Goldner, Matthew R.

Moving libraries to a Web scale

Matthew R. Goldner

MOVING LIBRARIES TO A WEB SCALE

Abstrakt:

V posledních deseti letech došlo k obrovskému zvýšení internetové konektivity v České i Slovenské republice. Vzhledem k tomuto posunu v oblasti připojení k vysokorychlostnímu internetu je tematika přechodu knihoven do Internetové sféry více než relevantní. Přihlédneme-li navíc k posunu v oblasti počítačové a informační vědy, knihovnám se nabízí velká příležitost ke změně správy sbírek a ke změně způsobu, jak jsou informace přístupny těm, kteří je hledají prostřednictvím Internetu. Tento článek se věnuje právě tomuto novému trendu, jaké možnosti nabízí knihovnám a jak OCLC Cooperative ve spolupráci s členskými knihovnami využívají výhod, které nový trend poskytuje.

Klíčová slova: *Cloud computing, Web scale, automatizace knihoven, spolupráce knihoven*

Abstract:

The last 10 years has seen a tremendous increase in Internet connectivity for both the Czech Republic and Slovakia.[1] This advance in high-speed connectivity makes the discussion of moving libraries to Web scale quite relevant for librarians in both countries. And the timing coincides with a shift in computing abilities and a shift in the information supply chain to offer libraries an unprecedented opportunity to change how collections are managed and how information is made discoverable and accessible to information seekers on the Web. This paper will examine what these shifts are, what opportunities they afford libraries, and how the OCLC Cooperative is working with member libraries to take full advantage of these shifts.

Keywords: *Cloud computing, Web scale, Library automation, Library cooperation*

1 The shift in computing abilities

Libraries were at the forefront 30 years ago in taking advantage of computing power to acquire, manage and circulate their collections. Those systems were built on the latest available technology for existing formats and collections. This resulted in systems that were distributed across libraries with each library duplicating software support, and data creation and maintenance - operating as a silo to provide their users with access to information. Also, as the Internet and Web exploded upon the

---

1 Internet and World Stats, Usage and Population Statistics, Czech Republic, accessed May 12,2011, http://www.internetworldstats.com/eu/cz.htm and Slovakia, http://www.internetworldstats.com/eu/sk.htm.

world and information moved from physical items to electronic formats, the existing library management systems where not capable of effectively managing these new formats and adapting to a new digital world.

As Andrew Pace has stated, "While libraries seemingly accepted the fate that the basic functions provided by an integrated library system would not change radically, the nature of their collections and associated workflow were themselves changing rapidly. … The inadequacy of the ILS was compounded by a desire among vendors and libraries alike to build new solutions with new technologies. Electronic Resource Management (ERM), Digital Asset Management (DAM), and Institutional Repository (IR) systems would be built with 21st century technologies to aid in these new library workflows. Paradoxically, as industry expert Marhsall (sic) Breeding points out, '[The process of evaluating library workflow] may be confounded by the fact that many libraries have adapted their workflows to match the limitations of their automation systems" (Breeding, 2007).'" [2] [3]

The advent of cloud computing solutions offers libraries in countries with a good Internet infrastructure an opportunity to address both of these issues: the inadequacies of library system functionality and the expensive redundancy of individual system deployment. The Gartner Group defines cloud computing as "a style of computing in which massively scalable and elastic IT-enabled capabilities are delivered as a service to external customers using Internet technologies." [4] In various presentations, KPMG breaks this into essentially four different types of cloud computing: infrastructure, platform, applications and services. To put this in more concrete terms, examples of each can be:

Type	What it is	Examples
Infrastructure	Buying space / time on external servers	Amazon A3 Bungee
Platform	An existing software platform to build your own applications on	Facebook
Applications	Software applications accessed with a Web browser	Google Docs Salesforce.com
Services	Ready-to-use services accessed with a Web browser	ADP

2   Pace, Andrew, "21st Century Library Systems," *Journal of Library Administration*, 49:641–650, 2009.
3   Breeding, Marshall, "It's time to break the mold of the original ILS," *Computers in Libraries*, 39–41, 2007.
4   Gartner Group, "Gartner Highlights Five Attributes of Cloud Computing," accessed June 23, 2009,
     http://www.gartner.com/it/page.jsp?id=1035013.

The above table illustrates why there are varying definitions of cloud computing. Many cloud services actually incorporate two or more of these types. For example, Google Docs provide infrastructure as well as applications.

But it is what cloud computing enables that is most important. By using cloud solutions any enterprise can:

1. Stop supporting all the computing power they need locally, essentially outsourcing their computing needs
2. Share data with other enterprises
3. Build reusable services
4. Move to Web scale solutions rather than siloed solutions.

All of these will be discussed in more depth when looking at how OCLC is employing cloud computing to build Web-scale solutions for libraries, but an explanation of what is meant by Web scale is warranted at this point. Chris Anderson, Editor of *Wired* magazine and author of "The Long Tail," stated, "The Web is all about scale, finding ways to attract the most users for centralized resources, spreading those costs over larger and larger audiences as the technology gets more and more capable. It's not about the cost of the equipment in the racks at the data center; it's about what that equipment can do."[5] I would extend Anderson's statement to say that Web scale refers to the ability to massively aggregate data; build useful services on that data which then attract a massively aggregated pool of users. This leads to a result where every user adds value for every other user.

Using an example from the United States, eBay has built a truly Web scale solution. First they massively aggregated data about items for sale. Then they built an auction service to allow anyone on the Internet to bid on these items, thus massively aggregating users, both sellers and buyers. The result is the ability to sell almost anything because of the huge pool of buyers, but just as important is that every buyer adds value for every other buyer. This is because the buyers rate the sellers allowing new potential buyers to have quantitative data to make a decision of whether to purchase an item from a specific seller.

---

5   Anderson, Chris, "Free! Why $0.00 is the Future of Business," *Wired Magazine* 16.03, Feb. 25, 2008, http://www.wired.com/techbiz/it/magazine/16-03/ff_free?currentPage=all, accessed April 13, 2011.

2 The shift in information supply chains

Before the Web, access to information was scarce and the attention needed to obtain information was abundant. Libraries held a monopoly on access to much of the world's information. However the Web changed this by moving information from siloed sources that had to be sequentially accessed to massively aggregated information with useful services built on top of it. Google did this for search—finding information on the Web. Amazon did this for purchasing books, other media, and finally as the store front for hundreds of thousands of businesses. Facebook did this for social networking. And each became dominate in the area of information they have aggregated and built services on.

If libraries are to be a force in an Internet world, then moving library operations to Web scale is critical. Standing alone, library resources are fragmented and difficult to find. And as the *OCLC Perception* reports published in 2005, 2007 and 2010 show, users do not think of the library website as the place to look for information, but start at search engines, social sites and other large aggregations of information and users. In the original report published 2005, 82% of users started information searches at search engines and only 1% started at a library website. By 2010 84% started with a search engine and 0% (not a single respondent) started at the library. [6]

Another critical change in the information supply chain for libraries is the one from physical to licensed electronic, open-access electronic and digital materials. As stated previously, the systems we use to manage these new information formats have grown up to meet single specific needs and disaggregated the management of collections into silos with little or no connection between them. Before the Web, libraries worked as individual enterprises because collaboration was time-consuming and expensive. While there was some movement towards collaboration using existing computer technologies—mainly in the areas of cataloging and resource sharing—much of the work of libraries had to be repeated by each library individually. Examples would be maintaining vendor records, maintaining serial publication patterns, maintaining interoperability with systems that were external to the library management system, maintaining knowledge bases of electronic resources, giving access to important local digital collections, and the list goes on.

With the advent of Web-scale solutions enabled by cloud computing, libraries need to reconsider how they manage their collections and look for solutions that offer rational efficient workflows and unify their collective presence on the Web. Otherwise two problems will continue:

---

6   OCLC, *Perceptions of Libraries, 2010: Context and Community*, accessed May 10, 2011,
     http://www.oclc.org/us/en/reports/2010perceptions.htm.

1. Libraries will continue to lose efficiency as they manage their collections in different silos

2. It will continue to be difficult to give users uniform delivery of access to the collections being managed.

The result will be an increased lack of relevance of library collections to researchers and information seekers.

3 Seizing the opportunity to change

Cloud computing and Web-scale solutions offer several opportunities for positive change in how libraries manage their collections and make them accessible. The first is simply sharing computing power in a cloud environment instead of each library maintaining more computing power than is needed for local services. When purchasing a server libraries must purchase enough computing power for peak load periods. This means the majority of the time the system is running well below capacity. The point of the Gartner definition that refers to "massively scalable and elastic IT-enabled capabilities" is what changes the game. Instead of purchasing enough power for peak capacity all the time, cloud computing solutions expand instantly to meet peak loads for any single institution as they are needed. This both offers the potential of reduced costs but also means more efficient use of computing power to make our services more environmentally friendly.

Next is the opportunity to reduce redundancy of data creation and maintenance. As a simple example, even though libraries have made substantial progress in sharing cataloging records to reduce the burden of original cataloging, they still store, back up and index the same records thousands of times.

Reducing data redundancy leads to many potential workflow efficiencies. If libraries share an aggregated bibliographic database, several possibilities unfold:

1. Cooperative authority control work

2. An opportunity to change copy cataloging as it is known today since records do not need to move to another system

3. Cooperative control of common licensing terms

4. Cooperative maintenance of publication patterns, vendor data, knowledge base data

5. Shared access to open access collections and important digital collections.

The critical end result will be improved access to information for users and a more visible Web presence for individual libraries and the library community.

4 OCLC's Web-scale Management Service

In the last two years the fact that the Internet has allowed new approaches to library management has resulted in the development of next generation library management systems. Two examples are Ex Libris' new library management system Alma[7] and the Kuali Open Library Environment (OLE) project.[8] Each offers promise of open systems which can manage today's hybrid library collections and have the flexibility for tomorrow's new collections. To get an in depth view of what a next generation system that takes full advantage of cloud computing and Web scale this article will look more deeply on one new solution, OCLC's Web-scale Management Service.

As members of the OCLC cooperative became aware of the potential of cloud computing and Web-scale solutions, they opened discussions with the staff of OCLC about extending WorldCat and other OCLC services to offer libraries a different future for the management of their collections. It had become apparent that:

1. Libraries standing alone could not compete against the move to Web-scale services being delivered by search engines and others

2. Economic conditions would lead to continued funding losses for libraries, requiring them to find ways to do more with less

3. Libraries needed a new approach to managing the disparate physical, electronic and digital collections of today's library

4. By working together libraries could define and shape a better future.

As a result, the cooperative embarked on creating the Web-scale Management Services as a cloud computing solution to streamline routine back office library services—cataloging, acquisitions, e-resource management and circulation—delivered via a Web browser using shared data and services. It would offer a 21st century user experience and put library resources where users do their work, on the Web and in all types of computing and mobile devices.

The first phase of development was WorldCat Local, which offers a unified Google-like search interface for library users.[9] The University of Washington went live as the first pilot library in April 2007. Since that time there has been continuous enhancement of WorldCat Local based on 20 rounds of usability testing, user feedback, data analysis and librarian feedback.

---

7   Ex Libris Web site, accessed 19 August 2011, http://www.exlibrisgroup.com/category/AlmaOverview
8   Kuali OLE Web site, accessed 19 August 2011, http://www.kuali.org/node/425
9   It should be noted that the Web-scale type solution that WorldCat Local offers has also been developed by other service providers such as Serial Solution's Summon service, EBSCO's EDS and Ex Libris' Primo Central.

The result of this type of solution is it:

1. Allows users to search all of a library's collections through a single search box.

2. Puts library collections where users are by allowing harvesting of WorldCat and access to WorldCat via APIs. This puts WorldCat into Google, Facebook, blog sites, etc.

3. Offers users the best method(s) of access to items they have discovered by use of a smart delivery resolver, which intelligently determines the item's availability physically and electronically and if it is available locally, through a consortium or at libraries around the world.

Public services staff benefit from several aspects of WorldCat Local as the public interface. They are able to organize resources into search groups that make sense to the public, such as subject areas. Physical, licensed and digital collections can be added to multiple search groups so that all resources are used to their greatest extent.

Because WorldCat Local, and similar solutions from other providers, is built on a single central index, which includes WorldCat plus citations from licensed databases, users not only discover the library's collections but also those of other libraries, museums and archives around the world. If the library is a member of a consortium for sharing resources, these libraries can be configured as a group so their items are presented to users following items from the library's own collection. And if the library chooses, then search results from around the world will be presented next, all in one unified search result.

St. Xavier University Library, Chicago, Illinois, began using WorldCat Local in 2009. According to Mark Vargas, Library Director, faculty and students had expressed a high level of frustration with the tools the library provided to do research. WorldCat Local provided the library with the ability bring together the multitude of databases, journals, catalogs and other resources the library offered, as well as eliminate the time spent teaching patrons how to use all of the different services and systems.

"WorldCat Local takes the drudgery out of discovery," he says. "It just simplifies things tremendously. It's intuitive, and students and faculty can run with it."

"We completely transformed the way we do instruction and we went from three quarters of the time spent on mechanical stuff to basically no time. We've eliminated the teaching of the tool and we can push learning. We can spend more time on what research is really supposed to be about, and that is evaluating information and critical thinking skills."[10]

---

10 Video of presentation American Library Association mid-Winter meeting, 2011.
    http://www.oclc.org/us/en/worldcatlocal/success/default.htm

Like other service providers working on next generation services for libraries, the next step for OCLC was the development of back office services. Before starting to design these services, the development team spent a year determining the correct architecture that would be responsive, scalable, fault-tolerant and secure. The architecture had to be open, allowing reuse of data and services by external systems. And it was determined it needed to abstract core library workflows out of the applications with a workflow engine to allow flexibility for current workflows but adapt for future workflows. Once this was complete, development began and in July 2009 a pilot was launched with academic and public libraries.

For one year, using rapid development, features were rolled out monthly to the libraries for testing and continuous feedback. Usability studies were conducted on-site with library staff for the circulation, acquisitions and e-resource management services. Then in July 2010 the back-office services were launched for early adopters with the first library going live in September of that year.

An advantage of this type of cloud based library management system is the ability to have rapid improvement. All the Web-Scale Management Services from OCLC continue to have monthly installs of minor enhancements and quarterly installs for major enhancements. Libraries take advantage of updates immediately since it is a cloud based service meaning all libraries share the same software via their Web browser with nothing to install locally.

One of the early adopters is Pepperdine University Libraries, Malibu, California. Michael Dula, Director for Digital Initiatives and Technology Strategy, Pepperdine University Libraries, shared his institution's experience with Web scale management services at the LITA National Forum 2010.

"We have so many systems, and just trying to get them to talk to each other was hard. If we upgraded one, it seemed like we broke something somewhere else … So, we were ready for WMS … It fit our technology goals. We wanted to be moving out of the business of managing servers, and we loved the chance to be in on the ground floor … With WMS, system costs will drop significantly—like 'getting another FTE' significantly. We're going to save a lot of staff time."[11]

5 What a cloud based management system can change for your library

Systems like WMS free IT staff from the routine tasks of doing data backups, system and software upgrades and troubleshooting problems. This gives libraries the opportunity to use these resources for creating locally needed services and

---

11 Video of presentation at LITA Forum, 2010,
   http://www.oclc.org/multimedia/2010/wms_cloudy_with_a_chance_of_cooperation.htm

reallocating resources to other critical projects needed to serve the library's local community.

It allows backroom staff to reconsider and redesign workflows, which traditionally have been created to match the capabilities of the systems being used. By bringing the management of all types of collections and formats back together, there is the opportunity rationalize workflows that match the library's needs instead of adopting workflows allowed by the system. This is further enhanced by the services being built on a workflow engine, which will allow libraries to order tasks and assign them to the correct staff members.

Sharing data with other libraries reduces redundant efforts to create, maintain and preserve data. A few examples are:

1. Libraries share a common vendor file, meaning all the core information related to a vendor is created and maintained once for everyone. The benefit is once interoperability to a vendor for electronic ordering and invoicing is put in place, it is there for all libraries to use. This removes the burden for each library to configure, test and maintain this function.

2. Licensing data that is common for all libraries is shared to simplify the maintenance of license terms. And because the system's workflow has been unified for all formats licensing data is integrated to the acquisitions workflow and subscription management.

3. Serial publication patterns are created once and then shared with all libraries on the system.

4. The global knowledge base of electronic resources is maintained in one place and then shared by all libraries for configuring to their locally licensed resources.

5. Open access materials are cataloged and managed in one place so any library can easily include them as part of their local collection.

As stated previously, OCLC's Web-scale Management Service has been created with an open architecture to allow re-use of the services within it. For example, the circulation module uses a suite of services including check-out, check-in, notifications, billing and patron management. These same services are externally exposed through documented APIs allowing anyone to connect an external service to the Web-scale Management Service or even reuse these services to enhance external services. This simplifies the process of talking to external vendor systems, institution financial systems, self check-out stations, etc.

A radical shift in the architecture of the service layer of the Web-scale Management Service was to turn it into a cooperative platform. Simply put this means that not only

OCLC staff can add services and applications to the Web-scale Management Service, but the platform is opened to the library community so that any library, vendor or other third party can add services and applications that can then be shared by all libraries. This truly allows libraries to have an active role in defining a new future for how collections are managed and disclosed to the world.

6 Cooperatively building a different future for libraries

From their inception, the OCLC Web-scale Management Service was built in cooperation with OCLC member libraries. Each part of the service, WorldCat Local, circulation, acquisitions and e-resource management, has been put through rigorous pilot testing and usability testing directly with users and library staff. By using rapid development methods OCLC has been able to quickly respond to lessons learned during pilots and ongoing testing following release of the services into production environments.

Since it is a cloud-based service, each time a new function or service is added every library immediately takes advantage of these features. Libraries are already experiencing the savings in staff time needed to maintain local systems since the burden of managing servers, back-ups and upgrades has been removed. Working closely with OCLC staff librarians are finding ways to improve their workflows to match to today's hybrid collections again resulting in savings of time and effort.

For libraries in countries which have not previously participated in the OCLC cooperative it does offer a choice going forward in how to manage library collections. OCLC has successfully partnered with national libraries in many countries to enrich the breadth of bibliographic data available to libraries and allow libraries around the world to benefit from the cataloging efforts of librarians in these countries. Once this first step is taken of participating in the cooperative the door is open to take full advantage of what a cloud computing Web-scale solution can bring to libraries in any country with a strong internet infrastructure.

The Web-scale Management Service does offer libraries around the world the opportunity to define and shape a different future. One based on 21st century technology, increased workflow efficiencies, reduction of redundant data management and an open system that allows the community to actively participate in the creation of new shared services. It truly takes advantage of what cloud computing and Web-scale solutions can offer and will provide libraries with a unified, global presence on the Web.

References

1. Anderson, Chris, "Free! Why $0.00 is the Future of Business," Wired Magazine 16.03, Feb. 25, 2008, http://www.wired.com/techbiz/it/magazine/16-03/ff_free?currentPage=all, accessed April 13, 2011.

2. Ex Libris Web site, accessed 19 August 2011, http://www.exlibrisgroup.com/category/AlmaOverview

3. Gartner Group, "Gartner Highlights Five Attributes of Cloud Computing," accessed June 23, 2009,   http://www.gartner.com/it/page.jsp?id=1035013.

4. Internet and World Stats, Usage and Population Statistics, Czech Republic, accessed May 12,2011, http://www.internetworldstats.com/eu/cz.htm and Slovakia, http://www.internetworldstats.com/eu/sk.htm.

5. Kuali OLE Web site, accessed 19 August 2011, http://www.kuali.org/node/425

6. Breeding, Marshall, "It's time to break the mold of the original ILS," *Computers in Libraries*, 39–41, 2007.

7. Pace, Andrew, "21st Century Library Systems," *Journal of Library Administration*, 49:641–650, 2009.

8. OCLC, *Perceptions of Libraries, 2010: Context and Community*, accessed May 10, 2011, http://www.oclc.org/us/en/reports/2010perceptions.htm.

9. Video of presentation American Library Association mid-Winter meeting, 2011. http://www.oclc.org/us/en/worldcatlocal/success/default.htm

10. Video of presentation at LITA Forum, 2010, http://www.oclc.org/multimedia/2010/wms_cloudy_with_a_chance_of_cooperation.htm