TECHNOLOGY, PRIVACY, CONFIDENTIALITY, AND SECURITY

As you've seen, most conflicts in the intersections of copyright policy, libraries, and technology come about because copyright policy constrains desirable technology or because copyright holders use technology to undermine copyright policy. Library policies play much lesser roles, except for the special case of preservation and archiving.

Some technology-policy issues work the other way around. New technologies and new applications of old technologies may be perfectly acceptable from legal and general social policy standpoints, but may conflict with library policies. What happens then?

With few exceptions, all libraries claim to protect user privacy and circulation confidentiality—and all libraries need to be concerned with the security of their collections (and their staff, but that's beyond the scope of this discussion).

The fundamental principal of user privacy means that a user's reading (listening, viewing) habits are strictly their own—that librarians don't concern themselves with those habits and strictly protect that information from others.

Circulation confidentiality is the same principle, but in reverse and on an aggregate basis. It's a relatively recent principle, at least in practice—after all, many public and other libraries used to use signature book cards, where past readership could be observed simply by reading the card.

Librarians care about collection security for obvious reasons: If the collection walks away on a regular basis, any library with a finite budget will cease to exist.

Balancing New Technologies, Privacy, Confidentiality, and Security

Innovative librarians keep on the lookout for new technologies that can improve library service. Companies develop new technologies and uses, then pitch them to libraries, pointing out the problems the new technologies can solve. That's as it should be; libraries have long been leaders in effective use of new technology, and should remain so.

Problems arise when new technologies and uses are implemented without considering the policy framework. Every technology, even seemingly minor ones, should receive at least a cursory policy scan.

If your library proceeds with a new technology that *does* affect privacy and confidentiality, and you haven't addressed those issues in advance, there's a good chance someone else will address them for you. That's particularly likely if you're on a coast, in a major urban area, or in any high-technology or upscale town or region.

When your users raise questions, you need to have answers. "We didn't think about that" generally doesn't serve very well as an answer.

The two examples that follow represent real-world situations, one where the technological development already exists in hundreds of libraries, and one where it's been suggested but rarely implemented. In neither case is the technology simply a bad idea. It's rarely that simple.

A third example considers technology that's been implemented in thousands of American libraries. Maybe you did a policy check when your library implemented that technology; maybe the policy check is still valid. Are you sure?

Radio Frequency Identification

Radio Frequency Identification (RFID) seems like a great idea for libraries, all the more so as the price of RFID tags keeps dropping. RFID may offer better security than existing systems but can also make circulation and returns faster, easier, and (particularly for returns) less likely to cause injuries to library staff.

Richard W. Boss summarizes some of the advantages:

> **Rapid charging/discharging.** The use of RFID reduces the amount of time required to perform circulation operations. The most significant time savings are attributable to the facts that information can be read from RFID tags much faster than from barcodes and that several items in a stack can be read at the same time . . .
>
> **Simplified patron self-charging/discharging.** For patrons using self-charging, there is a marked improvement because they do not have to carefully place materials within a designated template, and they can charge several items at the same time. Patron self-discharging shifts that work from staff to patrons. Staff is relieved further when readers are installed in bookdrops . . .
>
> **High-speed inventorying** . . . A hand-held inventory reader can be moved rapidly across a shelf of books to read all of the unique identification information. Using wireless technology, it is possible not only to update the inventory, but also to identify items [that] are out of proper order.
>
> **Automated materials handling** . . . This includes conveyer and sorting systems that can move library materials and sort them by category into separate bins or onto separate carts.[1]

Karen Schneider adds another indirect advantage:

> **Reduction in workplace injuries**. Workplace injuries caused by the repetitive motions related to flipping books and angling books under barcode readers cost libraries millions of dollars every year.[2]

This all sounds pretty good—good enough that at least 130 libraries in North America were already using RFID systems in August 2004, with hundreds more considering it.[3]

So What's the Problem?

Richard W. Boss says there isn't one:

> There is a perception among some that RFID is a threat to patron privacy. That perception is based on two misconceptions: (1) that the tags contain patron information and (2) that they can be read after someone has taken the materials to home or office.
>
> The vast majority of the tags installed in library materials contain only the item ID, usually the same number that previously has been stored on a barcode. The link between borrower and the borrowed material is maintained in the circulation module of the automated library system, and is broken when the

> material is returned. When additional information is stored on the tag, it consists of information about the item, including holding location, call number, and rarely author/title. The RFID tags can only be read from a distance of two feet or less because the tags reflect a signal that comes from a reader or sensor. It is, therefore, not possible for someone to read tags from the street or an office building hallway.[4]

If only it were so simple.

"Misconception" 2 is a simple fact: RFIDs *can* be read after someone has checked out the materials. That's not true of all RFIDs, to be sure. There are RFIDs that can be disabled permanently, for example RFIDs used as security devices in retail goods. Once they've been scanned by the right device, they should be inert.

Such RFIDs won't work for libraries. The whole point of a library RFID implementation is to use the same chip over and over, to charge *and* discharge an item, get it back to the right shelf, and assure it's where it should be. Library RFIDs are always readable: It's in the nature of the design.

Consider the last sentence in Boss's reassuring hand wave. Do you always walk more than two feet from the walls in an office hallway? (As a rough test, if you reach out your arm can you touch the wall? If so, you're probably closer than two feet.)

So you always keep a distance of more than two feet from any potential reader. The next-to-last sentence assumes that reader technology will never improve—that today's two feet won't be four feet, eight feet, or half a mile in another few years. That's a remarkably poor assumption, one that flies in the face of almost everything we know about improvements in transmitting and receiving technology.

A Little Paranoid Thought Experiment

Assume for the moment that Boss is right. He's certainly right on the first count: The RFID on a book does not, in and of itself, have any information on the patron. As he says, the link between borrower and the borrowed material is maintained in the circulation module of the automated library system.

The assumption that this link is broken when the book is returned is a bit facile, to be sure. Some systems retain that link either for a fixed period or until the next circulation, to allow time to check for damages. Some systems haven't been as strict about purging past circulation records as library policies should require.[5]

What's to stop a snoop (governmental or otherwise) from mounting a hidden reader just outside the library, near the "official" reader, or in a similar area where the two-foot limit is no problem? That gives the snoop a handy record of each item that enters or leaves the library. Combined with hidden cameras, it can identify who appears to have the item even without the use of patron identification.

Are such hidden cameras likely? There are already tens of thousands of full-time security cameras in use, with more to come. Their use is not only likely but probable.

For that matter, wouldn't it be convenient to use RFID chips in borrower's cards? After all, with RFID for items but barcodes for borrower's cards, a self-check station still needs both a laser scanner and an RFID scanner. Chip the library card and you've simplified the station.

After all, the patron's chip doesn't actually identify the patron (assuming your library uses patron numbers with no independent meaning). That link only exists in the library's database. How secure is your library's database? That's a significant question even without RFID chips in library cards, since it's the library's database that makes the item RFID meaningful by relating it to a bibliographic record. Without access to the database, the RFID information is useless.

Or is it?

Karen Schneider's Concerns

While discussing the advantages of RFID, Karen Schneider notes some issues. Skipping over those already discussed (such as library RFID tags *must* stay live, and computing and communications technology gets smaller, cheaper, and more powerful over time), consider her well-informed comments on several other issues—all of them policy issues that arise from this seemingly innocuous technology:

> 3. Libraries *should* only store barcode numbers on these tags, but we have yet to develop best practices profession-wide. At least one library in California has acknowledged that they store patron information on RFID tags . . .
>
> 4. Library databases are often maintained by library staff that "grew into" the job and may not have the training or expertise commonly associated with highly secure systems. It is dangerous to assume that library systems are so powerfully secured that they would be impervious to an organization seeking to probe databases in order to connect library barcodes with library records . . .
>
> 6. RFID cheaply and efficiently automates surveillance. . . . The promise of RFID is equal to its danger: It vastly reduces the labor overhead required to track items.
>
> 7. Reliance on features unique to library RFID is dangerous. . . . A truly privacy-friendly approach to RFID in libraries is to assume that all library RFID tags are world-readable, and work backwards from there.
>
> 8. Libraries nationwide have acknowledged that privacy concerns related to RFID are new territory . . .
>
> 10. Libraries have proved vulnerable to national agendas. Recent legislation . . . demonstrates that libraries have become highly porous battlegrounds for some of the larger privacy and public-forum debates in our society . . . With the PATRIOT Act, we have seen the government become increasingly inventive and aggressive in its efforts to track the reading habits of library users.[6]

Other Concerns

Potential problems don't stop there. David Molnar and David Wagner of the University of California, Berkeley, discuss ways library RFID chips may compromise reader privacy even *without* access to library databases. They discuss two dangers: tracking and hotlisting.

Tracking uses an item's RFID tag to follow the movements of that item—without knowing or much caring what the item itself is. To what purpose?

> Combined with video surveillance or other mechanisms, this may allow an adversary to link different people reading the same book. In this way, an adversary can begin profiling individuals' associations and make inferences about a particular individual's views, e.g., "this person checked out the same books as a known terrorist."[7]

Hotlisting? That's where someone compiles a list of items that it wants to recognize. Chances are, the RFID will contain the same number as the barcode on a book. What's to stop someone from going through the library

copying down barcodes for books of particular interest—or, for that matter, scanning the RFID tags to acquire *whatever* codes they contain, then relating those codes to the bibliographic information?

> Hotlisting is problematic because it allows an adversary to gather information about an individual's reading habits without a court order. For example, readers could be set up at security checkpoints in an airport, and individuals with hotlisted books set aside for special screening.[8]

Coping with RFID

None of this means libraries should shun RFID chips. It does mean that, as Schneider and Ayre both urge, libraries need to develop best practices and deep understanding of the possibilities. Ayre urges government privacy protections; other authors suggest a number of steps.

The first step is awareness and, subsequently, simple corollary steps, such as precluding the use of bar codes to search library catalogs. Molnar and Wagner offer specific technical options to improve the security of future library RFID systems; those options may not help existing installations, but—along with ALA best practices guidelines—they offer the likelihood RFID will be a less mixed blessing in the future.

One response of some futurists and technophiles to any question raised about privacy and confidentiality is there is no such thing as privacy, so you might as well get over it. That's not an acceptable answer—and would only become a true answer if libraries and other agencies choose to make it true.

Collaborative Recommendations and Similar Services

Why can't library catalogs be more like Amazon? Variations of that cry have risen in various quarters. Depending on what "more like Amazon" really means, one answer is that many of them already have—catalogs showing book covers, including tables of contents, linking to reviews.

What some people mean by "more like Amazon" is a collaborative filtering and recommendation technology that suggests new items for your consideration, based on some combination of your own buying patterns and combined patterns of other purchases. "People who purchased x also purchased y" represents a simple form of collaborative recommendation; the technology can go much further.

Since this isn't a discussion of Amazon, there's no point in considering whether Amazon's collaborative recommendation engine is unbiased. Some similar engines do appear to operate without bias (and to serve the company's aims in doing so), with Netflix being one of the most widely used. Netflix invites you to rate as many movies as you can. Based on those ratings, the records of what you've already viewed and liked, and similar records for a couple of million other viewers, Netflix can offer surprisingly apt suggestions for movies you might never have considered but will probably enjoy.

Wouldn't it be great if a library catalog could do the same—offer a personal reader's advisory that suggests some books (or CDs or DVDs) that you might really enjoy, based on your past borrowing and related borrowing records from other library users?

Given cheap disk storage and high-speed computing, the technology is feasible now. As far as I know, it hasn't been implemented in public libraries.

Confidentiality Issues

The problem with collaborative recommendations is that to work really well, they rely on stored knowledge of your past history and that of others. How do you provide such stored knowledge without compromising confidentiality?

There may be answers to that question, but those answers require testing and thought. At first glance, it seems problematic. You could achieve one level of collaboration by only coupling items taken out at the same time and storing those links with codes that can never be linked to an actual borrower. Thus, you could say that "someone took out book a, book b, book c, and DVD d at the same time."

If that pattern happens often enough, then you could suggest that someone else who checks out book a and book c *might* find book c and DVD d interesting. But that's a weak database—and it will keep recommending books a user has *already* read, which is likely to be more annoying than useful.

You'll have much stronger recommendations if the engine can track borrowing habits over time. I don't know how you could do that while maintaining confidentiality.

There is a way to avoid the problem of recommending an already-read book over and over, but it involves significant overhead. If records of a user's past circulation are only maintained on that user's own PC (or better yet, on a flash USB drive), stored in some encrypted manner that only the library database can relate to actual items, those records could be used on the fly to provide new recommendations without *necessarily* endangering privacy or confidentiality, assuming a secure link is used for the process.

These aren't trivial problems. They shouldn't be solved by asking users to acknowledge their reading history may not be private if they want new book recommendations. Library users don't generally have or need the same background or depth of awareness of privacy issues as librarians.

It's the job of librarians to maintain library principles, not to attract users to waive those principles by offering shiny new toys. I'm sure very few PC users want adware or spyware installed on their machines, but millions of them "signed" forms consenting to add such adware or spyware so they could achieve some desirable end.

In the case of RFID, the dangers may be limited and controllable compared to the benefits. It's not at all clear that the supposed benefit of automated reader's advisory outweighs the dangers, or that the dangers can be eliminated at reasonable cost. Before any such system comes into play, those issues need to be studied and resolved.

Online Access to User Records

This doesn't require much discussion. You probably offer Web access to your online catalog; most libraries do. There's a good chance you also allow library users to view their current records—to see what they have out and renew items online. Many libraries offer that service.

Are you sure you're not compromising privacy in the process?

Do you require that users register and create passwords before showing them their current item list? Probably not. Does the circulation information operate over a secure link? Again, probably not.

Does it matter? Possibly. If, by some chance, you allow users to login with *only* their card number, and if (worse) you then show them their record including name, then it certainly does. All an interested party needs to do is get a library card, figure out the range of numbers your library is using (and the check digit methodology, usually easy enough to determine), and the party can set up a harvester to associate all current circulation with the people holding the items.

But you probably don't make things that easy. My library doesn't use a secure link and doesn't use passwords, but it does require that you enter your name and card number. If anyone else gets that information, they can check on your current reading any time—but at least that's a smaller risk—assuming, of course, the database that links card numbers to patron names is truly secure. Is that a safe assumption?

Conclusion

Most new technologies don't raise major policy questions. Many new technologies raise more policy questions than the average librarian wants to consider. The first step in making sure that technology doesn't undermine policy is to consider the possibility.

Notes

www.ala.org/ala/pla/plapubs/technotes/rfidtechnology.htm

[1] Richard W. Boss,"RFID Technology for Libraries," ALA Tech Notes, May 14, 2004.

www.senate.ca.gov/ftp/SEN/COMMITTEE/STANDING/ENERGY/_home/11-20-03/karen.pdf

[2] Karen G. Schneider, "RFID and Libraries: Both Sides of the Chip," testimony presented at Committee on Energy and Utilities, California Senate, Nov. 20, 2003.

www.galecia.com/included/docs/position_rfid_permission.pdf

[3] Lori Bowen Ayre, "Position Paper: RFID and Libraries," draft chapter for *Wireless Privacy: RFID, Bluetooth and 802.11* (Boston: Addison-Wesley/Prentice Hall, 2005).

[4] Boss,"RFID Technology for Libraries."

[5] There is no question that some library systems have, at least in the past, chosen options within their integrated library systems that retain circulation records indefinitely: That's been personally confirmed by the author, albeit a few years ago. That choice may be useful for data mining, but in today's political environment, it's an extremely unwise decision. I assume all such decisions have since been reversed.

[6] Schneider, "RFID and Libraries."

www.cs.berkeley/edu/~dmolnar/library.pdf

[7] David Molnar and David Wagner, "Privacy and Security in Library RFID: Issues, Practices, and Architectures." Accessed October 19, 2004, www.cs.berkeley/edu/~dmolnar/library.pdf.

[8] Ibid.