Library Patron Data and Privacy: Cycles and Strategies

Becky Yoose
Library Data Privacy Consultant, LDH Consulting Services
Colorado State Library, 2/4/2020
Housekeeping

Resource list available in the handout

IANAL; Consult legal staff for legal advice

Exercises and Discussions - what to expect

Privacy measures are only as strong as the least-knowledgeable person working with patron data
Library Patron Data
Libraries and Patron Data

- Integrated Library Systems
- Database backups
- Print management systems
- Server logs
- Reference chat logs
- Public computer/wireless traffic logs
- SIP logs
- Security camera footage
- Card reader logs
- Library programs
  - Attendance logs
  - Feedback responses
- Staff email
- Vendor systems
- Paper documents
## Personally Identifiable Information [PII] and Library Data - NIST

### 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
Personal Information and Library Data - CO State

CRS Section 24-73-101 and 102
"Personal identifying information" w/r/t disposal and protection:

- Social security number
- Personal ID number
- Password or pass code
- Government-issued driver's license or ID number
- Biometric data
- Employer, student, or military ID number

CRS Section 24-73-103
"Personal information" w/r/t security breach (not encrypted):

- CO resident's first name or first initial & last name combined with one type of personal identifying info
- CO resident's username or e-mail address; password or security questions & answers
Library Patron Data And Regulations
“What happens if my library has...”

Medical-related data? HIPAA

Student data? FERPA

Data from minors (<13 years)? COPPA
CRS 24-90-119. Privacy of user records

(1) Except as set forth in subsection (2) of this section, a publicly supported library shall not disclose any record or other information that identifies a person as having requested or obtained specific materials or service or as otherwise having used the library.

(2) Records may be disclosed in the following instances:

(a) When necessary for the reasonable operation of the library;

(b) Upon written consent of the user;

(c) Pursuant to subpoena, upon court order, or where otherwise required by law;

(d) To a custodial parent or legal guardian who has access to a minor's library card or its authorization number for the purpose of accessing by electronic means library records of the minor.
Library Record Retention in Colorado

- Different libraries, difference schedules
  - Municipal Libraries - Schedule No. 70
  - Library Districts - SDRM Schedule 7 - 7.270 Program Records
- Both record retention schedules are optional - it is only legally binding if the library adopts the schedule as policy
- Retention periods in both schedules geared toward pre-automation libraries
Do we in Colorado need to worry about...

GDPR?
It depends.

CCPA?
Most likely not.
Library Patron Data Lifecycle Collection
What data are we and our vendors collecting?

Some systems/processes:

- ILS
- Server logs
- Web analytic software
- Social media pages
- Survey software
- Program release forms

Data collection includes:

- Patron & circulation information
- IP address/UUID
- Timestamps
- Search history
- Link clicks on sites and in marketing emails
What data are our vendors collecting?

Account Preferences: Borrowing History

Your public library does not keep records of your borrowing without your direction to do so. However, when you enable the Borrowing History feature, the BiblioCommons system will gather a list of the titles you borrow. The content on your Borrowing History page is visible only to you. The Borrowing History feature is not retroactive. It begins with the first item you return after you enable the setting.

OFF Your borrowing history is disabled.

Save Changes
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

- An investigative method from project management to determine the real need or cause of a problem
- Asking “Why” five times or until the person being asked cannot answer - whichever comes first
- Modified version with vendor can help with adjusting data collection practices with products or services
Collection Exercise - “Why’s that, again?”

“We need to collect Driver’s License Numbers!”

“Why?”

“We need to establish residency.”

“Why do we need to verify residency?”

“Otherwise they can’t use library resources if they aren’t residents in our service area!”

Are there alternative ways to achieve the business need of requiring patrons to verify that they live in the service area?
Library Patron Data Lifecycle

Storage and Retention
How long are we storing data?

... when no longer needed operationally?

... 1 year? Rolling year?

... in perpetuity?

... what about backups?

... what falls under Schedule No. 70? SDRM?
Where are we storing data?

Original systems, applications, and processes

Data extracted, exported, or otherwise taken from the original systems or processes
A VISUAL GUIDE TO PRACTICAL DATA DE-IDENTIFICATION

What do scientists, regulators and lawyers mean when they talk about de-identification? How does anonymous data differ from pseudonymous or de-identified information? Data identifiability is not binary. Data lies on a spectrum with multiple shades of identifiability.

This is a primer on how to distinguish different categories of data.

**DEGREES OF IDENTIFIABILITY**

- **EXPLICITLY PERSONAL**
  - Information containing direct and Indirect Identifiers.

- **POTENTIALLY IDENTIFIABLE**
  - Information which direct identifiers have been eliminated or transformed, but indirect identifiers remain intact.

- **NOT READILY IDENTIFIABLE**
  - Information which direct and indirect identifiers have been removed or manipulated to break the linkage to real-world identities.

**PSYDOMOUS DATA**

- **KEY CODED**
  - Direct and indirect identifiers have been removed or manipulated to break the linkage to real-world identities.

- **PSEUDONYMOUS**
  - Direct and indirect identifiers have been removed or manipulated to break the linkage to real-world identities.

**SAFE GUARDS and CONTROLS**

- **NOT FREQUENTLY USED in nature of data**
  - NOT FREQUENTLY USED in nature of data

- **LIMITED or MORE IN PLACE**
  - LIMITS on use or access

- **CONTROLS IN PLACE**
  - Controls for data breaches

**SELECTED EXAMPLES**

- Name, address, telephone number, SSN, government issued ID (e.g., Jane Smith, 123 Main Streeet, 555-555-5555)

- Unique device ID, license plate, medical record number, gender, race, birth date, etc.

- Social Security number, credit card number, bank account number, etc.

**ANONYMUS DATA**

- Direct and indirect identifiers have been removed or manipulated to break the linkage to real-world identities.

- For example, noise is added to a data set to hide whether an individual is present or not (differential privacy).
De-identification of Library PII Data

Obfuscation

- PII 1
  - Date of birth vs age

Truncation

- PII 1
  - Full address vs zip code
- PII 2
  - Call numbers

Aggregation

- PII 1
  - Age vs age ranges
- PII 2
  - Very high level call number ranges
Exercise - Raw vs De-identified Data

Raw data

- Date of birth - 2/24/1977
- 27 43rd St, Town, WA 92471
- NX180.I57 M275 2015
- FIC HARRIS 2018
- 11 Apr 2019 9:24 - 10:24

De-identified

- 42 years old
- 92471
- NX180
- FIC
- 4/11/2019, 01:00
Obligatory Disclaimers about De-identification

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.)

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
Vendors and Third Party Apps Considerations

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, contract addendums
Library Patron Data Lifecycle

Access
Who has access to what data?

Physical Access

- Desktop computers
- Laptops
- Mobile devices
- Server room/data center
- Offices and desks
- Flash drives
- File cabinets

Electronic Access

- ALL the User Permissions!
- Administrator account information
- Vendor access to local systems
- System log and database access
- Administrator site access
Some Data Access Best Practices

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

Reporting
Do you really need access to all the data for reports?

Reporting and publishing through...

- Views in databases
- Connections to data through Data BI tools (Tableau, Power BI, etc.)
  - Related - restrict access to data used in the report when published
- Dashboards and canned reports
Sometimes the only choice...  

... is to not release data.*

*except when required by law
Library Patron Data Lifecycle

Deletion
Evicting the Ghosts of Data Past

Electronic data
- Scrub backups and logs
- Deleting data vs wiping file/drive
- Look out for data living “outside” local and vendor systems

Physical data
- Paper - Shred → Proper disposal of shredded paper
- Properly dispose of disks, drives, other hardware
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?
Exercise - What Would You Do?

Library kept computer reservation logs on paper

Patrons would sign name and barcode to reserve a specific computer at a specific time of day

Library kept paper logs for a month

Library needs to track who was at which computer to charge for broken equipment or possible patterns of misuse of computer equipment

What would you advise the library to do to better protect patron privacy while still meeting their needs for reservations tracking?
A Data Inventory Starter Kit
Pick one piece of technology or process that your place of work is using or has used in the past.

- 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, paper documents, cloud/third-party products, 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?
Thank you

:-)