

Discovery versus Disintermediation: the new reality driven by today's end-user

Jane Burke
Vice-President, ProQuest, LLC
jane.burke@proquest.com

Abstract

Increasingly, libraries are viewed as irrelevant to the research process, leaving them vulnerable to being cut, both financially and from the mind of the end user. However, new ways of discovering content in library collections holds the promise of returning the researcher to the library. The author explores the impact of search technology including Discovery Layers (NGCs), Federated Search, and the emerging trend toward Web-Scale Discovery.

Discovery versus Disintermediation: the new reality driven by today's end-user

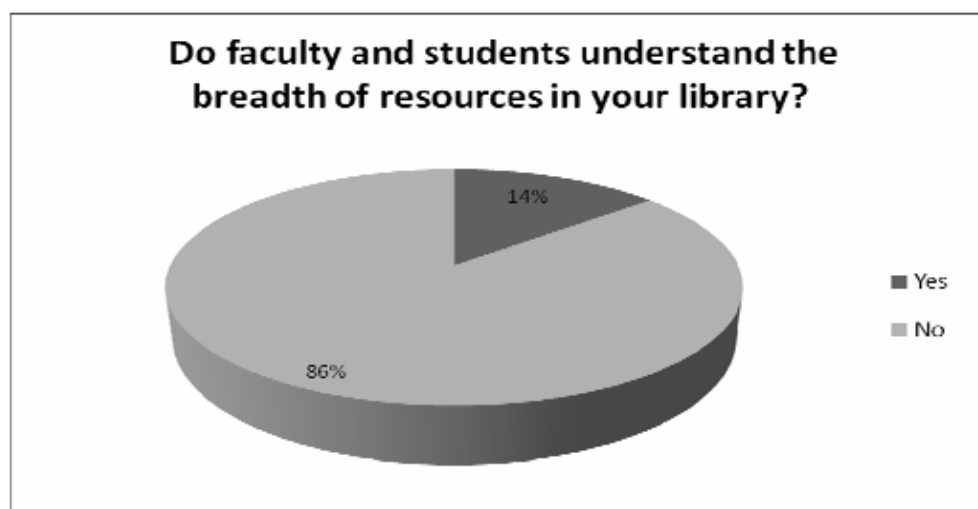
“Disintermediation” is the removal of intermediaries in a supply chain – cutting out the middleman. Increasingly, libraries are viewed as irrelevant to the research process, leaving them vulnerable to being cut, both financially and from the mind of the end user.

In 2001, the Pew Internet and American Life Project announced the results of a study that showed a growing majority of students preferred the Internet to the library for academic research. For those of us who work in and with libraries, it was hard evidence of a new reality: end-users were finding the allure of the Internet too powerful to resist. Said one student respondent, “Without the Internet you need to go to the library and walk around looking for books. In today's world you can just go home and get into the Internet and type in your search term. (Lenhart, 2001, p. 1)”

Modern Barriers to Library Use

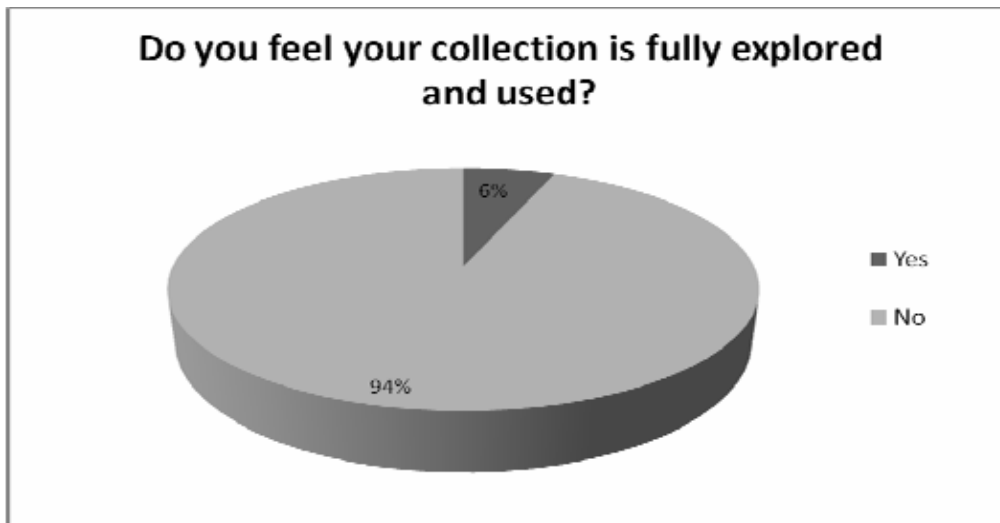
In the eight years since the release of that study, libraries have changed. Our collections are primarily digital now and accessible online, erasing the barrier of having to physically visit the library. However, as we have invested in expanded electronic collections – which now include hundreds of databases and thousands of ebooks – we have created a new barrier to use. Navigation is complicated, especially when compared to the simplicity of open Web searching. In a 2009 survey of 66 academic libraries, ProQuest found that 86 percent (see figure 1) of libraries feel that faculty and students do not understand the breadth of their collections, and 94 percent (see figure 2) think the collections are not explored to their fullest (ProQuest, 2009).

Figure 1



Source: ProQuest survey of academic libraries, August 2009

Figure 2

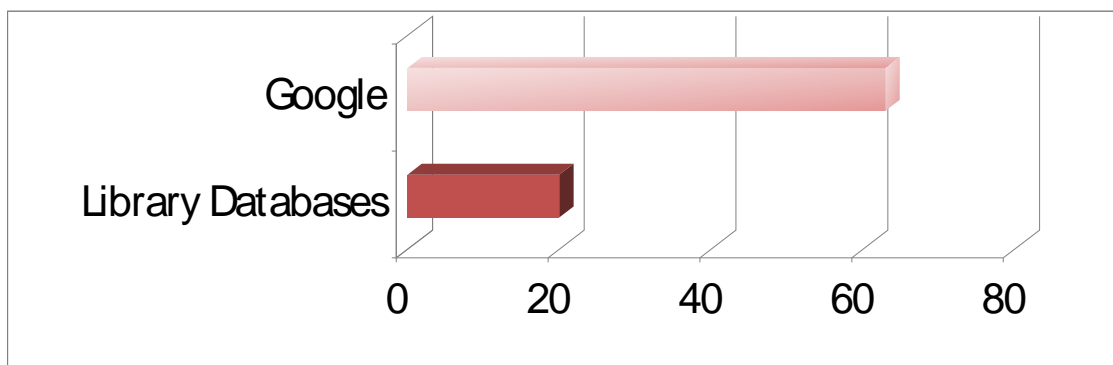


Source: ProQuest survey of academic libraries, August 2009

And we know what is getting in the way: Open-Web, advertising-driven search tools, such as Google, Yahoo and others, are easily accessed, simple and fast. A 2007 study of student research behaviour conducted by ProQuest and John Law (ProQuest, 2007), Serials Solutions' Vice-President of Discovery Services, found that more than 60 percent of students consider Google to be the easiest place to start research, compared to less than 20 percent who find library databases the easiest starting point (see figure 3). Why?

Figure 3

What is the easiest place to start research according to students?



Source: ProQuest survey of student research habits, 2007

Law's research found three primary barriers to full use of the library:

1. There is no clear and compelling starting place for research. Consider the main database page at most academic libraries: there are literally hundreds of potential starting places, confusing users: where do we start for *this* topic, they wonder.

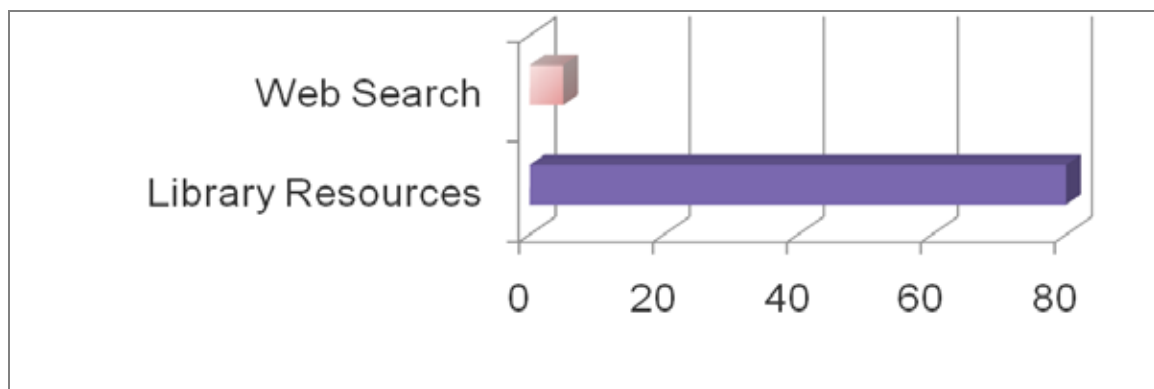
2. Difficulty identifying appropriate resources – Database names are often poor clues to their contents.
3. General lack of awareness of resources – Users simply are not aware of the range of resources available to them through the library gateway.

Libraries: the preferred content source

Despite these barriers, the library’s reputation as the source of the best, most credible content for classroom assignments remains intact. The current generation of undergraduate students has a full appreciation of the limitations of open Web data. The average undergraduate was in elementary school when the Worldwide Web became a standard for quick answers, and had plenty of exposure to media specialists who warned them of its questionable accuracy. Law’s research found that less than 10 percent of undergraduates found the Web to be a superior source of quality, credible content, while 80 percent define the library this way (see figure 4). Further, nearly 100 percent of students prefer library resources for classroom assignments (see figure 5).

Figure 4

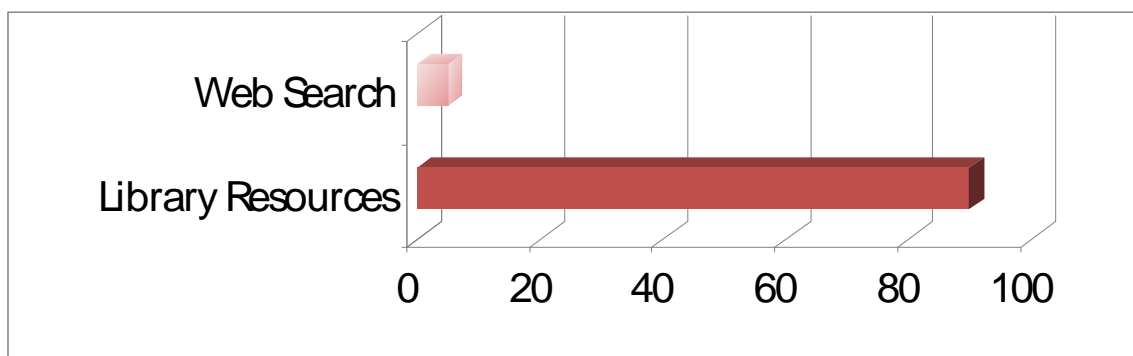
What is considered by students as the superior source for quality, credible content?



Source: ProQuest survey of student research habits, 2007

Figure 5

What is preferred for academic research and course assignments?



Source: ProQuest survey of student research habits, 2007

The Busy End-User

They *want* to use the library, but they seek simplicity. Project Information Literacy (PIL), a research project led by Alison Head and Michael Eisenberg at University of Washington's Information School, finds that students in the digital age are overwhelmed (Head and Eisenberg, 2008, p.2), . Head and Eisenberg note: "In general, students reported being challenged, confused, and frustrated by the research process, despite the convenience, relative ease, or ubiquity of the Internet. Frustrations included the effects of information overload and being inundated with resources. Participants also reported having particular difficulty traversing a vast and ever-changing information landscape. Specifically, participants' greatest challenges were related to finding the materials they desired, knew existed, and needed on a 'just in time' basis."

"Just in time" is the key phrase. According to PIL, students are pressed for time and seek research strategies that deliver quick results. In fact, 80 percent of these students define themselves as procrastinators, starting their research just days before their assignment is due and leaving little time for exploration. Their research showed that this "just in time" research strategy often employs short-cuts such as Wikipedia, which students use as a "survey" tool. They understand the risk of Wikipedia content, but value its ability to provide a topic overview that can jump-start their work on an assignment.

Conversely, the library presents the credible content that students want, but complicated navigation. Which resource to consult? Which pathway leads to the right research? A study by McClure and Clink explored the sources and navigation students used to locate materials – particularly online sources (McClure, 2009, pp. 115 -32). Among their findings is that libraries are perceived by students as both inconvenient and intimidating. The authors note that this is "especially and interestingly in its primary purpose – supporting student research and often assisting students in the identification, locations, and evaluation of sources. (pp. 121.)"

Indeed, Anne Prestamo, Associate Dean for Collection and Technology Services at Oklahoma State University, has documented the impact of the complexity that faces students. She spent a year analysing searches on her library's traditional OPAC. After backing out staff searches, she found that more than one third of searches by lay-users yielded zero or very few results (Prestamo, 2009, pp.12). With limited success, these harried, hurried students are turning to Google as their first stop for research. As McClure and Clink note, "These engines are easy to use, available to anyone with an Internet connections and quick and bountiful in their returns. (pp. 123)."

The growing disintermediation of the library in the research process used by students is compounded when we look at faculty use. Ithaka's 2006 "Studies of Key Stakeholders in the Digital Transformation in Higher Education" explored faculty perceptions of three roles of the library – purchaser, archive and gateway. While purchaser and archive remain solid and stable, the role of the library as a gateway to research is declining. "Faculty increasingly value electronic resources," note study authors Ross Housewright and Roger Schonfeld. "At the same time, while they value the library, they perceive themselves to be decreasingly dependent on the

library for their research and teaching and they anticipate that dependence to continue to decline in the future. There appears to be growing ambivalence about the campus library.”

The Value Gap

Carol Tenopir, University of Tennessee at Knoxville, has identified a growing problem that she has coined as “the value gap.” Library collections are both the biggest investment and greatest competitive advantage held by libraries. Yet, complexity and navigational barriers prevent this asset from being fully leveraged. As we continue to invest in collections that are ever more relevant to the campus community, those collections become ever more complex to navigate. As we add to complexity, we reduce the ability of our key stakeholders to use the library’s resources. In short, each successive investment has less value to the library than the one that preceded it. However, if we stop investing, we stop enhancing our greatest asset.

What libraries often fail to realise is that end users see most information objects as equal. A relevant article on a topic is as good as, if not better than, an entire book. While libraries count databases, those databases represent millions of objects to which end-users want, and need, simple access.

Dartmouth reports that in examining undergraduate use of its new discovery service, while the online catalogue has 2.2 million bibliographic records, the count of individual articles and other objects in the discovery service exceeded 300 million.

End users do not differentiate information by type or format. Yet libraries fail to realise this, and they continue to organise access to information in old ways. This is a holdover from the way in which we used to organise print collections. To be central to the research process, libraries must eliminate the containers – “break the silos” – and make all information accessible in a single search that is simple, easy and fast.

All hail “Discovery”

A number of libraries around the world are making strides in bringing the researcher – faculty and student – back to the library as a first stop. In the 2009 ProQuest survey, among the top strategies of the libraries who report their collections are being fully used and explored were investments in improved discovery.

The components of “discovery” centre on the end user:

- Compelling user interface
- Single search across multiple types of metadata
- Represent all of the library’s collections

Discovery services for libraries combine technology, content and community, making simple the complexity of information management.

Continuum of Discovery Solutions

The past three years have seen significant advancements in the technology that supports discovery and improves navigation of library collections. In the library market, “discovery services” are a continuum of solutions. This continuum is based on three categories of services, two of which are often combined:

- Discovery Layer (also known as “next generation catalogue”)
- Federated Search
- Web-scale Discovery

1. Discovery Layer

Traditional ILS vendors and Open Source collaborations are creating discovery layers (also called next-generation catalogues or “NGCs”) that enhance access to local collections. These tools include faceted searching -- clustering and cloud tools -- that allow users to iterate to ever-better search results from the library catalogue. These are natural starting points for improved discovery because we have control of the systems and we know and have control over the metadata in our local collection.

Discovery layer products are locally installed applications, following the traditional model of the ILS. Discovery layer provides functions not available in legacy OPACs, such as broad keyword searching, allowing users to search with natural language. Discovery layers also better utilise the detailed information libraries have built into the MARC records by creating facets from subfields and by utilising subject subdivisions as search terms.

Discovery layers also provide Web 2.0 features, such as tagging. Unlike the traditional OPAC, discovery layers encourage users to interact with the data – “the community effect.” By creating tags and lists, users add value to the library’s collection. Many discovery layer products also provide visualisation of results, engaging the end user.

An attractive element to libraries, discovery layer products offer a high degree of customisation. Libraries can choose from a wide variety of options to make the discovery layer represent their collection and its strong attributes. Customised facets, more obvious branding, and inclusion of local collections beyond the OPAC (such as the institutional repository) are just some of the available options. Some discovery layers offer innovative use of the library’s authority records to provide new ways of guiding users with the library taxonomies.

In the market today, there are two types of Discovery Layer solutions: those that are provided by commercial organisations and those that are open source solutions, created by libraries or groups of libraries. Commercial solutions include Encore from Innovative Interfaces, Primo from Ex Libris, Enterprise from SirsiDynix and AquaBrowser from MediaLab (a ProQuest company). These commercial solutions are offered via several different business models, including perpetual license (like the ILS) and annual subscriptions. Some are hosted, and others are installed on

servers at the institution. Maintenance and enhancements are supplied by the vendor.

Open Source discovery layer products are just that – open source applications. There are a number of these available, including VUFind developed at Villanova University in the US, Summa from the State and University Library of Denmark, and Blacklight from the University of Virginia. Another effort is XC, the eXtensible catalogue project at the University of Rochester and supported by partners and funding from the Mellon Foundation. Like all open source applications, these offerings are supported by the community of developers. There are no structured support organisations, although the community effect is strong.

Discovery layers are smart and laudable efforts, but they stop short of a complete solution. The full-text article or chapter content is truly the information object of choice for users, while NGCs point to the macro level -- databases, ebooks and journals where that content is held. With the many, many interfaces to that “commercial” content, Federated Search has gained wide use with varying levels of success as a tool to trawl the depths of the databases.

2. Federated Search

Federated search is a mature, proven technology that enables a single search across the OPAC and most library databases. It could be said that federated search, brought to the market in 1999, was the first discovery tool, and it provided for easy use across databases, particularly as the number of databases grew and grew.

Initially offered as separate technology, federated search is now provided in several models – independently as a software application (such as MetaLib from Ex Libris) and integrated into content services (such as EBSCO Integrated Search). OCLC has recently announced the inclusion of federated search into WorldCat Local. Federated search is also available as locally installed software or hosted (such as Serials Solutions 360 Search).

The basic premise of federated search is that the application launches searches against all of chosen databases, using connectors that are created for each data source. Connectors are available for many, many data sources, as fits a mature technology. Federated search is affordable, proven discovery technology. In a number of situations, federated search has been linked with the discovery layer to provide a more unified user experience. One of the first examples of this concept was the integration of AquaBrowser with 360 Search at the University of Western Sydney.

Pioneering as it was, federated search is based on a technology model that struggles to deliver “simple, easy, fast”. As collections grow, federated search must make more and more connections to offer a unified user experience, slowing the process with each connection. Many libraries have not promoted federated search because of this drawback. Libraries also express concern about how results are displayed when they are “homogenised” by the federated search engine.

3. Web Scale Discovery

Federated search's success relies on the weakest of its cascading series of search interfaces, often delivering slow results, with questionable relevance ranking. That drawback has given rise to the development of Web Scale Discovery. These efforts aim squarely at Google as the competitor and mimic that search engine's characteristics of simple, easy, fast. Like Open-Web searching, Web Scale Discovery enables single search-box access to a universe of content. In this case, it is the breadth of the library collection in all its formats – books, databases, articles... even artifacts from library museums. Unlike NGCs, Web Scale Discovery delivers results at the article level and unlike Federated Search, it requires only a single search of a pre-harvested set of metadata. Simple, easy, fast – exactly aligned with end-user expectations -- it creates a “best of both worlds” scenario: one-step searching paired with the superior content found in the library.

Longed for, Web Scale Discovery has taken time to reach the market because it requires an entirely new and labour-intensive technology model. As noted above, the vast majority of the library's collection is truly the articles, chapters and other objects within the commercial databases to which the library subscribes. Metadata must be collected from publishers to ensure accuracy, relevancy and speed, requiring agreements with thousands of publishers. The library's catalogue and other local collections must also be ingested. A large hosted platform means that the subscription content need only be indexed once (and subsequently updated) even as each institution's local collections are also being harvested.

The content must then be coordinated in a single index. Data changes every minute, so updating must occur constantly – a feat that libraries would be hard-pressed to accomplish. Incoming metadata varies in fields, quality, length and must be mapped into a rich, consistent schema for search and relevance ranking. The model is really that of Google for the library's scholarly content—one large pre-coordinated index that treats each information object equally and makes each one easily and quickly discoverable. There are no individual containers of data – all of the containers are broken open to create one index.

Getting the metadata to index requires discussions with many providers. Unlike Discovery Layer products, which re-index data that the institution holds locally, Web Scale Discovery provides its value from harvesting all of the subscription-based content plus the local collections. Publishers must provide access to their metadata for indexing into the Web Scale Discovery service. Even in the era of Google, publishers must understand the initiative and agree that their content can be included in the service. Partnerships with publishers deliver the metadata that will make their searches more robust without losing the simplicity of their own interfaces.

The actual ingestion of this volume of data, not only initially but on an ongoing basis, means that special programs are required. Just as Google indexes millions of objects every hour, a successful Web Scale Discovery service must do the same. Special programs and lots of hardware are needed.

In short, Web Scale Discovery requires a committed host with talents in both ingesting data and developing search technology, all supported by an underlying will to invest in big, expensive and uncharted architecture. Some would call Web Scale Discovery the “Googlisation” of the library’s collections – an apt description. The search model matches the search patterns our users have adopted, which is squarely where we need to be. Web Scale Discovery, although relatively new, offers the researcher a Google-like experience and retains the relevance of the library in the process. While it sounds like a pipe-dream, it is happening. Serials Solutions’ new Summon service mimics the simple searching strategy that users find on the open Web, with a single library-branded search box that instantly searches the library’s holdings – both physical and digital at the article level. The service aims to expose more of the library’s content in precisely the right context, reducing user frustration and delivering more value from the collection investment. WorldCat Local from OCLC has similar aims, with single access to the library’s collection, plus a view into the collections of other local library collections.

The trend toward Web Scale Discovery is gaining steam. At the time of this writing, both EBSCO and Ex Libris have announced intentions to develop large platforms for pre-indexed data to supplement their other offerings. EBSCO recently entered the federated search market, using federated search in conjunction with their own data on their platform and the company has similar plans for pre-indexed content. Ex Libris has announced Primo Central, a hosted repository of commercial metadata to augment its Discovery Layer, Primo. The technology will become further refined as library vendors learn from one another and leverage their own skill sets to improve discovery.

Is it soup yet?

When the first Web Scale Discovery service, Summon, was announced at the Online conference in Sydney in January, 2009, the response was big and welcoming. Good, quick understanding of the concept from libraries was evident, and among library vendors, there was immediate attention paid to how their own products could be modified to enhance and simplify discovery. Within nine months of the initial announcement, university libraries were launching single search-box sites to address the needs of their students and faculty. An interesting example is at University of South Australia, where the library is replacing Google™ Scholar with the Summon service in both the student and faculty search portals, driving library users back into the library collection rather than out to the Web. Other early-adopting libraries are combining next-gen catalogues with Web Scale Discovery to serve the multiple needs of their user groups. Since librarians see a very small proportion of their users in face-to-face settings, these intuitive user-friendly technologies level the playing field for novice searchers, ensuring they find relevant, superior content; a feat the open Web has yet to conquer on a consistent basis.

While there is compelling evidence of interest and promise, end-user surveys and studies are needed to confirm Web Scale Discovery’s ability to move the needle for libraries in winning back researchers. Coverage and currency of content will be determining factors in long-run success of Web Scale Discovery, but libraries will play an essential role, also. It will be up to the library, with support from vendors, to

market these new discovery tools, helping to shake entrenched habits of turning to the Open Web.

Will Web Scale Discovery be enough to close the value gap? Today, it is our greatest hope, addressing the most profound obstacles between the library and the end-user. Consider the experience of Grand Valley State University Libraries, the first commercial adopter of the Summon service. It has used the service to place the libraries squarely in the hero role, by teaching faculty members how to embed Summon widgets in course pages and syllabi. The strategy creates a win-win-win: *professors* benefit from reduced student complaints of “I can’t find anything on this topic;” *students* get fast access to quality information directly from their course page and the *library* directs usage back into its biggest investment, generating the kind of success metrics that make compelling budget arguments.

There’s additional reason to hope that the value gap will be closed: just as Pew Internet research revealed the movement of students away from the library in 2001, new research shows the pendulum swinging back as Generation Y rediscovers the library. Another Pew study conducted in 2007 found that, “Compared to their elders, Gen Y members were the most likely to use libraries for problem-solving information and in general patronage for any purpose. Furthermore, it is young adults who are the most likely to say they will use libraries in the future when they encounter problems: 40% of Gen Y said they would do that, compared with 20% of those above age 30 who say they would go to a library. (Rainie 2007, pp.3)”

If opportunity is indeed knocking, we must work together to ensure the door is easily identified and swings open wide enough to reveal the expanse of library riches.

References

- Berry, R. 2009. *Interview by Beth Dempsey*. Ann Arbor, MI: August 27, 2009.
- Dempsey, B. *South Australian students and faculty now using Summon™ web-scale discovery service to speed research*. Serials Solutions. Seattle, WA, 2009. Available from <http://www.serialssolutions.com/news-detail/michigan-south-australian-students-and-faculty-now-using-summon-web-scale-d/>.
- Head, A. and Eisenberg M. 2009. *Finding context: What today's college students say about conducting research in the digital age*. The Information School, University of Washington: Project Information Literacy Progress Report. Available from http://projectinfolit.org/pdfs/PIL_ProgressReport_2_2009.pdf.
- Housewright, R. and Schonfeld, R. 2008. *Ithaka's 2006 Studies of Key Stakeholders in the Digital Transformation in Higher Education*. Available from <http://www.ithaka.org/ithaka-s-r/research/Ithakas-2006-Studies-of-Key-Stakeholders-in-the-Digital-Transformation-in-Higher-Education.pdf>
- Law, J. 2007. *Observing student researchers in their native habitat*. Presentation at American Libraries Association Midwinter Conference. Seattle, WA.
- Lenhart, A, Simon M., and Graziano M. 2001. *The internet & education: findings of the Pew Internet & American Life Project*. The Pew Internet & American Life Project. Available from <http://www.pewinternet.org/Reports/2001/The-Internet-and-Education.aspx>
- McClure, R. and Clink, K. McClure portal: Libraries and the Academy - Volume 9, Number 1, January 2009, How Do You Know That?: An Investigation of Student Research Practices in the Digital Age. pp. 115-132.
- Prestamo, Anne. 2009. *Is Google winning? or how I learned to stop worrying and love the big gorilla*. Available from http://conference.amigos.org/sites/default/files/prestamo_presentation.pdf
- ProQuest, LLC. August 2009. Opinion study of 63 North American academic librarians.
- Rainie, Lee. 2007. *Information Searches that Solve Problems*. Pew Internet and American Life Project. Available at <http://www.pewinternet.org/Reports/2007/Information-Searches-That-Solve-Problems.aspx>
- Tenopir, Carol. The Value Gap, *Library Journal*. July 17, 2009. Available at <http://www.libraryjournal.com/article/CA6668190.html>