LIBRARY FREEDOM PROJECT

What is a Data Collection Audit?

Running a data collection audit is the first step in the process of developing or revising a privacy policy for your library. A data collection audit is a comprehensive look at all the ways the library collects and stores information from library users. Understanding the library's actual practices is a prerequisite for improving those practices, and for ensuring that library operations don't contradict the values and statements made in a privacy policy.

The **ALA's Data Life Cycles document** is a useful guide for grounding this process.

How do you do it?

Depending on the size of your library, this process may look different. In a large library system, you may need to sit down with staff across different departments to discuss actual practices for different library functions. In a small library setting, you may be familiar with all the processes, and can simply think through them yourself.

Make a list

Generate a list of any library operations that might involve collecting information. This will vary across different libraries, but some processes you might want to look at include:

- Library card sign-ups for adults and children
- Physical item circulation, including special items like Library of Things items
- Digital item circulation and other digital resource use
- Inter-library loan
- Item purchase requests and holds
- Holds shelves & paging
- Public computer use
- Reference questions and computer assistance
- Library website and OPAC use
- Copy/print/scan/fax services
- Library program registration & attendance
- Working with community partners & granting organizations
- Room reservations or appointments
- Library pass programs to museums & cultural institutions
- Art exhibits & library displays from community members
- Donation of library or archival materials
- Monetary donations, library merchandise purchases, and payments for fees/fines
- Print & email mailings
- Social media use
- Employee & volunteer records
- Phone calls, emails, web forms, and general correspondence
- Security cameras, incident logs, criminal trespass orders
- Lost & found items
- Reference materials that can only be used in the branch
- Community boards

Gather Info

Once you have generated your list of library operations, ask a series of questions for each one to determine how the library is collecting and managing the data involved in that process:

- What user data is collected?
- Is the patron informed about the data collection and given an opportunity to provide or deny consent?
- Why is this data being collected?
- Where is this data stored (all locations, physical/digital)?
- Who can access this data?
- Is this data anonymized? Can it be?
- How long is this data retained? What is the process for deleting it?
- Is this data backed up/duplicated in any systems or logs?

Take notes. Make sure that you are gathering information on how things are *actually working in your library*, not just how they are supposed to work in theory. Sometimes practice on the ground doesn't line up with intention, for lots of reasons. Be as honest as possible, because that is the only way to inform good policy in the end.

If discussing with other staff, stress that this is purely an information-gathering project, not intended to lay blame for practices that might not be aligned with library values around privacy. It may be helpful for non-supervisory staff to conduct the audit for this reason.

Common Pain Points

All libraries are different, but here are some common pain points related to privacy in public libraries:

- Patrons leave important private documents in photocopiers or printers
- Patrons are forced to share private information related to their computer tasks in a public computing area to get help
- Sign-up sheets for computers, rooms, or programs include the names and information of previous sign-ups visible to patrons
- User data is not properly erased from public computers
- Registration data for programs, appointments, and room bookings is retained long after the event has passed
- Holdshelf items do not anonymize patron data (have patron name and phone number instead of a hold alias or ID)
- Email interactions that involve patron information are archived or filed indefinitely
- Library staff leaving patron records open in ILS which leaves patron name, address, library card number, or phone number viewable by public
- Using third party systems, whose privacy settings you can't control, to collect or store patron information (eventbrite, Google Drive, etc)
- Not disposing of personal information correctly (using the recycling bin or trash can instead of the shredder)

Review & make recommendations

Once the data has been gathered for all the library processes that involve collecting personal information, review the notes and highlight areas where current practice is not aligning with library values around privacy.

Where are you over-collecting data that you don't really need? We should be striving to collect the minimum amount of information possible for library operations. There should be a clear reason that is necessary to library operations for collecting each piece of data.

Where are patrons not made aware of what their data is being used for? For example, it's common practice to add new patrons to print or digital mailing lists, but if this isn't disclosed during the card application process, the library is not respecting patrons' autonomy over their own information, even if they can opt out of the list later.

Where can you reduce access to or anonymize information? While some data must be kept for required reporting, consider where you can limit who has access to it, or anonymize it. This might look like moving files required to be kept in case of an audit into a locked drawer instead of an open cabinet, or deleting out fields with PII from a spreadsheet.

Where are you retaining data longer than necessary? Information should only be kept on file for as long as it is needed. Many institutions lack retention schedules for knowing when to delete certain types of information, and so default to keeping it indefinitely. Remember that personal information may end up stored in emails, downloaded files, calendar listings, etc.

Where practices are not aligned with library values, make recommendations and advocate for changes in operations to better protect patron (and staff!) privacy.

Use your findings to inform policy

Your library may or may not already have a privacy policy. If you do have one, it might be outdated. Whether you are developing a policy from scratch or trying to update an existing policy, use your findings from the data collection audit to guide you.

Present your findings to your Board as a way to be instructive about the needs for a privacy policy. Be careful about anything that seems to point to staff error, and rather framing issues in ways that illustrate that it is impossible to operate a public library without running up against certain pain points. This is an opportunity to present a case to the Board as to why a policy is necessary.

When creating or updating your policy, ensure that the policy does not misrepresent actual practices at the library (or, take the opportunity to change those practices before updating the policy), and that it does not contradict or unnecessarily duplicate other policies, like a Circulation Policy or Internet Access Policy.

Use the **Privacy Policy Template** from Library Freedom Project to get started if you are drafting a new policy, and adjust it as needed based on the results of your data collection audit.

Advocate

If you are struggling to implement changes in operations, you can cite policy as a reason to do so ("We really can't share that information with law enforcement without a warrant, because it explicitly states in our privacy policy that we don't do so"). Or, if you are finding it challenging to convince a board to adopt a policy, you can cite library operations ("We already do this in practice - I'm just asking that you codify it in our policy so we're in line with our professional values").

Libraries are unique in their commitment to privacy in the digital age, and it can be difficult to buck the trend of gathering as much data as possible and retaining it forever, especially when making the case to administrators or governing bodies without a background in libraries. Here are some arguments against common objections to good privacy practices that might work well when advocating to different audiences:

What if we need this information someday? Shouldn't we keep it just in case?

The data collection audit has already analyzed how long we need to retain each piece of information to provide library services. Keeping information for longer than it's needed increases the risk that it will be compromised in the event of a data breach. Additionally, all public libraries are required to report a majority of statistics such as program attendance in many categories, annually. There is no need to keep raw data beyond that point as we have access to all of our annual reports.

It's so much easier to do things the way we've been doing them. Why do we need to make this change?

Maintaining privacy is vital to fostering intellectual freedom. When people believe that they are being watched, they will self-censor their own choices. In order to fulfill our mission of connecting people with information, we need to make sure that we are protecting their privacy, and telegraph and communicate to patrons how important that is to us, even if it is less convenient.

Nobody else is doing this. Why should we care?

Privacy is a core value of librarianship. Libraries are uniquely positioned in that we are not trying to monetize people's data for profit, and we strive not only to protect patrons' privacy ourselves, but to equip them with the skills to better protect their own privacy in other situations. We must practice what we preach! There are also laws around library privacy [depending your state] that keep library records confidential.

The patrons don't even care about this. Why do we have to do it?

If individual patrons make the decision to opt out of protections for their privacy, that is their choice. But we should approach each interaction with the assumption that the patron does value their privacy. Consider a situation in which someone is trying to keep their information private from an abusive partner, family member, or government agency. We never know the situation someone might be in that would make their privacy very important to them.

Thank you to the Rose Foundation
for their generous support of our work.

ROSE FOUNDATION FOR COMMUNITIES & THE ENVIRONMENT