Anonymous In the Forest

Personally Identifying Information (PII) in Evergreen

Rogan Hamby (Equinox Open Library Initiative)

equinox
OPEN LIBRARY INITIATIVE

A Fireside Chat

This will be interactive discussion points.

Ask questions as we go. I'll do my best to watch chat.

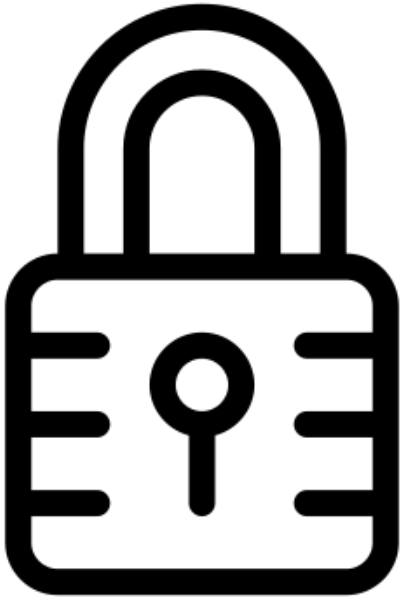
There will also be time for discussion at the end.

Privacy is a Necessary Problem ... like Laundry

"Data is the pollution problem of the information age, and protecting privacy is the environmental challenge."

Bruce Schneier
Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World

Security vs Privacy



VS

We Are Stewards

This isn't the library's data. This is the patron's data.

We own nothing.

What Are We Talking About

- The privacy landscape.
- Where Personally Identifying Information (PII) is stored and how it is accessed.
- The risks associated with that storage and retrieval.
- Options to mitigate risk.

The Order of Things

1. How little it takes to be PII.
2. The legal landscape.
3. Things everyone should do.
4. Things particular to Evergreen.

Section 1. How Little It Takes

What is PII

Information that can be used to **distinguish or trace** an individual's identity, either alone or when combined with other personal or identifying information that is linked or linkable to a specific individual.

What is PII

PII can be A) data about who someone is or B) what they are doing that alone or combined with other available information can identify someone.

How Little It Can Take

Latanya Sweeney, Carnegie Mellon University, "Simple Demographics Often Identify People Uniquely" (working paper)

T. Dalenius. "Finding a needle in a haystack – or identifying anonymous census record". Journal of Official Statistics, 1986.

Is this a Real Danger

- 1990 census summary data to a individuals with simple data including 5 digit zip, gender and date of birth
- You often need less, as little as a year of birth
- Card holders are a subset of population
- Low population areas can be most at risk and low birth year frequencies mean elderly and rural populations are high risk

Why Go Over This?

But what if need it one day?

Hoarding is as bad, especially when it comes to patron data. You can toss out the fifty year old government pamphlets and the closed circ records.

Section 2. The Legal Landscape

or

Section 2. "Kill all the lawyers."

So, the danger is real, what is next?

Your obligations will vary depending on where you live. However, these are a baseline, not goals.

Canada

https://www.priv.gc.ca/

The Privacy Act applies to federal institutions but each province and territory has distinct governance.

PIPEDA (Personal Information Protection and Electronic Documents Act) applies to private organizations.

European Union

GDPR - General Data Protection Regulation

https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en

Is extraterritorial.

China

PIPL - Personal Information Protection Law

Currently in draft form.

Vaguer than GDPR but similar. Also extraterritorial.

United States

Several federal laws and regulations apply to federal entities. Otherwise each state and/or municipality creates their own rules (or lack thereof).

United States - Example

- California Consumer Privacy Act (CCPA)
- https://oag.ca.gov/privacy/ccpa
- Defined as a series of rights.
- Vague about what PII is by saying anything that could be used to that effect.
- Many exemptions for businesses.
- Also extraterritorial.

Be Proactive

Libraries are often excluded from the dialogue in local governments. Talk to attorneys.

Section 3. Things Everyone Should do

DISCUSSION TIME

A best practice is to only expose the information needed for a specific task. What function in a library violates this principle with almost every single transaction?

The Circulation Desk

equinox
OPEN LIBRARY INITIATIVE

Step 1

1. **Have a policy.**

A Privacy Policy

- ALA – https://bit.ly/2T7CJRT
- Create a Privacy Policy.
- Make it a part of your technology plan so when you have to revise your technology plan you revise it as well.

If you take nothing else from it ....

- What to audit.
- Questions to ask.

The Bad - Some of the Technology Specific Components

Emerging Technologies include:

- Smartphones
- RFC
- Social Networks
- this one hurt a bit ...
- Interactive OPACS
- … my favorite
- Software (they call them apps but it is clear that any end user software is included)

Step 2

1. **Have a policy.**

2. **Create a procedure for handling disclosures.**

Forms of Disclosure

- The good : We are using your information to improve services to you with a third party.
- The bad : There has been a data breach and now your information is for sale along with the emails of various celebrities … just cheaper.
- The ugly : Your cousin who works at our branch decided to tell your spouse you checked out books on divorce.

Step 3

1. **Have a policy.**
2. **Create a procedure for handling disclosures.**
3. **Privacy must be a part of staff development.**

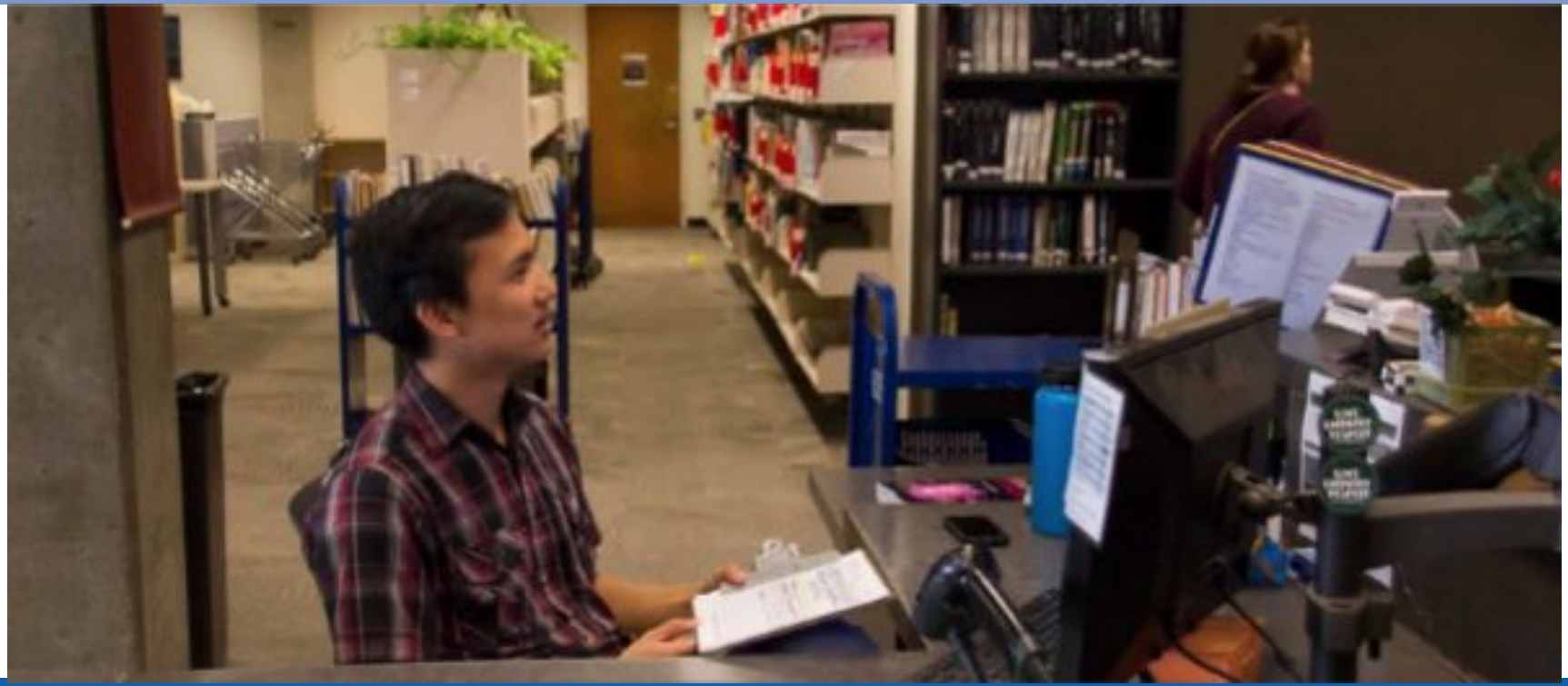
Step 4

1. **Have a policy.**
2. **Create a procedure for handling disclosures.**
3. **Privacy must be a part of staff development.**
4. **Ensure that your audit and review periodically.**

Section 4. Evergreen Itself

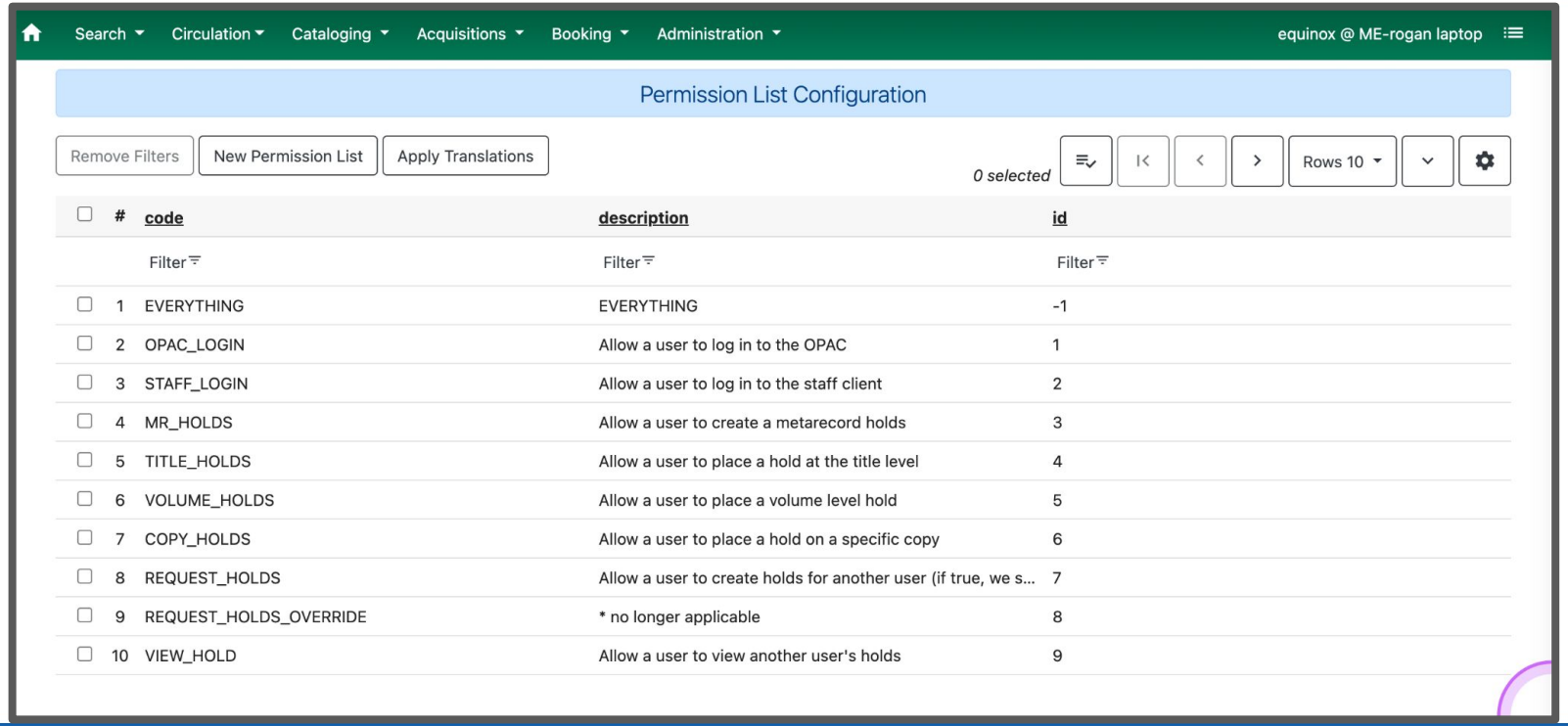
equinox
OPEN LIBRARY INITIATIVE

Our Biggest Risk, Again



equinox
OPEN LIBRARY INITIATIVE

What You Can Do



	#	code	description	id
		Filter	Filter	Filter
☐	1	EVERYTHING	EVERYTHING	-1
☐	2	OPAC_LOGIN	Allow a user to log in to the OPAC	1
☐	3	STAFF_LOGIN	Allow a user to log in to the staff client	2
☐	4	MR_HOLDS	Allow a user to create a metarecord holds	3
☐	5	TITLE_HOLDS	Allow a user to place a hold at the title level	4
☐	6	VOLUME_HOLDS	Allow a user to place a volume level hold	5
☐	7	COPY_HOLDS	Allow a user to place a hold on a specific copy	6
☐	8	REQUEST_HOLDS	Allow a user to create holds for another user (if true, we s...	7
☐	9	REQUEST_HOLDS_OVERRIDE	* no longer applicable	8
☐	10	VIEW_HOLD	Allow a user to view another user's holds	9

What You Can Do

A Few View Permission

VIEW_HOLD

VIEW_USER_FINES_SUMMARY

VIEW_USER_TRANSACTIONS

VIEW_USER

VIEW_TRANSACTION

VIEW_CIRCULATIONS

VIEW_COPY_CHECKOUT_HISTORY

DISCUSSION TIME

Has anyone here worked to limit staff permissions to limit exposure of patron information?

Storing Patron Data

We primarily store information about who patrons are in a schema called 'actor' which has tables that live under it.

Actor.usr

user names, family names, personal names, birth dates, preferred names, email addresses, phone numbers of various kinds and it goes on and on ...

Actor.usr

There are a few fields that are less obvious.

ident_values

photo_url

alert message

What You Can Do

- Evaluate what you store and how.
- Staff training.
- Do you need birth dates?
- Statistics you say? Maybe you only need the year then.
- Are those identifier values really necessary?

What You Can Do - Bonus Round

What have you done in the past that would violate current policies? - Sometimes it is a good idea to do some digital archaeology. *

* Don't forget the physical stuff you may have sitting around somewhere too.

equinox
OPEN LIBRARY INITIATIVE

DISCUSSION TIME

What have you seen that you have to remove and clean out? What scares you might be hiding in records?

So, we are all good now, right?

- barcodes (mostly if they are also usrnames)
- addresses
- and much more ...

Statistical Categories ... are dangerous.

- Patron Types
- Gender
- Age Group
- School District
- Municipality

Statistical Categories ... are good?

- Patron Types
- Gender
- Age Group
- School District
- Municipality

User Activity

You can define activity types and link patrons to them with dates.

More Free Text Fields

- Messages.
- Notes.

Family Relationships

- We're supposed to love linking data, right?
- Privacy waivers.
- Families.
- Guardians.

We are done, right?

equinox
OPEN LIBRARY INITIATIVE

Action Schema - The Usual Suspects

- Most of the action schema is information about what you did rather than who you are.
- Holds.
- Circulations.
- But ...

Action - the Unusual Suspects

- Curbside.
- User circulation history.
- Surveys.

Is There A Lesson in This?

There is such a thing as healthy paranoia when it comes to PII.

All of this goes back to those same questions ALA proposes when designing a privacy policy.

What Can You Do About it

1. Staff education.
2. Create a policy for removing inactive patrons.
3. Age circulations and holds.
4. Review all old content for removal.

Aging Transactions Bug

Aged transactions can be de-anonymized using post code and birth year

https://bugs.launchpad.net/evergreen/+bug/1861239

Community Wise

1. Thoroughly review purge patron function and make sure nothing is missed.
2. Consider tools for removing old data points in line with aging transactions and recommending them for review.

The Next One is Scary

Manage Folder Contents | Manage Folder

JWeston - Migrations: created by **equinox**

Create a new report from selected template | Submit

Limit output to 10 | Start | Prev | Next | Create a new Template for this folder

Select All None	name	description	docs	ui	create_time	owner
☐	006 values	no filter except tag=006		WebStaff	2021-03-11 14:35	equinox
☐	007 values	removed OU filter		WebStaff	2021-03-11 13:07	equinox
☐	Detailed Inventory List	List of titles and items filtered by shelving location and library. Displays detailed bibliographic information.		WebStaff	2021-01-26 14:23	equinox
☐	Count Items provided TO other Evergreen Libraries, choosing and showing Destination Libraries (converted from XUL) (clone)	Displays owning library, destination library, count of items sent. Filtered by send date range and owning library.		WebStaff	2021-01-28 16:50	equinox
☐	List of Holds Unfulfilled by Library (clone)	This report shows patron information for holds that remained unfulfilled during a designated time frame. This report is useful in showing long-term holds that remain unfulfilled. Displays: first name, barcode, pickup library, request date/time, hold id, hold exp date Filters: pickup library, hold request date/time		WebStaff	2021-02-04 16:12	equinox
☐	List of Org Units, Codes, and IDs			WebStaff	2021-03-11 12:58	equinox
☐	Record ID and TCN	for all records 03/11/2021		WebStaff	2021-03-11 15:22	equinox
☐	Record ID and TCN (clone)	for all records 03/11/2021		WebStaff	2021-03-23 14:22	equinox
☐	Record ID and TCN (limit results)	includes record type and delete flag		WebStaff	2021-03-23 14:56	equinox

Reports

RUN_REPORTS - Allow a user to run reports

SHARE_REPORT_FOLDER - Allow a user to share report his own folders

VIEW_REPORT_OUTPUT - Allow a user to view report output

CREATE_REPORT_TEMPLATE

What You Can Do

- Restrict reports to those who most need it.
- Create policies to manage how report output is distributed and stored.

What The Community Can Do

- Caveat personal opinion incoming …
- We should consider options for restricting reports by content per user.
- This would be a big change from our current model and would neither be easy or quick.

FINALLY

WE ARE DONE!

Uh, not quite.

Output and Back Ups

- Action Trigger Output
- Report Output
- Auditor Tables
- Data Backups
- Offline Transactions

NOW … WE ARE DONE?

Uh, not quite.

The OPAC

- "Interactive OPACs"
- We often install something like Google tracking and create our own privacy problem.

The Solution?

Use an alternative like Matomo.

2.2.3. Matomo Support

Support for the open source web analytics platform Matomo is now native to Evergreen. Support is on an org unit level so different libraries can have separate or no analytics..

- 3.6 release notes

Questions