Evergreen Conference

Opening Keynote

May 25, 2021

>> KATIE GREENLEAF MARTIN: Good morning everyone --- as you are filtering in, if you have questions you can put them in the chat. Thank you.

Good morning everyone and welcome to the Evergreen Conference. I see a lot of people introducing themselves in the chat and I noticed quite early on the west coast so we are thrilled to have everyone with us today. Let's see...Andrea, do you want to share your screen or should I leave this up?

>> ANDREA BUNTZ NEIMAN: I have some slides. You are looking at a beautiful large logo of the Evergreen international conference. Thank you, Katie and welcome everyone to the 2021 Evergreen International Conference I am Andrea Buntz Neiman. last spring reality of the pandemic crashed down on us and a 2020 conference which was in progress for over a year was one of the many events called off period

In an effort to snatch conference victory from the jaws of defeat there was a mad scramble on behalf of a motley crew of Evergreeners from the outreach committee in the event that we put on in six short weeks. This is something we should be proud of and I'm amazed we pulled that off last year.

Unfortunately by late 2020 it became apparent we could no longer reliably count on the safety of an in person 2021 event. The board in local planning committee made the difficult decision to cancel the planned conference.

The outreach committee was once again tasked to head up the online conference. And a slightly less motley group took up the reins. This time we had six months instead of six weeks and we were able to proceed with a bit more structure than 2020s mad dash. Personally, I'm really sad that we missed the gathering two years in a row and it says something may be that I need more of a social life but the Evergreen conference has always been one of the social highlights of my year truly and attended every single one since 2009.

I know we've all been struggling to navigate interpersonal relationships in a changed world and find this normal and I miss seeing every single one of you in person but that said there are definitely things that are easier with an online event. The cost of putting on an online event is significantly lower. Then an in person event. Things like live captioning and video recording and because prohibitive to do at an in person event of our size but online these are far easily achievable and preserves content for future community members and those who cannot attend. Lower costs means the overall attendee cost is lower and easier for people to attend who cannot get away for a week for an in person event.

This year am pleased to say we have our, some presenters presenting from the European Union, but I do not think would have been able to make it to an in person conference on thrilled we've been able to expand to reach this event thanks to being able to do it online. I hope we can gather in person again soon, but I also hope that the community continues to offer online offense either as part of in person offense or as their own thing --because it does provide a different way of providing information.

To return to that slightly less motley crew. At an in person event I would embarrass them all by making them stand up awkwardly so maybe this is another plus for an online event. Most every single one of these people on your screen are conference many monomers but some are also project board members, committee and interest group leaders, conference presenters or panelists and conference moderators but some of these people are several of those things. And those people do not like to sleep. I call myself chair but I want you to make no mistake that this was a group effort and these people all deserve a time of credit for figuring out how to put on a large scale online event including sponsorships, social events and an exhibit space so thank you to all the committee members for all your creativity and hard work. Thank you.

Speaking of people to think I would like to be among the first to thank our event sponsors, their support of the conference and Evergreen community is very appreciated. thank you to our champion sponsors Hopin has been sponsored by the Evergreen development initiative. Close captioning by Caption Access has been sponsored by Mobius.

yesterday conferences was sponsored by COOL, and Equinox.

thank you to both of these organizations as well. I want to thank our Allied sponsors Bibliomation, Emerald, Quippu, Statcourior and Unique.

I'd like to turn the floor over to Jason Boyer, for a few brief remarks. Jason?

>> JASON BOYER: Thank you, Andrea and for those of you who know me fairly well -- hello everyone welcome again to the 2021 Evergreen online conference. It is good to basically see you all here in a manner of speaking. I am excited for what the committee and presenters have put together for us all this week. I am certain this will be a great conference with a lot of valuable lessons to take away.

A lot has happened in the last year, even a couple of good things. You might've missed if you're distracted, the Evergreen project is a full 501(c)(3) entity as of September 2020. And we've also had some great releases with 35, 36 and this year's 37 bringing features like interfaces to the client, and entirely new OPAC scan, suggestions based on the records you actually have in your system and not spellcheck. Pretty nice stuff. I asked Andrew not to give me too much time to fill and she has kindly obliged but before I go I did want to reiterate how great it is to always see the community come together and share knowledge and explains online and in person. I do hope these online events continue possibly become more infrequent but after this week what I'll be looking forward to most is sitting at the crowded table in the hotel restaurant discussing database performance or arguing about type systems or maybe just finding out whether or not the pizza is any good at the place down the street.

Looking forward to seeing how this week goes and look forward to seeing you all in the future? Thank you all and have a great Evergreen Online Conference.

>> ANDREA BUNTZ NEIMAN: Great, thank you so much, Jason. I'm going to wrap up with a few practicalities about the conference and then the platform and then we will turn it over to the keynote.

First, a few notes about the schedule. All conference events are hosted here on Hopin where you are now, it will be helpful to think of sessions as rooms, so as physical spaces and then when you click on the session, that will let you access the individual presentations, there are sessions for track one and track two and special session for this opening keynote and then that will also be true for the closing --there's also sessions, separate sessions for lightning talks.

There are open discussions room for cataloging information. And then we will have exhibits saved through Thursday. and lightning talks will be today at 3:30 PM Eastern and tomorrow at 2:30 PM Eastern, the sign-up sheet will be posted in the Hopin reception room. There will be gaming and social time because what is an Evergreen conference without gaming you might've seen in the reception room, the social rooms built there and that will be Wednesday the 26th, tomorrow, 7-9 Eastern in the evening. So please join us to hang out and play games.

There's also the development and documentation hackfests which will be Friday the 28th all day. You can still sign up for those -- those are free to attend and those are on our website.

We are thrilled we are able to provide an online exhibit space. And thankful to our eight exhibitors listed here. I definitely encourage you to visit the Expo area of Hopin and visit the exhibitions. Today through Thursday 12-4 PM are the main time spread some exhibitors are also maintaining, doing demos at certain schedule time so please check the main conference schedule page on the Evergreen website for information about specific exhibitor hours and schedules.

I mentioned one of the things that we can do more readily in an online event is recording and captioning so be aware you were told this when you signed up at most sessions will be recorded. This included yesterday's pre-conferences as well as lightning talks. Keynote and conference presentations.

We will post these recordings within the next couple weeks on the Evergreen YouTube channel as well as the transcripts from the captioner --slides will be posted on the program description page of the conference website and just a nudge to presenters, please email your slides to the Evergreen conference list as you finish them will start nagging you again in a week or so about that.

Most events will also be live caption, the moderators will post live caption links at the beginning of each presentation.

Interest group meetings will not be captained or recorded, and hackfest rooms and social events will not be caption or recorded some other parameters there.

Next the code of conduct and video policy. This code of conduct applies to everyone associated with the event in every space is part of the event including discussion groups and social events in case you encounter an incident contact responder to let them know. Please be aware that all of the responders on the page are also presenters so you may need to email two of them at once or something like that if you have anything to report.

Hopefully we will not need that but we always want to make sure that is available at every event.

This is also subject to the community photo and video policy. Please note that the board did approve some changes to this policy this year because of online events we did talk about coming up with a granular red board around everyone's Avatar box but that became complicated. So know that cameras are always optional for all speakers and attendees and you do not ever have to share your camera if you do not want to.

In fact, in most moderated sessions you will need moderator approval to share your camera or speak.

Make sure that your careful screen sharing if you are screen sharing. I'm okay with you all singly slightly cluttered office behind me but just be aware of what is behind you if you are sharing your screen.

Finally, social media. If you are active on social media give us a follow and keep posting about the conference and use the official hashtag evgils21.

last but certainly not least, thinking to all of you, everyone who is speaking, presenting, moderating, attending, leading a discussion and all the other ways you make this a viable community. Thank you so much and thank you for being a part of this event.

I will not turn it over to Katie to introduce Becky Yoose our keynote speaker.

>> KATIE GREENLEAF MARTIN: Thank you, Andrea, and Jason for the welcome and remarks. I am thrilled this morning that we have Becky Yoose with us and she is the founder of Library Data Privacy Consultant for LDH to help navigate privacy and for over a decade Becky wrangled Library data in its various forms in academic and public library. She received her M LIS from the University of Wisconsin Madison in 2008. It has been a certified information privacy professional in the US with the international Association of privacy professionals since 2018. You can find her online yobj.net, and should follow her on Twitter at yo\_bj.

thank you to our keynote sponsor, Equinox Open Library Initiative in the Evergreen development initiative and Mobius.

I'm thrilled to turn it over to Becky for her keynote presentation, The Curious Case of Library Data Privacy.

>> BECKY YOOSE: Let me get my slide set up for you all. You should all be seeing my slide deck, so welcome everyone, and thank you so much for joining us today. I do see there are some West Coast folks who are along with me. I am in Seattle so I hope you have your coffee or tea ready, thank you for joining us.

Special thanks to the conference committee for the opportunity to speak to you today. planning and online conferences difficult and complex without a pandemic on top of everything, so thank you for the conference committee for making this conference a reality for today. While I'm sad that I cannot meet everyone in person I'm still happy to have the chance to see you all in the virtual world.

Unfortunately, speaking of pandemics, you probably hear a little construction for the pandemic has not stopped construction season in Seattle so I do apologize for the construction noises as they come and go during the session.

I live and work in the unseeded traditional land of the native people, the first couple of Seattle. The tribe were the first signatories of the treaty of Point Elliott in 1855. However, the tribe has been denied the rights established in the treaty for over 165 years to this day. United States federal government currently does not recognize this tribe, denying the tribe the rights and protections of federal recognition. The Duwamish, tribe are still here and they are living on their land. A portion of the speakers fee from the conference will be donated to the Duwamish, to provide financial compensation to the tribe for use of their land and resources.

The tribe has started a petition to send our state, to send your state congresspeople to create and support a bill in Congress that would grant the tribe federal recognition. The link to the petitions on the slide www.standwiththeduwamish.org.

there are a couple of housekeeping notes before we began, there will be issues pertaining specifically to the US version of librarianship which might not track with your expenses with librarianship where you are located since this is in an international conference, I use citations for the talk and you can find the citations and references for the stock on the URL on the slide which is is.gd/evgils21yoose.

this talk also contains references to abuse, racism, sexism, sexual assault and rape culture. Do what you need to do to take care of yourself during this talk. Even if it is just closing the browser right now and then coming back to the recording when you feel better prepared to deal with those topics.

Lastly, I will not be taking questions at the very end of the talk however, you can still reach me on Twitter or email if you have questions. And with that, let us begin.

Welcome to the obligatory origin story. One where I tell you that I loved reading as a kid and I practically lived in the public library. And living my lifelong dream of being a librarian.

Well, that is not my story. I was the kid would watch weather radar or peruse channels for hours --stuck into my brothers room to study the offer 1980s personal computer and dot matrix printer. I took over an hour for that printer to print a tree with three colors but I was going to watch that printer print from beginning to end. It was three colors, you don't understand, a treat with three colors, and the dot matrix printer.

I want to know how everything worked for the introduction to the World Wide Web in my teens was a game changer for a kid living on a farm in Wisconsin. I practically lived in Yahoo directory pages and delved into HTML, CSS and Basic but while the computers were located in the library, I didn't do much of this learning and library spaces.

There was that one time where I was asked to leave the public library, the library worker either didn't approve of me posting and admittedly teen angsty poem that I found in a student publication that I found need as a teenager to share to the world. Or more likely, I violated library computer use guidelines by posting to the discussion forum on a public computer. Now did I mention that my oldest sister, the one who supposed to be the librarian in the family, worked at the public library and was there when I was asked to leave.

So that was a little awkward. Fast forward a few years. I ended up working at my undergraduate university library. I started shelving books giving me an excuse to explore the collection. I have not been in a collection that big so to me it was like walking through that analog version of Yahoo directories. How did it become arranged this way, though? What were the processes? And structures that collected, organized and made these resources discoverable to patrons. So how do library actually function?

It was then I was introduced to technical services, and library metadata and from there the rest is history. Shortly after graduating with my comparative studies in religion degree, I went to the University of Wisconsin Madison for library school specializing in organization information and technology. I went from creating applications for catalogers and technical services to administering library systems to library IT management. But now most of my work deals with data privacy so you all might be wondering what is so special about library data and privacy that I have made a career choice to focus on that?

When you work in technical services or IT, you will quickly realize that you cannot have a library without metadata. There have been times where de-funding tech services prompted thoughts about the ILS database which would bring down the library catalog and other patient services. Now, while this thought was mainly therapeutic and not an actual plan to pull data plug, that's the official disclaimer, there was no actual plan, it does illustrate how the data is vital for libraries.

You need metadata to make your resources discoverable and accessible to your patrons said you need people to create, maintain and process metadata. You can have a library without decent metadata and this is a hill I will die on.

So here is another third truth, the current environment makes it impossible for operating without patron data. They rely on patron data for things like item circulation and providing access to electronic resources. What complicates this truth is that a patron cannot use the library nowadays without generating data. Without going through and avoiding generating data.

A patrons entrance into the physical library is captured by a door counter or card swipe, patrons movements are tracked through security cameras, if a patron has a smart phone that Wi-Fi might log the device or Bluetooth begin searching for phones in range in the library. We can complicate this even further by mentioning that the phone signal location is being tracked by that third, by a third party but we will get back to that.

A patrons computer use is captured by the computer reservation system and possibly the computer itself logging in applications, or use and web activity through specific browsers. Network activity generates data through system logs be it on the computer or through Wi-Fi and the print management system is recording patron activity through its own system logs. Did I mention that the security camera is recording that computer area including the computer screens. I am not done yet. The copiers memory storage retains the scanned image of a patrons document and the smart speaker that sits on the information desk and pick up conversations between patrons and library workers. And the heat sensors used for that big beautiful digital display to indicate which areas of the library are busy or crowded, capture body outlines. We can forget about those RFID tags in the books, films and audio materials that patrons check out at the self-checkout machine. There might be an RFID taken in that library card as well and we also have to mention there is a library event going on in one of those meeting rooms were a library worker is recording and taking pictures of the event if a patron decides to attend.

And then there will be handed a feedback form for the event and I will stop there.

This is just an example of what patron data can be generated from a visit to the physical library. Patrons using online labor surfaces generate data through visiting the library website, searching for catalog, accessing electronic resources etc. We can talk about web cookies, system logs, search logs, chat transcripts, emails, social media comments, however, there's only so much time that we have in this keynote talk.

So I will leave it there. All that that I mentioned in the last slide is compounded by the outsourcing of library services and functions to third parties. The trend towards data hosted services decreases the level of control on the library site and increases the vendor's ability collect and retain patron data, our libraries enable the generation of data. the trail of data created by patients through their use of the library reveals who we are in terms of our behaviors, interests and habits.

It's near impossible to separate the person from the data they generate. Even when we talk about data aggregation, or when we describe data through metadata, we can still get a relatively accurate picture of the person generating that data. So in essence, we are data and the library has a lot of it.

Some of you might be saying, Becky, we are librarians. And we are charged as a profession to protect patron privacy. That is fine, let's find out what that means.

We talk about privacy we first need to talk about intellectual freedom. The American Library Association or ALA describes intellectual freedom as a fundamental right for people to access information without restriction. The relationship between intellectual freedom and US librarianship with the adoptions of the ALA code of ethics in the Library Bill of Rights in 1939.

1939 was a busy year and the introduction to the Library Bill of Rights through the ALA Council gives us a clue to why that was. I called, today's indications in many parts of the world point to growing intolerance, suppression of free speech, and censorship affecting rights of minorities and individuals, end quote, the majority of these threats to free speech and censorship mostly overseas -- this doesn't mean that privacy only matters when government is involved but it's hard to untangle privacy from intellectual freedom due to the historical intertwining of the two.

Privacy in libraries is essential --if you ask library workers why it is critical, you are more likely to encounter intellectual freedom and protection against government intrusion as two reasons why that is.

There has been a shift in recent years regarding the boards of libraries protecting patient privacy and particularly patron data from exploitation by third parties. One example is article 7 of the ALA Library Bill of Rights. Adopted in 2019. Article 7 states that all people regardless of origin, age, background or views have a right to privacy and confidentiality in their library use. Library should advocate for, educate about and protect people's privacy. Safeguarding all library use data. Including personally identifiable information.

This is just one of the recent efforts around patron privacy from ALA's choose privacy every day and library privacy guidelines, to various efforts in organizations such as the library freedom project. This work led to changes in general awareness of issues of privacy --to privacy services to patrons. And these are all welcome developments in patron privacy but reality of patron data is that it is like litter, it is everywhere. And it is hard to wrangle.

The patron data generated through library use goes through a lifecycle, there are six stages in that cycle, collection, storage, access, reporting, retention and deletion. Each stage has a certain amount of risk attached. So let's take the first stage which is collection.

The data you decide to collect will determine what harm is caused if that data is leaked or breached. If you do not collect the data it cannot be leaked, reached or improperly used. Collecting driver’s license numbers in the patron record for example puts patients at risk for identity theft. We can mitigate risk by limiting data collection to just what is tied to a demonstrated business need. In this case, why do we need to collect driver’s license numbers?

Sometimes a data collection question reflects the need to review a policy or process. We do not need a government issued ID number on file if the business need of our process is to verify the patient's physical address and in fact, a driver’s license requirement will create a barrier for patients who do not have those licenses. Instead we can change the address verification process to creating a note in the patron record, stating that staff clarify the address without collecting additional data such as license number.

Some of this data that libraries collect is needed for library operations. At the same time, other data such as driver’s license numbers in the earlier example are not essential for operations. Unfortunately, libraries are not consistent in determining which is which.

This is partly because of the pressure to use data to prove one's worth or value to others. Without data it is hard for library to build and sustain support from the administration, government officials, taxpayers, donors and other decision-makers.

The need for data feeds into data FOMO or fear of missing data points that can be used to justify the existence of a project service or library. We tend to collect data. Because we think it might be helpful. So we do not have a solid case for collecting it now. So there's also the argument that patrons should not worry about data collection and processing. Because librarianship has a robust code of ethics pretty increasingly, we place ourselves in the role of information fiduciary claiming to have our patrons best interests in mind in our dealings with their data.

As long as we stay true to our ethics, there should be no problem. Correct? Data collected is data at risk. A typical example of this risk plays out and how this risk plays out in libraries is malware.

For example, for academic libraries out there, you have to worry about silence library and a malware attack reading fake library emails to phish library patients for user credentials providing the bad actors access to organization's internal networks and assets including sensitive data.

The vendors we work with collect patron data for their own purposes. Vendors go beyond web analytics software to elect patron data through the use of proxy services, third-party integrations, user accounts and so on, all this data can be used to create user profiles or resold to other third parties.

In some instances, like LEXIS-NEXIS contracting with ICE vendors that collect patron data put patrons at risk by creating and selling harmful products built on the data collected from those very patrons.

However, some of our own practices --in-house practices can be as invasive as a malware attack or a data hungry third-party app. We surveilled patrons through security cameras for example. Many library privacy state laws do not explicitly state if security camera footage of a patrons use of library resources such as recording that patrons Internet browsing session on a public computer screen is protected from disclosure like information in the patient circulation record.

We are replacing real-world connections with data. What traditionally has been a two way relationship between the library and community becomes a commodified one way relationship. Where library's market services divisions based on which segments patients belong to --vendors have been doing this for a long time. But more recent is the adoption of customer relationship management systems or CRMs combining patron data with external data sets mostly from data brokers giving levers access to personal data about their patrons well beyond what is asked for on the library registration process. Such as income, education level, race, ethnicity, gender identity, religion, political affiliation, and sexuality.

Libraries can choose from a variety of CRM's and data analytics products from library vendors and from libraries purchasing these products this purchase is a response to being stuck in reality that demands data to do our work but we can justify this -- as long as ethics there should not be a problem.

The Santa Cruz civil grand jury investigated the Santa Cruz public library system and the use of a data analytics and market segmentation product. The jury found that the library violated patron privacy expectations as well as professional ethics and standards in their use of the product. What makes this report worthy of attention in the library world is that this investigation was not done by library workers. It was not done by ALA, it was not done by any other library professional library body. But it was done by people who can very well be patrons of that public library.

So, getting called out on privacy issues by patrons is not the best look for libraries to keep talking about library privacy and ethics. Libraries keep saying that privacy in the library ultimately serves a patron. I do not know that as a profession, if we are in consensus of that.

This is partly due to the ethics we keep referring to when we talk about library privacy. When we talk about library ethics, we talk about the various statements from the ALA code of ethics be like article 3 about protecting the patient's right to privacy while using library materials. But I think there's a better way, to start figuring out what we mean when we talk about library ethics that goes beyond going line by line through each article.

The following sentences come from the opening of the ALA code of ethics. We significantly influence or control the selection, organization, preservation and dissemination of information. in the political system, grounded in an informed citizenry, we are members of our profession explicitly committed to intellectual freedom and the freedom of access to information.

This follows the same arguments and emphasis on intellectual freedom and knowing what would be in the patient's best interest. But in addition, we must practice a high level of neutrality concerning our selection and distribution, again, in the best interest of the patrons, at least we endanger our professional duty, our ability to create informed citizen patrons.

And here, my audience, we meet that one term that has defined library ethical discussions and that is neutrality. The code of ethics does not shy away from neutrality as a core ethical standard. Such as article 7 stating that personal convictions should not conflict with professional duties. So if we oversimplified ethics, they need to create an informed citizenry that is vital to our modern democracy. So, like most of you be -- might not be happy without summary, it doesn't really clearly define library ethics.

But perhaps that is by design. Ethics are notoriously hard to define. What might be ethical to one person might not be ethical to another. Library ethics is not immune from this. Nevertheless, humans like having binaries, tell us what is ethical and what is not.

So we are going to do that, let us talk about what ethics are and what they are not. Ethics are not values or goals or vision statements. Ethics is as former US Supreme Court Chief Justice Potter Stewart puts it, knowing the difference between what you have the right to do and what is right to do.

Ethics is doing the right thing even when no one is looking. Ethics is recognizing the differences between the consequences of what you consider as right and what others consider as right. Ethics is moral justice.

Equally as importantly, ethics is harm reduction. Bringing this back to library ethics. You might start to notice why we cannot have a conversation about patron privacy without acknowledging the harm created by our actions. Libraries are a product of society. For example, the early public lepers in the US served in part as the place for "uneducated immigrants" can be educated in the proper social norms of the dominant social class. Libraries aid in the patriotic mission of cultivating that informed citizenry, but we exclude marginalized groups, libraries or defenders of intellectual freedom it's not only allowed but welcomed into library spaces.

We provide access to a variety of information and claim to provide access to a variety of information, but this information that we provide access to does not reflect our communities diversity. We hyper focus on one way to organize information at the exclusion of equally valid knowledge systems.

It is important to be aware that libraries reflect the society in terms of the norms and beliefs of privileged and dominant groups. As a concept, neutrality serves as keeping that status quo for those who are privileged or in the dominant groups.

Nevertheless, we still need to suss at the library privacy question regarding who benefits when libraries choose to use privacy practices and technologies. Suddenly you might be stopping me here because you would probably be saying at this point the issue is how we use that technology and not the technology itself.

In essence, technology is neither ethical or unethical. It is neutral. This is where I introduce you to that wall of research stating that technology is not neutral. The creators of many algorithms and technologies argue that the systems are objective and unbiased. Ruha Benjamin counters this technology as objective narrative with the concept of the new Jim code defined as the employment of new technologies that reflect and reproduce existing inequities that are promoted and perceived as more objective or progressive abandoned the discriminatory systems of a previous era.

Pick any book or movie listed on this slide and you will be flooded with examples of how facial recognition software can't accurately, can identify nonwhite faces. Search engines return them images of mugshots when searching for groups of black teenagers and when a search for groups of white teenagers produces little to no mugshots in those results. You will learn that personal assistants like Siri were programmed to make jokes and snappy comebacks. But they weren't programmed to help someone searching for help after a sexual assault or an episode of domestic violence.

You will find countless stories of how algorithms denied people jobs, insurance, desperately needed medical care and housing because the algorithms and data used to train the algorithms are biased towards certain races, ethnicities, gender identities, sexualities and disabilities --will find that even someone's identity, someone's name will disqualify them from opportunities if the name is not common white sounding name or a typical man's name --even if the person is more qualified than their white counterparts, the technologies we create reflect who we are and what we believe and what we value.

Open source is also a product of society and carries the biases, beliefs and values of those who shaped the movement. Open source is collaborative but also meritocratic product to a fault. People can participate in many ways.

Many want to be a part of the open source community so the community's actions have made it clear they are not welcomed ranging from micro-aggressions to community talents of abusers, we cannot create neutral technology for neutral libraries because neither is neutral. Open-source communities cannot release a product to the world and absolve themselves of any harm in the use of the product. This is another method of neutrality in and how the tech is used as the user's problem and not the creators.

As companies and businesses are scrutinized and theoretically faced consequences if their products are used for harm open-source communities can and should be subject to the same scrutiny and consequences but so what happens when we remove that rotting deadweight that has neutrality from ethics in tech, it forces us to start thinking about how we affect the health, safety and well-being of people in our communities. And in that sense, many people and open-source do make that effort.

Open-source communities art knows stranger to advocacy. Centered around ideas and values such as digital equity, accessibility and privacy rate in terms of library privacy open-source communities in librarian there are systems created that do not trade privacy.

But we can only go so far with ethics to guide open-source communities. One path to harm reduction is licensing that requires developers and users to abide by ethical standards and principles. This is illustrated either current efforts of, by the organization for ethical source. Currently, there are seven principles guiding this effort and you see this on the slide -- ethical source, common good ethical privacy, accessibility, community safety and compensation. These licenses can only go so far in harm reduction.

When we get into the practicalities of legal enforcement of licenses is costly. It is also very hard to get a community to get a document to list all uses that are and are not allowed, even if they are guidelines. It's notoriously hard to get that to be agreed upon.

The ethical source movement is heading in a good direction. But for some of us we might need something more concrete or something more immediate. Here we can turn to the integration of ethics in the design process by integrating ethics into each stage of the development lifecycle we have greater control to create technology and processes that align with ethics. The ethical design guide provides various tools and resources about developing technology-based on inclusive design principles and reducing harm embedded in the systems design. This guide is a great source to learn about ethics and concepts that otherwise are not represented in our own codes.

The ethical OS toolkit is another resource developers can use to identify ethical issues through eight risk zones, surveillance, bias, data control and monetization. By identifying potential ethical problems with the technology early on in the development process, developers can remedy these issues before deployment. They do include these URLs on the slide but there are also in -- and the library that I linked earlier in the beginning of the presentation.

Nonetheless we still need to adjust how libraries use this technology within the data lifecycle. We have to acknowledge that the practices around technology are as important as the creation of technology and so it is both the creator and the user that need to be involved in this. We also need to talk about inheritance. At the beginning of library automation the digital library systems mirrored analog systems that we created to organize information. Straight down to the formatting from the cards in the library card catalog. And to a certain extent, we still will live with that legacy, it is Conway's Law all the way down. Conway law says the systems rebuild is it reflects the automatic systems. organizations are shaped by values, beliefs, values and ethics. What we decide to adopt from other industries or from tech fields also has their own inheritances. Their own Conway law of reflecting the human structures and biases of its creators.

Libraries adherence to neutrality as ethical standard leads or systems to create harm and an example of this inherence and bias problem is library discovery systems. Our profession treats information retrieval as a neutral act which ignores effective systems have.

various examples of bias and library discovery systems in this recent book Mass by Trust Bias in Library Discovery. When you have search results about hearsay in the US for search about rape in the US or you get culture results when you specifically searched for rape culture in the US. It is hard to argue that our discovery systems cannot cause harm. another inherence problem we have is the way our systems generally deal with data trade

For example, why in the year 2021 are we still finding vendor identity data fields in the ILS patron records? Why? Why do libraries believe that having the gender identity of the patron is essential for the patron to use the library? Why do we have systems and processes in place that make it impossible for patrons to avoid creating a massive amount of data exhaust even with the smallest of library interactions?

It is clear that this is a multifaceted problem, so how do we realistically mitigate harm to our patrons?

Welcome to the world of data ethics. In the privacy world, we talk about different breaches such as security and data breaches. Data breaches failed to handle data with organizational or professional values. Data ethics can minimize the risk of an ethics breach --general definition of data ethics is the behavioral norms that inform decisions and actions in all data collection, management and use. It's a holistic approach in determining impacted data systems. There are transparency and decisions around data use and following the industry use and best practices such as data minimization, data ethics frameworks share the charge to put context around data management, to focus on how this management impacts individuals behind this data.

Including different impacts on different populations. Data ethics considers the power dynamics between the user and the organization. And the organizations -- they are in a position to take advantage of user data in ways that harm users. So here is a quick litmus test that you can do at your own library.

Tell your patrons what your actual data practices are. Do not show them your privacy policy or point them to the library bill of rights or the code of ethics, but clearly explain what actually happens in the patron data lifecycle at your library, including patient data that is fed into vendor systems.

If your patrons respond with -- you did what? With my data? You have a possible ethics breach. This is what happened at the Santa Cruz public library. And why they were investigated by the civil grand jury. Even if Santa Cruz used an open source version of the CRM or had complete control over the data lifecycle, the act of creating that user profile can still be considered an ethics breach. This still falls under the fallacy that libraries will default on doing what is right for the patron, instead of what the library has the right to do.

Data ethics requires us to confront the truth of how easy does to center the organization's needs and rights over the rights of the people. The data fields and market segmentation or analytics product, waiting to be analyzed. It is also easy to not question why you believe data would be better than talking to your patrons if you want to understand them. And it is also even easier to forget that your patient data reflects who is using the library and not the entire community that you are supposed to serve as a library.

We forget that data is not neutral. And our analysis will reflect the biases of the technology and processes used to collect and analyze data create we forget patron data analysis isn't really that far removed from surveillance. We can acknowledge the truth about the harms we can create through analytics but we can still be complicit. At my former workplace we had a data warehouse for reporting for analytical purposes. We did our best in minimizing harm through risk assessments should be included in the warehouse and strict data security measures, excluding raw patron data and projecting re-identification tests on identification methods. The risks were still there but a combination of having a big enough service population where the risk of re-identifying de-identifying data is limited and having the time and resources to identify and integrate risk mitigation controls made the data warehouse far less problematic, but not wholly unproblematic.

And this is for a large urban library. Many libraries have demographic outliers where they have small service populations which make many de-identification processes not viable. Re-identification is possible if the data has enough lingering patron data. Particularly data about patient activity and is combined with other data sets. So yes, you can identify a patron with only the location, time of checkout or use and the resource used to checkout. In a collection use report. From there I can look for patterns of checkouts of related topics or subjects location and time of day to create a rough profile.

So if I want to get more specific I can also connect those data points to other data points in different data sets. Suppose I gain access to geolocation data from that mobile phone carrier mentioned at the beginning of the talk. Or even access to the IP address or other unique data from the electronic device so I can create a digital fingerprint. In that case, not only can I identify the patron, but I note what they are reading. To the degree that I can exploit this knowledge through nonconsensual behavioral marketing or re-selling this newly constructed user profile to data brokers.

We never seem to make it past that point in the discourse. It is like the Ever Given boat into the Suez Canal, the privacy are trained to onboard that, we have people out there in the field who are tirelessly working to get the profession moving out of this stock position around data advocacy and practices. The library freedom project and the cohorts from the library freedom Institute objects such as data groups like DLF privacy and ethics technology and even the efforts of several ALA sub committees and working groups, they are all working towards more ethical and privacy preserving data library management module.

The ship isn't moving unless there is a coordinated movement between tugboats. It takes only a handful of tugboats to steer the largest of ships as we see Ever Given I do not have definite answers but I have ideas of how the resource committee session can help the librarianship boat to get it moving again.

First, let's focus on the software process. Privacy by default takes privacy by design asking for privacy to be considered at each stage of a processor project and it makes it explicit and defaulting any process or system to the highest level of privacy or mitigation strategies. For example, patrons could change user settings to the level proxy they are comfortable with. All user accounts would start with the highest level of privacy possible in that system. Speaking of patron accounts -- we have to acknowledge the reality, -- we want to have save searches or safe alerts to personalize their library experience.

Here we can't look at zero knowledge --record system such as the ILS, while there might be encryption of data in transit the library can still access the stored and encrypted data if the library has a copy of that encryption key. Zero knowledge is one decryption key. And that key stays with the user. And not the library. Any encrypted data stored in the system can only be decrypted by the user. This would remove the libraries access to patron data and eliminate several privacy risks around data access, retention and use. Zero knowledge architecture brings up an excellent question for library technology.

What data in the systems do libraries really need access to? And what data can exist in their systems? Without having to be out in the open for the library to access?

We've covered some ethical design tools and frameworks in this presentation that can be incorporated into planning, development and maintenance cycles. These resources can also be used to review existing technologies through audits, maintenance or enhancement cycles. You have a lot of flexibility with these resources in terms of integration into processes, so go forth, play, research, try what works and if it doesn't try another.

Nevertheless, we can talk about development frameworks, encryption, privacy controls, all of this can only go so far. We need community mechanisms in place. Which are not in place in the librarianship at this point. We need mechanisms in place that provide transparency and accountability -- this includes how our labor open source systems meet existing library ethics and privacy standards. What should happen if the system fails to live up to those ethics or standards?

This could include a governance group or working group that sets the ethical and privacy standards for community work. Regularly reviews the ethical and privacy locations of systems developed by the community and oversees the investigation of possible ethics breaches.

At my 2015 talk about the intersections of libraries and technologies, I asked the question to the library technology community, what exactly is the world that we are making? And does it match the world we want to make? And so six years later, I will ask that same question to you all. What is the world we want to make? Is it a world where we can lead the way in creating privacy preserving systems? Or is it one were we collaborate with library privacy advocates and create privacy and data ethics standards and accountability systems that are so desperately needed for our profession.

As a collective do we push back against this narrative the data is a commodity to be mined and exploited to ensure the institution survival? Or do we dare to have all three worlds in one?

Special thanks to Gail and Dorothy for help shaping this talk. Again you can find me at my email and at Twitter if you would like to continue the conversation. I want to thank you all for joining us today. And thank you so much for listening.

I hope you enjoy the rest of the conference and I look forward to continuing the discussion online. Again, thank you.

>> KATIE GREENLEAF MARTIN: Thank you so much, Becky, that was wonderful and am actually looking forward to going back and doing a recording again after I have had a little time to think through and digest. So the talk will be up on YouTube in a few weeks because Becky has graciously agreed to include that in her licensing --so you can look forward to that.

The exhibit halls are opening today at noon so you can head over there and check out our exhibitors and we will see you back for sessions this afternoon. Thank you, everyone, and have a wonderful rest of your conference.