The impact of technology on user-centric discovery

Timothy M. McGeary[*][*]

Associate University Librarian for Digital Strategies and Technology, Duke University Libraries, Box 90193, Duke University, Durham, NC, USA
E-mail: tim.mcgeary@duke.edu; ORCID: http://orcid.org/0000-0001-9694-7650

Abstract. The amount of resources and information available to users seems infinite, but what is actually discoverable and accessible is much less in comparison. Among the many longstanding roles of libraries is guiding users to the resources that they need, and although technology has expanded the possibilities of this service, it has also exposed new risks and challenges. Libraries also have a unique position within the technology-driven marketplace, as both a content consumer and a content provider. They also have conflicting objectives - protecting users while competing for users. This paper explores these dynamics that result from the impact of technology on library user-centric strategy and solutions.

Keywords: User-centric discovery, discovery services, libraries and information discovery

1. Introduction

User-centric discovery has been the foundation of libraries since their origin. Nearly everyone who works in a library has some impact, no matter how large or small, in enabling user access to information. Throughout centuries of library history, technology has had a major impact on libraries and library users, from new media or formats of publishing to the creation of MARC to the Internet. Regardless of the paradigm or catalyst of change, the user has always been in the center, the focus of improvement and service. As services improve, users respond. As users see improvements in other service contexts, they demand improvements in library services. The rate of consumer-based services has demanded a lot from libraries, and libraries have responded admirably and with strong commitment to their users.

Yet in the rapid pace of technological change during the Internet era, the user is not only the focus of improvement, but also a target. The user is not just a consumer of information, they are information. The user does not merely use or share data, they are data. Users expect both personalization and privacy, regardless of their behavior or whether they are truly informed of consequences of their behavior. Service providers, including libraries, want to make data-driven decisions for continuous improvement. Libraries are one of the strongest advocates of users and user privacy, yet libraries are competing with service providers that blatantly leverage user data for their own benefit or profits. Can we continue to develop user-centric discovery without compromising our deepest held values about privacy? Can libraries enforce user

---

[*]E-mail: tim.mcgeary@duke.edu.

privacy in the face of technology advancement? This article will explore this tension through the impact technology has made on user-centric discovery over the past few decades.

To adequately discuss the impact of technology today on user-centric discovery, we should do a brief review of the last three decades of discovery innovations within libraries. The pace of technology innovation continues to accelerate, which increases pressure on the ability of libraries to develop and incorporate discovery services to users. As we will explore throughout this discussion, financial resources and considerations are paramount, while libraries have struggled to emerge from the constraints placed during the 2008 recession.

2. Brief and selective history of discovery innovations in libraries

The advent of web-based online public access catalogs, or OPACs, in the 1990's was one of the first steps taken by libraries to implement user-centric discovery in the dawn of the commodity Internet. It was a major bridge to empower the average library user with the ability to search on their own on the open Internet anywhere on or off campus, using expanded keyword searching directly from the same data librarians and library staff used to manage and inventory their collections, and ultimately giving the user a greater awareness of the entirety of the library's collection. While this would take some time to become clear, the implementation of web-based OPACs initiated a new path of prioritization, starting to tip the scales to the needs of the user over the needs of the library, though nearly every college and university library have stories of a few faculty that refused to use the OPAC or decried the removal of the card catalog.

Web-based OPACs were the standard for twenty years, with incremental updates and improvements that represented the limited pace of technology advancement afforded by the rigid backend systems. In the early to mid-2000's, it became clear that development of these web interfaces would be stifled by systems primarily developed in the 1960's through 1980's. So while library technologists sought to overcome these limitations for the good of the end user, librarians and library vendors began to re-examine the basic architecture of library management systems and developed blue-sky ideas for the good of the library staff user. As Google began to take the lead in search-engine brand recognition and retail companies implemented improved web search interfaces by giving users options to limit their searches by facets, some libraries, notably the North Carolina State University Libraries, saw this changing environment as an opportunity.

The North Carolina State University Libraries acquired the proprietary Endeca discovery software to improve the library, both as a benefit to their users and ultimately creating new challenges for their own staff development. This improved discovery interface implemented new matching techniques, multiple relevance ranking strategies, and significant other user-centric tools [1].

The response to the early success and public presentations of NCSU's Endeca discovery interfaces was strong, yet many libraries also found the cost of the software to be out of reach. Through his work on developing software for digital collections with XML-based metadata, Andrew Nagy, a developer in the Villanova University Library, developed a strategy to use Apache Solr and Lucene to develop a PHP-based open-source version similar to Endeca called VuFind [2]. Within the same time frame, the University of Virginia developed a parallel project called Blacklight, using Ruby-on-Rails with Apache Solr and Lucene [3]. Open source software development within libraries was merely experimental and incubating, but these innovations became catalysts for rapid growth in community participation of open source projects and significant investment by vendors, such as Innovative Interfaces Encore and SirsiDynix Enterprise Portal,

both launched in 2006, while Ex Libris Primo and WorldCat Local launched shortly after in 2007. The slow pace of incremental change to user-centric discovery was over.

OPACs and faceted discovery interfaces were effective because the primary source for discovery was encapsulated completely in integrated library systems. But during this time subscriptions to electronic journals were increasing exponentially, content aggregation platforms were being developed, and new publishing practices emerged. Libraries were purchasing and licensing the same content in different formats, a practice which the ILS was unable to manage completely. So while proprietary and open source discovery interfaces competed for implementation of discovery of local resources, searching for licensed electronic content was severely lacking. Users were able to search for titles of journals and databases in local discovery interfaces, but they could not search article titles or metadata about the articles. When Google Scholar was released as an active beta, it looked promising, but appeared to be insufficient to meet user needs. OpenURL Resolvers built atop of e-Resource Knowledge Bases were instrumental in connecting users to a resource, but they were not developed for discovery. Thus in 2009–2010, the index-based discovery services, Proquest Summon and EBSCO Discovery Service, were launched. Each leveraged the proprietary Knowledge Bases that that the individual companies had built for electronic resource management services.

3. Competition: The single search box

During the first half of this decade, Google clearly solidified its stake as the trusted market leader for search engines, many more libraries were implementing external or vendor-provided faceted discovery interfaces, and even more libraries were investing in index-based discovery services. With three potential starting points for search, libraries astutely assessed that the competition of discovery would be problematic for user-centric design, resulting in inconsistent research experiences, and make it difficult to develop core information literacy and research techniques.

Moreover, the need for effective staff management software for electronic resources often resulted in financial incentives to bundle electronic resource management software with index-based discovery services which may not necessarily provide the best user-centric discovery solution [4].

Libraries have spent a significant amount of effort in researching and comparing local discovery services, index-based discovery services, and Google for the purpose of developing a coherent user-centric strategy while also attempting to reinforce the library as the gateway to information, specifically seeing Google as both a source of inspiration and arch-competitor [5]. Comparing EDS, Summon, and Google Scholar on the basis of relevance ranking demonstrate equality between EDS and Summon. And while Google Scholar is equivalent to both for known-item searches, in topical searches Google Scholar excels [6]. A report of a 2018 survey of librarians demonstrates that the search for the best discovery tool to serve all users is elusive: difficulty with combining metadata from multiple sources, rampant lack of interoperability, and balancing the needs of library-centric management with user-centric discovery are just a few of the challenges that persist [7].

While the Google single search box seems simple, the algorithms and sources Google uses are complex, and furthermore Google incorporates proprietary methods for providing consumer-driven personalization (see [4]). A comparison is often made to the complexity of the library systems environment of local physical collections, licensed electronic collections, digital resources, and the library website [8]. And thus libraries have been attempting to present a single search box as a user-centric strategy and as a demonstration of confidence that the library's search tools can fulfill the user's needs [9]. What is

becoming clear, however, is how unlikely it is that an open access discovery interface will ever fully compete against commercial offerings due to the financial and human resource investment required of the library community (see [3]).

4. What do users want?

But what do library users want? Studies in North America [10] (see also [8]) the United Kingdom [11], and Australia [12], for example, consistently reflect the expectation to search in one place and find all of the resources that they need. Additionally strong is the desire for personalization and a seamless user experience, including integration with social media and other resource tools that support collaboration. The Sero Consulting study of UK users (see [11]) describes in-depth the user demands based on the "digital native" paradigm", including:

- Users expect to search and access library resources online anywhere: on campus in the library, on campus outside of the library, off campus, and on any device;
- Users expect the ability to select different discovery tools to meet different requirements, yet ultimately attain successful results;
- Users expect modern digital services, like mobile apps, from the library to empower them with just-in-time support and services based on their registered profiles.

Most notable in the Sero Consulting study is the Venn diagram visualizing *where* users access library content, with 32.5% of users selecting locations other than the library premises. Thus, there are many tensions to work through for libraries in order to implement user-centric discovery. Among the most significant are:

- Users increasingly want to access resources outside the library and from off-campus locations;
- Users want the ability to select the discovery tools;
- Users want the personalization to which they have become accustomed when using commercial consumer services while libraries are committed (and often required) to protect anonymity and intellectual freedom;
- Library-supported open source discovery interfaces will likely never overcome commercially-produced discovery tools, especially Google, in a sustainable manner.

The Library community must develop a roadmap for managing user data and privacy, develop internal library services that center around user needs, and negotiate external library services that serve users without turning them into a commodity.

5. User data and privacy

It is important to address these issues head on and identify both the technology challenges and potential outcomes. The most critical elements that weave through the list above are user data and privacy. How do libraries reconcile user-expressed desire for library service with library values to protect user privacy? While at times speaking of privacy as a sort of "sacred cow" to protect at all costs, libraries drive its users to services like Facebook and Instagram that are based entirely on user data, and as such share, sell, and

sometimes exploit user data and user behavior. This would be ironic if not for the fact that every one of us gives up our privacy in the consumer services we choose to use [13].

Privacy and confidentiality lay the foundation of free thought, free speech, and free association. Monitoring and data mining may intellectually constrain what a patron searches for, reads, and ultimately, thinks. Our library users expect this privacy for intellectual freedom, or simply, "the right to be left alone - the most comprehensive of rights, and the right most valued by a free people" as Supreme Court Justice Louis Brandeis wrote in 1928 (Olmstead v. U.S.) [14].

Is there a middle ground to be found to develop user-centric discovery solutions that allow libraries to preserve its honor protecting the privacy of its users? Are there ways that libraries need to reasonably compromise in their strict adherence to privacy for the goal of serving its users, and empowering users to make informed choices? Can there be trust between libraries and commercial organizations when it seems every news cycle presents details of another data breach, secretive corporate behavior, or never-ending discoveries of security vulnerabilities?

Universities and libraries now realize what commercial and consumer providers have long known: the data about their users is deep, untapped, and full of potential. While libraries have long aimed to protect this data from being used harmfully, the era of constrained financial support, especially for libraries, requires much more intentional data-driven decision-making. Information Security Offices have been more successful in wielding this double-edged sword to protect users from harm and leverage data to build high-performing technology services not only to support, but also to enhance research, teaching, and scholarship. It is time for libraries to take stock of the data it could be collecting, yet promptly discarding for the sake of privacy, and investigate how to leverage this data to build more user-centric solutions, while also maintaining our values of protecting privacy and intellectual freedom. Libraries are data controllers, and data privacy laws, such as the General Data Protection Regulation (GDPR) [15], provide guidelines for both consensual and contractual use of data, as well as user rights and protection.

When it comes to the externally-licensed content of library collections, the context of user data and privacy is more complex. A foundational issue is that content is licensed by the library for a negotiated set of users of that library, thus these users are not themselves customers of the content providers. Furthermore, based on library privacy laws and policies, such as the American Library Association (ALA) Library Bill of Rights [16], library users have a right to privacy, physical or virtual, without being monitored, scrutinized, or impeded from open inquiry of information.

Yet how do we reconcile the growing consumer preferences of personalization, access from anywhere, and convenience in services with our fundamental commitment to user and data privacy? As discussed above, if nearly one-third of access to library services is happening outside of the provided networks of libraries and campuses, how do we deliver the same service with all of the conveniences users now expect in web-based content and applications? There are three similar applications, Kopernio (powered by Clarivate's Web of Science; see: https://kopernio.com/), Lean Library (a Sage company; see: https://www.leanlibrary.com/), and LibKey Discovery (a product of Third Iron; see: https://thirdiron.com/libkey-discovery/) that aim to provide browser-based, seamless access to library licensed content, leveraging the library proxy service when users are off-campus. Another solution in development is the RA21 project (see: https://ra21.org/), which has created a very clear user experience for users to select the library or organization they affiliate with and login with that organization's single sign on system, eliminating the need for the library proxy.

These technology solutions are indeed user-centric in their focus. But what are the risks to users? Each of these solutions has the potential for user activity to be tracked, the resources they access to be monitored, and user data to be mined for other purposes, including the potential for it to be shared or sold. As seen

in social media, user activity data is used for targeted advertising, ultimately generating revenue thanks to the user. As these services are gaining traction, library administrators, information security professionals, and general counsel officers are examining the risks and liabilities.

Within the past few months, some content providers have approached libraries attempting to require user attributes be shared in order to maintain access to the resource as part of a transition away from proxy-based authentication to a single-sign-on. This attempt to request or require is counter to the trusted relationship between the user and their library, as the home organization is beyond the negotiated licensing relationship between the library and the content provider, and in many states is a violation of state law. North Carolina General Statute 125-19 [17], for example, states "A library shall not disclose any library record that identifies a person as having requested or obtained specific materials, information, or services, or as otherwise having used the library" except when necessary for the reasonable operation of the library, upon written consent of the user, or pursuant to subpoena, court order, or otherwise required by law". The fact that users of the library do not have the ability to give informed or expressed consent is critical, and information security and general counsel offices have routinely rejected click-through workflows as agreeable or acceptable.

An alternative would be for content providers to request users setup profiles outside of the library access. This would certainly be consistent with other commercial consumer web or networked services, but it runs the risk of jeopardizing the user experience unless the personal user profile can be connected with the organizational user profile. How would the content provider determine if the user remains an authorized user of the organization? Are there limitations to the user access if the user logs in with their personal profile rather than through the library access workflow? Could they maintain their personal profile benefits when logging in through their organizational access workflow? These questions, and many more, are technical challenges to providing the user-centric discovery and access benefits that many users crave and providers often want to provide.

6. User benefit versus user profiling

It is critical, however, that users have the opportunity to fully understand the benefits that they are gaining as a result of providing their personal and activity data versus the risks to which they are being exposed. It seems unlikely that data security would not be a major topic of consumer product launches after so many incidents have been revealed, such as user data breaches within financial institutions, undisclosed user data sharing incidents by Facebook to Cambridge Analytica, and the news that Ring, an Amazon company, has been using human contractors rather an Artificial Intelligence to review and label video footage of its security door-bell systems [18]. But in fact, user data security was barely mentioned at the most recent Consumer Electronics Show in Las Vegas. For example, LG released a new platform called ThinQ (see: https://www.lg.com/us/lg-thinq) that connects its products and appliances to Artificial Intelligence (AI) to learn your habits and adapt its products to your lifestyle. You can imagine the user-centric benefits of your kitchen appliances working based on your scheduled routine it has learned through your activity. But that can also be called user profiling, and that data, if compromised, could be maliciously used [19].

There are similar hypothetical library examples of user profiling. While libraries could learn the habits of its students, such as when they enter the library, how long they stay, where they study, with the objective of ultimately providing recommender services, what if librarians actually inserted themselves instantly when a user initiates a search on the library website? What if a librarian could see the active searches and

inform users of alternate search terms? What if a librarian saw a user accessing a controversial journal from a known predatory publisher and steered them to other journals having better reputations or a known commitment to open access? While privacy seems to be deteriorating in our online activity, privacy still remains a library core value and we need to find ways to advocate for responsible public and commercial policy (see [11]).

In this drive to develop technology for user-centric discovery, it is critical that we do so without treating user data as a commodity. User data should not be traded, sold, or shared, nor should it be used as a bargaining chip or negotiation tactic. It is indeed powerful information, and while holding it, we must accept the responsibility to protect it. At the core, user-centric discovery should begin with the expectation that the absolute minimum user data will be available. In the context of RA21 for licensed content access, the absolute minimum data would be a binary data point indicating if the user successfully or unsuccessfully authenticated as an authorized user. In user-centric resource sharing, it would be the absolute minimum data required for the transaction to be completed successfully from discovery to delivery to return.

Beyond the starting point, we must be willing to adapt to the user expectations of personalization, options, and convenience. We should explore what additional benefits we can provide and measure the value of these benefits against costs, liabilities, and risks. We should aim to assess our services with the data that we have at our fingertips. We should aim to serve our users better with the responsible use of data. We should be willing to go further than we have dared before in collaborating and obtaining informed consent from our users.

7. Conclusion

Thus, it is more critical than ever that technology innovations include enforceable policies to protect both users and technology providers. The core of the entire library community should be based on trust, from users to libraries to commercial partners, thus we should aim to develop policies that enhance the technology that we create and serve the best interest of all parties. The innovations, the services, the policies, and the enforcement should all be communicated in plain language, and it should be a priority to update and actively maintain this communication as we develop new solutions. After all, is anything more user centric than open communication?

What is absolutely clear is that the technology impact to user-centric discovery has created a new paradigm for libraries. Gone are the days when everything discoverable within a library is literally and physically within the libraries; we must depend on partnerships with other libraries and content providers to completely serve our users. Gone are the days when users had to physically come into the library for access; we must enable, authorize, and collaborate on information and content delivery from any location, from any device, and without any assistance. Yet what remains steadfast is our commitment to our users: to give them access to as much information that they seek; to make their experience simple, satisfying, and complete through continuous improvement; and to protect their intellectual freedom and privacy as our most foundational value, and the core to user-centric discovery.

About the Author

Timothy M. McGeary provides leadership for digital strategies programs and information technology services and operations within the Duke University Libraries, including management of Data and Visualization Services, Digital Collections and Curation Services, Assessment and User Experience, and three departments within Information Technology Services. On behalf of the Duke University Libraries, Tim is the primary liaison to the Vice Provost of Research and the Office of Information Technology (OIT), and also participates on a bi-monthly committee for Corporate Relations. Tim has presented nationally and regionally on technology development, multi-institutional collaboration, managing electronic resources, open source solutions, and the Open Library Environment (OLE). He is also published on topics of library technology and managing electronic resources as an investment portfolio.

Tim came to Duke University from the University of North Carolina at Chapel Hill, where he served as the Director of Library & Information Technology. Prior to that, he served as the Team Leader of Library Technology for Lehigh University. He received a B.A. in Music and an M.S. in Information and Systems Engineering from Lehigh University.

References

[1] K. Antelman, E. Lynema and A.K. Pace, Toward a twenty-first century catalog, *Information Technology and Libraries* **25**(3) (2006), 128. doi:10.6017/ital.v25i3.3342.

[2] J. Houser, VuFind implementation at Villanova University, *Library Hi Tech* **27**(1) (2009), 93–105. doi:10.1108/07378830910942955.

[3] E. Sadler, Project Blacklight: A next generation library catalog at a first generation university, *Library Hi Tech* **27**(1) (2009), 57–67. doi:10.1108/07378830910942919.

[4] M. Breeding, *Index-Based Discovery Services: Current Market Positions and Trends*, 8th ed. Vol. 54, ALA TechSource, Chicago, IL, 2018.

[5] R.C. Schonfeld, Does Discovery Still Happen in the Library?: Roles and Strategies for a Shifting Reality, September 24, 2014, https://doi.org/10.18665/sr.24914, Accessed June 16, 2019.

[6] K. Ciccone and J. Vickery, Summon, EBSCO discovery service, and google scholar: A comparison of search performance using user queries, *Evidence Based Library & Information Practice* **10**(1) (2015), 34–49. http://dx.doi.org/10.18438/B86G6Q, Accessed June 16, 2019.

[7] D. Allison and M. Mering, *"Use of Discovery Tools in ARL Libraries" Faculty Publications*, Vol. 366, University of Nebraska Libraries, 2018. https://digitalcommons.unl.edu/libraryscience/366, Accessed June 16, 2019.

[8] L. Dempsey, Reconfiguring the library systems environment, *Portal: Libraries and the Academy* **8**(2) (2008), 111–120. doi:10.1353/pla.2008.0016.

[9] C. Lown, T. Sierra and J. Boyer, How users search the library from a single search box, *College & Research Libraries* **74**(3) (2013), 227–241. doi:10.5860/crl-321.

[10] X. Niu, B.M. Hemminger, C. Lown, S. Adams, C. Brown, A. Level, T. Cataldo et al., National study of information seeking behavior of academic researchers in the United States, *Journal of the American Society for Information Science and Technology* **61**(5) (2010), 869–890. doi:10.1002/asi.21307.

[11] H. Harrop, D. Kay, O. Stephens, S. Schmoller and J. Kay, We love the library, but we live on the web: Findings around how academic library users view online resources and services, 2015, https://www.coursehero.com/file/15960897/UK-Survey-Report-July-2015/. Accessed June 16, 2019.

[12] D. Wells, Library discovery systems and their users: A case study from curtin university library, *Australian Academic & Research Libraries* **47**(2) (2016), 92–105. doi:10.1080/00048623.2016.1187249.

[13] B. Fister, What's so sacred about privacy? Peer to peer review, *Library Journal* (2014), Retrieved January 03, 2019, from http://lj.libraryjournal.com/2014/05/opinion/peer-to-peer-review/whats-so-sacred-about-privacy-peer-to-peer-review/. Accessed June 16, 2019.

[14] Olmsted vs. United States, https://en.wikipedia.org/wiki/Olmstead_v._United_States, Access June 16, 2019.

[15] See https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en, Accessed June 16, 2019.

[16] *Privacy* American Library Association, 2006. http://www.ala.org/advocacy/intfreedom/librarybill/interpretations/privacy (Accessed January 24, 2019) Document ID: 5c653c23-920b-b254-d94c-6dcf4ccd86c6. Accessed June 16, 2019.

[17] See: https://www.ncleg.net/EnactedLegislation/Statutes/HTML/ByChapter/Chapter_125.html, Accessed June 16, 2019.

[18] J. Morse, Ring's 'AI' reportedly involved people in Ukraine watching customer videos, Mashable, January 10, 2019, see: https://mashable.com/article/amazon-ring-ai-privacy-videos/, Accessed June 16, 2019.

[19] P. Pachal, CES 2019 had nothing to say about the Biggest Conversation in Tech, Mashable, January 11, 2019, Retrieved January 15, 2019, from https://mashable.com/article/ces-2019-consumer-data-privacy/ and Pachal, P., Privacy Concerns Largely Ignored at Annual Consumer Electronics Show, Interview by S. Simon, January 12, 2019, Transcript. Retrieved January 15, 2019, from https://www.npr.org/2019/01/12/684748454/privacy-concerns-largely-ignored-at-annual-consumer-electronics-show, Accessed June 16, 2019.