OLAQ

Handcrafted or Mass Produced: What Are you Willing to Pay and is it Worth it?

Gary Houk
OCLC

Alane Wilson

Handcrafted or Mass Produced:

What are You Willing to Pay and is it Worth it?

by Gary Houk

Vice President
Cataloging and Metadata Services
OCLC Online Computer Library
Center, Inc.
houkg@oclc.org

and by

Alane Wilson
alane.wilson@att.net

The Industrial Revolution gets bad press. The phrase tends to conjure up images of fields and teams of oxen giving way to factories belching smoke into the once clear rural skies. Dehumanized workers toiling at boring, repetitive tasks while they dream of the bucolic, agrarian past.

So, the theme of this issue of *OLAQ* must be considered somewhat provocative. For the artisans and craftspeople known as catalogers *Industrializing the Work Flow: New Trends in Technical Services* is likely to suggest that a new trend in technical services is an increasing dependence on bland bibliographic records, not well made, received from some cataloging factory thousands of miles away. And you can have those records in any color as long as it's black.

As a long-time staff member of what to some is a big cataloging factory, I'd like to suggest that industrializing workflows in technical services would bring many benefits, just as the first and second Industrial Revolutions did. Without the Industrial Revolutions, for example, this journal might not exist, OCLC definitely wouldn't exist, and there's a very good chance that the public library wouldn't either. Before mechanization, factories—even libraries—were cottage industries. Work processes were generally carried on by means of hand labor and simple tools.

> Without the Industrial Revolutions, for example, this journal might not exist, OCLC definitely wouldn't exist, and there's a very good chance that the public library wouldn't either.

Mechanical inventions changed how textiles were made, and how library catalog cards were produced.

In the world of libraries, the mechanization of the production of a library catalog was revolutionary, changing fundamentally the nature of the catalog. The first library catalog goes back to the Great Library of Alexandria. The Alexandrians from Callimachus onwards tried to keep track of what the Library owned by means of a subject catalog. In this they followed Aristotle's divisions of knowledge. The first recorded Librarian was Zenodotus of Ephesus, holding that post from the end of Ptolemy I's reign until 245 B.C.E. His successor Callimachus of Cyrene was perhaps Alexandria's most famous librarian, creating for the first time a subject catalog in 120,000 scrolls of the Library's holdings, called the *Pinakes* or Tables (Bevan, 1968). And for about 2,000 years, things pretty much stayed the same. A librarian would record by hand the information about items in a library's collection creating, in essence, an inventory list. The first Revolution in Cataloging allowed the hand crafted catalog to give way to the typewritten one, in 1901 when the Library of Congress began its card program thus extending, multiplying and leveraging the work of the individual cataloger, and ushering in the era of shared cataloging.

The second Revolution in Cataloging came a relatively short time later but it was another giant leap forward. With the birth of the Ohio College Library Center's online-shared cataloging service in 1971, librarianship was poised to reach another level of bibliographic efficiency.

In 1965, when Fred Kilgour proposed an online-shared cataloging system, you must appreciate the fact that computerized library systems did not exist, networked computers would not exist until 1972, and there was no agreed-upon standard for communicating bibliographic data. There were no cathode ray tube terminals with lower-case characters and there were no retrieval

systems that could retrieve single entries from an online catalog.

It is no exaggeration to say that the OCLC Online Union Catalog and Shared Cataloging System pioneered the computer revolution in libraries. It enabled libraries to rapidly and efficiently catalog books and print customized catalog cards. The database was not only an electronic card catalog; it was an electronic union catalog that provided location information for the materials listed in the catalog by participating libraries. It was a new library tool that was dynamic (Smith, 1994).

The shared cataloging system made it unnecessary for more than one library to originally catalog an item. The system made copy cataloging not only practical, but also widely available. Presently, most libraries have to do original cataloging for only about six of every 100 items they acquire. The shared cataloging system also increased productivity of catalogers. For example, Ohio University reported that the first year it used the OCLC system, it was able to increase the number of books cataloged by a third, while reducing its staff by 17 positions through attrition (Smith, 1997).

For the year ended June 30, 2002, libraries cataloged 49.4 million items on the OCLC system and added 2.7 million records to the OCLC database. Imagine the cost of originally cataloging 49.4 million items! At this writing in 2003, WorldCat contained more than 51 million records and more than 884 million location listings.

However, more than 30 years after the introduction of shared cataloging, the second Revolution in Cataloging has yet to fully impact the work of most catalogers. Cataloging is still a mostly mechanized cottage industry. "Hand crafted one at a time to last a lifetime." Although this is a phrase from a modern advertisement for handmade furniture, it could very well be the motto of many a cataloger working in these early years of the 21st century. Automation in cataloging in particular still is and has been used as a tool to get old

> Automation in cataloging in particular still is and has been used as a tool to get old tasks accomplished more productively, rather than as a tool to create more productive ways of getting things done.

tasks accomplished more productively, rather than as a tool to create more productive ways of getting things done. In other words, collectively, we've done a fine job of using machines to share, extend, leverage and multiply the work of individual catalogers. The cataloging tools and services OCLC has provided to catalogers for decades allowed WorldCat to become the huge, rich metadata repository that it is, and all librarians should take great pride in that accomplishment.

But OCLC founder Fred Kilgour asked in 1977 "are we automating nineteenth-century librarianship?" Phrased another way, have we extended, multiplied, and leveraged human mental abilities in cataloging? Probably not. Cataloging is still a labor-intensive activity, focused on the physical manifestation of a printed work. Even "copy" cataloging often is not. Local fields and data are added by local employees to address the perceived requirements of the local communities served. This is as if the employees of John Smith's Ford Dealership in Columbus, Ohio took each shipment of Explorers that arrived on their lot and added a window on the control panel that displayed Ohio State University football scores, as well as replaced the rear window with an extra long tailgate for tailgating parties. Do all Explorer buyers in Columbus, Ohio want these local features? No. Does it add to the cost of the Explorers in Columbus? Yes.

The cost of cataloging, whether it's copy cataloging or original cataloging, is not just the cost of a record from OCLC or RLG. It is also the cost of training and paying catalogers, office space, computers, networks, and materials. Studies show that the total cost of cataloging is around $30 per title and even higher for non-book, non-English titles. With the number of catalogers declining rapidly, the number of trained catalogers graduating from library schools declining, and with most libraries' budgets seeing drastic reductions, the cost of cataloging is, or will be, on the minds of library administrators. And using Z39.50 in an Internet scavenger hunt to locate "free" records will not solve the budget crisis because the record is a very small part of the total cost.

The silver lining to this cloudy situation is an environment ripe for change. Lack of expertise and lack of money will drive decision makers to seek effective ways of doing the same things (the first Industrial Revolution) but will also encourage those visionary decision makers to reshape not only the workflow but also to reshape library services to their communities (the second Industrial Revolution). And libraries' partners like OCLC must be prepared to offer services and tools that extend services to users beyond the library system.

Cataloging is a means to an end; it has evolved over time and must continue to evolve so that libraries can meet the economic and competitive challenges they now face. In the past, descriptive cataloging helped users discover authoritative knowledge resources held by a particular library, but now cataloging must also help connect them to those authoritative resources, whether held by the library or by some other provider. The purpose of the catalog is no longer just a form of inventory control (classification is a particularly sophisticated form of telling people where a particular package is located in the warehouse). Librarians have been good at training library users how to read rich, complex inventory records but

our communities of users are more discerning consumers now. They have used many inventory control systems with user interfaces designed for the user not the warehouse manager. Users can find books and clothes using simple search queries at Amazon and Lands End. They can read several pages of a book, and virtually try clothes on before buying. However, library users are, for the most part, expected to parse a sophisticated set of metadata in order to make a "purchase" decision about the invisible content.

Amazon, Lands Ends and Google are among many institutions outside of the library community that, arguably, have the same public purpose as libraries: to connect people to things they want and need. To meet these needs, they have developed non-library metadata to facilitate the discovery and fulfillment process. New metadata standards and formats for a variety of resources have been born, won't go away, and must be dealt with by the library community. The Web has connected previously disparate market places in a shared space that is much bigger than WorldCat, bigger than all the combined library catalogs, and one we all increasingly take for granted as an always-present information space. Interoperability between the library space and these other

> The Web has connected previously disparate market places in a shared space that is much bigger than WorldCat, bigger than all the combined library catalogs, and one we all increasingly take for granted as an always-present information space.

market spaces must enhance the discovery and delivery of resources. The value of library metadata will be enhanced when it is a common part of the shared, global information ocean millions of people swim in every day

For 32 years OCLC has provided tools designed to produce an electronic version of a written catalog card for inclusion in local library systems. There are now approximately 8,250 governing members of OCLC—institutions that contractually agree to contribute cataloging data to WorldCat and that continue to find shared cataloging to be cost-effective. These governing members have used OCLC cataloging tools to extraordinary effect. We think it's time, however, to embrace the Industrial Revolution and leave behind the cottage industry our member libraries and we created and have sustained. We've relied on our members to handcraft the bibliographic records contributed to WorldCat because there were no more reliable sources of metadata than the catalogers who had the physical items before them as they cataloged. We've devoted hundreds, if not thousands, of person years developing and maintaining sophisticated cataloging tools because we needed catalogers to take the raw materials, work in their own "homes" and return the finished articles. We rely on and are dependant on people in our cottage industry correcting their work, deleting their work, and returning their holdings. But our skilled workers are diminishing in number and the demands for access to content are increasing, and so we must plan for a future for WorldCat where there are fewer and fewer catalogers to contribute high quality metadata. As the responsible steward of WorldCat, concerned about the growth and quality of the database built by generations of catalogers, OCLC is investigating ways to embrace the revolution and industrialize the cataloging workflow.

We can now work directly with authors, publishers and materials vendors in order to capture metadata and work with them to create high quality cataloging records earlier in the publication cycle, thus driving down the total costs of cataloging. This enhances the patron experience with the library catalog, and it provides libraries with new opportunities to lower their cataloging costs. Libraries can automatically receive catalog records at the same time they order materials, thus speeding materials into circulation, again improving service levels to patrons. These services, however, are really about building a better mousetrap, not about rethinking the whole process of cataloging.

OCLC's Office of Research is working on a set of experimental services that would change the cataloging process by harvesting metadata automatically by pulling metadata from different repositories (library catalogs, institutional repositories, publishers, content creators). The mechanics of harvesting are becoming routine and well understood, and it is not a stretch at all to imagine for some types of resources cutting out the humans in the exporting and importing of metadata processes. OCLC software could periodically scan repositories of metadata and retrieve new and changed items. This metadata will often not have been created within a framework of consistent practice; approaches to subjects or names will be different for example, and this introduces the interesting challenge of effectively fusing and recombining metadata dynamically so that it is useful to diverse communities.

The library community and OCLC also need to figure out how to make our existing investments in structured metadata work harder by mining, developing, and exposing relationships across documents and other resources. The people using search services like library catalogs and Google are not engaged in searching, they are engaged in finding. As libraries compete with the web and with bookstores, and as libraries expand their collections to include electronic resources and digital archives that they either own or just point to, it is crucial that we collectively find

ways to drive down the costs of technical services in favor of improving the library patron experience. Cooperation will continue to be a key success factor, but the universe of cooperation must expand beyond libraries to include all of the organizations that are in the supply chain for information resources.

Libraries must define their role in this process: to focus on being the transparent middle layer, assembling content in a seamless way, invisible to the users of their services. Industrialization does not mean only mechanizing processes and producing goods and services more efficiently. It also means becoming ubiquitous, part of the infrastructure, so integral that the users of your services and goods cannot imagine life without what you produce and provide. It is true that many librarians and library staff could not imagine their work lives without WorldCat but if the WorldCat "factory" closed down next week, would anyone but the "factory workers" miss it? As large and ubiquitous as WorldCat seems to many, the records are the equivalent of fine furniture: the hands of skilled craftspeople make every one.

When we buy furniture we have choices. Our choices are dictated by taste, price and availability. We can buy mass-produced furniture, or we could buy custom-made. It's likely most of us have mass-produced furniture that we're quite satisfied with. Few of us could afford both the cost and the wait for a houseful of custom-made furniture. Before mass-production, just as now, very few people could afford custom-built furniture. Those who couldn't, owned very little furniture and what there was might be poorly made. Mass-production of furniture allows anyone, not just the wealthy, to furnish whole houses. Mass-produced furniture is good enough for most people. The library world must find a way to mass-produce "good enough" metadata that is available to millions of people. Continuing the cataloging cottage industry guarantees the market for our handcrafted, expensive products will be the information wealthy. We must learn from successful industrialization and mass-produce good quality metadata that will furnish the empty rooms of the millions of people living in houses built by Google.

The WorldCat of today is a late 20th century knowledge map. It is essentially an electronic version of the card catalog, which itself represented the apex of early 20th century knowledge mapping. Now, we are on the verge of creating a 21st century knowledge map, one that builds on technologies that were only dreamed of when Fred Kilgour hooked up the first terminal to WorldCat 32 years ago at the beginning of the second Revolution in Cataloging. It is time to build on the knowledge and skills of the past 32 years and truly industrialize the workflow.

History teaches us that the great revolutions aren't started by people who are utterly down and out, without hope and vision. They take place when people begin to live a little better—and when they see how much yet remains to be achieved.

HUBERT H. HUMPHREY
(1911 to 1978)
U.S. Democratic politician, vice president. Speech, April 2, 1966, Durham, N.C.

References

Bevan, E., 1968. *The House of Ptolemy.* Argonaut Inc., Chicago.

Smith, K., 1994. OCLC: Past and Future. In D. Williams et al (Eds.), *For the Good of the Order: Essays in Honor of Edward G. Holley*. JAI Press, Greenwich, CT.

Smith, K., 1997. OCLC Online Computer Library Center, Inc., Furthering Access to the World's Information for 30 Years. The newcomer Society of the United States, Princeton Academic Press.

17