of cyberpunk once wrote that "The future is already here. It's just not evenly distributed yet." As things go, Gibson doesn't write as much

cyberpunk these days because, if you look around, you'll notice that cyberpunk really isn't fiction anymore. Zoom in and look at libraries and you'll see we're living in an age where libraries are a ripe target for cybercriminals, and government agencies and interference. Copyright is in flux if for no other reason than billionaires want it to be. And the Library of Congress is playing with AI. There are all kinds of artificial intelligence things that can do stuff for you. I mean hey, remember that AI demonstration they did at the last IUG? With the cartoon wizard?

Neither do I.

QUOTE ON SLIDE



*Plane train auto hauling space age cargo
Bizzaro technology, every day is tomorrow*

~Aesop Rock. From the song Mindful Solutionism off the album *Integrated Tech Solutions*

You'll notice a few themes running through this talk, but I'll establish one right here at the beginning. Beloved, I believe that you can learn a lot about library technology from music. All you need to do is listen to the right songs. And yes, you'll find links to those through the URL too.

So let's ease into this with something that's easy to talk about and isn't controversial at all.



AI

Oh dammit, my slides are out of order. Welp... we'll press on.

AI is big. That's not a shock. And it's a big problem, and I don't think that's a shock either. There are plenty of reasons to use it, because it's a tool just like any other. And there are plenty of reasons *not* to use it because it's a tool *unlike* any other. It's all over the news, it's all over your favourite software and products, and it's all over your *once* favourite software and products because those products turned to crap after someone shoved AI into it.

I mean, Logitech forced ChatGPT into its mouse drivers. Yeah, I don't know about you but, when I look at my mouse I'm thinkin', ya know... it's just not smart enough.

Problems Within AI



Hackers and mindjackers are on the front lines
While cryptocurrency
is our gold mine
We barely trust each other but we have to
Because we know full well we can't trust you.

~Mike Dynamo, Funan Beat Empire. From the song Megacity, off the album *Retrofuturism*

I have so many cautionary tales to share.

To begin with, let's look at the problems within AI and ask yourself, and others; *please* ask others and see what they say. But ask yourself, why are people willing to give AI a break? AI is so often wrong, so often *hilariously* wrong, that it becomes the butt of a joke. And we're not talking small failures here either, we're talking highly public failures where ChatGPT is a guy standing at a podium and his pants fall down. Look, when you have a host on a popular television show like *The View* talking about how President Woodrow Wilson pardoned his brother-in-law, Hunter deButts... ?

Yeah, right? I *said* AI was the butt of a joke.

There was that time that ChatGPT would crash and refuse to acknowledge the mention or existence of David Mayer. Who's David Mayer? I have no idea, but back in early December of last year, ChatGPT was "unable to produce a response" if you asked it anything about David Mayer. Even better, ChatGPT would terminate that session and you'd have to open up a new chat to ask something else.

And, apparently, people are okay with this?

Here's a scenario you'll vibe within: you sit down at your desk, and you need to write a quick email. As you write, the spell-check in your email software is pointing out your spelling mistakes. One problem, though: You're not making any. You're reasonably certain that you've spelled the word 'librarians' correctly and you're absolutely sure that you've nailed the spelling of the town your library is in. Additionally, it's checking your grammar -- it's never done *that* before -- suggesting that you change certain phrases in a way that is not compatible with English.

Two questions: How long do you put up with that? And how much trust do you have in your spell-check going forward?

AI is a fascinating misnomer in that artificial intelligence is neither artificial nor intelligent. It's a complicated guessing machine that applies mathematics to language. And check my

sources, there's a hell of an article by a guy named Baldur Bjarnason and he talks about how the intelligence in AI is something *we* see because we want to. He calls it The LLMentalist Effect and he likens it to the same process you'd see with the classic psychic con artist.

You have a self-selecting audience. Chances are good that, if you don't believe in psychic abilities, you're not going to spend any time with someone who claims to have them. People who believe in the intelligence of AI *want* to believe in it.

Now, that in mind, your psychic GPT sets the scene and the first thing it does is manage your expectations. In the same way a psychic might tell you that their spirit guide doesn't always speak clearly, or that things are hazy in the ethereal mists. You're told "this is new technology, it's not perfect" or that "it's early days" or "this is a beta." That way you'll understand that, ya know, it's not going to be right all the time. And it's heavily implied that these things will get better with time.

With a psychic, they're going to ask you what you want to know or who you'd like to communicate with. There aren't too many tales of a psychic walking up to some rando at the 7-11 and saying "Hey, Brad, your mom misses you too, but she thinks you need to move beyond your grief and live a full and happy life. Also, she stashed a couple thousand dollars under the hardwood floor beneath the bed. Cheers, lad." No, the first thing the psychic needs to know is, well, what do you want to know? And that's the first thing an AI needs too. Sometimes, probably most of the time, you ask your question, get a response and move on.

But that's not always the case, and sometimes, and this applies to me... maybe we need to "talk" with the AI to learn more. Here's where you have to be careful because you are not having a conversation with *anything*. You are embarking upon a game where you supply information and the AI will try to guess what you want. It's like 20 Questions but with exponents. Many times it guesses correctly but all too often, it doesn't. The thing is, that we've been handed a word to describe this that, for some reason, is different from the word we use to describe anything else like this in software. If you find that your ILS is doing something wrong, what do you call it?

It's a bug! When you put in a ticket through the Supportal what choice do you select? You select Defect / System Issue! But not so with AI, no we call that a *hallucination*. We say that ChatGPT or Claude or Bard or whatever *hallucinated* the wrong information. Folks, that anthropomorphizes and imparts a humanity onto a mathematical model in a manner that it does not deserve. It's not a hallucination, it's defective.

And in a recent study, AI researchers and computer scientists at New York University discovered that an LLM only needs to have .001% of its training model corrupted with wrong information or misinformation in order to produce wrong answers with high confidence levels. Silicon Valley has literally created a digital Dunning Kruger machine and they're

marketing it as being the next big thing. It's a standard belief in library work, in research, and in intelligence work that it's better to have no information than it is to have wrong information. Current AI products not only venture outside this foundational philosophy, they blow right through it.

Show of hands, how many of you have far less trust in Google than you did two years ago? I mean, not only is it the case that someone who quickly finds what they're looking for doesn't spend as much time looking at ads, but also that generative AI and LLMs are poisoning the wells of knowledge and ideas with information that's not just wrong but, in some cases, *horribly* wrong. Like putting glue on your pizza wrong, or potty training your kids by smearing feces on a balloon.

One of the biggest problems with AI, at least in the realm of "how it works," comes down to the expertise problem. It's a kind of a paradox in that, if you wish to study something, you must first learn everything about it. When you're dealing with an AI, you're asking questions. But we all know that the AI will just straight up lie to you and deliver the wrong answer. The problem is, you need to know enough about your subject to spot it. I'd say at least 80% of my queries are about coding. How do I do something in Python, how do I create this function in PHP, how can I automatically calculate the first and last day of the previous month in SQL? I've asked all that stuff and I have gotten back some BS, mes amies.

But I know enough about Python, PHP, and SQL to see the answer is wrong. And often, even when the AI is wrong, I gain an insight into what I need, and often that provides the answer anyway, at least by accident. I would never ask ChatGPT about, I don't know, recurring headaches because I'm not a neurologist and neither is ChatGPT. I don't know enough about the subject to know if I'm looking at a wrong, and possibly dangerous, answer.

But lots of people will absolutely believe whatever Chatty G drops on them.

Now, I look out among you and, lo and behold and goddess above, I see a lot of women out in the audience. Ladies, I'm not even going to ask you this question because I already know the answer and that answer yeah, hell yeah, you've put up with some misogyny in your lifetime haven't you? Well, I have bad news, because AI has a problem with women. It turns out, that if you train your little demon in the framework with a bunch of data generated mostly by males? Problems arise.

I don't need to tell you about the shortage of women working in STEM fields. I bet y'all already know something about that. And what this means is that you have a lot of *dudes* imparting *their* opinions and *their* wills upon these mathematical models. The training data alone, and we're gonna talk about that training data soon enough, but the training data alone shows heavy biases against women and against the GLBTQ community. Amazon found itself

in the news when it got out that its AI based resume screening service had a real problem with putting women to work.

And in case that *wasn't* obvious, let me tell you a short story about three women working in AI. Their names are Tasha McCauley, Helen Toner, and Dr. Timnit Gebru. It's a real short story and here's how it goes: Once upon a time three women voiced their concerns about how AI is being used and developed and then they were fired.

From the data to the delivery, AI is unwell and those diseases are called misogyny, homophobia, racism, and ageism. A tutoring company was sued by the US EEOC because its AI powered recruiting system automatically rejected women 55 and older and men 60 and older. A healthcare AI failed to flag Black patients when assessing the need for high-risk medical care management. An investigation by the Washington Post discovered that at least eight people have been falsely arrested based upon an ID provided by a facial recognition AI system. All of these people were Black.

It's 2025, and people think machines are intelligent, but if they are, look what they were taught.

I'm old, okay? The grey hair is a dead give away. But how many of y'all are familiar with a basic tenet of computer science called GIGO?

Seriously, they used to teach this stuff in schools. In the third grade we had to learn about this before the teacher would even let us touch the Apple IIe. GIGO is an easy to understand concept and it means Garbage In, Garbage Out. You put garbage data into any computer system from an Apple IIe to an LLM, you're gonna get garbage data back.

One of the funniest things about AI is that it *violates* this tenet. I know, I mean, I *know* my website has been scraped by every AI crawler out there. And, since I've been using this image as my logo for, oh, twelve years now, I wanted to see if it ChatGPT might give me my own logo back if I ask for it. So I ask Chatty G "Hey, can you create a logo for a podcast called Cyberpunk Librarian?"

It gave me this.

Folks, I spelled "Librarian" correctly. It was part of the *input!* It took good data and made garbage. And this kinda thing? Happens all the time.

Speaking of which, where did they get that data? The training data, I mean? Carl Sagan said "If you wish to bake an apple pie from scratch, you must first invent the universe." If you want to create an LLM or a GPT, you need to feed it. That's gonna require a lotta data and apple pies. So... how do you get it?

Well that's the easy part, beloved. You *steal* it.

AI & Copyright

I don't know about you, but I have literally watched the IIS log files update as OpenAI and Anthropic's Claudebot crawled the OPACs of the libraries I work for. They're harvesting all of your cataloging data from the titles to the summaries. From the subject headings to the call numbers. I guess we might rejoice just a little, because, after all these years, an information system that lives outside the library world is *finally* interested in MARC records!

Which reminds me... I wonder how OCLC feels about AI companies slurping all those records?

But there's something that I don't think is being talked about enough and it's that, if you... as a librarian... as an information professional... if you are using a commercial AI and I don't care which one it is...

You're trading in stolen goods.

Now, dear hearts, I'm a shadow librarian. I have no problem with this. But there exists a wild and crazy group of librarians who still believe that copyright has some kind of sanctity and if that is you, then you need to understand that every commercial AI could not exist... *cannot* exist, without the largest theft ever perpetrated upon humanity. If OpenAI or Anthropic or Microsoft had to pay for the content they stole, well... I can imagine numbers that big, but I've never seen that many of anything. Every single commercial AI vendor you have ever heard of is being sued. Because Silicon Valley stole everything; blogs, movies, newspapers, music, magazines, books, scholarly journals, YouTube, the data of your library collection and every other library collection. They stole it all.

And they're fine with it. Indeed, Mark Zuckerberg went on record and said that most creators overvalue their content. Their stuff isn't as important as they think it is. Zuckerberg opined that there needs to be a discussion about copyright and a relitigation of how it works. "When you put something out in the world, to what degree do you still get to control it and license it?" Well, last I checked, Mr. Zuckerberg, I get to control it for the rest of my life and then for 70 years thereafter.

Last year, Mustafa Suleyman, Microsoft's CEO of AI, said that anything on the open web was "freeware" and that anyone could copy it and use it without permission, and certainly without compensation. And Sam Altman of OpenAI said that content creators should be happy! They should see the theft of their work as a "significant upside benefit" and they should be proud their work is powering artificial intelligence.

And then these tech bros go crying to the federal government and to the press when DeepSeek came along and did everything as well as, if not better than, ChatGPT's models.

And the AI CEOs had the audacity to be angry when it looked like DeepSeek yanked their training data. DeepSeek came along and stole from *them!*

It's very clear. The tech bros believe that the only copyright that matters, are *their* copyrights and the copyrights that belong to their business partners. Your copyright? Not so much. You're probably overvaluing it anyway.

You need to understand and expect that everything you put online; the bibliographic records in your catalogues, the events in your event calendar, your social media posts, the articles posted on your library website, the photos you've shared on Instagram... Anything and everything you put online will be taken by an AI company and used without your permission.

Period.

Some people will tell you that copyright was broken in 1998 with the passage of the Sonny Bono Copyright Term Extension Act. I'll tell you that copyright was broken on April 16, 1971 with the invention of FTP and then again in 1986 when BBSes started using the ZMODEM data transfer protocol. With these two inventions, it was laid out that, if it could be digitized, then it could be copied and shared.

Meta, OpenAI, Microsoft, Apple, Anthropic, and others. They downloaded everything they could and no one could say anything about it. By the time we realized what was stolen from us, they already had the data they wanted. In 2013 a young man took his own life, a life that was destroyed by a prosecutor and a legal system determined to make him *pay* for his crimes against copyright laws.

His name was Aaron Swartz. He helped develop RSS, Markdown, and was involved in the early stages of the development of Creative Commons. His crime, if you'd like to call it that, was that he downloaded a lot of articles from JSTOR. JSTOR let it go, but the federal prosecutors called it theft and went after him with so many charges that other prosecutors and judges criticized their methods.

To my knowledge, no one from any AI company has been prosecuted for anything under the Computer Fraud and Abuse Act.

Teresa Nielsen Hayden is a sci-fi editor, author, and essayist. Cory Doctorow mentioned her in one of his articles, specifically quoting something she said. "Just because you're on their side, it doesn't mean they're on your side." Last year the Library of Congress undertook a study to see what would happen if they used AI to catalogue digital materials. The results weren't all that surprising. When it came to LCCNs and ISBNs, the AI had around an 85% success rate. Makes sense, because LCCNs and ISBNs are fairly predictable in their structure

and design. But, when it came to figuring out subject headings, genres, and more in-depth stuff? At or around 50%.

Amikoj, that's a *coin flip*.

But the question you should be asking is hey, why is the LC using looking at AI instead of, you know, cataloguers and librarians? Just because you're on their side, doesn't mean they're on yours.

And before we move on, can we just, for a moment, acknowledge the horrific environmental impact of AI? The University of Illinois Urbana-Champaign projects that the water usage for AI will lie between 4.2 and 6.6 billion cubic meters by 2027. That's *fresh* water usage, by the way. I hope we don't need any of that water for, ya know, other things.

But it's okay, Microsoft is going to be firing up a new energy source for their AI efforts. They're proud to announce that, soon, they'll be drawing upon the electricity produced by the Crane Clean Energy Center. Which is a brand new name for an old facility originally called Three Mile Island, the site of the worst commercial nuclear disaster in the history of the United States. You'll notice here that only one set of cooling towers is in use. The other two haven't been used since 1979 for... reasons.

Two years ago, Microsoft and Apple were jumping up and down to make sure people knew about their eco-initiatives and all of the things they were doing to go green. I bet you don't hear a single word about that at Build or WWDC this year. And if you do, you *point and laugh*.

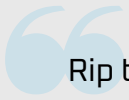
So, manage your AI use responsibly. There's more riding on it than a rewritten email with a professional tone and fewer profanities. And keep in mind, every single major AI product you've ever heard of is deeply, and embarrassingly unprofitable. Sooner or later, someone is going ask about all that smoke from the money burning machine.

Copyright

So let's go back to copyright for a minute. This, too, is big problem, and it's only getting bigger. I won't go throwing around conspiracy theories. However, I will point out the incredibly apt timing between the Internet Archive losing its battle with Hatchett and it falling under a cyberattack from a so-called hacktivist group that didn't seem to have a whole lot of reasons to launch such a coordinated attack. I mean, literally, it was less than a month between the two events. Maybe it was serendipity, who knows?

However there are groups of people out there more organized than a hacktivist group and their pockets are much deeper. These groups are quite interested in removing media and content, effectively black holing it.

We call them corporations.



Rip the heart from a community
Their culture surely dies
They call it progress but who moves ahead?
The people who control the wealth

~The Unseen. From the song Cultural Genocide, off the album *So This is Freedom?*

Corporations are responsible for the loss and destruction of more media than you can imagine. I'm sure you've heard of some of this, like how NBC erased early episodes of Johnny Carson's *Tonight Show*. Or maybe how the BBC did the same thing with the first episodes of *The Avengers* where John Steed played second fiddle to Dr. David Keel. Honor Blackman didn't show up until season two. But that's not it, friends. The rise of personal streaming video on the Internet was a massive event in the history of broadcasting. Who knew that watching someone play *Minecraft* would be so entertaining?

Pity so much of it has been deleted by Twitch and YouTube.

Speaking of which, how many of y'all circulate video games? Anyone having troubles with some of these games being rendered unplayable? Either through online multiplayer server shutdowns or server side distribution systems going dark? Because you may know by now that, often, when you buy a game disk... the game may not be on it. No, there's just a bit of code that tells the console where to go to download the game. And if that game had a multiplayer component, powered by online servers? And if the company who runs those servers decided "Heyo, that's enough of that! Shut 'em down!" Well, you can't play that game anymore.

Remember what Ms. Hayden said about being on someone's side? It doesn't mean that they're on your side? In 2024, the US Copyright Office, a division of the Library of Congress, decided against a petition from video game archivists and researchers. They wanted a DMCA exemption so they could archive, study, and provide examples of games, even after the companies shut them down. Needless to say, this denial makes it much harder to create an archive or a library of video games. The Entertainment Software Association praised the decision, saying:



With today's decision, the U.S. Copyright Office confirms that the current level of video game preservation is appropriate and recognizes that video games present unique copyright concerns that should always be taken into consideration.

Let's be honest. The current level of video game preservation at the corporate and ownership level is functionally zero. There is only one company that regularly does anything with its legacy games, and that's Nintendo.

That aside, the response from the Video Game History Foundation is far more interesting:



The game industry's absolutist position—which the ESA's own members have declined to go on the record to support—forces researchers to explore extra-legal methods to access the vast majority of out-of-print video games that are otherwise unavailable.

Extra-legal methods.

Welcome to the wonderful world of anarchivism, something you're likely to see more and more of in the coming months and years. So much so that your librarians working in the front-of-house may find themselves drawing upon the work of anarchists to provide information and media for your patrons. Anarchists share a few things in common with AI companies in that, sometimes, if you want to save something for later, you need to appropriate it without asking permission. Or, to put it another way, we make a copy of it. Or, to put it another way, we steal it. But you see, the difference between anarchists and AI companies is that, when an AI company does it, it's progress and technological development. When an anarchist does it, it's a felony. I guess the distinction lies in that, you need to steal so much of it that no one can count the cost.

There are plenty of other reasons something can disappear and, as things go, music is one of the biggest triggers. For instance, take the original *WKRP in Cincinnati*. For the younger folks, this was a sitcom set in a radio station that played rock music. Like, real rock music, on the show. So you'd hear Pink Floyd, Rolling Stones, Heart, Boston, Janis Joplin and so many more. I love this show, it's one of the reasons I'm a podcaster and one of the reasons I run an Icecast server. Now, do you have any idea how hard it is to clear music rights for a television show about playing music on the radio? One that ran for four seasons and 90 episodes? By the time this show came out on physical media, some of the original songs didn't come with it. So if you wanted to watch this show, in its original form, you had to rely on those extra-legal methods. The same exact thing happened to *Beavis and Butthead* and *Daria*. In some cases, episodes of all of these shows were re-edited by fans to re-include the original music.

And it's not just the entertainment that disappears is it? In the months leading up to the Trump administration, anarchists - some working in groups like the ArchiveTeam and others running solo - got to work copying, downloading, and distributing data deemed to be "at risk." Things like weather data and modeling, the kind of stuff you'd use to track climate

change, which is something the current administration denies the existence of. Things like security camera footage from The US Capitol on January 6th, 2021. The National Institutes of Health are a focus for preservation as well since they are an organization previously targeted by the first Trump administration. And along with that, resources and data for the GLBTQ+ community. It could well be that, if someone needs that data, well, they're going to find it on a completely different website than where it originated.

I won't pretend the appropriation, indeed the copying and redistribution of this data, isn't stealing. Of course it is. It's the same basic act as that time I wrote a Python script to scrape PowerPACs for MARC records. Totally a proof of concept, because, hi, have you met me? I hate MARC. As Hunter Thompson once said, "I'm a relatively respectable citizen. Multiple felon perhaps, but certainly not dangerous."

Copyright means nothing to the tech and AI companies. And when it comes to archivists, we're having a hard time trying to figure out why it should matter to us, especially in situations where culture, science, art, literature, and data are in danger of vanishing forever.

Ownership

Ownership is an important thing in the library world, but I'd argue that the most important thing you own right now, and the thing that you need to work the hardest to protect, *isn't* your collection. It's not your physical items. It's not your digital content and DLC. It's not your facilities and its certainly not your policies.

It's your data.

Your data is powerful. It's *valuable*. And it's incredibly valuable to cybercriminals. Please note, I'm not using the word "hacker." Hackers aren't cybercriminals, and if you need proof of that I invite you to talk to one of the folks who participated in the IUG hackathon. I deplore the use of that word to describe someone who commits crimes using computers and networks, so I won't be doing that.

No I'm talking about people looking to make a dollar and they're more than happy to leverage your data to do it. Find someone who runs an SSH server and ask them about how often its being scanned and probed. It's all automated. You can script intrusion, and it's not even that hard. And for all y'all out there running Polaris, always always remember that Windows Server is a well known, and widely understood attack surface.

So besides you, who else is your data valuable to? Darknets? Sure. Ransomware groups? Yes. Three letter agencies? Absolutely. How many of you have had to comb your data because of a subpoena or a search warrant? Take a moment, and reflect on just how much personally identifying information you have. Not only names, addresses, emails, and phone

numbers, I'm talking circulation history and, I don't know about you, but circulation histories seem to be the most interesting thing.

One of the worst things that can happen to your library is that it becomes the main character on social media. In 2023, the British Library suffered an intrusion and ransomware event. Not only that, but 600GB of data was stolen by a group of criminals called Rhysida, who were also responsible for successful attacks on SeaTac International Airport, Insomniac Games, the Chilean Army, and the City of Columbus, Ohio. That same year, Toronto Public library was attacked.

In 2024, Calgary Public Library, Seattle Public Library, Delaware Libraries, Solano County Library, and the public library of London, Ontario were attacked by various cybercriminals. These days, ransomware is the big thing, but data theft is a good side hustle. Not only that, but Library of Congress saw an email breach in 2024 and I think most of us know about the attacks on the Internet Archive.

Which brings me to two questions: What do you own? And do you really need to own it?

As MC Frontalot once said:



...the only way that you could pray to make impression on the era ahead is if, instead of being notable, you make the data describing you undecodable

~MC Frontalot. From the song *Secrets from the Future*, off the album *Secrets from the Future*

Because, you can't hide secrets from the future with math. Now, the easiest way to make your data undecodable, is to not have it at all. You have a special power in the ownership of your data that you can decide that you don't want to own it anymore.

There has never been a better time to look at your data, especially your patron and circulation data, and ask yourself "Do we really need that?" If you take a look at the screen here, you'll see a Polaris SQL query that delivers the entirety of a patron's circulation history, from holds to checkouts. All of that information is tucked away in the PolarisTransactions database and it's just sitting there waiting for someone to misuse it. I'm bringing this up here because I worked with a library, a very large library that you all know that I will not name, and they had no idea this kind of data was being stored. I watched some eyes grow real wide when I ran through this query on a Zoom call and showed them how much data is being retained.

I can't speak to Sierra because I don't know that much about it. But I can tell you that Polaris has the ability to purge patron circulation history from the Transactions database. If you're not using any of that functionality, then I invite you to investigate that, and I invite you to bring it up in every single meeting you have with administrative people. There are so many strategies to getting the data you want, and then getting rid of the source of that data in order to protect and maintain the privacy and security of your patrons. As of Polaris 7.7, this is the standard list of details added to any check out. I know some folks worry about demographics and where patrons live vs what they check out. As you can see, the patron's PostalCodeID is right down here. You can still get that demographic data while anonymizing the patron themselves.

But let's step back just a little bit. I'm *not* going to ask for a show of hands on this but I bet, I just *bet* that some of you have seen a certain phenomenon in the upper levels of administration and maybe at the branch management level too. It's a phenomenon I call "The library with the largest patron count wins."

Purge your patrons, people. Seriously. That patron who hasn't used their card in the last five years? They've haven't set foot in a library in the last 1,826 days? You know, day number 1,827 probably isn't the day they're coming back either. Why are you keeping that data? That's data you not only don't need, but you don't *want* it. At best, it's out of date. At worst, it *isn't* and, if your library is unlucky enough to make the news for the wrong reasons, that's data that can be stolen.

And believe me, I work with 24 different library systems. I know the challenges inherent in convincing some librarians and administrators to purge patron data, and all of the whereas's and wherefore's that go with it. What if they owe money? How long should we wait? What if they have associated patrons? I get it, but your patrons need your protection and one of the ways to protect them is to let their data go. If, by some miracle, on day 1,828 that patron comes strolling back into the library and their account was purged? You know what you do?

You make 'em a new library card! Welcome back! Now you have good data, and a sign of life. That's something you didn't have before and, if you're purging patrons on the regular, no one else can have it either.

Now, hard truth time, and the clue is in the documentation: **Removing Patron ID Information from Transaction Files**. Transaction, as in PolarisTransactions, specifically PolarisTransactions.Polaris.TransactionDetails. Those who fiddle with the Polaris databases will remember that there is a primary Polaris database and a Results database too. Circulation information with patron identifying data might be retained in Polaris.Polaris.ItemRecordHistory or Polaris.Polaris.ItemRecordHistoryDaily. Did a patron owe a fine on an item? That will be recorded in Polaris.Polaris.PatronAccount. Did a patron receive, say, a hold notification that their item was ready? Well, you'll likely find that

mentioned in Results.Polaris.NotificationHistory and
PolarisTransactions.Polaris.NotificationLog and perhaps even
PolarisTransactions.Polaris.EmailLog depending on how the notification went out.

But at the very least, you can anonymize that data.

Purge your inactive patrons. Sanitize your circulation data. For your active patrons, collect only what you need. The base minimum for them to operate within your library policies and to provide the desired data. When I hear about a library who wants to add a driver's license field to their Polaris UDFs, I die a little inside.



Deleting every record, you know what I mean
Puttin' mind on my money, downloadin', flee the scene, clean

~Rockwell Knuckles. From the song *Urgent, Important, Please Read*, off the album *Our Pathetic Age*

Conclusion

Things are hard out there, beloved. And I think they're gonna get harder. There's gonna be more and more AI slop in the media and the metadata than ever before. And, folks, that stuff is like toxic waste. You can't easily get rid of it and it contaminates everything. You'll need to question everything now, and there may be no easy answers.

Libraries are under attack in a way they haven't been since 1933. But if librarians love something above reading and metadata and catalogues and media... it's talking to other librarians. Hell, that's why we're here, isn't it? Keep talking, and keep supporting each other. I've worked in libraries for a long time now, and having things like the IUG Discord and the Forums? I'm not throwing out a plug for these services, but I'm damn sure going to praise them. Having a place to go, where you can find solutions to the problems that plague your job and hamper your patrons. Where you can get the latest info and find out about weird bugs and even weirder features. A place where everybody knows your name. It's like *Cheers* but you have to bring your own beer. That's not just a resource, it's a support system.

Find your support systems. Keep looking for answers. Help those who need it. Run the data, share your knowledge, share the media... and be kind.

Well, I've got a piece of news, son
That might make you change your mind

Your life is historically meaningful
And spans a significant time

~Bad Religion. From the song Slumber, off the album *Stranger Than Fiction*

Extra Notes

Add Coe's quest to links