OCLC: Collaboratively Building Webscale Services With Libraries

Jay Jordan
president and CEO, OCLC

Abstract

....

In 2000, OCLC shared with its membership a new strategy to "weave libraries into the web, and the web into libraries." The plan called for OCLC to transform WorldCat from a bibliographic database and online union catalog to a globally networked information resource, interweaving the physical and electronic collections of the world's libraries. Salient features of the strategy were to extend the cooperative, build a new technological platform, transform WorldCat, introduce new services, and, ultimately, make libraries a visible and viable presence on the Worldwide Web. Since then, OCLC has made considerable progress, culminating in the launch in 2012 of its computer cloud-based WorldShare Management Services. The concept of Webscale is discussed within the context of six trends affecting libraries and their users.

....

In 2012, OCLC began implementing Webscale services for libraries based on cloud computing technologies. This was the culmination of a technological evolution

that began in 2000 when OCLC shared with its membership a new strategy to "weave libraries into the web and the web into libraries."

The plan called for OCLC to transform WorldCat from a bibliographic database and online union catalog to a globally networked information of text, graphics sound and motion. The new WorldCat would interweave the World Wide Web with the physical and electronic collections of the world's libraries, archives and museums. It would help information professionals better manage their collections and services. More important, it would help people navigate the world's constantly expanding body of knowledge and find the information they need."

The salient features of OCLC's strategy were to: extend the cooperative; build a new technological platform; transform WorldCat from a bibliographic database and online union catalog to a globally networked information resource; introduce new services; and, ultimately, weave libraries into the Web.

Extend the cooperative

The members of OCLC have always wanted to make the cooperative as inclusive as possible, with the more libraries attached to the network generating economies of scale and benefits for both libraries and their users. In 2008, the membership adopted a new governance structure for the OCLC cooperative that was designed to extend participation in the OCLC cooperative to an increasing number of libraries and cultural heritage institutions around the world.

Under the new structure, a member of the OCLC cooperative is any library, archive or museum that contractually agrees to contribute intellectual content to the OCLC cooperative or share resources with it. A member can participate in governance of the OCLC cooperative through one of three Regional Councils: OCLC Asia Pacific; OCLC Europe, the Middle East and Africa; and OCLC the Americas. The Regional Councils send Member Delegates to the OCLC Global Council, which in turn elects six members of the 15-member OCLC Board of Trustees.

European libraries have participated in the extension of OCLC's governance. In 1998, OCLC EMEA had only a single representative on the then OCLC Users Council (Global Council). In 2001, as a result of a study of OCLC's governance, the Users Council became the Members Council, and OCLC added six international transitional delegates to seed the global effort, two of who were from the EMEA (Europe, the Middle East and Africa) region. In 2012, the 48-member Global Council had 11 delegates from the EMEA region, including Denmark, Finland, France, Germany, Netherlands, South Africa, Sweden and United Kingdom. The President of the 2011-2012 Global Council is Berndt Dugall, Direktor/Librarian, University of Frankfurt, Germany.

The 2012-2013 President of Global Council will be Chewleng Beh, Senior Director Library & Professional Services & Director of SILAS, National Library Board, Singapore.

At the 2011 EMEA Regional Council meeting in Frankfurt, Germany, March 2-3, 2011, there were participants from 89 institutions and 24 countries in Europe, the Middle East and Africa. Clearly, there is a strong and active community of librarians in the EMEA region of the OCLC global cooperative.

It is worth noting that European library directors have also been elected to the OCLC Board of Trustees, including Christine Deschamps, University of Paris V; the late Ian Mowat, Edinburgh University, U.K.; and Elisabeth Niggemann, Director of the German National Library.

Alliances

A series of alliances between OCLC and the Pica Foundation in the Netherlands has extended the OCLC cooperative to new libraries and users in Europe, the Middle East and Africa. The Pica Foundation was created in 1969 as a cooperative, not-for-profit organization by the Royal Library of the Netherlands and a number of university libraries. In 1978, the Royal Library of the Netherlands signed OCLC's first international agreement, whereby OCLC provided the Royal Library with 750,000 records from WorldCat (with holdings removed) for nine Dutch libraries. That same year, 1978, Pica's central online database was established to reduce library cataloging costs. Over the next 20 years, Pica implemented new services in local library systems, reference and end-user access. Pica also moved beyond the Netherlands, extending its services to hundreds of academic, public and other libraries throughout Europe.

In 1999, the Pica Foundation and OCLC established Pica B.V. to better serve the European library community. In 2002, Pica B.V. merged with OCLC Europe, the Middle East and Africa to form OCLC PICA, and in 2007, OCLC acquired full control of the organization, which is now called OCLC Europe, the Middle East and Africa (OCLC EMEA).

From some 400 libraries in EMEA in 1998, the OCLC cooperative today connects more than 7,000 libraries in the region with a global portfolio of services, many of which are developed and supported from OCLC's European offices.

Build a new technological platform

In 2000, all OCLC services became Internet-based. Libraries were now using Internet protocols to communicate with each other, and the OCLC cooperative became truly part of the global, digital network of the World Wide Web.

That same year, OCLC began to move from an environment in which it had built and maintained its own proprietary system for over 30 years to one of hardware and licensed software with widespread industry adoption. The new system was based on Oracle database technology with open architectural models that allowed better interoperability within OCLC services and also with external services. In building the new platform, OCLC used rapid application development techniques and continuous improvement based on user feedback. This new approach would prove invaluable as OCLC pursued its strategy to weave libraries into the Web and, ultimately, build Webscale services with libraries.

The OCLC computer network has evolved constantly since the OCLC online system began operation on August 26, 1971. The online union catalog and shared cataloging system began with

a Xerox Sigma 5 computer; a telephone network of four dedicated, synchronous lines running at 2400 bits per second, and 54 cathode-ray-tube terminals specially designed to handle the ALA character set. When OCLC moved into its present main office in Dublin, Ohio, U.S. in 1981, the computer facility comprised 44,400 square feet on three stories and housed 17 mainframe computers. At the time, the computer facility was built to reclaim heat produced by the mainframes and use it to help heat the rest of the building. By 1994, the last mainframe had been decommissioned, and OCLC had to install a gas-fired boiler to heat the building.

At this writing, OCLC operates a large server farm with 760 units and one mainframe. The equipment occupies about 1,400 square feet on one floor. The Dublin data center supports 600 terabytes of storage. In 1998, OCLC systems were handling about 3.4 million transactions per day. In 2012, OCLC handles over 23 million transactions per day with response times less than a second.

As OCLC launches new services, it must have an IT infrastructure in place so that its library services are always on and always accessible. In 2008, OCLC implemented a second data center in Westerville, Ohio, about 15 miles from the Dublin center. In 2011, OCLC implemented a data center in London, United Kingdom. In 2012 OCLC is scheduled to open data centers in Australia and Canada to support its new Webscale Management Services (see below) and identity management systems. These centers will employ state-of-the art technologies to ensure high levels of performance, reliability, scalability and cost-effectiveness. Each center will provide services 24/7 and will be backed up by uninterruptible power systems and redundant heating, ventilation and air conditioning systems. And most importantly, the new centers will enable OCLC to comply with data privacy requirements around the globe.

Besides operating data centers in support of users, OCLC also provides a global development platform that meets the needs of eight engineering centers in Australia, Germany, the Netherlands, the United Kingdom and the United States.

Transform WorldCat

WorldCat has enjoyed continued success as a bibliographic database. On August 26, 2011, the OCLC cooperative observe the 40th anniversary of the start of WorldCat as the OCLC Online Union Catalog. As WorldCat enters its fifth decade, it continues to be a rich and valuable resource for libraries and their users even as the OCLC cooperative is working to transform it by moving beyond bibliography to the information itself.

The transformation of WorldCat has entailed two goals. First, it was intended to make the database ever more useful for professional catalogers as well as reduce the costs of library technical processing. This included making WorldCat available for machine-to-machine interactions, such as look-up and linking. Second, the transformation was designed to make the database more accessible and useful for the end-users of libraries.

In engineering the transformation, OCLC developers were aware of dichotomies in perceptions of WorldCat.

Librarians use WorldCat for cataloging, resource sharing and other management activities. They expect online catalogs to help staff carry out work responsibilities; to have accurate, structured data; and to exhibit library principles of organization.

End-users (students, teachers, professors, business professionals) also use WorldCat, but they expect it to look and behave like popular

Web sites or search engines. They expect to see summaries, abstracts, tables of contents, and links to needed information.

The WorldCat that emerged in 2005 on the new technological platform contained the following information in addition to the bibliographic information of a MARC record:

- book jackets
- tables of contents
- articles
- reviews
- readers' advisories
- eBooks
- OPAC/Open URL links
- electronic holdings data
- graphics, sound, motion

The new platform supports not only MARC 21, AACR 2, but also Dublin Core and other standards such as OpenURL, LDAP, EAD, SOAP, XML, SRU, OAI-PMH and Shibboleth. The platform has reusable components that can be embedded in local library applications (see below). It can create custom views of holdings for groups of institutions. A user in WorldCat can now link to evaluative content, digital documents and objects in other knowledge repositories.

Going forward, the new technological platform provides a solid foundation for sustained growth and innovation. These new capabilities enable OCLC to be more inclusive and to support more institutions in more countries and languages around the world. As will be seen below, OCLC has introduced a series of new services and capabilities that were not possible on its proprietary, legacy systems.

The Unicode standard has enabled Worldcat to support access in a number of languages and character sets in the vernacular. WorldCat now supports 12 language scripts: Arabic, Bengali, Chinese, Devanagari, Greek, Hebrew, Japanese, Korean, Latin, Tamil and Thai. This has made it practicable for an increasing number of international organizations to merge their national union catalogs or other large files with WorldCat via automated, batch processes.. In addition, OCLC has made technological enhancements and workflow improvements to its batchloading processes to accommodate the loading of large files into WorldCat.
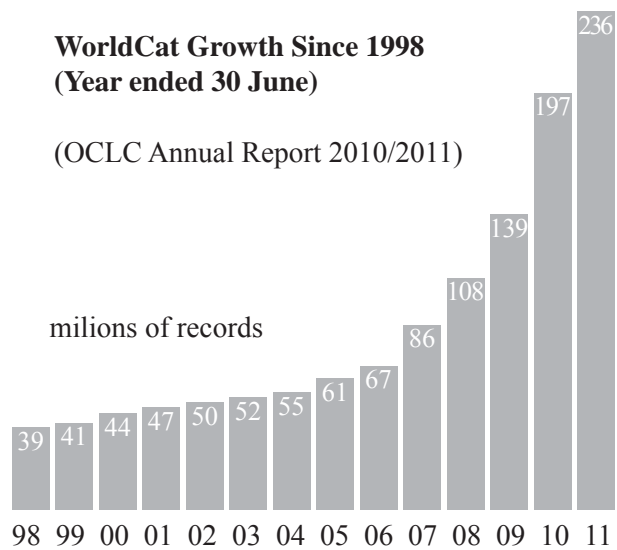
As the list below illustrates, libraries in the EMEA region have made significant contributions to WorldCat via batchloading in the past five years.

Institution	Records	Holdings	Unique Records
Dutch Union Catalogue	9.3 million	21.7 million	4.6 million
LinkUK	3.7 million	19.1 million	1.2 million
University of Göttingen	1.1 million	801,833	368,418
National Library of Education, Denmark	447,217	431,025	210,239
National Union Catalogue, Poland	528,236	514,158	429,626
Bavarian State Library, Germany	22.5 million	19.8 million	6.5 million
National Library of Sweden	2.1 million	2 million	1.9 million
HeBIS, Germany	17.1 million	32.2 million	3.7 million
German National Library	12.5 million	12 million	3.7 million
Museum of Natural History, London	440,831	226,135	103,628

Museum of Natural History, London	440,831	226,135	103,628
GBV, Germany	43.3 million	72.8 million	11.6 million
BSZ, Germany	10.2 million	25.8 million	4.3 million
UnityUK Live	33 million	233.4 million	373,332
National Library of Scotland	4.4 million	4.1 million	1 million
Cambridge University, UK	5.7 million	1.4 million	489,437
National Library of Israel	3.7 million	3.6 million	1 million
University of Sheffield, UK	1.3 million	969,601	167,596
University of London, UK	1.1 million	627,952	57,606
British Library	20.8 million	12.8 milllion	5.8 million
Danish Union Catalog	14.8 million	17.8 million	7.6 million
IDS Verbundkoordination, Germany	37.4 million	56.1 million	6.4 million
Bibliothèque nationale de France	18 million	14.8 million	10.7 million
ABES, France	19.9 million	57.1 million	5.3 million
BVB, Germany	28.8 million	39 million	6.2 million
National Library of Spain	6.2 million	5.6 million	3.2 million
COBISS.SI-IZUM	3.3 million	4.5 million	3.1 million
University of Valladolid	1.7 million	1.6 million	588,436
Edinburgh University, UK	1.3 million	1.3 million	211,975
BVB, Germany	11.7 million	26.7 million	3.1 million
RERO, Switzerland	5 million	8.4 million	3.5 million
BZB, Germany	2.6 million	3.5 million	807,233
University of Manchester	1.4 million	1.4 million	19,767

The OCLC Annual Report, 2010/2011 stated that a new record was added to WorldCat approximately every 1.2 seconds. As of June 30, 2011, there were more than 236 million bibliographic records in WorldCat and 1.74 billion holdings attached to the records. The acceleration in growth that started in 2007 is attributable to the batchloads of large files mentioned above.

**WorldCat Growth Since 1998
(Year ended 30 June)**

(OCLC Annual Report 2010/2011)

milions of records

236 | 197 | 139 | 108 | 86 | 67 | 61 | 55 | 52 | 50 | 47 | 44 | 41 | 39

98 99 00 01 02 03 04 05 06 07 08 09 10 11

Jay Jordan

WorldCa t Growth Since 1998
(Year ended 30 June)

(OCLC Annual Report 2010/2011)

As of June 30, 2011, the percentage of records in languages other than English in WorldCat was 58.5 percent. The chart below shows the distribution of records for the top 50 languages.

	Records		Records		Records
English	97,715,495	Arabic	1,086,067	Bulgarian	213,623
German	32,961,653	Finnish	715,534	Yiddish	198,179
French	21,289,166	Korean	464,931	Tamil	196,428
Spanish	9,668,452	Indonesian	446,615	Afrikaans	175,087
Japanese	7,523,190	Catalan	408,807	Urdu	174,773
Chinese	5,836,099	Norwegian	407,910	Greek (ancient)	134,025
Italian	3,908,303	Turkis	407,536	Slovak	115,389
Dutch	3,504,621	Hungarian	394,749	Bengali	110,814
Russian	3,216,507	Croatian	363,384	Malay	101,189
Latin	3,164,850	Greek (modern)	335,850	Sanskrit	81,179
Swedish	2,052,693	Thai	318,920	Lithuanian	79,895
Danish	2,026,406	Serbian	304,556	Welsh	75,361
Portugese	1,997,032	Persian	266,317	Icelandic	71,257
Slovenian	1,487,107	Hindi	235,122	Armenian	70,303
Polish	1,468,281	Ukrainian	229,344	Telugu	68,685
Hebrew	1,340,111	Vietnamese	222,575	Gujarati	62,293
Czech	1,231,903	Romanian	221,763		

It should be noted that the National Library of Serbia had agreed to add its records and holdings to WorldCat along with those of other about 136 other Serbian libraries. There are approximately 2.3 million records in the Serbian union catalog and another 1.5 million in the national bibliography.

The Montenegro Union Catalogue of 255,000 records, with 26 contributing libraries, is also being scheduled for loading into WorldCat under the aegis of IZUM, the Institute of Information Science in Slovenia.

In 2012, IZUM and OCLC were exploring the establishment of a strategic partnership to develop national library information systems in Albania, Bosnia and Herzegovina, Bulgaria, Croatia, Kosovo, Macedonia, Montenegro, Serbia and Slovenia. The following activities were under consideration:

- integrate OCLC products and services with those from IZUM to provide and effective management and information discovery environment for libraries within the COBISS.Net network.
- develop, localize and implement webscale management systems and COBISS applications.
- assist OCLC in loading records from the union catalogues of participating countries into WorldCat and from WorldCat to the catalogues of the libraries within the COBISS.Net network.
- distribute OCLC products and services in the target countries specified above.

The Montenegro Union Catalogue of 255,000 records, with 26 contributing libraries, is also being scheduled for loading into WorldCat under the aegis of IZUM, the Institute of Information Science in Slovenia.

At this writing, the CIPE consortium (Cooperazione Interuniversitaria Periodici Elettronici, Interuniversity Cooperative for Electronic Periodicals) of 11 public university libraries in Italy has agreed to load 11 million records from their collections into WorldCat.

Quality Control

Quality control of records in WorldCat is an ongoing effort that involves both the OCLC membership and OCLC itself. Libraries voluntarily invest their time and expertise in making improvements to WorldCat through the following programs. The Expert Community Program, established in 2009, involves an average of 1,000 institutions a month that. voluntarily correct or enhance records in WorldCat. The Enhance Program, established in 1983, today includes.229 institutions (359 OCLC hold-

ing symbols) in the U.S., United Kingdom and South Africa.

Removing duplicate records from World-Cat optimizes searching effectiveness and improves productivity in libraries. OCLC began running new Duplicate Detection and Resolution (DDR) software in WorldCat in February2010. DDR was developed as an extension of the WorldCat matching algorithms to identify and remove duplicate bibliographic records from WorldCat. OCLC has made significant improvements over the original DDR software previously employed. For the year ended June 30, 2011, the DDR software had removed a total of 8.8 million records.

WorldCat on the Web

On August 8, 2006, OCLC launched the WorldCat.org Web site. The site offered a search box that people can download and use to search all the records in WorldCat and identify libraries that hold an item.

This was one of the signal achievements in the history of the OCLC cooperative. It was a tribute not only to the vision of OCLC Founder Frederick G. Kilgour, but to the perseverance and hard work of catalogers and librarians who have built WorldCat record by record since 1971. The OCLC database had begun life as a cataloging and resource sharing tool, but Kilgour and the OCLC pioneers always dreamed that one day, it would be widely available to the general public.

Subsequently, OCLC has added a series of enhancements to WorldCat.org, all designed to help libraries create a compelling user environment. For example, WorldCat Identities creates a summary page for the more than 25

million personal and corporate authors mentioned in WorldCat. Each Identities page presents a summary for the individual identified, including: total works, genres, roles, classifications, a publication timeline and an audience level indicator. Additional features enable users to create lists, tag records with their own category descriptions, build bibliographies and install search plug-ins on Facebook and Firefox.

WorldCat: Representing the Collective Collection

OCLC is also extending WorldCat to represent the collective collection of the OCLC cooperative, including physical holdings such as books and journals, licensed digital content and the growing array of special collections that are being digitized.

At June 30, 2011 the collective collection of the holdings of institutions in the OCLC cooperative accessible through WorldCat.org and WorldCat Local comprised approximately 2 billion items, including:

- 236 million bibliographic records
- 1.7 billion holdings
- 417 databases represented
- 531 million article records
- 35 million institutional repository records
- 15 million archival records
- 8 million records from Google,
  Hathi Trust

OCLC continues to work with libraries, Google and the HathiTrust to derive new MARC records that represent these digital collections based on the rich collection of print records contributed to WorldCat by the OCLC membership over the last 40 years.

Introduce New Services

As OCLC pursued its overall strategy of weaving libraries into the Web, it continued to maintain and enhance existing services as well as introduce new ones. Here is a partial list of new services introduced by OCLC since 1998:
- OCLC Connexion cataloging service (2002)
- QuestionPoint virtual reference service (created with Library of Congress) (2002)
- OCLC began distributing CONTENTdm digital collection management software (2002)
- WebJunction online community for public libraries; funded by Bill and Melinda Gates Foundation (2003)
- WorldCat Collection Analysis, Terminologies service, WorldCat Registry (2006)
- WorldCat Local, which provides a single interface to a library's collection (2008)
- OCLC Developer Network (2008)
- WorldCat on mobile devices (2009)
- WorldCat Digital Collection Gateway (2009)
- OCLC acquires EZproxy authentication and access software (2008)

From this list, it is useful to elaborate on WorldCat Local and the Developer Network to demonstrate the evolution of OCLC's strategy from weaving libraries into the Web to that of building Webscale services for libraries.

WorldCat Local provides a single interface to the collections of a library. It interoperates with locally maintained services such as circulation, resource sharing and resolution to full text to create an integrated experience for library users. WorldCat Local searches the entire WorldCat. org database and presents local and group library holdings at the top of the results list, as well as ownership details for WorldCat libraries outside the local library and consortium.

With the implementation of WorldCat Local,

OCLC began providing a local connection to cloud computing. It is a service provided to the library across the Internet that eliminates costs to the library for hosting, operating and maintaining software. WorldCat Local is "activatable," which means that library staff configure the service from their library using their library's policies. This configuration is done using an online questionnaire that ideally requires the involvement of staff from both public services and technical services.

In 2008, OCLC established the WorldCat Developers Network by inviting a small group of developers from OCLC cataloging institutions in North America and Europe to use the WorldCat API (Applications Programming Interface) to build applications that would guide people from the Web to library services. These developers could then link WorldCat information to Internet applications as well as presentations, blogs, and emails.

The Developers Network sponsored events such as the WorldCat Hackathon held at the New York Public Library in 2008 and a Mashathon in Amsterdam in 2009. These events bring developers together in a creative, collaborative environment. This open-source, code-sharing infrastructure improves the value of OCLC data for all users by encouraging new Web services uses.

Web services enable applications to interconnect over the Web through machine-to-machine interfaces. They cover a wide range of activities that let people tap into the computing power on the Web. In addition to WorldCat.org, WorldCat Local and the WorldCat Identities mentioned above, OCLC has also introduced these Web services in the past five years.

The xISBN service, developed by OCLC Research, supplies International Standard Book Numbers (ISBNs) associated with an individual intellectual work, based on information in the WorldCat database. It finds all related editions of a book, including paperback, hardback, audiobook, foreign and out-of-print. Easily incorporated into library catalogs, the service is available free to OCLC cataloging members and for a fee to others.

The WorldCat Registry enables a library to manage its institutional identity more efficiently. On a secure Web platform, a library can create and maintain a single profile that includes information of use to the library's consortium members, technology vendors, e-content providers, funding agencies and other partners. This access enables the library to automate routine tasks such as activation of a new subscription service or renewal of an existing one. The Registry includes more than 120,000 institution records for OCLC and non-OCLC members.

Through another Web service, OCLC and Google exchange data to facilitate the discovery of library collections through Google search services. OCLC member libraries participating in the Google Book Search™ program, which makes the full text of more than one million books searchable, may share their WorldCat-derived MARC records with Google to better facilitate discovery of library collections through Google. Google links from Google Book Search to WorldCat.org, which drives traffic to library OPACs and other library services. Google shares data and links to digitized books with OCLC, which makes it possible for OCLC to represent the digitized collections of OCLC member libraries in WorldCat.

Integrated Library Systems

With the merger with Pica BV in 1999 and the formation of OCLC EMEA, OCLC reentered

the local systems business. Pica was a well-established local systems provider, primarily in Europe with its LBS and CBS systems. LBS is an integrated local library management system that supports acquisitions, cataloging, and circulation. The CBS system (Central Library System) provides the infrastructure for the creation and management of union catalogs and tools for regional interlibrary loan. There are CBS installations at the Royal Library of the Netherlands, the German National Library and the National Library of Australia, and at ABES (Agence bibliographique de l'enseignement supérieur) in France, to name a few.

OCLC EMEA also develops and maintains four other local library systems acquired through acquisitions. The OLIB (originally developed by Fretwell Downing Informatics) library management system is installed in about 250 organizations worldwide. The SunRise library system (originally developed by Sisis) is installed in more than 150 libraries, primarily in Germany, Switzerland and the Netherlands; the Bavarian State Library is a SunRise user. The Amlib system has some 500 installations in Australia, Africa and the United States. In 2011, OCLC acquired BOND, a provider of library management systems to German public libraries. As of June 30, 2011, OCLC supported integrated library systems at 5,035 institutions, primarily in Europe, the Middle East, Africa and Australia.

OCLC's long-term strategy for the ILS is to create an infrastructure to deliver Webscale solutions (see below) to manage library business processes. The design of the future library management environment will comprise a number of important aspects. It will be an evolutionary process, with new service implemented in a building block process that will eventually become a unified solution. The new services can also be integrated into a library's existing infrastructure and

current ILS. The notion is to preserve the functionality of the ILS, but at the same time enhance it by placing it in a network environment, and extending it to manage print, licensed and digitized material with one solution

Six trends affecting libraries and their users

With WorldCat.org and WorldCat Local, OCLC was indeed weaving libraries into the Web, and it began to make course adjustments to its strategic directions. At the same time, the pace of technological change continued unabated. Six trends that are affecting libraries and their services became apparent.

Trend 1: student-produced apps

The first trend is student-produced apps. There is a large population of students out there who are frustrated by their institutions, and they are doing something about it. They are creating their own apps and sharing them with their friends. It is interesting to note that the following applications were developed by students: Yahoo!, Pubget, Napster, EasyBib, Google Scholar Books, Facebook, reddit and RedLaser.

Students at The Ohio State University had developed apps that show the availability on campus of parking spaces and washing machines. At North Carolina State University, innovative students have created apps that identify what computer and conference rooms are available in the library. Another app has a webcam that shows students how long the line is at the coffee shop in the library.. In California, students developed an app that provided an "interlibrary loan" service among students' personal libraries. Fortunately, librarians are also developing apps for students Librarians at the University of Tennessee at Chattanooga have an app that will guide a student to

a book in the stacks. As noted earlier, the OCLC Developer Network has created 70 applications in the last two years, and these apps are getting more than 20 million calls a month.

Student-produced apps are evidence that consumers have taken control of their information-seeking. Librarians should be working with students and enabling them to help where they are the experts in what they need.

Trend 2: information habits of young people

The second trend is the information habits of young people, K-12 and college students. To stay informed about what else is going on in the library space, OCLC does a lot of market research that it shares with the members. The reports are based on surveys of the general public and library users around the world. Recent findings indicate that the library is not the information source of first resort for most people:

- 84 % use search engines to begin an information search
- 2 % begin an information search on a library Web site
- 90 % are satisfied with search engines

Clearly, these finding do not bode well for libraries if they do not change.

In the U.S., the Institute for Museum and Library Services funded projects involving researchers at OCLC, The Ohio State University and Rutgers University. The projects wanted to find out:

- How individuals find information to meet their needs
- Why information seekers do not choose to use library services first for their information needs
- How libraries can develop services and systems to meet the needs of information seekers.

Here are some of the comments they encountered from so-called "Millennials" (born 1979-1994).

- "The library is a good source if you have several months."
- "It is hard to find things in a library catalog."
- "I don't step in the library anymore… better to read a 25-page article from JSTOR than a 250-page book."

Much of the findings can be summed up in this quotation:

"The first thing I do, is, I go to Google… I don't go into the [library] system unless I have to because there's like 15 logins, you have to get into the research databases. Then it takes you out of that to [the local consortium]…"

Clearly, we need to understand the changing information habits of Millennials and the next generation of learners that we seek to serve.

Trend 3: evolution of search

The third trend is the evolution of search. It would be difficult to discuss trends in search engines without mentioning Google. In the U.S., from October 2010 through September 2011, Google had an 80.2 percent share of the search engine market. In second place was Yahoo!, with 9.5 percent. In third place was Microsoft's Bing with 8.6 percent. (Statistics are from StatCounter Global Stats—a Web market research firm). There are, however, other search engines besides the Big Three in the U.S.

Wolfram Alpha refers to itself not as a search engine, but as a "computational knowledge engine." It generates output by doing computations from its own internal knowledge base, instead of searching the web and returning links. Wolfram|Alpha introduces a fundamentally new way to get knowledge and answers—not by

searching the web, but by doing dynamic computations based on a vast collection of built-in data, algorithms, and methods. Its components include:

- linguistic analysis with new kinds of algorithms for 1000+ domains
- curated data:10+ trillion pieces of data from primary sources with continuous updating
- dynamic computation: 50,000+ types of algorithms and equations
- computed presentation: 5,000+ types of visual and tabular output

According to *Wikipedia*, users submit queries and computation requests via a text field. Wolfram|Alpha then computes answers and relevant visualizations from a knowledge base of curated, structured data. Alpha thus differs from semantic search engines, which index a large number of answers and then try to match the question to one. Wolfram Alpha supports Apple's Siri for factual question answering on the iPhone.

Yebol is another new search engine. It uses association, ranking and clustering algorithms to analyze related keywords or web pages. Yebol presents as one of its goals the creation of a unique "homepage look" for every possible search term.

Libraries are also getting involved in alternatives to search engines. Since 2008, with grant funding from the John D. and Catherine T. MacArthur Foundation, researchers and developers from OCLC, and the information schools of Syracuse University and the University of Washington have been working on what they call a "credibility engine." This innovative approach will deliver search results based on the citations and recommendations of reference librarians. The researchers have just received another round of funding--$350,000– to continue their work on the Reference Extract Project and the credibility engine.

Trend 4: increase in digital books
The fourth trend is the steady increase in digital books and the need for libraries to manage them. The terminology in this area is unclear. There is a tendency to use 'ebooks' for materials available for license from external providers, and a tendency to use 'digital books' for materials digitized from library collections. Ebooks may evoke an environment currently fragmented by provider platforms, with restrictions on use, and managed in a licensed e-resource workflow. They are for reference, information, reading. Digital books may evoke a digital library environment, an aspiration to provide higher level research services based on text mining, entity identification, and so on, and various funding and cooperative initiatives which aim to increase the corpus. This discussion of digital books involves the second category.

Providing easy access to the digital collections for researchers is a necessity. For example, HathiTrust is a very large-scale collaborative repository of digital content from research libraries including content digitized via the Google Books project and Internet Archive digitization initiatives, as well as content digitized locally by libraries. Founded in 2008, the HathiTrust partnership includes over 50 research libraries across the United States and Europe, and is based on a shared governance structure. Costs are shared by the participating libraries and library consortia. The network effects of mass digitization are sure to promote increased use of digitized books, which will place new demands on libraries and librarians. It is interesting to note that working with the Hathi Trust, OCLC introduced a WorldCat Local prototype user interface for the discovery of items through the Hathi Trust Digital Library.

At the same time that the University of Michigan is supporting the Hathi Trust, it is also engaged in the making of print books. The library

has an Espresso Book Machine that automatically prints, binds, and trims—on demand—library-quality paperback books with four-color covers indistinguishable from their factory-made versions. It will print, bind and trim a 300 page book in about 7 minutes .With so much effort going into getting the library's print books into digital form, the question might be asked, why is the library investing in a machine that gets digital books into print form? Leadership at the library states that they believe that high-quality, cost-effective print is an important part of the digital future. The Espresso Book Machine helps make it possible to connect their users with the information they need, when they need it and in the form they want it.

While academic libraries must deal with mass digitization of the collective collection, public libraries face the challenge of making their e-book collections accessible to their patrons. The public library may subscribe to ebooks from Overdrive or Ebrary or other providers. Their patrons walk in with a variety of readers, such as Amazon Kindle, Barnes and Noble Nook, or the Apple iPad. The challenge is to help patrons get the ebooks on to their ebook readers. Clearly, managing digital books is going to be a long-term trend for libraries.

Trend 5: cloud services

The fifth trend is cloud services. Since 2008, OCLC has been discussing with its membership its plans and activities to build Web-scale services with libraries in the Internet "cloud." This means Web-based applications with shared data and services. Webscale, however, is more than simply cloud computing. Web-scale also means concentrating computer resources, applications and data to deliver benefits to large numbers of users through the Web. Libraries can move their hardware and software to the cloud, where OCLC or somebody else will operate them.

It is interesting to note that OCLC Founder Frederick G. Kilgour anticipated today's cloud computing back in 1967. His original design for the OCLC online system called for six subsystems that would connect libraries to a centralized computer resource:

- online union catalog and shared cataloging
- serials control
- technical processing (acquisitions)
- interlibrary loan
- retrieval by subject
- remote catalog access and circulation control

Libraries would access these subsystems from remote workstations. The costs of the hardware and software in the cloud—mainframes owned by the OCLC cooperative—would be shared by the members based on their use of the subsystems.

Trend 6: linked open data

The sixth trend is linked open data. Open data is the idea that certain data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control. The Semantic Web is a "web of data" that enables machines to understand the semantics, or meaning, of information on the World Wide Web. It extends the network of hyperlinked human-readable web pages by inserting machine-readable metadata about pages and how they are related to each other, enabling automated agents to access the Web more intelligently and perform tasks on behalf of users.

Many librarians at major institutions have recognized that a key to the bibliographic future lies in migrating their data from library silos into an open, global pool of shared data. This will require that library metadata work well with non-library datasets and vice versa. There are a number of linked open data projects getting under way.

Jay Jordan

For example, the GeoNames geographical database covers all countries and contains over eight million place names that are available for download free of charge. The data is accessible free of charge through a number of web services and a daily database export. GeoNames is already serving up to over 20 million web service requests per day.

DBpedia.org is a community effort to extract structured information from Wikipedia and to make this information available on the Web. Dbpedia is a project aiming to extract structure content from the information created as part of the Wikipedia project. This structured information is then made available on the Web. DBpedia allows users to query relationships and properties associated with Wikipedia resources, including links to other related datasets. DBpedia enables the user to make sophisticated queries against Wikipedia and to link other datasets on the Web to Wikipedia data. The DBpedia knowledge base currently describes more than 3.4 million things in up to 92 different languages.

OCLC itself has large files of linked data: the Virtual International Authority File (VIAF), the *Dewey Decimal Classification* and FAST

In 2003, the Library of Congress, Bibliothèque nationale de France, the German National Library and OCLC began working together to combine multiple name authority files into a single name authority service, thereby lowering the cost and increasing the usefulness of library authority files. Using software developed by OCLC Research, VIAF matches and links the authority files of national libraries and groups all authority records for a given entity in to a merged "super" authority record. The long-term goal was to include authoritative names from many libraries into a global service that will be freely available via the Web to users worldwide. VIAF now in-

cludes these contributors:
Library of Congress [United States] representing NACO*
Deutsche Nationalbibliothek [Germany]
Bibliothèque nationale de France
National Library of Australia
National Library of the Czech Republic
Bibliotheca Alexandrina [Egypt]
Getty Research Institute [United States]
National Library of Israel
Istituto centrale per il catalogo unico
(Central Institute for the Single Directory of the Italian
Libraries and Bibliographic Information)
National Library of Portugal
National Library of Spain
National Library of Sweden
Swiss National Library
Vatican Library
National Union Catalog of Poland
Library and Archives Canada
National Szechenyi Library [Hungary]
RERO (Library Network of Western Switzerland)
BruNO (Brussels network of Flemish public libraries)
Systeme Universitaire de Documentation (Sudoc) [France]
*NACO (National Authority Cooperative Program) includes:
Biblioteca Nacional de México
British Library
National Agricultural Library [USA]
National Library of New Zealand (Te Puna Mātauranga o Aotearoa)
National Library of Medicine [USA]
National Library of Scotland
National Library of South Africa
National Library of Wales

In 2012, the VIAF will become an OCLC service and will also be offered as linked open data. By linking disparate names for the same

person or organization, VIAF provides a convenient means for a wider community of libraries and other agencies to repurpose bibliographic data produced by libraries serving different language communities.

It should be noted that VIAF data is helping to populate the ISNI database along with data from 15 rights management societies and trade and professional associations. ISNI stands for International Standard Name Identifier, an ISO-certified global standard. The ISNI database contains information on authors, actors, publishers, researchers and performers. OCLC is a founding member of the ISNI International Agency, a London-based not-for-profit organization that is responsible for ISNI's administration and governance. The ISNI service runs on an OCLC CBS metadata management system based in Leiden, the Netherlands.

OCLC also makes the *Dewey Decimal Classification* (DDC) available as linked open data. The DDC is presented as a small terminology service that responds to regular HTTP requests with either a machine- or a human-readable presentation of Dewey classes. The data is reusable by anyone for non-commercial purposes.

OCLC has recently made another resource available as linked open data. FAST is now available under the Open Data Commons Attribution License. FAST stands for Faceted Application of Subject Terminology. It is an enumerative, faceted subject heading schema derived from the Library of Congress Subject Headings. It was created through a multi-year collaboration of OCLC Research and the Library of Congress to make the rich LCSH vocabulary available as a post-coordinate system in a Web environment.

In summary, these six trends are relentlessly

pushing the OCLC cooperative in the direction of organizing and managing our libraries to provide information to users when and how they need it, in a form they want.

Against that backdrop, in 2009, OCLC announced its strategy to move library management services to Webscale

Weaving libraries into the Web evolves into building Webscale

By 2008, OCLC had indeed succeeded in weaving libraries in to the Web and the Web into libraries. The next step was to leverage that progress.

Accordingly, in 2009, OCLC announced that it would collaboratively build Webscale services with libraries. The goal was to use cooperation to reduce the costs of library management functions such as circulation and acquisitions while enhancing the library user's experience.

At the same time, OCLC would continue to develop and support its existing library management systems in Europe and Asia Pacific. OCLC will accelerate efforts to create robust data-exchange capabilities between OCLC library management systems and the WorldCat platform. Libraries and partners using current OCLC library management systems would be able to participate in this new development by adding Web-based services to their local solutions to extend their services for end users.

Webscale for OCLC meant more than Internet-hosted solutions of current library services. It also meant new system architectures and workflows engineered to support extremely high transaction rates and cooperation on a truly grand scale.

The concept of Webscale is strongly aligned with OCLC's historic mission of operating a

computer network and infrastructure that creates economies of scale, enabling more libraries to reduce costs and share resources. The network effects of the thousands of OCLC members working together can create a large gravitational hub for the library community and its users.

Libraries are critical contributors to the information landscape, and yet, libraries are still optimized to provide the services at the level of the institution or the communities they serve— the university, college, city, township, school or company. Libraries are naturally "Institution scale" But libraries need to increasingly prove their relevance in a "webscale" information ecosystem.

For example, Amazon uses the power of a cloud-computing infrastructure to provide a 24/7 online experience to online purchasers, as well as providing an "on-demand" provider of this computing infrastructure to others. Amazon provides a flexible platform on which developers and programmers can innovate and experiment using Amazon Web Services. Amazon leverages the Web environment to build, maintain and evolve a variety of relationships with others in order to provide an optimum user experience. Amazon recognizes the importance of being where their users are, rather than expecting the user will only use the Amazon portal. In short, Amazon understands the power of the cloud.

In the library ecosystem, a variety of organizations are working to build platforms that can deliver operational value for libraries. They are building shared infrastructure to reduce costs and increase efficiencies around the different elements of library workflow. Organizations like Duraspace and Portico are building shared infrastructures for preservation.

OCLC and others are creating platforms that allow libraries to manage collections across for-

mats (print, electronic and digital) in order to reduce the complexity and cost of back-office workflows. The important element with these efforts will be to ensure that the different platforms created can interoperate to maximize the network effects for the libraries using them.

In late 2011, OCLC launched the new OCLC WorldShare Platform and brand. OCLC World-Share provides a Web-based platform for collective innovation with shared services, integrated applications and a streamlined approach to managing library workflows.

The first services built on this new technical infrastructure are Webscale Management Services, which have been rebranded as OCLC WorldShare Management Services, and include circulation, acquisitions and license management applications. Over time, OCLC will bring together additional OCLC services and applications under the OCLC WorldShare name, including resource sharing, consortial borrowing, metadata management and additional applications. OCLC's currently deployed library management solutions will continue to be maintained and enhanced in line with libraries' ongoing requirements under their current brand names.

Together with WorldCat, WorldShare helps the world's libraries connect in new ways to operate, innovate and collaborate at Webscale. WorldCat represents the shared data about the world's libraries, connecting library data to other information providers on the open web, ensuring information seekers can find and get the materials and services available in the world's libraries. The OCLC WorldShare platform is the shared infrastructure—that will allow libraries to create, collect, manage and share their resources in new and more efficient ways at Webscale.

The WorldShare Platform is a result of that strategy put in motion in 2000 that included open

architecture, agile development, and continuous improvement based upon user feedback. The platform is built on a shared cloud-computing infrastructure and leverages the aggregated data in WorldCat, the WorldCat knowledge base and the WorldCat Institution Registry. The platform provides a flexible and broadly accessible environment. It supports applications developed by third parties. Going forward, it will provide a robust framework of Web services enabling other organizations, library developers and partners to create, configure and share a range of applications that deliver new functionality and value.

WorldShare platform partners at this writing included these commercial third parties: Atlas Systems, EasyBib, EBSCO and Google Books. Most recently, Relais International will use the OCLC platform for resource sharing and document delivery, including the WorldCat Search API and WorldCat Resource Sharing API. Sustainable Collection Services will use WorldCat and the platform to support its toolkit that helps academic libraries manage the de-selection and drawdown of low-use print monograph collections while supporting shared print archiving efforts.

The WorldShare Platform includes a toolbox for developers who want to use OCLC services and data to build new or enhance existing applications. There is an app gallery through which libraries can easily share apps created by the community. For example, the Alibris app allows libraries to upload Alibris order spreadsheets directly to the Acquisitions ordering module. The Amazon app allows libraries to place orders for materials directly with Amazon from WMS Acquisitions. The New York Times best seller app allows libraries to integrate the current NYT best seller list with data from WorldCat, WMS Circulation Holds and the WMS Acquisitions order module.

This shared infrastructure must, of course, be delivered globally. As mentioned earlier, OCLC is opening data centers in the United Kingdom, Australia and Canada to support its new Worldshare Management Services.

To date, more than 171 libraries have committed to WMS or the WorldShare platform, and 31 had gone live by early 2012. Libraries outside the U.S. have also signed up. The Norwegian consortium BIBSYS signed an agreement with OCLC to provide its new Library System based on OCLC's Web-scale Management Services. This will involve more than 100 libraries in Norway, including the National Library. In addition, Tilburg University Library in The Netherlands, the University of New Brunswick Libraries in Canada, and the University of Aruba have agreed to use WorldShare Management Services. In early 2012, the University of Delaware in the U.S. agreed to use WorldShare Management Services, becoming the first member of the Association of Research Libraries (ARL) to do so.

Conclusion

OCLC's mission and public purpose are clearly aligned with the concept of Webscale. From the beginning of online shared cataloging in 1971, OCLC has been providing infrastructure and services that have enabled members to build efficiencies in the management of libraries and to increase the visibility and availability of library resources through rapidly changing technology. „OCLC WorldShare provides even more of the cooperative services and systems that will allow libraries to leverage their activities and data together, at Webscale."