Privacy and User Experience in 21st Century Library Discovery

Shayna Pekala

ABSTRACT

Over the last decade, libraries have taken advantage of emerging technologies to provide new discovery tools to help users find information and resources more efficiently. In the wake of this technological shift in discovery, privacy has become an increasingly prominent and complex issue for libraries. The nature of the web, over which users interact with discovery tools, has substantially diminished the library’s ability to control patron privacy. The emergence of a data economy has led to a new wave of online tracking and surveillance, in which multiple third parties collect and share user data during the discovery process, making it much more difficult, if not impossible, for libraries to protect patron privacy. In addition, users are increasingly starting their searches with web search engines, diminishing the library’s control over privacy even further.

While libraries have a legal and ethical responsibility to protect patron privacy, they are simultaneously challenged to meet evolving user needs for discovery. In a world where “search” is synonymous with Google, users increasingly expect their library discovery experience to mimic their experience using web search engines. However, web search engines rely on a drastically different set of privacy standards, as they strive to create tailored, personalized search results based on user data. Libraries are seemingly forced to make a choice between delivering the discovery experience users expect and protecting user privacy. This paper explores the competing interests of privacy and user experience, and proposes possible strategies to address them in the future design of library discovery tools.

INTRODUCTION

On March 23, 2017, the internet erupted with outrage in response to the results of a Senate vote to roll back Federal Communications Commission (FCC) rules prohibiting internet service providers (ISPs), such as Comcast, Verizon, and AT&T, from selling customer web browsing histories and other usage data without customer permission. Less than a week after the Senate vote, the House followed suit and similarly voted in favor of rolling back the FCC rules, which were set to go into effect at the end of 2017. The repeal became official on April 3, 2017 when the President signed it into law. This decision by U.S. lawmakers serves as a reminder that today’s internet economy is a data economy, where personal data flows freely on the web, ready to be compiled and sold to the highest bidder. Continuous online tracking and surveillance has become the new normal.

Shayna Pekala (shayna.pekala@georgetown.edu) is Discovery Services Librarian, Georgetown University Library, Washington, DC.
ISP are just one of the many players in the online tracking game. Major web search engines, such as Google, Bing, and Yahoo, also collect information about users’ search histories, among other personal information. By selling this data to advertisers, data brokers, and/or government agencies, these search engine companies are able to make a profit while providing the search engines themselves for “free.” In addition to profiting off of user data, web search engines also use it to enhance the user experience of their products. Collecting and analyzing user data enables systems to learn user preferences, providing personalized search results that make it easier to navigate the ever-increasing sea of online information.

The collection and sharing of user data that occurs on the open web is deeply troubling for libraries, whose professional ethics embody the values of privacy and intellectual freedom. A user’s search history contains information about a user’s thought process, and the monitoring of these thoughts inhibits intellectual inquiry. Libraries, however, would be remiss to dismiss the success of web search engines and their use of data altogether. MIT’s preliminary report on the future of libraries urges, “While the notion of ‘tracking’ any individual’s consumption patterns for research and educational materials is anathema to the core values of libraries...the opportunity to leverage emerging technologies and new methodologies for discovery should not be discounted.”

This article examines the current landscape of library discovery, the competing interests of privacy and user experience at play, and proposes possible strategies to address them in the future design of library discovery tools.

BACKGROUND

Library Discovery in the Digital Age

The advent of new technologies has drastically shaped the way libraries support information discovery. While users once relied on shelf-browsing and card catalogs to find library resources, libraries now provide access to a suite of online tools and interfaces that facilitate cross-collection searching and access to a wide range of materials. In an online environment, many paths to discovery are possible, with the open web playing a newfound and significant role.

Today’s library discovery tools fall into three categories: online catalogs (the patron interface of the integrated library system (ILS)), discovery layers (a patron interface with enhanced functionality that is separate from an ILS), and web-scale discovery tools (an enhanced patron interface that relies on a central index to bring together resources from the library catalog, subscription databases, and digital repositories). These tools are commonly integrated with a variety of external systems, including proxy servers, inter-library loan, subscription databases, individual publisher websites, and more. For the most part, libraries purchase discovery tools from third-party vendors. While some libraries use open source discovery layers, such as Blacklight or VuFind, there are currently no open source options for web-scale discovery tools.
Outside of the library, web search engines (e.g. Google, Bing, and Yahoo), and targeted academic discovery products (e.g. Google Scholar, ResearchGate, and Academia.edu) provide additional systems that enable discovery. In fact, web search engines, particularly Google, play a significant role in the research process. Both students and faculty use Google in conjunction with library discovery tools. Students typically use Google at the beginning of the research process to get a better understanding of their topic and identify secondary search terms. Faculty, on the other hand, use Google to find out how other scholars are thinking about a topic. Unsurprisingly, Google and Google Scholar provide the majority of content access to major content platforms.

The Data Economy and Online Privacy Concerns

In an information discovery environment that is primarily online, new threats to patron privacy emerge. In today’s economy, user data has become a global commodity. Commercial businesses have recognized the value of data mining for marketing purposes. Bjorn Bloching, et. al. explain, “From cleverly aggregated data points, you can draw multiple conclusions that go right to the heart and mind of the customer.” Along the same lines, the ability to collect and analyze user data is extremely valuable to government agencies for surveillance purposes, creating an additional data-driven market.

The increasing value of user data has drastically expanded the business of online tracking. In her book, Dragnet Nation, journalist Julia Angwin outlines a detailed taxonomy of trackers, including various types of government, commercial, and individual trackers. In the online information discovery process, multiple parties collect user data at different points. Consider the following scenario: a user executes a basic keyword search in Google to access an openly available online resource. In the fifteen seconds it takes the user to get to that resource, information about the user’s search is collected by the internet service provider (ISP), the web browser, the search engine, the website hosting the resource, and any third-party trackers embedded in the website. The search query, along with the user’s Internet Protocol (IP) address, become part of the data collector’s profile on the user. In the future, the data collector can sell the user’s profile to a data broker, where it will be merged with profiles from other data collectors to create an even more detailed portrait of the user. The data broker, in turn, can sell the complete dataset to the government, law enforcement, commercial businesses, and even criminals. This creates serious privacy concerns, particularly since users have no legal right over how their data is bought and sold.

Privacy Protection in Libraries

Libraries have deeply-rooted values in privacy and strong motivations to protect it. Intellectual freedom, the foundation on which libraries are built, necessarily requires privacy. In its interpretation of the Library Bill of Rights, the American Library Association (ALA) explains, “In a library (physical or virtual), the right to privacy is the right to open inquiry without having the subject of one’s interest examined or scrutinized by others.” Many studies support this idea,
having found that people who are indiscriminately and secretly monitored censor their behavior and speech.\textsuperscript{18}

Libraries have both legal and ethical obligations to protect patron privacy. While there is no federal legislation that protects privacy in libraries, forty-eight states have regulations regarding the confidentiality of library records, though the extent of these protections varies by state.\textsuperscript{19}

Because these statutes were drafted before the widespread use of the internet, they are phrased in a way that addresses circulation records and does not specifically include or exclude internet use records (records with information on sites accessed by patrons) from these protections. Therefore, according to Theresa Chmara, libraries should not treat internet use records any differently than circulation records with respect to confidentiality.\textsuperscript{20}

The library community has established many guiding documents that embody its ethical commitment to protecting patron privacy. The ALA Code of Ethics states in its third principle, “We protect each library user’s right to privacy and confidentiality with respect to information sought or received and resources consulted, borrowed, acquired or transmitted.”\textsuperscript{21} The International Federation of Library Associations and Institutions (IFLA) Code of Ethics has more specific language about data sharing, stating, “The relationship between the library and the user is one of confidentiality and librarians and other information workers will take appropriate measures to ensure that user data is not shared beyond the original transaction.”\textsuperscript{22} The library community has also established practical guidelines for dealing with privacy issues in libraries, particularly those issues relating to digital privacy, including the ALA Privacy Guidelines\textsuperscript{23} and the National Information Standards Organization (NISO) Consensus Principles on User’s Digital Privacy in Library, Publisher, and Software-Provider Systems.\textsuperscript{24} Additionally, The Library Freedom Project was launched in 2015 as an educational resource to teach librarians about privacy threats, rights, and tools, and in 2017, the Library and Information Technology Association (LITA) released a set of seven privacy checklists\textsuperscript{25} to help libraries implement the ALA Privacy Guidelines.

\textit{Personalization of Online Systems}

While user data can be used for tracking and surveillance, it can also be used to improve the digital user experience of online systems through personalization. Because the growth of the internet has made it increasingly difficult to navigate the continually growing sea of information online, researchers have put significant effort into designing interfaces, interaction methods, and systems that deliver adaptive and personalized experiences.\textsuperscript{26} Angsar Koene, et. al. explain, “The basic concept behind personalization of on-line information services is to shield users from the risk of information overload, by pre-filtering search results based on a model of the user’s preferences... A perfect user model would...enable the service provider to perfectly predict the decision a user would make for any given choice.”\textsuperscript{27} The authors continue to describe three main flavors of personalization systems:

1. content-based systems, in which the system recommends items based on their similarity to items that the user expressed interest in;
2. collaborative-filtering systems, in which users are given recommendations for items that other users with similar tastes liked in the past; and
3. community-based systems, in which the system recommends items based on the preferences of the user’s friends.28

Many popular consumer services, such as Amazon.com, YouTube, Netflix, Google, etc., have increased (and continue to increase) the level of personalization that they offer.29 One such service in the area of academic resource discovery is Google Scholar’s Updates, which analyzes a user’s publication history in order to predict new publications of interest.30 Libraries, in contrast, have not pressed their developers and vendors to personalize their services in favor of privacy, even though studies have shown that users expect library tools to mimic their experience using web search engines.31 Some web-scale discovery services do, however, allow researchers to set personalization preferences, such as their field of study, and, according to Roger Schonfeld, it is likely that many researchers would benefit tremendously from increased personalization in discovery.32 In this vein, the American Philosophical Society Library recently launched a new recommendation tool for archives and manuscripts that uses circulation data and user-supplied interests to drive recommendations.33

Opportunities for User Experience in Library Discovery

A major challenge in today’s online discovery environment is that the user is inhibited by an overwhelming number of results. This leads to users rely on relevance rankings and to fail to examine search results in depth. Creating fine-tuned relevance ranking algorithms based on user behavior is one remedy to this problem, but it relies on the use of personal user data.34 However, there may be opportunities to facilitate data-driven discovery while maintaining the user’s anonymity that would be suitable for library (and other) discovery tools. Irina Trapido proposes that relevance ranking algorithms could be designed to leverage the popularity of a resource measured by its circulation statistics or by ranking popular or introductory materials higher than more specialized ones to help users make sense of large results sets.35 Michael Schofield proposes “context-driven design” as an intermediary solution, whereby the user opts in to have the system infer context from neutral device or browser information, such as the time of day, business hours, weather, events, holidays, etc.36 Jason Clark describes a search prototype he built that applies these principles, but he questions whether these types of enhancements actually add value to users.37 Rachel Vacek cautions that personalization is not guaranteed to be useful or meaningful, and continuous user testing is key.38

DISCUSSION

There are several aspects to consider for the design of future library discovery tools. The integrated, complex nature of the web causes privacy to become compromised during the information discovery process. Library discovery tools have been designed not to retain borrowing records, but have not yet evolved to mask user behavior, which is invaluable in today’s data economy. It is imperative that all types of library discovery tools have built-in functionality to
protect patron privacy beyond borrowing records, while also enabling the ethical use of patron
data to improve user experience.

Even if library discovery tools were to evolve so that they themselves were absolutely private
(where no data were ever collected or shared), other online parties (ISPs, web browsers,
advertisers, data brokers, etc.) would still have access to user data through other means, such as
cookies and fingerprinting. The operating reality is such that privacy is not immediately and
completely controllable by libraries. Laurie Rinehart-Thompson explains, “In the big picture,
privacy is at the mercy of ethical and stewardship choices on the part of all information
handlers.”39 While libraries alone cannot guarantee complete privacy for their patrons, they can
and should mitigate privacy risks to the greatest extent possible.

At the same time, ignoring altogether the benefits of using patron data to improve the discovery
user experience may threaten the library’s viability in the age of Google. Roger Schonfeld explains,
“If systems exclude all personal data and use-related data, the resulting services will be one-
dimensional and sterile. I consider it essential for libraries to deliver dynamic and personalized
services to remain viable in today’s environment; expectations are set by sophisticated social
networks and commercial destinations.”40 Libraries must find ways to keep up with greater
industry trends while adhering to professional ethics.

RECOMMENDATIONS

While libraries have traditionally shied away from collecting data about patron transactions, these
conservative tendencies run counter to the library’s mission to provide outstanding user
experience and the need to evolve in a rapidly changing information industry. As the profession
adopts new technologies, ethical dilemmas present themselves that are tied into their use. While
several library organizations have issued guidance for libraries about the role of user data in these
new technologies, this does not go far enough. The NISO Privacy Principles, for instance,
acknowledge that its principles are merely “a starting point.”41 Examining the substance of these
guidelines is important for confronting the privacy challenges facing library discovery in the 21st
century, but there are additional steps libraries can take to more fully address the competing
interests of privacy and user experience in library discovery and in library technologies more
generally.

*Holding Third Parties Accountable*

Libraries are increasingly at the mercy of third parties when it comes to the development and
design of library discovery tools. Unfortunately, these third parties not have the same ethical
obligations to protect patron privacy that librarians do. In addition, the existing guidance for
protecting user data in library technologies is directed towards librarians, not third party vendors.
The library community must hold third parties accountable for the ethical design of library
discovery tools. One strategy for doing this would be to develop a ranking or certification process
for discovery tools based on a community set of standards. The development of HIPAA-compliant
records management systems in the medical field sets an example. Because healthcare providers are required by law to guarantee the privacy of patient data, they must select Electronic Health Records systems (ERMs) that have been certified by an Office of the National Coordinator for Health Information Technology (ONC)-authorized body. In order to be certified, the system must adhere to a set of criteria adopted by the Department of Health and Human Services, which includes privacy and security standards. Another example is the Consumer Reports standard and testing program for consumer privacy and security, which is currently in development. Consumer Reports explains the reason for developing this new privacy standard, “If Consumer Reports and other public-interest organizations create a reasonable standard and let people know which products do the best job of meeting it, consumer pressure and choices can change the marketplace.” Libraries could potentially adapt the Consumer Reports standards and rating system for library discovery tools and other library technologies.

*Engaging in UX Research & Design*

Libraries should not rely on third parties alone to address privacy and user experience requirements for library discovery tools. Libraries are well-poised to become more involved in the design process itself by actively engaging in user experience research and design. The opportunities for “context-driven design” and personalization based on circulation and other anonymous data are promising for library discovery but require ample user testing to determine their usefulness. Understanding which types of personalization features offer the most value while preserving privacy is key to accelerating the design of library discovery tools. The growth of User Experience Librarian jobs and the emergence of User Experience teams and departments in libraries signals an increasing amount of user experience expertise in the field, which can be leveraged to investigate these important questions for library discovery.

*Illuminating the Black Box*

When librarians adopt new discovery tools without fully understanding their underlying technologies and the data economy in which they operate, this does not serve users. Librarians have ethical obligations that should require them to thoroughly understand how and when user data is captured by library discovery tools and other web technologies, and how this information is compiled and shared at a higher level. Not only do librarians need to understand the technical aspects of discovery technologies, they also need to understand the related user experience benefits and privacy concerns and the resulting ethical implications. As technology continues to evolve, librarians should be required to engage in continued learning in these areas. Such technology literacy skills could be incorporated in the curriculum of Library and Information Science degree programs, as well as in ongoing professional development opportunities.

*Empowering Library Users*

Because information discovery in an online environment introduces new privacy risks, communication about this topic between librarians and patrons is paramount. Librarians should
proactively discuss with patrons the potential risks to their privacy when conducting research online, whether they are using the open web or library discovery tools. It is ultimately up to the patron to weigh their needs and preferences in order to decide which tools to use, but it is the librarian’s responsibility to empower patrons to be able to make these decisions in the first place.

CONCLUSION

With the rollback of the FCC privacy rules that prohibit ISPs from selling customer search histories without customer permission, understanding digital privacy issues and taking action to protect patron privacy is more important than ever. While privacy and user experience are both necessary and important components of library discovery systems, their requirements are in direct conflict with each other. An absolutely private discovery experience would mean that no user data is ever collected during the search process, whereas a completely personalized discovery experience would mean that all user data is collected and utilized to inform the design and features of the system. It is essential for library discovery tools to have built-in functionality that protects patron privacy to the greatest extent possible and enables the ethical use of patron data to improve user experience. The library community must take action to address these requirements beyond establishing guidelines. Holding third party providers to higher privacy standards is a starting point. In addition, librarians themselves need to engage in user experience research and design to discover and test the usefulness of possible intermediary solutions. Librarians must also become more educated as a profession on digital privacy issues and their ethical implications in order to educate patrons about their fundamental rights to privacy and empower them to make decisions about which discovery tools to use. Collectively, these strategies enable libraries to address user needs, uphold professional ethics, and drive the future of library discovery.

REFERENCES


amended January 22, 2008,
http://www.ala.org/advocacy/proethics/codeofethics/codeethics.

22. “IFLA Code of Ethics for Librarians and other Information Workers,” International Federation of Library Associations and Institutions, August 12, 2012,

23. “Privacy & Surveillance,” American Library Association, approved 2015-2016,
http://www.ala.org/advocacy/privacyconfidentiality.


28. Ibid., 168.

29. Ibid.


31. Schonfeld, Meeting Researchers Where They Start, 2.


35. Ibid.


39. Laurie A. Rinehart-Thompson, Beth M. Hjort, and Bonnie S. Cassidy, “Redefining the Health Information Management Privacy and Security Role,” *Perspectives in Health Information Management* 6 (2009): 4s


