Protecting Privacy in the Library Patron Data Lifecycle

Training One of the PLP Data Privacy Best Practices Training for Libraries Project

January 2020
This project was supported in whole or in part by the U.S. Institute of Museum and Library Services under the provisions of the Library Services and Technology Act, administered in California by the State Librarian. The opinions expressed herein do not necessarily reflect the position or policy of the U.S. Institute of Museum and Library Services or the California State Library, and no official endorsement by the U.S. Institute of Museum and Library Services or the California State Library should be inferred.

This document does not constitute legal advice, and is for informational purposes only. Please consult an attorney or other legal counsel for legal advice.

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Cover Page Photo Credit - Privacy by Nick Youngson CC BY-SA 3.0 Alpha Stock Images

Handbook developed by Becky Yoose, LDH Consulting Services
becky@ldhconsultingservices.com
# Table of Contents

- Patron Data Lifecycle Quick Reference Map ............................................................... 1
- Workshop Exercises ........................................................................................................ 2
- Slides and notes ............................................................................................................... 9
- Your First Data Inventory Template ................................................................................ 31
- References and Further Reading ...................................................................................... 35
### Patron Data Lifecycle Quick Reference Map

<table>
<thead>
<tr>
<th>Phase</th>
<th>Question</th>
<th>Best Practice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collection</td>
<td>What patron data are you collecting, and why?</td>
<td>Only collect data needed for demonstrated business cases</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Practice “The Five Whys”</td>
</tr>
<tr>
<td>Storage</td>
<td>Where is the patron data stored?</td>
<td>Limit number of data storage areas</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Limit storage of patron data in local and vendor systems</td>
</tr>
<tr>
<td>Access</td>
<td>Who has access to the patron data?</td>
<td>Limit physical and electronic access to data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Audit vendor security and privacy practices</td>
</tr>
<tr>
<td>Reporting</td>
<td>What patron data is published to staff and to the public?</td>
<td>Aggregate data and control access to data through dashboards, database views, and other data reporting tools</td>
</tr>
<tr>
<td>Retention</td>
<td>How long is patron data kept?</td>
<td>Follow local and state retention regulations</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ensure backups and logs are covered in retention policies</td>
</tr>
<tr>
<td>Deletion</td>
<td>How is patron data deleted?</td>
<td>Properly dispose of physical and electronic media that contained patron PII</td>
</tr>
</tbody>
</table>
Workshop Exercises

Discussion – Where does patron data live at your library?

Your library collects, stores, and uses patron data, but do you know where the data lives at your library? Here’s a list to get you started:

- Integrated Library Systems
- Database backups
- Print management systems
- Server logs
- Reference chat/desk logs
- Public computer/wireless traffic logs
- Interlibrary Loan requests
- Security camera footage
- Card reader logs
- Meeting room reservations
- Authentication system logs
- Library programs
  - Attendance logs
  - Feedback responses
- Paper forms
- Staff email

Where else can you find patron data in your library?

Spend one minute listing other places where you can find patron data at your library.

Discuss with the entire group. Were there any places that you missed when you first created your list? If so, write them down so you can use this list as a check list for your library!
Exercise – Identifying Patron PII

Identify all PII data in the email below. For extra credit, identify which PII is either PII 1 [data about a person] or PII 2 [data about a person's activities].

To: helpdesk@plp
From: roseHelix@gmail
Date: January 14, 2010 4:27 PM
Subject: HELP!!!

I can't login to renew THE ONLY WOMAN IN THE ROOM with my PIN 1111. Can you renew the book for me? My barcode is 12345678910.

Thank you! Rosalind Franklin

Follow up question to consider
What patron PII would be included in staff internal communications, including email, request tracking, and so on?
Exercise - “Why’s that, again?”

You go to your boss after finding out that they want to collect driver’s license numbers (DLNs) in the patron record, and you’re not sure if this is the best idea in the world. You decide to practice your newfound “Five Whys” method, only to stop after the second why when your boss is interrupted:

Boss: “We need to collect Driver’s License Numbers.”
You: “Why?”
Boss: “We need to establish residency.”
You: “Why do we need to verify residency with Driver’s License Numbers?”
Boss: “Otherwise they can’t use library resources if they aren’t residents in our service area!”

You won’t be able to talk to them for the rest of the day. You suspect that your third “Why?” would have been met with “Because I said so!” - it’s been that type of day in the library. But perhaps this is enough information for you to work with in determining if you really need to collect those numbers.

Are there alternative ways to achieve the library’s business need of requiring patrons to verify that they live in the service area that doesn’t involve storing DLNs in the patron record?

1. Spend one minute to think of alternative ways.
2. Spend two minutes sharing what you find with another person.
3. Spend four minutes sharing what you and your partner discussed with another pair.
4. Report your discussions to the entire group.

What alternatives did you come up with? Are there other factors that need to be addressed in this conversation?
Exercise – De-identifying Data

Below a selection of raw PII data for you to de-identify. Using the methods of obfuscation and truncation, de-identify the raw data in the left column. Note that there could be multiple ways of de-identifying data depending on your needs.

<table>
<thead>
<tr>
<th>Raw data</th>
<th>De-identified data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date of birth - 10/24/1977</td>
<td></td>
</tr>
<tr>
<td>NX180.I57 M275 2015</td>
<td></td>
</tr>
<tr>
<td>11 Apr 2020 9:24 - 10:24</td>
<td></td>
</tr>
<tr>
<td>27 43rd St, Town, WA 92471</td>
<td></td>
</tr>
<tr>
<td>FIC HARRIS 2018</td>
<td></td>
</tr>
</tbody>
</table>
Exercise – What can possibly go wrong?

Divide into groups of four. Read the scenario assigned to your group and discuss the questions in each scenario. Assign a note taker and a person to report out to the entire group.

Be prepared to report the following to the entire group:

• Your scenario
• The main two to three threats to patron data privacy in the scenario
• Up to four strategies to mitigate the threat to patron data privacy
Exercise – Patron Data Requests

Think back to a time at your library where someone requested patron data. On your own, answer these three questions:

1. Who made the request, and what was requested?
2. What was your response to the request?
3. What informed your response? Library policy, regulations, ethics, request circumstances, etc.

Share your answers with the person next to you. What were the similarities and differences between the two request scenarios?
Exercise - What Would You Do?

A library was subpoenaed by the local authorities for their library's computer reservation logs as well as security camera footage. The library kept the logs on paper, and patrons would sign their name and barcode to use a particular computer at a particular time that day. The library kept the paper logs for about a month, give or take a few days if staff remember to throw them away. Using the historical logs, the library could find the patron who was using a particular computer at a particular time of day in case the library needed to charge for a broken keyboard, mouse, or other computer equipment, and keeping a log can make it easier to discover potential patterns of computer misuse.

The library has security cameras recording various parts of the library, including the public computer area, the circulation desk, the children's area, and the entrance of the building. These picture-only recordings are stored on an external hard drive connected to the security camera server for 90 days, where then the files are deleted from the system.

The subpoena asked for computer reservation logs and recordings from the last 30 days, and after consulting with legal counsel, the library had to comply and hand over the logs and recordings.

A library staff person reached out to you do ask what they could do better to avoid this from happening again.

Keeping in mind what you learned in the workshop about the data lifecycle, and how privacy factors into the cycle, what would you advise the library to do to better protect patron privacy while still meeting their needs for tracking reservations?
Protecting Privacy in the Library Patron Data Lifecycle

Becky Yoose
Library Data Privacy Consultant, LDH Consulting Services
Pacific Library Partnership, January 2020

This project was supported in whole or in part by the U.S. Institute of Museum and Library Services under the provisions of the Library Services and Technology Act, administered in California by the State Librarian. The opinions expressed herein do not necessarily reflect the position or policy of the U.S. Institute of Museum and Library Services or the California State Library, and no official endorsement by the U.S. Institute of Museum and Library Services or the California State Library should be inferred.

Workshop Housekeeping – Guidelines

- All responses and questions are valid.
- Assume good intent.
- When you disagree, challenge or criticize the idea, not the person.
- Be mindful of the time.
- One speaker at a time.
- Speak from your own perspective.
- Help protect others' privacy by observing the Chatham House Rule.
Workshop Housekeeping - Logistics
IANAL; Consult legal staff for legal advice
Exercises and Discussions - what to expect
General Overview vs Deep Dive
Privacy measures are only as strong as the least-knowledgeable person working with patron data

Workshop Schedule
9:00 – 9:20: Welcome and housekeeping
9:20 – 10:00: Training and exercises
10:00 – 10:10: Break #1
10:10 – 11:00: Training and exercises
11:00 – 11:10: Break #2
11:10 – 11:45: Training and exercises
11:45 – 12:00: Wrap up

Introduce Yourself!
1. Name
2. Job title and where you work
3. List one thing that you do to protect your personal privacy
Section One:
Library Patron Data Overview

Libraries and Patron Data - Where It Is In Your Library
- ILS
- Database backups
- Print management systems
- Server logs
- Reference chat/desk logs
- Public computer/wireless traffic logs
- ILL requests
- Security camera footage
- Card reader logs
- Meeting room reservations
- Authentication system logs
- Library programs
  - Attendance logs
  - Feedback responses

Libraries and Patron Data – The “Big Three”
Vendor and third party application data
Paper forms
Staff email
Discussion –
Where does patron data live at your library?

Personally Identifiable Information [PII] and Library Data

PII 1 - Data about a patron
● Name
● Physical/email address
● Phone number
● Date of birth
● Patron record number
● Library barcode

PII 2 - Activity that can be tied back to a patron
● Search & circulation histories
● Computer/wifi sessions
● Reference questions
● Electronic resource access
● IP Address
● Program attendance

Exercise –
Identifying Patron PII
Exercise Answer Key - Identifying PII

### PII 1
- Rosalind Franklin [Name]
- roseHelix@gmail.com [Email]
- 12345678910 [Barcode]
- 1111 [PIN]

### PII 2
- THE ONLY WOMAN IN THE ROOM [Checked out book title that needs renewing]
- January 14, 2020 4:27 PM [Timestamp]

---

### Section Two:
Library Patron Data And Regulations

“What happens if my library has…” (Federal Laws)

Medical-related data? HIPAA
Student data? FERPA
Data from minors (<13 years)? COPPA
Library Patron Data in California

California Gov Code § 6267
"All patron use records of any library ... shall remain confidential and shall not be disclosed" except when a person is acting within scope of their duties, written authorization from patron, or by court order.

Patron use records definition
- Written or electronic record identifying patron
- Written or electronic transaction of patron's use of library resources

California Gov Code § 6254
Disclosure exemption for library circulation records that identify "borrower of items available in libraries" (not including records of fines imposed on borrower)

Do libraries need to worry about CCPA?
- California Consumer Privacy Act of 2018 (CCPA)
- Regulates the sale of personal information by covered businesses
  - Gives California residents:
    - Right to access what personal information is collected and shared with service providers and other third parties
    - Right to request deletion of information
    - Right to opt out of sale of personal information
- Areas of concern for libraries:
  - Special requirements for minors
  - Household information as part of personal information definition

What about GDPR?
- Most US libraries are not under scope
- Exceptions - academic libraries whose institutions have EU presence and collect/ process EU resident personal data
- GDPR rights to individuals to access, change, limit processing, and delete data
- GDPR best practices to adopt - Privacy by Design (PbD)
Section Three:  
The Library Patron Data Lifecycle

Library Patron Data Lifecycle  

Collection

What data is our library collecting?

Example systems:
- Integrated Library Systems
- Server/Application logs
- Web analytic software
- Social media pages
- Survey software

Data collection includes:
- Patron & circulation information
- IP address/UUID
- Timestamps
- Search history
- Link clicks on sites and in marketing emails
Vendors and patron data collection – defaults

If you do not have a demonstrated business need to explain why you are collecting a data point, then you should not collect that data.

The Five Whys Method

Why are we doing X?
Because A.

Why A?
Because B.
Exercise –
“Why’s that, again?”

Library Patron Data Lifecycle

Storage

Where are we storing data?

Native systems and applications

Data extracted, exported, or otherwise taken from the native system and applications
### De-identification of Library PII Data

<table>
<thead>
<tr>
<th>Obfuscation</th>
<th>Aggregation</th>
</tr>
</thead>
<tbody>
<tr>
<td>● PII 1</td>
<td>● PII 1</td>
</tr>
<tr>
<td>○ Date of birth vs age</td>
<td>○ Age vs age ranges</td>
</tr>
<tr>
<td>Truncation</td>
<td>● PII 2</td>
</tr>
<tr>
<td>● PII 1</td>
<td>○ Very high level call number ranges</td>
</tr>
<tr>
<td>○ Full address vs zip code</td>
<td>○ Call numbers</td>
</tr>
<tr>
<td>● PII 2</td>
<td></td>
</tr>
</tbody>
</table>

### Pseudonymization

Some considerations:

- Algorithms
- Hashing and salt

### Differential Privacy

All the mathematical equations!
Exercise – De-identifying Data

Exercise - Raw vs De-identified Data

<table>
<thead>
<tr>
<th>Raw data</th>
<th>De-identified – one way</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date of birth – 10/24/1977</td>
<td>42 years old</td>
</tr>
<tr>
<td>27 43rd St, Town, WA 92471</td>
<td>92471</td>
</tr>
<tr>
<td>NX180.157 M275 2015</td>
<td>NX180</td>
</tr>
<tr>
<td>FIC HARRIS 2018</td>
<td>FIC</td>
</tr>
<tr>
<td>11 Apr 2020 9:24 - 10:24</td>
<td>4/11/2020, 02:00</td>
</tr>
</tbody>
</table>

Obligatory Disclaimer De-identification of Data (Part One)

Data de-identification methods do not provide adequate privacy protection for these types of data:

- Outliers in service population
- Small overall service population or subset (degree program, etc.)
Vendors and Third Party Apps Considerations

CRMS and data analytics products areas for privacy negotiations during the acquisition process:
- Request for Proposals
- Functional Requirements
- Contract Negotiations
- Contract Addendums

Data storage: What, Where, Who Has Access

Data sharing: What, Where, Who, Raw vs De-identified vs “Anonymized” vs Aggregated

Data privacy policy: what to negotiate, theirs (if they have one) or yours

Library Patron Data Lifecycle

Who has access to what data?

<table>
<thead>
<tr>
<th>Physical Access</th>
<th>Electronic Access</th>
</tr>
</thead>
</table>
| ● Desktop computers
| ● User permissions |
| ● Laptops
| ● Administrator account information |
| ● Mobile devices
| ● Vendor access to local systems |
| ● Server room/data center
| ● System log and database access |
| ● Offices and desks
| ● Administrator site access |
| ● Flash drives
| |
| ● File cabinets
| |
| ● Security camera terminal and tapes
| |

Access

Who has access to what data?
Exercise – “What can possibly go wrong?”

Some Data Access Best Practices

● Lowest/most restrictive level of access to meet operational needs
● Lock everything - restrict access to hardware through physical barriers
● "Lock" everything
  ■ Require login for systems and physical devices
  ■ Enable multi-factor authentication (MFA) if possible
  ■ Encryption of hard drives and mobile devices
  ■ Remote wipe for mobile devices
● Regular audits of...
  ○ Account access to systems
  ○ Keys (physical and electronic)

Library Patron Data Lifecycle

<table>
<thead>
<tr>
<th>Reporting</th>
</tr>
</thead>
</table>
Do staff really need access to all the data for reports?

Consider giving staff the following:

- Database views for report queries and building
- Connections to data through Data BI tools (Tableau, Power BI, etc.)
  - Related - restrict access to data used in the report when published
- Dashboards
- “Canned” reports with some parameters for customization (date, location, etc.)

How much data to provide in the report?

Report on the highest level of data that will meet the core need of the report audience... … except when release of entire data sets is expected.

Obligatory Disclaimer De-identification of Data (Part Two)

Data de-identification methods are subject to varying re-identification risks, primarily through PII 2 data:

- Identifying patterns
  - Example - AOL
- Fuzzy matching
  - Example - Taxi Cab Data
Example of fuzzy matching – Library edition

<table>
<thead>
<tr>
<th>Data set #1</th>
<th>Patron barcode match between two data sets = possible reidentification and tracking of patron activity tied to unique real world individual</th>
</tr>
</thead>
<tbody>
<tr>
<td>● Patron physical address</td>
<td></td>
</tr>
<tr>
<td>● Patron barcode</td>
<td></td>
</tr>
<tr>
<td>● Patron age</td>
<td></td>
</tr>
<tr>
<td>Data set #2</td>
<td>Accuracy of reidentification increases with additional data sets (including Open Data and data broker data sets)</td>
</tr>
<tr>
<td>● Call number</td>
<td></td>
</tr>
<tr>
<td>● Subject headings and topics</td>
<td></td>
</tr>
<tr>
<td>● Patron barcode</td>
<td></td>
</tr>
</tbody>
</table>

Sometimes the only response...  

... is to not release data.*

*except when required by law

Exercise – Data Patron Request
How long are we storing data?
... when no longer needed operationally?
  ... 30 days?
  ... 1 year? Rolling year?
  ... in perpetuity?
  ... what about backups?
  ... what about vendor systems?

Library Patron Data Lifecycle

Retention
—

Library Patron Data Lifecycle

Deletion
—
Deleting data

Electronic data
- Backups and logs - when and how are they deleted?
- Wipe the drive after electronic data deletion
- Look out for data living “outside” local and vendor systems

Physical data
- Shred paper and dispose of shredded paper properly
- Properly dispose of disks, drives, other hardware, including degaussing or otherwise physical destruction of the disk or drive

Vendor considerations
- How does the vendor delete physical and electronic copies of your patron data?
- When you leave, can you take your data with you?
- Can your patrons take their data with them?
- Can your patrons request their data to be deleted with a vendor?

Library Patron Data Lifecycle Roundup
Exercise – What Would You Do?

Section Four: A Data Inventory Starter Kit

Tips for your first data inventory and beyond

- Start small – select one system to focus on
- Ask around – sometimes you learn things that you wouldn't otherwise
- Vendor hosted systems – focus on what you can control when recommending actions based on inventory
- Privacy and security audit integrations with data inventory
- Setting up a schedule for inventories
Pick one piece of technology (locally or vendor hosted) that your place of work is using or has used in the past.

Possible questions to ask:
- What patron data are you collecting?
- How is the patron data being used?
- Where is the patron data being stored? Don't forget backups, log files, etc.
- How long are you keeping patron data?
- How are you deleting patron data when it's no longer needed?
- Who has electronic and physical access to the patron data?

Section Five:
Wrap up

What is one thing from this workshop that you can put into practice or discussion at your library when you return?
Thank you
:-)
Your First Data Inventory Template

Starting out on your first data inventory can be daunting. This data inventory template can help your library navigate their way through a data inventory of one system or application used by your library. Once you have gone through the inventory with one system, you can review and revise the template to suit your library's needs.

System/Application Name:__________________________________________________________

Date of inventory: _____________________________

Inventory taker(s):

Collection

What data are you collecting?

What patron Personally Identifiable Information (PII) is collected? Include both data about the person (name, email, address, ID number, etc.) and data of activity that can be linked to the person (web searches, accessing materials, etc.)

Why is the data being collected? What business need is being met by data collection?
How is the data being used?

Storage

Where is the original/raw data being stored?

What versions of the primary data exist internally? Examples include reports, shared drives, email, export files, etc.

What formats are the data stored in?

Retention

How often is the data collected and/or updated?
How long is the data kept, including in backups and logs?

Access

Who has electronic access to the data, including user accounts, logs, and backups?

Who has physical access to the data or the physical storage that contains the data?

Reporting

What data is shared with external audiences? How is this data shared?

What data is shared with internal/org audiences? How is this data shared?
Deletion

How is the data deleted? Include both electronic and physical formats and mediums.

Administration/Policy

Who is responsible for the administration of the system?

Who decides what data is collected by the system?

What internal policies and procedures apply to the collection, retention, and reporting/disclosure of data?

What local, state, or federal laws apply to the collection, retention, and reporting/disclosure of data?
References and Further Reading

General


AASL Resources


ALA Resources


California Law Resources

https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?sectionNum=6254.&lawCode=GOV.

https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?sectionNum=6267.&lawCode=GOV.

GDPR/CCPA Resources


**Industry Standards**


Section 6254 (excerpt)

6254. Except as provided in Sections 6254.7 and 6254.13, this chapter does not require the disclosure of any of the following records:

... 

(j) Library circulation records kept for the purpose of identifying the borrower of items available in libraries, and library and museum materials made or acquired and presented solely for reference or exhibition purposes. The exemption in this subdivision shall not apply to records of fines imposed on the borrowers.

(Amended by Stats. 2019, Ch. 25, Sec. 1. (SB 94) Effective June 27, 2019.)
6267. All patron use records of any library which is in whole or in part supported by public funds shall remain confidential and shall not be disclosed by a public agency, or private actor that maintains or stores patron use records on behalf of a public agency, to any person, local agency, or state agency except as follows:

(a) By a person acting within the scope of his or her duties within the administration of the library.
(b) By a person authorized, in writing, by the individual to whom the records pertain, to inspect the records
(c) By order of the appropriate superior court.

As used in this section, the term “patron use records” includes the following:

(1) Any written or electronic record, that is used to identify the patron, including, but not limited to, a patron's name, address, telephone number, or e-mail address, that a library patron provides in order to become eligible to borrow or use books and other materials.

(2) Any written record or electronic transaction that identifies a patron's borrowing information or use of library information resources, including, but not limited to, database search records, borrowing records, class records, and any other personally identifiable uses of library resources information requests, or inquiries.

(3) This section shall not apply to statistical reports of patron use nor to records of fines collected by the library.

(Amended by Stats. 2011, Ch. 80, Sec. 1. (SB 445) Effective January 1, 2012.)