

Matt Goldner, Product & Technology Advocate

OCLC, Inc.

Paper presented at Stellenbosch Symposium / IFLA Presidential Meeting 2010

Knowledge creation, dissemination and implementation: The librarian's role in today's knowledge economy

Reading the history of the library at Stellenbosch University it was no surprise that the library's central importance increased in the early decades of the University. This is a story repeated over the last centuries at universities and colleges around the world. Many universities and colleges placed the physical library building as a central anchor of the campus. It was placed there to emphasize its position at the core of education and scholarly activity carried on by the institution. Of note is the explanation of the history of the library at Stellenbosch University.

“This building occupies the unique position of being underneath the centrally situated Jan Marais Square. The reason for this unique position is that in planning a new library it was found that, apart from the Jan Marais Square, no centrally situated building sites were available on campus. ... It was therefore decided to build underground.”ⁱ

As we know this core position of the library is in a state of great flux and challenge.

A changing information landscape

To understand what the future role of the university library is in today's knowledge economy we need to understand how knowledge is being created and communicated in today's world. There has been a great deal of research on scholarly information, scholarly communication, scholarly publishing and research behaviors. Analysis of this wealth of research has yielded some high-level statements on actions libraries should be considering. I intend to draw on key points from this research to find the strengths and gaps in current library practice and hopefully challenge us to consider new forms of librarianship and urge us toward new ways to participate in knowledge creation.

Our starting point must be the information landscape we find ourselves in. Specifically, I refer to the Web and the Internet, which have dramatically changed how information is created and disseminated. One of the most thought-provoking works I've read concerning the impact of the Web on knowledge is David Weinberger's *Everything is miscellaneous*.ⁱⁱ One of Weinberger's

key premises is that with the Web information is now all piled in a heap, and the way we order information to achieve knowledge is radically changing.

As libraries evolved, we worked hard to classify and categorize information. We earned the trust of our university colleagues as the preferred organizers of the world's knowledge. We started with books arranged in some manner on shelves, which Weinberger calls the first order of information, ordering the objects themselves. Then metadata about the books was used to organize the collection in new ways and we had card catalogues, or Weinberger's second order of information. While we see strength in this standard metadata creation and enhancement, he sees the weakness of this second order in the fact that it must have a predetermined organization, such as an alphabetical or classification scheme. If we tried to emulate his third order of information, which sees all the information simply put in a heap, a rather useless pile of library cards would result. This is where the Web comes into the picture. Instead of cards in a heap we have our data in the binary heap of the Internet. While we librarians have worked hard to classify and categorize our collections, the Web allows anyone to sort information in any form they find personally useful.

This is possible because users find and interact with information on the Web. They link to items, they tag items, they write reviews, they create lists and they share items with others directly and on social and collaborative sites. This activity enables machines to act on what users have done to create new relationships not based on any predetermined order.

This new information environment allows the scholarly world to employ new methods of research and collaboration that exclude many traditional library services, such as collection development, cataloging and archiving. Researchers and scientists can find, gather and organize information in ways that are useful to them and share directly with their community of peers. One of the earliest examples of direct sharing was arXiv.org started in 1991 as a repository for preprints in physics. It has grown to include preprints in mathematics, nonlinear sciences, computer science, quantitative biology, quantitative finance and statistics, and as of February 5 had over 585,000 preprints. This one example of shifts in scholarly communication behavior is being replicated in other fields of research around the world.

What does this mean for us as information professionals? As we move from just-in-case collection building and management to just-in-time delivery of information is that the end of our role? Are we here simply to ensure access to static information? Or do we need to be pursuing a more active role in the creation of new knowledge and not just access to that knowledge.

How might libraries respond?

I'd like to draw on a framework for discussion from the paper commissioned by OCLC Research and published in January 2009, *Scholarly Information Practices in the Online Environment* by Palmer, Tefreau and Pirmann.ⁱⁱⁱ Their research was organized on five core scholarly activities: searching, collecting, reading, writing and collaborating.

Over the last 20 years, our community has focused a great deal of attention to improve searching of our collections. Though needed search enhancement resulted and progress has

been made, I would ask if we have actively engaged the community we are serving frequently enough to fully understand their behaviors and expectations. Equally important have we stayed too insular in our approach to creating new search services and failed to look beyond our own capacity and framework to see how we integrate to the large information landscape that scholars use in their research. One simple example is how well we connect our collections into services like Google Scholar and Google Books. Drawing from my own direct experience, while Worldcat.org is an impressive catalog of the world's library holdings, its greatest success is not driven by researchers starting a search in WorldCat. Instead about 40 percent of the monthly traffic, or around 8 million referrals, actually come from the various searches on Google sites, the bulk of that from Google Books. Another 15 percent is driven by links on other Web sites, such as book review sites, blogs, etc. That traffic combined with all other sources then drives over 1 million clicks to a library service every month. In this case success is driven by matching user behavior in searching and inserting ourselves into their workflow.

This however begs the question of our role once more. Are we spending too much time and money on just improving our ability to serve up information? In the report commissioned by the Research Support Library Group *Researcher's Use of Libraries and Other Information Sources*^{iv} it was found that across all disciplines "finding information electronically was deemed easiest to do; accessing the information was more difficult and using it more difficult still." Here I would posit *use* is the crux of where we should be looking to define our role as information professionals moving forward. We should be providing tools not just for the discovery of information but which enable its use and reuse leading to the creation of new knowledge.

How do we do this?

Weinberger uses a wonderful analogy of the heap of information on the Web. He likens user behavior to that of sweeping together piles of leaves from the heap. Each user will end with a different pile of leaves though they started their collecting from the same heap. Going back to the example of WorldCat on Google, we need to start by getting our leaves into the heap. But if they are there then we have to maintain persistent access to the information. That is not so difficult when they are kept in our local framework. But as researchers gather and collect information from various sources into their own personal pile we must consider how we work with them to enable gathering, maintain access, and be participants in collaboration and support dissemination.

At this point there are many topics that could be set forth for consideration. We could look at the success of embedded librarians in the classroom or the changing role of bibliographers to work in closer collaboration with faculty and research projects. The value of widgets, which allow access to information delivered from libraries and the reuse of that information in other work areas, can be considered. We could explore the importance of tools that can intuit citations in resources and build links to the resources automatically for scholars to browse through an area of research. The role that linked data with persistent URIs, such as the prototypes done with the Dewey classification and the Virtual International Authority File (VIAF), could be studied to find if it brings value to the scholar. All of these ideas, and many not mentioned, can impact both the future of librarianship and the role we take in the creation of knowledge.

Instead of looking at these types of specifics I would like to turn our thoughts to a broader context, which I believe can enable us to pursue the ideas set out above. At a conference last August I heard a speaker make the statement “community trumps technology.” She was referring to a grassroots community that started small in New York and has grown to include the major research libraries throughout that state. Its sole original purpose was to expedite the delivery of materials, either physically or electronically, between college and university campuses. Though technology played a major role in accomplishing this goal it was actually two other factors that brought success. First was the collaborative agreement among the colleges and universities to create this service. Second was the service had access to aggregated data and users. To state it another way, technology enabled the parties to collaborate in a new way, but it was their innovative use of technology that brought success, not the technology itself.

Value of Web-scale solutions

As I previously stated we must consider the information landscape we find ourselves in, this Web of information. In his work, *The Long Tail*, Chris Anderson^v states that the Web is all about scale, finding ways to attract the most users for centralized resources. Tim O’Reilly, in speaking about Web 2.0, states that “[There is] a major theme of Web 2.0 that people haven’t yet tweaked to. It’s really about data ... [and who] gives the best access to a class of data.”^{vi}

Outside our library space, businesses have discovered the value of Web-scale solutions. Solutions that build services on massively aggregated data that attract massively aggregated users, with every user adding more value to the service for the next user. The point I wish to make is libraries need to build on the value of Web-scale solutions if they want to participate with the scholarly community that lives and works on the Web.

A good example of major research libraries joining forces to do this is the Hathi Trust. It has already grown to include:

California Digital Library	University of California, Riverside
Columbia University	University of California, San Diego
Indiana University	University of California, San Francisco
Michigan State University	University of California, Santa Barbara
Northwestern University	University of California, Santa Cruz
The Ohio State University	The University of Chicago
Penn State University	University of Illinois
Purdue University	University of Illinois at Chicago
University of California, Berkeley	The University of Iowa
University of California, Davis	University of Michigan
University of California, Irvine	University of Minnesota
University of California, Los Angeles	University of Wisconsin–Madison
University of California, Merced	University of Virginia

From their Web site we read, “HathiTrust was conceived as a collaboration ... to establish a repository for these universities to archive and share their digitized collections. Partnership is

open to institutions worldwide who shared this grand vision. ... You've heard of other digital libraries. This one is different in concept and scale. Its greatest promise—and challenge—rests in defining how to serve researchers in the digital age. Together we will develop and refine the services needed to search and use such a large digital collection, and realize collectively our greatest potential as a library community.”^{vii}

Several factors make this an important move by these academic libraries. First is the very fact they are building a Web-scale service for researchers around the world instead of each trying to replicate the same effort locally. Working together, each library is focusing on achieving the widest possible distribution of their locally created content. Second they extend beyond even their own capacity to create a valuable resource for scholars. Where copyright allows the Hathi Trust system delivers the full digital content instantly. If the item is in copyright then they have extended their service to link to libraries worldwide through WorldCat. A single click can locate libraries closest to the researchers that hold the copyrighted item. Finally the commitment made to develop and *refine* services needed to search and *use* such a large digital collection. This is both a recognition that any service today must evolve and that its purpose is to support use of materials, not simply discovery of them.

Going back to the paper by Palmer, Teffeu and Pirmann, in their conclusion they set forth new services that libraries should look at collaborating on for researchers. Some of these are:

- Tools for browsing and exploring sets of e-texts
- Clearinghouse for tools for collection, storage and sharing resources
- Support for downloading, storing and organizing e-texts
- Tools to foster collaboration across institutions
- Tools to support document reviewing^{viii}

These recommendations get to the heart of two main issues. Researchers need simpler ways to use and reuse the information we deliver to them and the library community is more likely to succeed if we work to build tools to accomplish this through collaboration with each other and our communities.

Technology is a great thing, when it is applied through innovative ideas to solve real problems. What the Web shows us is that the larger the community participating, the greater the value of the solution. But Chris Anderson went on to say more about Web scale. He also noted that Web-scale solutions allow us to spread “... costs over larger and larger audiences as the technology gets more and more capable.” We can not only create a more valuable service by collaborating on Web-scale solutions, but we can reduce the expense to our own institutions. This means we can reallocate some of the capacity we have so we can meet the specific local needs of our scholarly community.

This could then allow very localized innovation, such as mentioned previously with embedded librarians or greater collaboration with faculty and bibliographers. It also allows for local innovation that can spread virally, such as greater integration with local scholarly repositories, tools to help scholars take digital content and reuse it, widgets that enable broader dissemination of new ideas and research across the Web, etc.

Conclusion

To conclude I would ask us to consider how the Web has changed the way we discover, use and create knowledge. It has shown us that scale matters if we are to be active participants in this new world. In the conclusion of the paper by Palmer, Teffeu and Pirmann they state:

“In determining priorities for development, there are two kinds of service contributions that seem most worthy of investment:

1. services that are most likely to actually advance the conduct of research, either by simplifying difficult tasks or by supporting new kinds of analysis with digital content; and
2. services that provide economies of scale across institutions, disciplines or genres of information.”^{ix}

I would humbly state I believe the second point must precede the first for us to succeed. We should, wherever appropriate, work together to build the necessary scale of data and users to make a service work for our scholarly communities. This will give us the platforms we need to increase our productivity, reduce costs, collaboratively build shared resources and define our role as information professionals in the future of scholarship and knowledge creation.

ⁱ http://library.sun.ac.za//eng/about/gen_History.htm (viewed 2/10/2010).

ⁱⁱ Weinberger, David, *Everything Is Miscellaneous: The Power of the New Digital Disorder*, New York: Times Books, 2007.

ⁱⁱⁱ Palmer, Carole L., Lauren C. Teffeu and Carrie M. Pirmann, 2009, *Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development*, (report commissioned by OCLC Research), published online at: <http://www.oclc.org/programs/publications/reports/2009-02.pdf>.

^{iv} Carpenter, J., M. Wallis, and N. Smith. *Discovering Research Resources: Researchers' Use of Libraries and Other Information Sources*. Cultural Trends. 43/44 (2003): 1-22.

^v Anderson, Chris, *The Long Tail*, London: Random House Business, 2006.

^{vi} O'Reilly, Tim, http://www.wired.com/techbiz/people/news/2007/04/timoreilly_0413 (viewed 2/10/2010).

^{vii} <http://www.hathitrust.org/> (viewed 2/10/2010).

^{viii} Palmer, Teffeu and Pirmann, pp. 37-41.

^{ix} Palmer, Teffeu and Pirmann, p. 42.